



SELF-ORGANIZING MAP WITH NGUYEN-WIDROW INITIALIZATION ALGORITHM FOR GROUNDWATER VULNERABILITY ASSESSMENT

Maureen Nettie Linan ¹⁾, Bobby Gerardo ²⁾, Ruji Medina ¹⁾

¹⁾ Graduate Programs, Technological Institute of the Philippines, 1109 Cubao, Quezon City, Philippines, maureenettie@gmail.com, ruji.medina@tip.edu.ph

²⁾ College of Information and Communication Technology, Lapaz, 5000 Iloilo City, Philippines, bgerardo@wvsu.edu.ph

Paper history:

Received 25 April 2019

Received in revised form 23 October 2019

Accepted 17 February 2020

Available online 31 March 2020

Keywords:

Clustering algorithm;

DRASTIC;

Groundwater assessment;

Self-organizing map;

Small island.

Abstract: Assessment of groundwater vulnerability to contamination plays a vital role in the utilization and protection of groundwater resource. In this study, a vulnerability map for Boracay Island, Philippines was developed using a modified self-organizing map algorithm to determine groundwater vulnerability in light of massive tourism developments in the island. Self-organizing map using the Nguyen-Widrow initialization algorithm was used to cluster DRASTIC data which were pre-processed using data cleaning normalization schemes. The vulnerability map developed showed that groundwater resource in the island is susceptible to contamination as confirmed by groundwater quality analysis. The result of the study demonstrates the effectiveness of the improved SOM algorithm as a tool for assessment of groundwater vulnerability and is comparable with the traditional DRASTIC method. The developed methodology allows grouping of datasets into clusters that represent the level of vulnerability to contamination of the groundwater. Further, this approach can be applied to other islands to ensure the balance between tourism developments and ecological integrity of the scarce groundwater resource.

Copyright © Research Institute for Intelligent Computer Systems, 2020.

All rights reserved.

1. INTRODUCTION

Groundwater quality is a primary factor in resource management. It is critical since contamination may render groundwater unsafe to use with environmental and health consequences. Contaminants, though naturally filtered through soil and rock formations, still threaten the water resource quality. The gravity of vulnerability assessment of groundwater is a priority, especially for small islands, whose sustaining ability to maintain groundwater is at risk.

Several approaches were developed to evaluate groundwater vulnerability [1, 2]. The most common among them is the DRASTIC Method, an overlay and index method frequently adopted in many studies. It employs seven parameters to estimate the vulnerability index – depth to water (D), net recharge (R), aquifer media (A), soil media (S), topography (T), vadose zone impact (I) and hydraulic conductivity (C). The vulnerability index

is calculated by giving relative ranks and weights to these seven parameters [3–5]. Its limitations lie in the subjectiveness of assigning numerical values to the descriptive entities and relative weights of different attributes [6]–[8]. Moreover several studies modified and tested the validity of the DRASTIC model, i.e. using artificial neural networks (ANN) [9,10] and clustering algorithms [11–13].

Clustering algorithms group set of data into multiple clusters so that objects within a cluster have high similarities. The study adopted the Self-organizing map (SOM) clustering algorithm to assess groundwater vulnerability. SOM, an artificial neural network model, is one of the algorithms used to cluster and visualize results through reduction of high-dimensional data [14]. It provides visualized results essential in studies like water quality assessment in hydro-geochemically complex areas, water chemical composition assessments, hydro-geochemical characterization, and water pollution mapping [15,16]. Also used for spatial analysis of

groundwater quality, SOM defined a classification based on the hydrochemical characteristics of groundwater. SOM algorithm successfully classified and characterized groundwater in terms of quality [17]. When combined with DRASTIC to produce a Hybrid SOM-DRASTIC model, SOM was able to classify groundwater vulnerability utilizing weighted ratings of DRASTIC parameters as input and vulnerability classes as output and using GIS software to develop the vulnerability maps. It proved useful for managers and land use planners by providing a robust alternative tool towards developing vulnerability-based classification and land use planning strategy [18].

