CAN GROUND WATER SAMPLING STRATEGIES BE IMPROVED BY INCORPORATING FUZZY LOGIC IN A GIS?

Barnali Dixon
USF St. Petersburg
210 Davis Hall
Department of Geography
St. Petersburg, FL 33701

1. INTRODUCTION

Contamination of ground water is a major concern for citizens as well as local, state, and federal agencies that are involved with the management, quality and quantity of water and human health. However, the generation of vulnerability maps or delineations of vulnerable zones in a map for monitoring purposes are difficult since ground water contamination depends upon numerous, complex interacting parameters with inherent spatial and temporal uncertainty. Therefore, development of an affordable but reliable technique to generate a ground water vulnerability map at the watershed or county scale will be a useful ground water management tool. Environmental managers could use this tool to prioritize and guide sampling strategies.

An integrated system of advanced information technologies such as 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. (4). 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 (8) and assessing an aquifer's pollution potential using a methodology called DRASTIC (3, 7).

The overall goal of this research is to improve the methodology for the generation of vulnerability maps by using detailed landuse and soil information in conjunction with selected parameters from the DRASTIC model. The specific objective of this study is to incorporate soil structures and landuse parameters with the fuzzy rule-based model in a GIS to generate ground water vulnerability maps.

DRASTIC, an overlay and index method developed for the Environmental Protection Agency (EPA) by the American Water Well Association (1) is a widely used model. This model assesses vulnerability of an area to pollution. The DRASTIC index (DI) was calculated by Aller and uses the most important hydrogeologic factors that affect the potential for groundwater pollution. (1).

$$DI = DwDr + RwRr + AwAr + SwSr + TwTr + IwIr + CwCr \qquad [1]$$

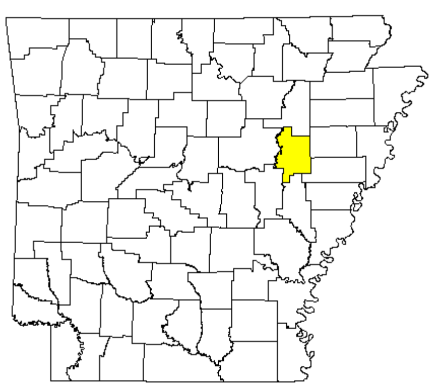
Where, D = Depth to water table; w = weight, r = rating
R = Net recharge; T = Topography
A = Aquifer media; I = Impact of vadose-zone media
S = Soil media; C = Hydraulic conductivity

The DRASTIC model does not provide absolute answers; it only differentiates highly vulnerable areas from less vulnerable areas. The nature of the DRASTIC model is additive. Every parameter in the model has a fixed weight indicating the relative influence of the parameter in transporting contaminants and water to ground water. The parameter rates are variable, which allow the user to calibrate the model to suit a given region. In this paper the term vulnerability will mean contamination potential.

The theory of fuzzy logic based on fuzzy sets was proposed by Zadeh (17), and states that a complex system will be represented better by descriptive variable of linguistic types than by the traditional representation of differential equations (5). (For more detailed discussion on fuzzy logic please refer to 16, 12, 6, 3,7).

This approach is unique because it attempts to improve sampling strategies by incorporating soil structure and landuse information that are critical to the transmission of water and contaminants through the profile. The DRASTIC model does not include soil structure or landuse parameters. The parameter S (soil media) of the DRASTIC model was developed using soil texture information. It has been recognized that textural properties of soil are far less dynamic than soil hydraulic properties. Therefore, relatively dynamic soil properties such as soil moisture or macro pores should be incorporated in a model that attempts to predict ground water contamination potential. Studies have shown the importance of soil structure in determining water and contaminant transport through preferential flow (14, 9, 11). Soil structure, viz. pedality of a soil, indicates a soil's ability to transport water and contaminants through the soil profile as it alters a soil's pore size distribution and pore continuity and connectivity (15). Therefore, in this study the preliminary vulnerability map generated by incorporating selected parameters from DRASTIC was modified using soil structure (pedality) and landuse parameters.

