Vulnerability of Shallow Aquifers of the Conterminous United States to Nitrate: Assessment of Methodologies

Karthik Kumarasamy
Utah State University

VULNERABILITY OF SHALLOW AQUIFERS OF THE CONTERMINOUS UNITED

STATES TO NITRATE: ASSESSMENT OF METHODOLOGIES

by

Karthik Kumarasamy

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Civil and Environmental Engineering

Approved:

Dr. Jagath Kaluarachchi Major Professor	Dr. Gilberto Urroz Committee Member
Dr. Darwin L. Sorensen Committee Member	Dr. Byron R. Burnham Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2007

ABSTRACT

Vulnerability of Shallow Aquifers of the Conterminous United States to Nitrate:

Assessment of Methodologies

by

Karthik Kumarasamy, Master of Science

Utah State University, 2007

Major Professor: Dr. Jagath J. Kaluarachchi
Department: Civil and Environmental Engineering

Groundwater is an important natural resource for numerous human activities, accounting for more than 50% of the total water used in the United States. Groundwater is vulnerable to contamination by several organic and inorganic pollutants such as nitrate, heavy metals, and pesticides. Assessment of groundwater vulnerability aids in the management and protection of limited groundwater resources.

The focus of this thesis is to (1) statistically compare two groundwater vulnerability assessment models; modified DRASTIC (Acronym for Depth to water, net Recharge, Aquifer media, Soil media, Topography, Impact of vadose zone, and hydraulic Conductivity of aquifer) and ordinal logistic regression for NO_3^- contamination of shallow groundwater of the US, (2) analyze any discrepancies in the predictability of each of these models, and (3) discuss the advantage of each of the above-mentioned models with respect to performance, data requirement, and its ability to predict vulnerability. Analysis of NO_3^- concentration in groundwater allows for a reliable comparison of the two models.

The results from the OLR model indicate a better correlation between the observed and average predicted probabilities. A very low R^2 value was obtained between the modified DRASTIC and nitrate concentration, indicating poor prediction capabilities and need for high resolution data. Limitation with respect to requirement of more data with respect to prediction is seen in both the methods.

(88 pages)

ACKNOWLEDGMENTS

I would like to thank my major professor, Dr. Jagath Kaluarachchi, for providing me the guidance, excitement, and much needed support and inspiration during my studies. I would also like to thank my committee members for their useful comments and patience with me. I would like to thank the faculty of the Department of Civil and Environmental Engineering for the classes they offered, as the courses were pivotal in strengthening my foundation in the field. My sincere thanks go to Kerie J. Hitt of the USGS for specifically preparing several datasets at my request.

Further thanks go to the staff of the Utah Water Research Laboratory and the office of the Department of Civil and Environmental Engineering for their help and guidance at different stages of my thesis. Finally, my thanks go to my family for their love, support, and inspiration, without which I would not be here.

This work has been supported by the Utah Water Research Laboratory.

Karthik Kumarasamy

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Groundwater is an important source of water for diverse human activities such as agriculture, industry, drinking and various other municipal uses. Protection of this resource has become a major endeavor since the late 70's with the public attention drawn to incidents of contamination (U.S. Environmental Protection Agency, 1990). Industrial wastes contributed to numerous pollution problems. Environmental impacts caused by many of the chemicals that were produced were not known until much later. Some of these chemicals have penetrated into the subsurface causing contamination of groundwater (Bedient, Rifai, and Newell, 1999).

Protection of groundwater resources is always cheaper than remediation and restoration of the aquifer, and in most cases it is very difficult to remediate an aquifer to its original state. One of the tools supporting decision-making in aquifer protection is the evaluation of shallow aquifer vulnerability. The concept is based on the assumption that all areas are not equally vulnerable; thereby aiding in the implementation of appropriate land management practices at local and regional scale. The maps produced by aquifer vulnerability assessment models will aid in efficient groundwater management strategies.

Natural attenuation capacity varies widely at different locations. Instead of imposing restrictions everywhere it is economically viable to apply restrictions to certain areas. This is the general principal underlying the concept of aquifer vulnerability and its mapping (Foster, 1987). There is a clear distinction between aquifer vulnerability and pollution risk. Pollution risk depends both on the aquifer vulnerability and the existence of significant pollutant loading entering the subsurface to produce high enough concentrations to affect public health. This implies that an aquifer can be highly vulnerable; but have no pollution risk, if there is no significant contaminant loading.

Methods to estimate vulnerability of an aquifer can be broadly classified into three categories: overlay and index methods, process-based simulation methods, and statistical methods (NRC, 1993). Overlay and Index methods involve combining various physiographic factors to obtain a final vulnerability score. These methods are popular because of the minimum data requirement. These methods demand expert judgment in their usage rather than the controlling physical processes. Process-based methods are the most accurate, but demand substantial data and are computationally costly. They require robust computer systems for their assessment. The ability of the statistical methods to accommodate the uncertainty of data is better than other methods. Statistical methods are also more flexible compared to the other two categories (Twarakavi and Kaluarachchi, 2005).