The accuracy of the clustering result of SOM is dependent on weight initialization [19, 20], thus, proper initialization of cluster centers is important and critical. Since weights in SOM are randomly initialized, the quality of results together with learning speed is greatly affected [21]. To address the issue of clustering in this study, the Nguyen-Widrow initialization algorithm was implemented. Nguyen-Widrow initialization algorithm is a method for initialization of the weights of neural networks to reduce training time. In weight initialization, some small number of random values for weight is assigned for the operation of backpropagation networks [22]. Hence, this study introduces the use of the Nguyen-Widrow initialization algorithm in the SOM algorithm to assess groundwater vulnerability to contamination in Boracay Island, Philippines. The study presents the use of the initialization algorithm in weights initialization for SOM clustering. The dataset of DRASTIC ratings is normalized and trained prior to clustering. A groundwater vulnerability map is developed from the extracted clusters.

2. STUDY AREA

Boracay is a small island with a total land area of 1,006.64 hectares located between 11°57'–12°00' latitude and 121°56'–121°57' longitude off the northwestern tip of Panay island in the Western Visayas, Philippines. The land mass measures 6.8 kilometers long by 3.3 kilometers at its widest distance, rising to elevations ranging from 50 to 105 meters above mean sea level. The depth to the water table as measured from existing wells in the three administrative units called barangays varies from 0.58 to 4.096 m based on the report of Mines and Geosciences Bureau (MGB) Regional Office VI

[23]. The position of Boracay as the Philippines' top tourist destination has placed immense pressure on its groundwater resources owing to overwithdrawal and contamination from human activities. Over the years, tourist influx has increased dramatically and local immigrants have also settled in the island due to economic opportunities from tourism [24] such as hotels, restaurants, and other establishments openings. The increase in population and proliferation of residential and commercial establishments are negatively impacting the quality of the island's groundwater resources, thus the need to evaluate the vulnerability to contamination of its resources.

3. METHODOLOGY

The SOM algorithm is an unsupervised learning algorithm widely used for clustering large sets of data. SOM is made up of two layers. The input layer is one-dimensional with each data item associated with an n-length vector of elements, while the output layer consists of radial units typically organized in one or two-dimensions [14]. In this study, SOM is consists of 7-input vectors and 4-output neurons. Each input node is connected to the output neurons. The SOM architecture of the study is presented in Fig. 1.

The process flow for the study outlined in Fig. 2 utilized data from available maps of Boracay Island and the thematic map of the DRASTIC parameters developed through GIS. The dataset includes depth to water (D), net recharge (R), aquifer media (A), soil media (S), topography (T), vadose zone impact (I) and hydraulic conductivity (C). Ratings, as defined in the DRASTIC method, were used. The dataset was divided into two parts: 70% of the data used for training, 30 % for clustering.

The data sets were pre-processed using data cleaning normalization schemes. With the normalization process, better clustering result is achieved and the negative effects produced by noise and outliers can be avoided. The data were normalized using the equation (1):

$$X = \frac{X_i - \min(X_r)}{\max(X_r) - \min(X_r)} \quad (1)$$

where X_i is the raw data, $\min(X_r)$ the smallest value in X_r and $\max(X_r)$ the largest value in X_r .