FIGURE 1
LOCATION OF THE STUDY AREA



1.1 LOCATION OF THE STUDY

The study area, Woodruff County, is located in the Mississippi River Valley region of Arkansas. This county was selected to conduct the pilot project for the entire Mississippi Delta region of Arkansas because it represented two of the major land resources areas (MLRAs), MLRA 131 and MLRA 134, of the region (Figure 1). About 80 percent of the study area is drained by the Cache-Lower White-Big River. The upper most aquifer system in the study area is part of a much larger sedimentary system known as the Mississippian embayment alluvial aquifer system.

2. METHODOLOGY

2.1 DIGITAL DATABASE

The primary data layers used in this study are soils, elevation, potentiometric surface, recharge, clay confining units, geology and landuse (Table 1). Detailed descriptions of the generation of data layers for the parameters D, R, A, S, T, I and C can be found in Dixon et. al. (7). The fuzzy rule-based model was developed using selected parameters from the DRASTIC model specifically the parameters D, R, S and I. The parameters A, T and C varied little within the county, therefore, and therefore, they were not included.

The generalized soils map and soils permeability values, estimated from the soil texture information, were obtained from the SOILS5 database to create the S layer. Once the preliminary fuzzy rule-based map was developed, it was modified by using soil structure data. Soil structure information for each soil series found in the study area was obtained from Official Soil Series Description (OSSD), an on-line database of Iowa State (http://www.statlab.iastate.edu/soils/nsdaf/; viewed 1/16/02). The series level soil map was developed by reclassifying soil map units from the county level SSURGO data. The soil series map then was reclassified according to the soil structure information with pedality points outlined by Lin et al. (11) to indicate water and contaminant transporting properties of the soils. All of the reclassification was performed using the r.reclass command in GRASS, a GIS software. In this study, pedality of the top horizon was included in the development of the vulnerability map, with the assumption that water and contaminant transport properties of the A horizon ultimately control the amount of contaminant and water transmission through the profile.

TABLE 1

PRIMARY DATA LAYERS USED IN THE STUDY

Attributes	**Scale/Resolution**	**Source**
Soils	1:24,000	NRCS
Elevation	30 m	USGS
Potentiometric Surface	20 ft contour interval	USGS
Recharge	Point data	USGS
Clay Confining Unit	Point data	USGS
Geology	1:24,000	AGC
Landuse	30 m	USGS
Pesticide Application Rate	Tabular	UAEX
Well Water Quality	Tabular	ADEQ
Well Location	GPS	ADEQ

NRCS = Natural Soil Conservation Services, USGS = US Geological Survey, AGC = Arkansas Geological Commission, UAEX = University of Arkansas, Extension Services, ADEQ = AR Dept. of Environmental Quality.

The landuse data used in this study were obtained from Landsat Thematic Mapper (TM) 1992, using images from two seasons, spring and summer. The image was classified into 4 level I classifications, such as urban, forests, water, and agriculture. Then the images were further classified for agricultural crops. The average pesticide application rates surveyed by University of Arkansas Extension Services (UAEX) for 1992 was used with landuse data layers to create the layer for pesticide loading index (PI). Summation of application rates of all pesticides applied to a given crop was used in PI. In a double cropped field, e.g. soybean and wheat, the PI was adjusted by summation of pesticide rates applied to both crops. While generating the data layer for PI, pesticide application rates were set to 0 for the landuse categories pasture, forests, water, and urban.

