Can an integrated ground water vulnerability mapping tool facilitate sensitivity analysis in a spatial domain?

B. Dixon
Dept. of Environmental Science Policy and Geography,
University of South Florida, St. Petersburg

Abstract

Modeling ground-water vulnerability reliably and cost effectively from non-point source pollution at a regional scale remains a major challenge. In recent years, Geographic Information Systems (GIS), neural networks and fuzzy logic techniques have been used in several hydrological studies; however, very few of these research studies undertook an extensive sensitivity analysis. Therefore, the overall objective of this research was to examine the sensitivity of neuro-fuzzy models in a spatial context by integrating GIS and neuro-fuzzy techniques. The specific objective was to assess the sensitivity of neuro-fuzzy models by varying input parameters to the models. The research reports a case study of Woodruff County, located in the Mississippi Delta part of Arkansas, US. This county was selected for its extensive agricultural landuse and the presence of underlying sand and gravel (alluvial) aquifer.

The neuro-fuzzy models were developed using NEFCLASS-J software in a JAVA platform and were loosely integrated with a GIS. Various plausible parameters that are critical in transporting contaminants in and through the soil profile to the ground-water included depth to ground water, recharge of the ground water, thickness of the claycap, soil drainage class, soil hydrologic group, soil structure, and landuse. Water quality data from 55 wells were used for validation of the models. Bentazon was the most commonly found contaminant. In order to validate the model predictions, coincidence reports were generated among model inputs, model predictions and well contamination data for pesticides. A total of 4 neuro-fuzzy models were developed. The sensitivity analysis showed that neuro-fuzzy models were sensitive to the input data layers used in the models.

Keywords: GIS, GPS, pesticides, modeling, neuro-fuzzy.

1 Introduction

Modeling ground-water vulnerability reliably and cost effectively from non-point source pollution at a regional scale remains a major challenge. An integrated system of advanced information technologies such as Geographic Information Systems (GIS), GPS, remote sensing and fuzzy logic could provide a framework from which real-time or simulated assessment of non-point source (NPS) pollution can be made [8]. Loague and Corwin [17] have shown that integrated process-based computer simulations with GIS could be a useful tool in regional scale assessment of NPS contamination of ground water. Integration of GIS and vulnerability indices that allows generation of sensitivity maps could improve management of water resources and landuse [7, 10, 12]. Research has shown that fuzzy rule-based models are capable of producing comparable results using about 40 percent fewer variables [2]. The fuzzy rule-based approach has been used in solute transport studies [14] and assessing an aquifer's pollution potential [6, 10]. Coupling between GIS and fuzzy rule-based techniques are particularly useful when modeling fuzzy inputs common to hydrogeologic parameters because they tolerate imprecision and uncertainty and show marked reduction in information loss when used with simple GIS techniques [22, 4, 21, 5]. For more detailed discussion on fuzzy logic please refer to [23, 18, 9, 6, 10].

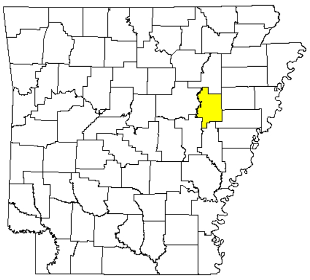


Figure 1: Location of the study area.

Neuro-fuzzy modeling is an approach where the fusion of NN and fuzzy logic find their strengths and complement each other [9, 10, 16, 19]. A key disadvantage of fuzzy logic based approach is inability to meet pre specified accuracy and lack of self-learning and generalization capability. Neuro-fuzzy approach employs heuristic learning strategies derived from the domain of NN theory to support the development of a fuzzy system. A marriage between NN and fuzzy logic techniques should help overcome the shortcomings of both techniques discussed at length by [19]. A neuro-fuzzy technique can learn a system's behavior from a sufficiently large data set and automatically generate fuzzy rules and fuzzy sets to a pre-specified accuracy level. They are capable of generalization, thus overcoming the key disadvantages of fuzzy logic based approach. A fusion of NN and fuzzy logic provides a system that usually requires less computational power but has the ability to generalize and learn through the convergence of net. This research reports a case study of Woodruff County, Arkansas (Figure 1). This county was selected for its extensive agricultural landuse and the presence of underlying alluvial aquifer.