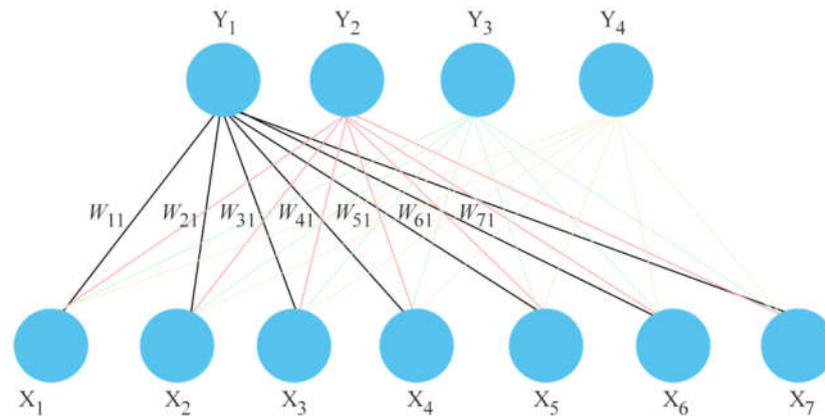


Figure 1 – The SOM architecture

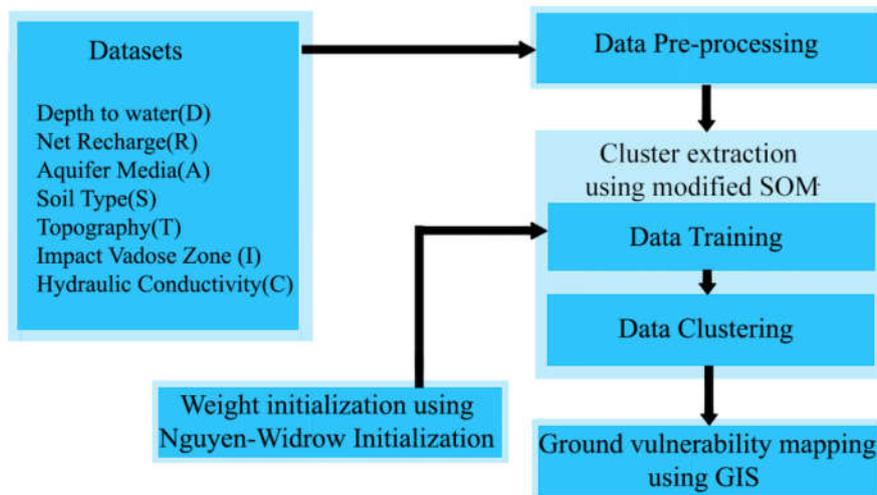


Figure 2 – The methodology of the study

Clustering of data using the modified SOM algorithm is achieved mainly by two steps: 1. training the data with the initialized parameters; and 2. clustering of data. At the start of the learning process, weights were initialized to small numbers using the Nguyen-Widrow initialization algorithm.

Input vectors are then randomly selected from the given dataset. Winning neuron, also called the Best Matching Unit (BMU), was determined using Euclidean distance. To increase the similarity with the input vector, weights are then adjusted. The process will be repeated until all the vectors are trained. The final values of these weights are then used for clustering.

Clusters were extracted using the SOM algorithm with Nguyen-Widrow for initialized weights. The pseudocode for the modified SOM algorithm is listed:

ALGORITHM: SOM algorithm with Nguyen-Widrow initialization algorithm [25]

1. Initialize weight using the Nguyen-Widrow initialization algorithm

- a. Initialize all weights w_i of hidden layers with random values over the interval $[-0.5$ to $0.5]$
- b. For each hidden layer, calculate beta value β

$$\beta = 0.7 * H^{(1/n)} \quad (2)$$

where n is the number of input units and H is the number of hidden units

- c. For each synapse
 - i. For each weight $\{$
 - Adjust weight w_{ij} by dividing by norm of weight for neuron and multiplying by the beta value

$$w_{ij} = \beta * \frac{w_{ij}(random)}{\|w_i(random)\|} \quad (3)$$

where $random$ is between $[-1.0$ to $1.0]$

- 2. Randomly select an input vector $x(t)$.
- 3. Find “Best Matching Unit (BMU)” c , using the equation:

$$c = arg(\min_{1 \leq i \leq mn} \{\|w_i(t) - x(t)\|\}), \quad (4)$$

$\| \cdot \|$ is the Euclidian distance measure, $x(t)$ and $w_i(t)$ are the input and weight vector of neuron i at iteration t respectively.