2.2 MODELS

In this study, vulnerability was generated using three different methods: DI_{fuzz}, VI_{fuzz} and VI_{fuzz_ped}. First, the preliminary fuzzy rule-based model DI_{fuzz} was developed using four parameters from the DRASTIC model. The Fuzzy Inference Engine (13) was modified and loosely coupled with the GIS software GRASS. A fuzzy model is comprised of fuzzy sets and fuzzy rules. The input parameters for the model DI_{fuzz}, were D, R, S and I. The output parameter DI_{fuzz} indicated the vulnerability or contamination potential. Four fuzzy sets were defined for each of the input parameters except I. Three fuzzy sets were created for the input parameter I. The output parameter, DI_{fuzz}, also had four fuzzy sets. The rules were of the form '*if* D is H *then* DI is L', where D and DI are linguistic variables while H and L are their linguistic values. Examples of fuzzy rules used in the model are presented in Table 2. In the fuzzy rule-based model the concepts of rates and weights used in DRASTIC were reflected through the number of fuzzy sets and rule bases. A slight change in a highly influential parameter had more effect on the output parameter of the rule base than a less influential parameter.

TABLE 2

EXAMPLES OF ACTUAL FUZZY RULES USED IN DI_{FUZZ}

Input fuzzy sets				**Output fuzzy set**
If D = high and	R = low and	S = low and	I = thick then,	DI_{fuzz} = low
If D = low and	R = high and	S = high and	I = thin and then,	DI_{fuzz} = high

Once the preliminary fuzzy-rule based DI_{fuzz} was developed, the fuzzy rule-based model was refined by adopting/modifying methods proposed by Lin et. al (10). This includes calculations of the model parameters for aquifer sensitivity index (SI) and PI. PI was calculated from Landsat TM images by multiplying average application rate of pesticides to the major crops of the region (10). Equations 2 and 3 show methodologies proposed by Lin et. al, (10), whereas equations 4 – 6 show methodologies used in this study.

$$VI = \frac{(SI * PI)}{VI_{max}} * 100 \qquad [2]$$

Where, VI = Relative ground water vulnerability index

SI = Aquifer sensitivity index

PI = Pesticide loading Index

VI_{max} = Maximum value of VI calculated from DRASTIC

$$SI = (DI / DI_{max}) * 100 \qquad [3]$$

Where, DI = DRASTIC Index calculated by methodology proposed by (1)

DI_{max} = The maximum DRASTIC Index

$$VI_{fuzz} = \frac{(SI_{fuzz} * PI)}{VI_{fuzz\_max}} * 100 \qquad [4]$$

Where, SI_{fuzz} = Aquifer sensitivity index

PI = Pesticide loading Index

VI_{fuzz_max} = Maximum value of the fuzzy rule-based vulnerability (255)

$$SI_{fuzz} = (DI_{fuzz} / DI_{Fuzz\_max}) * 100 \qquad [5]$$

Where, DI_{fuzz} = Fuzzy rule-based vulnerability model

$DI\ f_{uzz\_max}$ = Maximum value of fuzzy rule-based vulnerability (255)

$$VI_{fuzz\_ped} = DI_{fuzz} * Pedality_{top} \qquad [6]$$

Where, VI_{fuzz_ped} = vulnerability from the fuzzy rule-based model and soil structural information

DI_{fuzz} = Fuzzy rule-based vulnerability model

$Pedality_{top}$ = Soil structural information (Ped shape + Ped size + Ped grade)

2.3 COINCIDENCE REPORTS

Once the vulnerability maps were generated using the DI_{fuzz} , VI_{fuzz} and VI_{fuzz_ped} methodologies, field data were used to generate coincidence reports to evaluate the performance of these methodologies. 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. The ADEQ provided an Excel workbook with GPS locations as well as water quality data for the 55 wells. Seven out of 55 wells were contaminated with at least one type of pesticide. Out of the 55 wells, 11 wells were sampled more than once. Therefore in our final analysis, only the 44 wells with single visits were used. The command r.coin was used in GRASS to generate the coincidence reports.

3. RESULTS AND DISCUSSION

In this study, an attempt was made to refine the preliminary vulnerability map developed with parameters from the DRASTIC model using fuzzy logic in a GIS by incorporating soils and landuse parameters. Three models: DI_{fuzz}, VI_{fuzz} and VI_{fuzz_ped} were developed and resultant vulnerability maps were compared with the field data.