2 Objectives

The specific objective of this study was to assess the sensitivity of neuro-fuzzy models by varying input parameters (Figure 2).

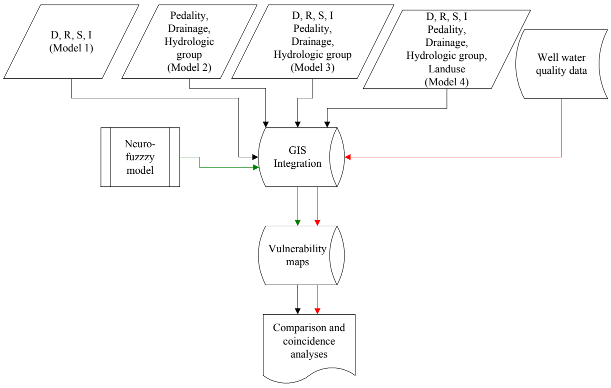


Figure 2: Flowchart showing objective.

3 Methodology

3.1 Development of GIS data layers

This study uses various combinations of eight data layers as input to the neuro-fuzzy models including selected parameters from the DRASTIC model [1]. The input data layers are D (depth to ground water), R (recharge of the aquifer), S (soil media) and I (thickness of the clay cap as it indicates impact of the vadose zone), pedality, drainage, hydrologic group and Landuse. Selected variables are combined to determine sensitivity of the models to the input parameters. Hydrogeologic data such as D, R and I were obtained from US Geological Survey. See [10] for more details. Landuse (LULC) data was derived from classified images from Landsat TM5. The soils data viz. soil drainage class (D_r) and hydrologic group (H) were obtained from the Natural Resource Conservation Service (NRCS). Pedality was calculated from the soil structure information following the methodology described in [9]. GIS software GRASS was used to create the GIS layers and ERDAS Imagine was used for image processing.

Water quality data from 55 wells were used for validation of the models. The water quality data was analyzed by the Arkansas Department of Environmental

Quality (ADEQ) for 61 pesticides/herbicides and degradation products. ADEQ provided an Excel spreadsheet concentration of contaminations, well ID with locations of wells collected using a Global Positioning System (GPS). GRASS command s.menu was used to create site files for the wells. The wells were then reclassed into 2 categories: contaminated wells and non-contaminated wells. A well was considered to be contaminated if concentrations of pesticides/herbicides were above the detection limit. Please refer to [10] for details of the well water quality data.

3.2 Development of the neuro-fuzzy model and GIS integration

In this study the neuro-fuzzy software NEFCLASS-J (NEuro Fuzzy CLASSifier) for JAVA platforms was used [19]. Information on this software is available at the website: http://fuzzy.cs.uni-magdeburg.de/nefclass/nefclass.html (viewed 07/02/03). NEFCLASS-J used a supervised learning-like algorithm based on fuzzy error backpropagation. The fuzzy sets and the linguistic rules, which perform this approximation and define the resulting NEFCLASS-J systems, were obtained from a set of examples provided in the training data sets. The example of training data came from the GIS-based raster maps. Since NEFCLASS-J is written in JAVA, the output function was customized in JAVA to integrate the model output with a GIS.

An essential part of fuzzy logic based approaches are fuzzy sets defined by membership functions and rule bases. Shapes of the fuzzy sets are defined by the membership functions. Membership functions allow representation of a linguistic variable to a fuzzy set as a matter of degree. NEFCLASS-J allows users to select types of fuzzy sets and strategy of rule base generation during the learning processes. In this study, trapezoidal fuzzy sets were used along with 10 fold cross validation and ‘best per class’ rule learning strategy (Table 1). For a detailed discussion of fuzzy logic, fuzzy sets, membership functions and rule bases please refer to [23]. Discussion of neural networks basics such as learning algorithms, training, perceptrons, epochs can be found in [15].