- Update the weight vector of the neurons using the following equation:

$$w_i(t + 1) = w_i(t) + h_{c,i}(t)[x(t) - w_i(t)] \quad (5)$$

where $h_{c,i}(t)$ is a Gaussian neighborhood function.

- Repeat steps 2 to 5 for all the input data.

The extracted clusters were used to develop a vulnerability map using GIS software.

4. RESULTS AND DISCUSSION

Initial weights were generated using the Nguyen-Widrow algorithm from which final weights were obtained for SOM clustering. Three groundwater clusters were determined, as shown in Table 1.

Table 1. Cluster Mean

Attribute	Clusters		
	1	2	3
Depth to Water (D)	10.0	10.0	10.0
Net Recharge (R)	8.0	8.0	3.0
Aquifer Media (A)	8.0	9.0	6.0
Soil Type (S)	9.7	3.0	3.0
Topography (T)	9.6	9.6	4.4
Impact Vadose Zone (I)	8.0	9.0	6.0
Conductivity (C)	10.0	4.0	1.0

All DRASTIC parameters play a role in determining vulnerability indices. However, four parameters, - depth to water, net recharge, topography, and hydraulic conductivity - are significantly more influential than the other parameters in assessing groundwater vulnerability [7, 25]. Hence, these parameters are used in discussing the resulting clusters using modified SOM algorithm.

Clusters 1, 2, and 3 are assigned to areas with very high vulnerability, high vulnerability, and moderate vulnerability, respectively. From the indices, Cluster 1 is made up of areas with high net recharge, topography, and hydraulic conductivity, all of which contribute to very high vulnerability. A higher rating in topography represents the lowest slope which implies that these low lying areas tend to retain water longer due to its almost uniform slope allowing greater infiltration of recharge and greater potential for contaminant migration. Referencing geological data in areas with very high vulnerability are observed to dominantly contain unconsolidated calcareous sand and silty clay deposits. This soil structure has large pores between them permitting

more rapid permeability of recharge and faster contaminant transport.

Cluster 2 includes areas with high net recharge and high topography rating but with moderate hydraulic conductivity. This high vulnerability cluster is observed in areas with coralline limestone, calcareous sandstone, siltstone, shale, and basal conglomerate which is a medium-textured soil and has an intermediate rate of soil permeability.

It was observed that Clusters 1 and 2 are lowland areas with a slope from 0-6%. On the other hand, Cluster 3 areas had a low net recharge and hydraulic conductivity rating but with moderate topography. Cluster 3 is located in hilly areas within the island, with a slope ranging from 6-18%.

Across the three clusters, depth to water table had a rating of 10.0 showing that the island is highly susceptible to contamination from surface pollution. This was confirmed in the groundwater data which shows a water table 0-5 meters deep [23].

The clusters generated using the modified SOM were used to develop the vulnerability map of Boracay Island as shown in Fig. 3. The clustering results were confirmed with the water quality analysis of Boracay Island presented in Table 2. Chloride analysis revealed that groundwater in all barangays contained high concentrations of chlorides in the groundwater compared to the standard of 250mg/L maximum. Groundwater in barangays Manoc-manoc, Yapak, and Balabag contained 420 mg/L, 1,999 mg/L, and 3,129 mg/L, respectively. Results of bacteriological analysis also showed that water is unsafe for drinking based on the elevated number of bacteria present, with values ranging from 2.20 to 16.00 MPN/100 ml where the standard value is 0.00 MPN/100 ml. Meanwhile, coliform test results show that groundwater from Brgy. Balabag has fecal and total coliform values of 16 MPN/100ml while those in Brgy. Yapak and Brgy. Manoc-manoc values excess 16 MPN/100ml. The groundwater samples all exceed the standard of 0 MPN/100 mL and show that the water is unsafe for drinking [23]. The water quality results are consistent with the vulnerability mapping developed using clustering. This shows that using the clustering method for assessing groundwater vulnerability is an effective method for assessment of groundwater vulnerability to contamination.