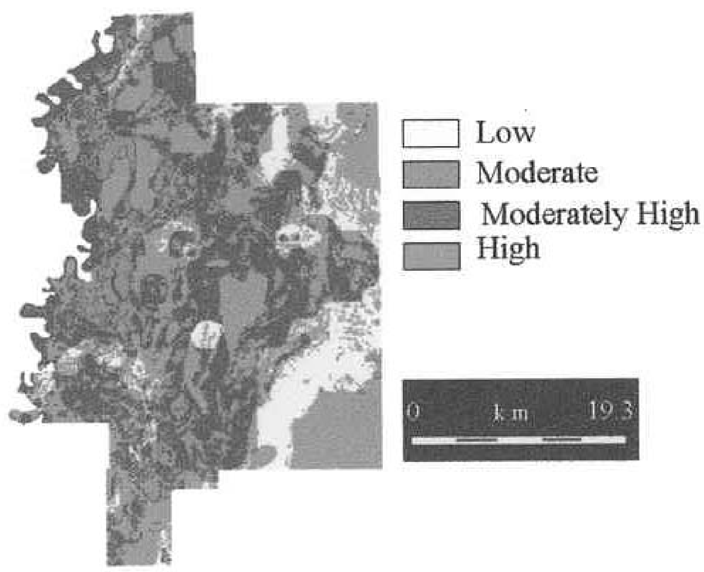
Figure 2 shows the spatial distribution of DI_{fuzz}. The western part of the county has more areas classified as highly vulnerable than the eastern part of the county. Table 3 summarizes the spatial distribution of vulnerability categories generated by the preliminary fuzzy rule-based model DI_{fuzz}.

The vulnerability map from the model VI_{fuzz} was created from PI and SI_{fuzz}. The spatial distribution of VI_{fuzz} is shown in Figure 3 and summarized in Table 4. The parameter PI was used in generation of vulnerability map from the model VI_{fuzz}. Spatial distribution of the PI is summarized in Table 5. The parameter SI_{fuzz} was created from the DI_{fuzz} model. Spatial distribution of SI_{fuzz} is summarized in Table 6. Average application rate for the pesticides surveyed by the UAEX is presented in Table 7. Incorporation of landuse data did not improve the model prediction markedly. Poor correlation was noted between landuse data and well contamination data. This could be attributed to the point source nature of the contamination such as presence of the mixing sites or spills.

FIGURE 2

DI_{FUZZ}:

SPATIAL DISTIBUTION OF VULNERABILITY CATEGORIES



Finally, a vulnerability map VI_{fuzz_ped} was created by incorporating pedality data for the top soil horizon with the preliminary fuzzy rule-based model DI_{fuzz}. The spatial distribution of the VI_{fuzz_ped} is shown in Figure 4 and summarized in Table 8. The majority of highly vulnerable areas occurs in the western part of the county. Comparison among the three vulnerability maps showed that DI_{fuzz}, and VI_{fuzz_ped} resulted in similar spatial trends, whereas VI_{fuzz} showed considerable spatial variation. The model VI_{fuzz} was generated using pesticide loading information, and did not improve the prediction, mainly due to the presence of '0 values' in the PI layer.

3.1 COINCIDENCE ANALYSIS

Locations of contaminated wells are shown in Figure 5. Table 9 shows the pesticides