Table 1: Parameters for neuro-fuzzy models.

Learning Rate	0.01
No. of fuzzy sets	4 TRAPEZOIDAL for each variable
Rule learning strategy	Best per class
Stop control	Maximum number of epochs=1000 Minimum number if epochs= 100
Validation mode	10-fold cross validation

Use of a neuro-fuzzy model requires similar steps as neural networks. Development of a neuro-fuzzy model is comprised of three steps: training (also known as learning), validation (also known as testing) and application. The training data sets were obtained for the entire watershed using the GRASS command r.stats. This command generated a table showing all possible combinations of the input parameters used for a given model. The training and application data set for the Woodruff County is consist of 408 and 4,093,760 rows, respectively for all models except the model consist of pedality, drainage and hydrologic group. See [9, 11] for details of generation of training and application data with integrated GIS approach.

3.3 Sensitivity analysis

First model was run using D, R, S and I parameters. This model from here on will be referred to as Model 1 (M1). The second model was run using parameters such as pedality, drainage, and hydrologic group. This model will be referred to as Model2 (M2). The Model 3 (M3) was created with parameters D, R, S, I, pedality, drainage, and hydrologic group. The final model was created using all the parameters from M3 and landuse.

3.4 Coincidence analysis

Once the vulnerability maps were generated using the above methods, field data were used to generate coincidence reports to evaluate their performances. Water quality data for 55 wells were used in this study. The Arkansas Department of Environmental Quality (ADEQ) sampled and analyzed the wells for 61 pesticides and degradation products. Seven out of 55 wells were contaminated with at least one type of pesticide. Bentazon was the most common pesticide found in the contaminated wells followed by Metachlor. A set of coincidence reports was generated between vulnerability maps and well contamination data to compare actual contamination with potential contamination (or vulnerability) generated by our models.

Table 2: Characteristics of the training data for the models.

Parameter Name	Mean	SD	Minimum	Maximum	Missing
D	8.07	1.93	3.00	10.00	0
R	3.60	2.79	1.00	10.00	0
S	3.77	2.44	0.00	9.00	0
I	7.09	1.82	3.00	10.00	0
Pedality	40.22	20.89	14.00	100.00	0
Drainage	3.56	1.61	1.00	8.00	0
Hydrologic group	2.33	0.94	1.00	5.00	0
Landuse	2.53	0.75	1.00	4.00	0

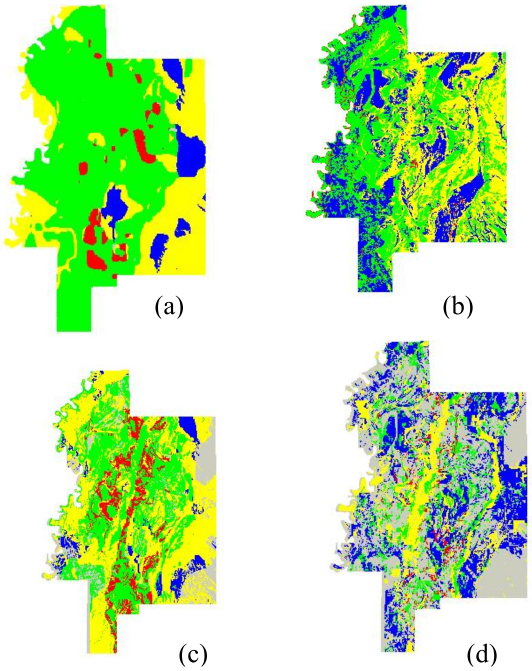
4 Results and discussions

4.1 Training of the models

Characteristics of the training data is presented in Table 2. The parameter ‘pedality’ showed maximum SD. Although the model M1 showed maximum number of misclassifications (87) the model M2 showed higher percentage of misclassification (Table 3). The models M1, M2, M3 and M4 generated 76, 7, 204 and 272 rules, respectively.