To verify the validity of the modified SOM method used in groundwater vulnerability assessment, T-test was used to determine if data results for the cluster means and DRASTIC means are significantly different from each other. The P-values calculated for each cluster show that the results are not significantly different for each method implying that groundwater vulnerability assessment using modified SOM is comparable with the traditional DRASTIC method. Table 3 shows the result of the T-test.

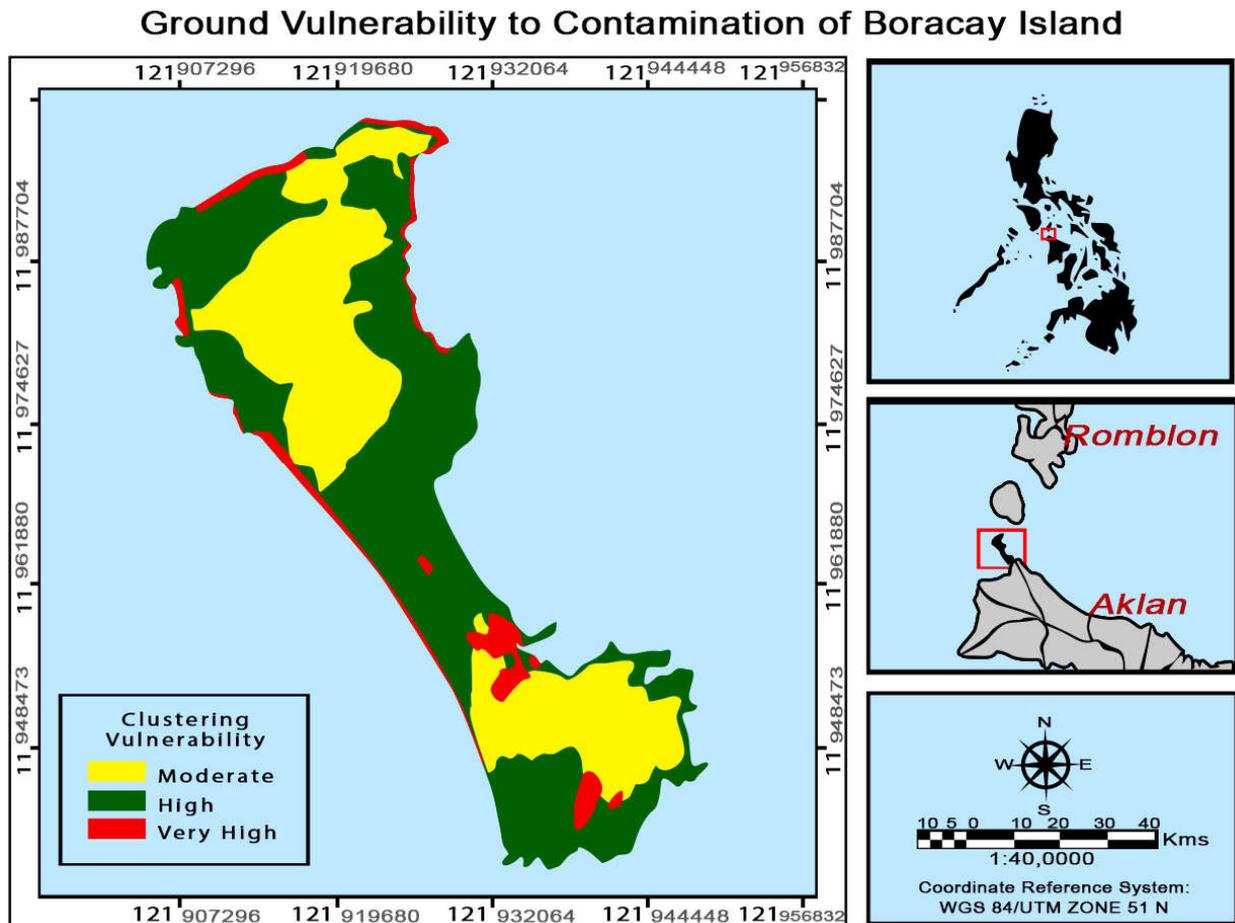


Fig. 3: Groundwater vulnerability map of Boracay Island using modified SOM algorithm