FIGURE 3
VI_{FUZZ}:
SPATIAL DISTIBUTION OF VULNERABILITY CATEGORIES

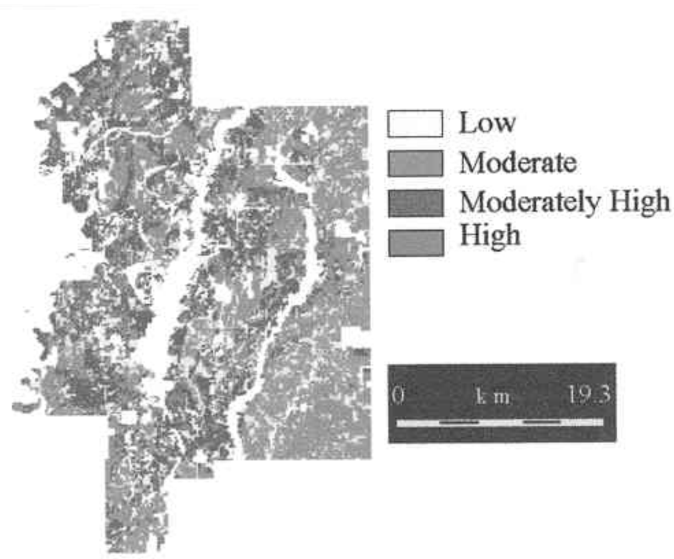
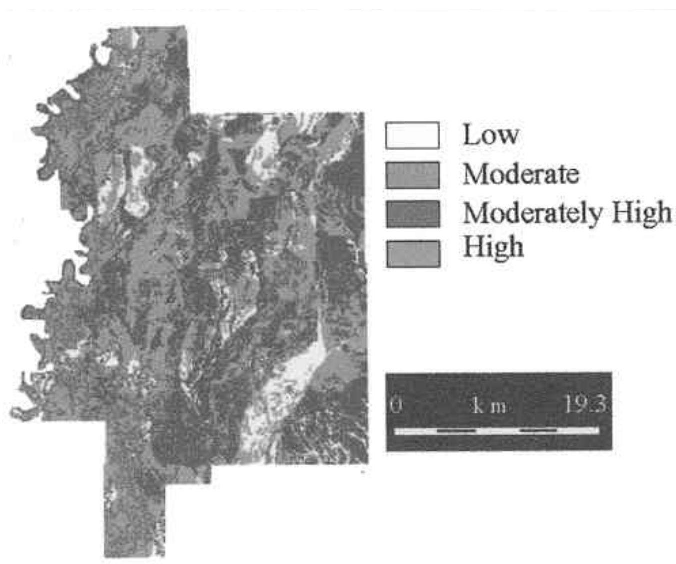


FIGURE 4
VI_{FUZZ_PED}: SPATIAL DISTIBUTION
OF VULNERABILITY CATEGORIES



detected in contaminated wells. Bentazon was the most common pesticide found in the contaminated wells. Four out of seven contaminated wells had Bentazon, a highly soluble pesticide. One well out of seven contaminated wells had more than one pesticide. This coincidence analysis provided the answer to our question “Can ground water sampling strategy be improved?” A good model should not only be able to predict contaminated wells with elevated vulnerability, but also shows most of the non-contaminated wells classified in the low vulnerability category.

The model DI_{fuzz} predicted three out of seven contaminated wells as highly vulnerable. The same model predicted two of the contaminated wells in moderately high vulnerability category and one in moderately vulnerable category. This prediction is better than the direct DRASTIC model discussed in (7). When the DI_{fuzz} model was compared with the wells that are not contaminated, the result was not promising. Thirteen of the non-contaminated wells were classified in the highly vulnerable category, whereas about 14 non-contaminated wells were classified as moderately vulnerable. Only two wells that were not contaminated coincided with the low vulnerability category.