Table 3: Performance of the training data for each model.

Models	Patterns	Misclassifications	Error	Estimated number of misclassifications for unseen data		
				Mean	SD	Bars
DRSI (M1)	408	87	172.75	0.3183	0.0672	+/- 0.057871
Pedality, D_r, H (M2)	27	10	14.69	0.7	0.2211	+/-0.190153
DRSI, Pedality, D_r H (M3)	408	76	146.74	0.5046	0.0681	+/-0.058636
DRSI, Pedality, D_r H, LULC (M4)	408	61	134.56	0.5759	0.0678	+/-0.058352


(a) (b)

(c) (d)

Figure 3: Spatial distribution of models.

4.2 Sensitivity of the models

4.2.1 Vulnerability maps

Vulnerability maps generated by various models are presented in Figures 3 a – d and spatial distribution of vulnerability categories are summarized in Table 4. The models M1 and M2 do not show any non-classified category, however, M3 and M4 showed 12 and 46% of the total area as non-classified, respectively. All models showed comparable areas (in percentage) under high vulnerability categories but spatial location of this highly vulnerable category varies from model to model. Compared to other models, M1 showed higher percentage of area under moderately vulnerable category. The model M2 and M4 showed equal percentage of area under low vulnerability categories.

The model M2 was generated using input variables such as pedality, D_r and soil hydrologic group, whereas the model M4 was generated using input variables from the M2 model and D, R, S, I and LULC. The models M3 and M4 showed considerable area as non-classified. We need to fine tune these rules.

Table 4: Spatial distribution of vulnerability categories in percentage.

Vulnerability Categories	**Area Coverage by the Models (%)**			
	M1	M2	M3	M4
Not classified	0	0	12	46
Low	6	24	5	24
Moderate low	28	34	42	16
Moderate	61	40	33	13
High	4	2	8	2
Total	100	100	100	100

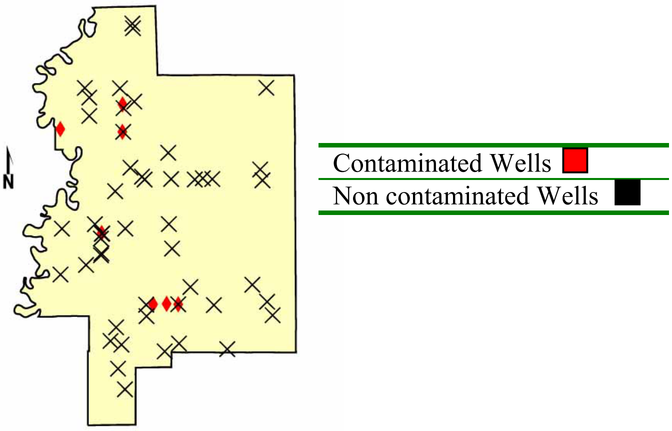


Figure 4: Location of wells in the study area.

4.2.2 Coincidence reports

Coincidence reports between vulnerability maps and well data were generated using GRASS raster files (Figure 4). M3 performed reasonably well in predicting contaminated wells – it did not perform well while classifying non-contaminated wells (Figures 5 and 6). M3 predicted only 2 wells that were not contaminated in low contamination category. The model M1 predicted 1 contaminated well as highly vulnerable and 1 in low vulnerability category. The same model predicted about 5 non-contaminated wells in moderately low vulnerability category. The comparison of point data to a spatial data for accuracy assessment is not the ideal way. A well performing model should be able to classify contaminated wells in high category and non-contaminated wells in low or moderately low vulnerability category. The coincidence analysis was performed to get an idea of how the models are performing in a relative sense. These values should not be used as absolute indicators of the suitability of models. Models 3 and 4 show considerable number of wells that are coinciding with non-classified category.