Table 2. Water quality analysis

Baran gay	Bacteriological analysis (Standard: 0 MPN/ 100ml)	Chloride Analysis (Standard: 250 mg/L)	Total Coliform (Standard: 0 MPN/100ml)	Fecal Coliform (Standard: 0 MPN/100ml)
Manoc- manoc	2.2 to 16 MPN/ 100ml	420 mg/L	>16 MPN/100 ml	>16 MPN/100 ml
Yapak	2.2 to 16 MPN/ 100ml	1999 mg/L	>16 MPN/100 ml	>16 MPN/100 ml
Balabag	2.2 to 16 MPN/ 100ml	3129 mg/L	16 MPN/100 ml	16 MPN/100 ml

Table 3. T-test Result

	T-value	P-value
Very High	-0.07706	0.469925
High	-0.07875	0.223144
Moderate	0.68811	0.252236

5. CONCLUSION

Modified SOM clustering is an applicable method for groundwater contamination vulnerability assessment. Further, the use of modified SOM algorithm methodology allows grouping of datasets

into clusters that represent the level of vulnerability to contamination of the groundwater. The vulnerability map developed out of the clustered data clearly revealed that the groundwater resources in Boracay Island were vulnerable to contamination

as confirmed by the water quality analysis. Finally, the output of the clustering method is comparable with the traditional DRASTIC method and is an effective method for assessment of groundwater vulnerability to contamination.