TABLE 3
DI_{FUZZ} : DISTIBUTION OF VULNEARBILITY CATEGORIES

DI_{fuzz}	Hectares	Percent Cover
Low	20,467	13.3
Moderate	67,866	44.2
Moderately-high	45,523	29.6
High	19,914	12.9
Total	153,770	100

TABLE 4
VI_{fuzz}: DISTRIBUTION OF VULNERABILITY CATEGORIES

VI_{fuzz}	Hectares	Percent Cover
No pesticides	35,531	23.1
Low	31,895	20.7
Moderate	43,493	28.3
Moderately high	35,611	23.2
High	7,237	4.7
Total	153,770	100

TABLE 5
DISTIBUTION OF LANDUSE USED FOR PI

Landuse	Hectares	Percent Cover
Urban	2,101	1.4
Water	1,797	1.1
Forests	31,505	20.5
Soybeans	42,681	27.8
Rice	25,969	16.9
Cotton	4,517	2.9
Sorghum/Corn	19,094	12.4
Pasture	1,322	0.9
Bare Soil	24,784	16.1
Total	153,770	100

TABLE 6
SI_{FUZZ}: DISTRIBUTION OF AQUAIFER SENSITIVITY

SI_{fuzz}	Hectares	Percent Cover
31	17,170	11.0
38	2,484	1.6
41	814	0.5
50	67,864	44.2
58	2	0
61	2885	1.9
65	42,637	27.8
85	17,475	11.4
100	2,439	1.6
Total	153,770	100

TABLE 7
AVERAGE APPLICATION RATE OF PESTICIDE APPLICATION FOR MAJOR CROPS IN 1992

Crops	Application (lbs/acre)
Soybean	8.825
Rice	6.313
Cotton	8.7
Wheat	1.625
Grain Sorghum	10.25
Corn	10.1

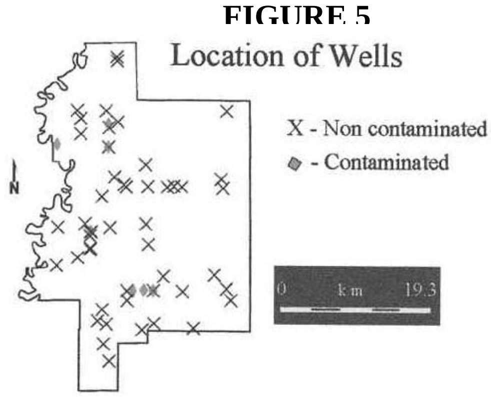
TABLE 8
VI_{FUZZ_PED}:
SPATIAL DISTRIBUTION OF VULNERABILITY CATEGORIES

VI_{fuzz_ped}	Hectares	Percent Cover
Low	16,053	10.3
Moderate	68,265	44.5
Moderately - high	42,584	27.7
High	26,868	17.5
Total	153,770	100

TABLE 9
NAME AND OCCURRENCE OF PESTICIDES

Pesticide	Occurrence
Bentazon	4
Metalochlor	2
Acifluorfen	1
Fluometron	1
Metribuzin	1