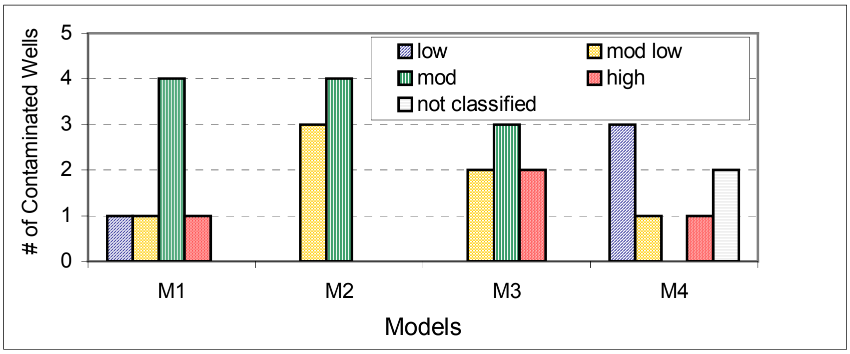


Figure 5: Coincidence between contaminated wells and vulnerability categories.

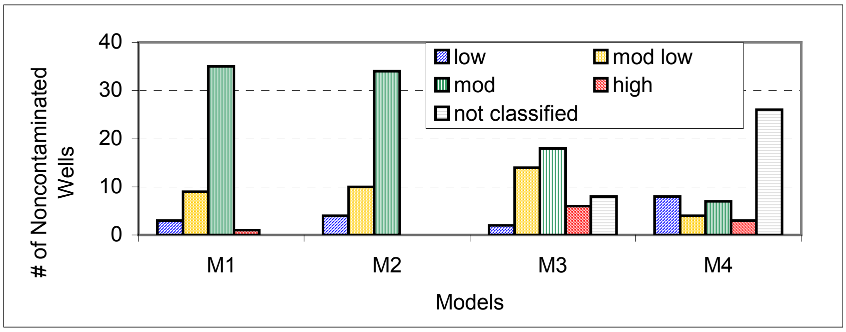


Figure 6: Coincidence between contaminated wells and vulnerability categories.

5 Conclusions

Integration of GIS and neuro-fuzzy model allowed sensitivity analysis of the models and facilitated display of results in the spatial context. Sensitivity analyses were conducted by varying input data layers. Coincidence reports with well water quality data did not yield conclusive results, nor should they be used as absolute indicators. The fine-tuning of the contradictory data is needed. Further application of the same methodologies to other regions with similar as well as different hydrogeologic settings is required to determine the transferability of these techniques. This study is only a beginning of the new era of ground water vulnerability modelling at the regional scale where vulnerability indices are integrated with a GIS as suggested in [8]. In the future, vulnerability maps should be generated from multiple approaches including simple fuzzy rule-based system and geostatistical methods, and then all of these maps should be compared in a GIS to identify ground-water vulnerability zones.

Acknowledgements

Project supported by: CSREES Prime agreement #2001-35102-10830 subcontract # UA.AES2001-118. I would also like to thank Ms. Nivedita Candade for her assistance with the project.

References

[1] Aller, L. T., J.H. Bennett, R. Lehr, J. Petty, and G. Hackett. 1987. DRASTIC: A Standardized System for Evaluating Ground Water Pollution Potential Using Geo-Hydrogeologic Settings. U.S. Environmental Protection Agency Report. EPA600/2-87/036.

[2] Bardossy A. and M. Disse. 1993. Fuzzy Rule-Based Models for Infiltration. Water Resources Research 29(2): 373-382.

[3] Burrough, P.A., R.A. MacMillan, and W. Deursen. 1992. Fuzzy classification methods for determining land suitability from soil profile observations. J. Soil Sci. 43:193-210.

[4] Burrough, P. A. 1996. Opportunities and Limitations of GIS-based Modeling of Solute Transport at the Regional Scale. In *Application of GIS to the non-point source pollutants in the vadose zone*, SSSA, Special Publication # 48, D. Corning and K. Loague eds., Madison, WI.