6. REFERENCES

- [1] S. Stevenazzi, M. Masetti, and G. P. Beretta, "Groundwater vulnerability assessment: from overlay methods to statistical methods in the Lombardy Plain area," *Acque Sotterranee – Italian Journal of Groundwater*, vol. 6, no. 2, pp. 17-27, Jun. 2017.
- [2] A.V. Deshpande and S.N. Patil, "Assessment of groundwater quality by using statistical analysis from kopargaon taluka, Ahmednagar, India," *International Journal of Advanced Geosciences*, vol. 4, no. 2, pp. 15–20, Jun. 2016.
- [3] A. Z. Abiy, A. M. Melesse, Y. M. Behabtu, B. Abebe, *Groundwater Vulnerability Analysis of the Tana Subasin: An application of DRASTIC Index Method*, in: A. Melesse, W. Abteu (Eds.), *Landscape Dynamics, Soils and Hydrological Processes in Varied Climates*, Springer Geography, Springer, Cham, 2016, pp. 435-461.
- [4] C. Jaseela, K. Prabhakar, and P. S. P. Harikumar, "Application of GIS and DRASTIC Modeling for Evaluation of Groundwater Vulnerability near a Solid Waste Disposal Site," *International Journal of Geosciences*, vol. 7, no. 4, pp. 558–571, 2016.
- [5] T. O. Abdullah, S. S. Ali, N. A. Al-Ansari, and S. Knutsson, "Groundwater Vulnerability Using DRASTIC and COP Models: Case Study of Halabja Saidadiq Basin, Iraq," *Engineering*, vol. 8, no. 11, p. 741-760, Jan. 2016.
- [6] S. Javadi, N. Kavehkar, K. Mohammadi, A. Khodadadi, and R. Kahawita, "Calibrating DRASTIC using field measurements, sensitivity analysis and statistical methods to assess groundwater vulnerability," *Water International*, vol. 36, no. 6, pp. 719–732, Oct. 2011.
- [7] S. Saidi, S. Bouri, H. Ben Dhia, and B. Anselme, "Assessment of groundwater risk using intrinsic vulnerability and hazard mapping: Application to Souassi aquifer, Tunisian Sahel," *Agricultural Water Management*, vol. 98, no. 10, pp. 1671–1682, Aug. 2011.
- [8] S. F. Tavassol and G. S. Gopalakrishna, "Pb contamination and analysis of Aquifer in Karaj Plain, Alborz Province, Iran using GIS-based DRASTIC Model," *Bulletin of Environment, Phramacology and Life Sciences*, vol. 3 (Spl issue II), pp. 263-271, 2014.
- [9] R. Barzegar, A. A. Moghaddam, and H. Baghban, "A supervised committee machine artificial intelligent for improving DRASTIC method to assess groundwater contamination risk: a case study from Tabriz plain aquifer, Iran," *Stochastic Environmental Research and Risk Assessment*, vol. 30, no. 3, pp. 883–899, Mar. 2016.
- [10] M. A. Baghapour et al., "Optimization of DRASTIC method by artificial neural network, nitrate vulnerability index, and composite DRASTIC models to assess groundwater vulnerability for unconfined aquifer of Shiraz Plain, Iran," *Journal of Environmental Health Science and Engineering*, vol. 14, no. 13, pp. 1-16, Aug. 2016.
- [11] S. Javadi, S. M. Hashemy, K. Mohammadi, K. W. F. Howard, and A. Neshat, "Classification of aquifer vulnerability using K-means cluster analysis," *Journal of Hydrology*, vol. 549, pp. 27–37, Jun. 2017.
- [12] A. A. Nadiri, M. Gharekhani, R. Khatibi, and A. A. Moghaddam, "Assessment of groundwater vulnerability using supervised committee to combine fuzzy logic models," *Environmental Science and Pollution Research*, vol. 24, no. 9, pp. 8562–8577, Mar. 2017.
- [13] H. Zou, Z. Zou, and X. Wang, "An Enhanced K-Means Algorithm for Water Quality Analysis of the Haihe River in China," *International Journal of Environmental Research and Public Health*, vol. 12, pp. 14400–14413, Nov. 2015.
- [14] T. Kohonen, "Essentials of the self-organizing map," *Neural Networks*, vol. 37, pp. 52–65, Jan. 2013.
- [15] T. T. Nguyen, A. Kawamura, T. N. Tong, N. Nakagawa, H. Amaguchi, and R. Gilbuena, "Clustering spatio-seasonal hydrogeochemical data using self-organizing maps for groundwater quality assessment in the Red River Delta, Vietnam," *Journal of Hydrology*, vol. 522, pp. 661–673, Mar. 2015.
- [16] K. Nakagawa, H. Amano, A. Kawamura, and R. Berndtsson, "Classification of groundwater chemistry in Shimabara, using self-organizing maps," *Hydrology Research*, vol. 48, no. 3, pp. 840–850, Jun. 2017.
- [17] L. Belkhiri, L. Mouni, A. Tiri, T. S. Narany, and R. Nouibet, "Spatial analysis of groundwater quality using self-organizing maps," *Groundwater for Sustainable Development*, vol. 7, pp. 121–132, Sep. 2018.
- [18] F. Rezaei, M. R. Ahmadzadeh, and H. R. Safavi, "SOM-DRASTIC: using self-organizing map for evaluating groundwater

potential to pollution,” *Stochastic Environmental Research and Risk Assessment*, vol. 31, no. 8, pp. 1941–1956, Oct. 2017.