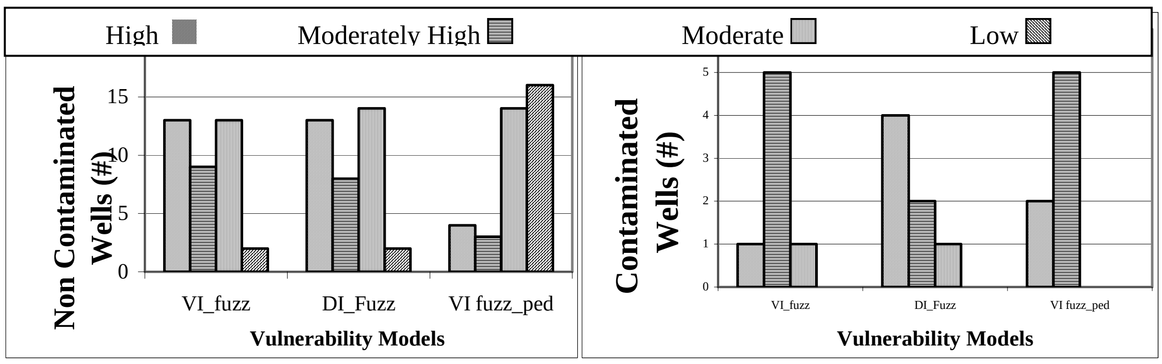
FIGURE 5



From this study it also becomes obvious that vulnerability predicted by the VI $_{fuzz}$ model was not adequate. Five out of seven contaminated wells were classified in the moderately high vulnerability category (Figure 6). When the same model was compared to the non-contaminated wells (Figure 6), 13 of the non-contaminated wells coincided with highly vulnerable category. If a watershed manager were to sample based on the map generated by this model, the result may not have been very useful. The model DI $_{fuzz}$ showed higher coincidence with contaminated wells, however, both the models (DI $_{fuzz}$ and VI $_{fuzz}$) showed similar performance when compared with non-contaminated wells.

FIGURE 6

MUTUAL OCCURRENCE OF WELLS AND VULNERABILITY CATEGORIES IN DIFFERENT MODELS



The vulnerability map (VI $_{fuzz\_ped}$) generated by incorporating pedality information with the DI $_{fuzz}$ model presented in the Equation 6, generated a relatively useful map. Coincidence reports between contaminated wells and the vulnerability map indicated that five out of seven wells coincided with moderately high vulnerable categories and two coincided with highly vulnerable area. None of the contaminated wells coincided with moderately vulnerable areas. When the non-contaminated wells were coincided with the vulnerability map generated by the model VI $_{fuzz\_ped}$, 16 wells coincided with the low vulnerability category, and 14 wells with the moderate category. Only 3 and 4 wells that were not contaminated coincided with the moderately high and high vulnerability categories, respectively (Figure 6).

4. CONCLUDING REMARKS

The vulnerability maps were generated in three different ways. The vulnerability map from the model VI $_{fuzz\_ped}$ showed better results when compared with the field water quality data. This map was generated using a fuzzy rule-based model and soil structure information. Environmental managers in Woodruff County may use the model VI $_{fuzz\_ped}$ to prioritize sampling of the wells. The preliminary fuzzy rule-based model DI $_{fuzz}$ generated by incorporating four parameters from the DRASTIC mode, was not adequate. The model VI $_{fuzz}$ was generated using PI information, which did not improve the prediction. Incorporation of landuse and pesticide use information needs to be explored further. In this study, pedality of the top horizon was used as it ultimately controls the amount of contaminants and water available to be transmitted though the profile. In the future, the model VI $_{fuzz\_ped}$ will be used for other counties in the region to assess its performance.

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

4. 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.

5. Cox, E. 1994. The Fuzzy Systems Handbook. London: Academic Press.

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

7. 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 (In press).

8. 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

9. Lin, H.S. and K.J. McInnes. 1995. Water Flow in Clay Soil Beneath a Tension Infiltrometer. Soil Science 159:375-382.

10. Lin, H.S., H.D. Scott, and J.M. McKimmey. 1996. Identification of Optimal Locations for Sampling Ground Water for Pesticides in the Mississippi Delta Region of Eastern Arkansas. University of Arkansas: Arkansas Water Resources Center. MSC-185.

11. Lin, H. S., K. J. McInnes, L.P. Wilding, and C. T. Hallmark. 1999. Effects of Soil Morphology on Hydraulic Properties: I. Quantification of Soil Morphology. Soil Science Society of America Journal 63:948- 953.

12. 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.

13. Numata, K. 1991. General Purpose Fuzzy Reasoning Library. Version 3.0. Tokyo University, Japan..

14. Quisenberry, V. L., B. R. Smith, R. E. Phillips, H.D. Scott and S. Nortcliff. 1993. A Soils Classification System for Describing Water and Chemical Transport. Soil Science 156: 306-315.

15. Scott, H.D. 2000. Soil Physics: Agricultural and Environmental Applications. Ames, Iowa: Iowa State University Press.

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

17. Zadeh, A. L. 1965. Fuzzy Sets. Information and Control 8:338-353.

Acknowledgement: Funding provided by USDA CSREES Subcontract. The prime USDA contract # 2001 – 35102 10830