[5] Burrough, P. A. and R. McDonnell. 1998. Principles of Geographical Information Systems. Oxford University Press: New York. p346.

[6] Cameron, E. and G.F. Peloso. 2001. An Application of Fuzzy Logic to the Assessment of Aquifers' Pollution Potential. Environmental Geology 40:1305-1315.

[7] Connell, L. D. and G.V. D. Daele. 2003. A Quantitative Approach to Aquifer Vulnerability Mapping. Journal of hydrology 276:71 – 88.

[8] Corwin, D. L, K. Loague and T. R. Ellsworth. 1996. Introduction to Non-Point Source Pollution in the Vadose Zone with Advanced Information Technologies. In: Assessment of Non-Point Source Pollution in the Vadose Zone, ed. D. L. Corwin, K. Loague and T. R. Eliisworth. Geophysical Monograph 108. AGS, Washington D.C

[9] Dixon, B. and H.D. Scott. 2001. Application of Fuzzy Logic to Predict Ground Water Vulnerability in Northwest Arkansas. AWRC-USGS Report. MSC # 240.

[10] Dixon, B., H.D. Scott, J.C. Dixon, and K.F. Steele. 2002. Prediction of Aquifer Vulnerability to Pesticides Using Fuzzy Rule-Based Models at the Regional Scale. Physical Geography 23:130 – 152.

[11] Dixon, B., H. D. Scott, J.C. Dixon and J. V. Brahana. 2003. Applicability of Neuro-fuzzy techniques in predicting ground water vulnerability: A sensitivity analysis [in review: Journal of Hydrology].

[12] Dixon, B. and Candade, N. 2004. Comparison of Neural Network and Neuro-fuzzy Techniques in Ground Water Vulnerability Mapping: A Case Study. In proceedings of *AWRA's 2004 Spring Specialty Conference Geographic Information Systems (GIS) and Water Resources III Water*. Nashville, May 2004

[13] De Gruijter, J. J, D. J. J. Walvoort, and P. F. M. Van Gaans. 1997. Continuous soil maps - a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma. 77:169-195.

[14] Dou C., W. Wolt, and I. Bogardi. 1999. Fuzzy Rule-Based Approach to Describe Solute Transport in the Unsaturated Zone. Journal of Hydrology 220(1-2): 74-85

[15] Fausett, L. 1994. Fundamentals of Neural Networks. Prentice Hall: New Jersey. p. 460.

[16] Khan, E. 1999. Neural fuzzy based intelligent systems and applications. In *Fusion of neural networks, fuzzy sets, and genetic algorithms: Industrial applications*, edited by L.C. Jain and N.M. Martin. CRC Press, Washington, D.C. p. 331.

[17] Loague, K. and D. Corwin. 1998. Regional Scale assessment of non-point source ground water contamination. Hydrological Processes. 12:957-965.

[18] Mitra, B., H. D. Scott, J.C. Dixon and J.M. McKimmey. 1998. Application of Fuzzy Logic to the Prediction of Soils Erosion in a Large Watershed. Geoderma 86:183 - 209.

[19] Nauck, U. and R. Kruse. 1999. Design and Implementation of a neuro-fuzzy data analysis tool in JAVA. Manual. Technical University of Braunschweig. Germany.

[20] Smith, P. A., Scott, H.D., and Fugitt, T. (1994). Influence of geographic database scale on prediction of groundwater vulnerability to pesticides. *Journal of Soil Contamination*, Vol. 3, 285-298.

[21] Sui, D.Z. 1992. A fuzzy GIS modeling approach for urban land evaluation. Computers, Environment, and Urban Systems 16 (2): 101-115.

[22] Wang, F., G. Hall and Subaryono. 1990. Fuzzy information representation and processing in conventional GIS software: database design and application, Int. J. GIS. 4:261-283.

[23] Yen, J. and R. Langaria. 1998. Fuzzy Logic: Intelligence, Control and Information. New Jersey: Prentice Hall.