- [19] E. Z. Moattar, M.-R. Feizi-Derakhshi, and M. Bakhshi, “Improvement of Self-Organizing Maps Algorithm with Weighting Optimization,” *International Journal of Advances in Electronics and Computer Science*, vol. 3, no. 4, p. 74-76, Apr. 2016.
- [20] A. A. Akinduko, E. M. Mirkes, and A. N. Gorban, “SOM: Stochastic initialization versus principal components,” *Information Sciences*, vol. 364–365, pp. 213–221, Oct. 2016.
- [21] D. Jude Hemanth, J. Anitha, *Computational Intelligence Techniques for Pattern Recognition in Biomedical Image Processing Applications*, in: S. Kulkarni (Ed.), *Machine Learning for Problem Solving in Computational Applications: Intelligent Techniques*, Information Science Reference (imprint of IGI Global), USA, 2012, pp. 195-209
- [22] S. Aisyah, M. Harahap, A. M. Husein Siregar, and M. Turnip, “Optimization of training backpropagation algorithm using nguyen widrow for angina ludwig diagnosis,” *Journal of Physics: Conference Series*, vol. 1007, 012050, 2018. [Online]. Available: <https://iopscience.iop.org/article/10.1088/1742-6596/1007/1/012050/pdf>
- [23] E. L. Linan, V. B. Ella, and L. M. Florece, “GIS-Based Assessment of Groundwater Vulnerability to Contamination in Boracay Island Using DRASTIC Model,” *Journal of Environmental Science and Management*, vol. 16, no. 2, pp. 19-27, Dec. 2013.
- [24] “Population of Region VI - Western Visayas (Based on the 2015 Census of Population) | Philippine Statistics Authority.” [Online]. Available at: <https://www.psa.gov.ph/content/population-region-vi-western-visayas-based-2015-census-population>. [Accessed: 24-Apr-2018].
- [25] M. N. N. Linan, B. D. Gerardo, and R. P. Medina, “Improving self-organizing map with nguyen-widrow initialization algorithm,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, no. 1, pp. 535–542, Jul. 2019.
- [26] A. Nawafleh, M. Awawdeh, and E. Salameh, “Assessment of Groundwater Vulnerability to Contamination in Irbid Governorate, North Jordan,” *DIRASAT: Pure Sciences*, vol. 38 no. 2, pp. 122-133, Jul. 2011.



Maureen Nettie N. Linan is a CHED-FDP II scholar currently taking her Doctor in Information Technology degree at the Technological Institute of the Philippines, Quezon City. She received her Bachelor’s degree in Information and Computer Science from the Cebu Institute of Technology University, Philippines and her Masters in Information Systems from the University of the Philippines Open University as CHED-MIS scholar. Presently, she is an Associate Professor at the Iloilo Science and Technology University, Iloilo City, Philippines. Her research interest includes database management, software engineering, and data mining.



Bobby D. Gerardo, a Professor VI of the College of ICT, is currently the Vice President for Administration and Finance of West Visayas State University, Iloilo City, Philippines. His dissertation is on *Discovering Driving Patterns using Rule-based intelligent Data Mining Agent (RiDAMA) in Distributed Insurance Telematic Systems*. He has published more than 100 research papers in national and international journals and conferences. He is a referee to international conferences and journals such as in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Knowledge and Data Engineering*, *Elsevier Journal on Telematics*, *Future Generation Computer Systems* and on *Bioinformatics*. His research fields are in the area of distributed systems, telematics systems, CORBA, data mining, web services, ubiquitous computing, and mobile communications.



Ruji P. Medina is Dean of the Graduate Programs of the Technological Institute of the Philippines in Quezon City. He holds a Ph.D. in Environmental Engineering from the University of the Philippines with sandwich program at the University of Houston, Texas where he worked on the synthesis of nanocomposite materials. He finished his MS in Environmental Engineering from the Mapúa Institute of Technology, graduating Summa Cum Laude. He obtained his Bachelor’s degree in Chemical Engineering from the University of the Philippines in Diliman, Quezon City. His research interests include urban mining, electronic wastes, and nanomaterials. He counts among his expertise environmental modeling and mathematical modeling using multivariate analysis.