The use of aquifer vulnerability techniques assists in the decision-making processes. It is to be noted that the use of these methods is not intended to replace on-site investigations or to substitute any type of practice. These procedures do not reflect the suitability of a site for a particular land use activity. The advantage of these techniques is their ability as a screening tool, or their use in combination with other assessment techniques. The most appropriate use is to provide assistance in resource allocation and prioritization of the many types of groundwater related activities.

Groundwater is vulnerable to many chemicals including nitrate. The primary sources of nitrate are inorganic fertilizer and animal manure. The chemical formula of nitrate anion is NO_3^-. NO_3^- is soluble in water and can easily leach through the soil. NO_3^- can persist in shallow groundwater for decades (Nolan, Hitt, and Ruddy, 2002).

Ingestion of NO_3^- through drinking water by infants and some susceptible can cause low oxygen levels in blood, a condition called *methaemoglobinaemia*. This condition mainly affects babies less than six months old or while in the womb. Effective delivery of oxygen to different parts of the body does not occur at exposure to higher levels of NO_3^-. The result being, infants may have blueness around the mouth, hands and feet (hence the name *blue baby syndrome*). This condition is potentially fatal (Spalding and Exner, 1993). This condition led the US Environmental Protection Agency (US EPA) to establish a maximum contaminant level (MCL) of 10 milligrams per liter (mg/L) nitrate as nitrogen (NO_3^- - N) (U.S. Environmental Protection Agency, 1995).

Other evidence for adverse health effects associated with NO_3^- include a case study in Indiana where 19-29 mg/L of nitrate in a rural, domestic well was believed to be the cause of eight spontaneous abortions among four women during 1991-1994 (Centers for Disease Control and Prevention, 1996). In Nebraska, nitrate concentration of 4 mg/L or more in water from community wells have been associated with increased risk of non-Hodgkin's lymphoma (Ward et al., 1996). The concentration of NO_3^- - N in natural groundwaters is commonly 2 mg/L or less (Mueller and Helsel, 1996).

The NO_3^- ion is the highly oxidized form of N, with the oxidation state of +5. NO_3^- concentrations are usually reported in units of milligrams per liter (mg/l) with the mass representing either the nitrate-N or the total mass of nitrate ion in water (nitrate-NO_3^-). The molecular weight of nitrate is 62; the molecular weight of N is 14, so the ratio of a concentration measured as nitrate-NO_3 to an equivalent concentration measured as nitrate-N is 4.43. The MCL 10 mg/l of nitrate-N is equivalent to 44.3-mg/l of nitrate-NO_3^-.

Motivation

Aquifer vulnerability determination is an important management tool to protect groundwater resources. There are a number of evaluation procedures to assess the vulnerability of groundwater resources. Each of these methods has its benefits and limitations. DRASTIC (acronym for Depth to water table, net Recharge, Aquifer media, Soil media, Topography (Slope), Impact of vadose zone, hydraulic Conductivity) is a groundwater vulnerability assessment technique with widespread use in the US and around the world. The method is simple to use, but the computation of the final DRASTIC score is very subjective. Availability of data is not always in the form as described in the procedure. Limitations and benefits with respect to data and knowledge have profound effect on the final DRASTIC score. A more recent approach to evaluate vulnerability is ordinal logistic regression. The weights in the form of coefficients are statistically determined and are more universal in the computation of vulnerability. This method requires advanced statistical understanding to obtain the results and thereby lacks the simplicity of DRASTIC.

This proposed study will compare the relative performance of the two methods, modified DRASTIC and ordinal logistic regression, at national-scale by using NO_3^- - N concentration as a performance indicator. Different research groups conducted several studies to compare various methods such as, DRASTIC, EPIK, German method, GOD, and ISIS (Gogu, Hallet, and Dassargues, 2003) and these studies have compared a range of methods but no work has yet been done involving modified DRASTIC and ordinal logistic regression method. It is also observed that national-scale comparison of these two models not has been done so far. The impetus for this study comes from the idea that there are several different models available to assess vulnerability and no common methodology to understand or compare the result of each of these procedures. Literature also suggests that there is considerable interest in developing the criteria and procedures to evaluate and map groundwater vulnerability and this study aims to contribute to that ultimate goal.

Approach

This study is divided into four sections with respect to the objective, namely, (1) Assess the distribution of NO_3^- across the conterminous US (CONUS), (2) Develop a ground water vulnerability map using the modified DRASTIC, (3) Develop a ground water vulnerability map using ordinal logistic regression, (4) Statistically compare both the models. To achieve this objective, the research was divided into five different chapters, namely, literature review, NO_3^- analysis, model development of modified DRASTIC and ordinal logistic regression, comparison, and summary.

CHAPTER II

LITERATURE REVIEW

Groundwater Vulnerability

Groundwater is a major source of water supply, both for domestic and industrial uses and normally requires minimal treatment. In view of the extensive reliance on groundwater resources as an economical and safe source of drinking water, aquifer protection to minimize the deterioration of water quality should receive significant attention. Remediation of polluted aquifer resources is always expensive and protracted, and is often abandoned, leading to loss of valuable resources at a considerable economic cost. These are the motivating factors for protecting zones, which are more vulnerable with respect to others.

The term vulnerability in hydrogeology was first used in the late 1960's by the French hydrogeologist J. Margat. It has been used more widely since the 1980's (Haertle, 1983; Aller et al., 1987; Foster and Hirata, 1988). Presently, the term is commonly used all over the world. A common definition of groundwater vulnerability is still not agreed upon, and various definitions of vulnerability have been proposed with similar meanings. An often-used definition from NRC (1993) is as follows: 'Groundwater vulnerability is the tendency of or likelihood for, contaminants to reach a specific position in the groundwater system after introduction at some location above the uppermost aquifer.' The US EPA definition is, 'Probability that a specific contaminant (usually surface-derived) will be detected at or above a specified concentration in the subsurface at a specific location.'

Aquifer vulnerability can be subdivided into two semi-independent components (1) the penetration capability of pollutants in a hydraulic sense and (2) the attenuation capacity. The unsaturated zone plays an important role as the first line of natural defense against groundwater pollution. The conditions present in this zone are of considerable importance for the fate and transport of the contaminant. Though the fate processes occur in the saturated zone, the rates at which these occur is relatively low. It is, therefore, of great significance to understand the role of the unsaturated zone and fully consider it in the computation of vulnerability. In cases of more persistent contaminants the unsaturated zone merely introduces a large time lag before the contaminant can arrive at the water table, without any or insignificant attenuation. The pollutant penetration rates in case of fissured formations increases by orders of magnitude compared to most other formations. This condition leads to greater chances of the pollutant reaching groundwater (Foster, 1987).

Active pollutant elimination and attenuation occurs at much higher rates in the soil zone. Higher clay mineral and organic content and a very large bacterial population contribute to these increased rates. In this perspective, it is of importance to judiciously include this parameter in the computation of vulnerability.   Scientifically, it is more appropriate to evaluate vulnerability to specific contaminants rather than a generic contaminant. However, due to insufficient and inadequate resources or data, this ideal condition cannot usually be achieved. Hence, vulnerability mapping is less refined, more generalized and is used at a reconnaissance level (Haertle, 1983; Aller et al., 1987).

Overview of Vulnerability Assessment Models

Vulnerability assessment models consider a range of parameters to evaluate vulnerability. Some of the methods of wide usage are as follows; (1) DRASTIC, (2) AVI, which is the acronym for Aquifer Vulnerability Index, (3) GOD, which is the acronym for groundwater occurrence, Overall aquifer class in terms of degree of consolidation and lithological character, and depth to groundwater table, (4) SEEPAGE, System for Early Evaluation of Pollution potential of Agricultural, and (5) logistic regression. This section describes, in detail, two of the most commonly used methods, namely, modified DRASTIC and ordinal logistic regression.

Modified DRASTIC model

DRASTIC is an empirical model developed by the National Water Well Association in conjunction with the US EPA to determine aquifer vulnerability at a regional-scale. Although DRASTIC is physically based, the final DRASTIC index is a numerical index. This method was created to evaluate aquifer vulnerability of any area. This model can only be used for areas of more than 100 acres. Due to the wide variability of pollutants a generic pollutant was selected. It is assumed that the pollutant has the mobility of water. This model does not readily assess the condition of leaky aquifers or confined aquifers.

This system is neither designed nor intended to replace on-site investigations or any particular methodology or practice. The vulnerability index given by DRASTIC does not reflect a site's suitability for any particular land use activity. This procedure is a means of determining the relative vulnerability of groundwater for a particular area with respect to the other. The most appropriate charge of this methodology is to provide assistance in the decision-making process and to be used in combination with other evaluation tools.

The system encompasses two portions, namely, the hydrogeologic settings and the relative ranking of the hydrogeologic parameters. In the hydrogeologic settings are physical characteristics, which affect the pollution potential of groundwater. The parameters that are considered in the DRASTIC model are ***d****epth to water table,* ***r****echarge,* ***a****quifer media,* ***s****oil media,* ***t****opography (slope),* ***i****mpact of vadose zone media, and (aquifer hydraulic)* ***c****onductivity* (DRASTIC).

This study involves the comparison of modified DRASTIC that integrates on-ground N loading along with other DRASTIC parameters. The numerical ranking system consists of three significant parts: weight, range, and rating. Each parameter in the modified DRASTIC procedure has been assigned a weight based on its relative importance with respect to other parameters. The weights range from 5 to 1, with the most significant parameter having the weight of 5 and the least having a weight of 1, as shown in Table 1. According to the authors (Aller et al., 1987), these weights are constant and cannot be changed.

Each of the factors in DRASTIC have been divided into either ranges or into significant media types. The ratings of this system vary from 1 to 10. All the factors in the modified DRASTIC evaluation method have one rating per range except for Aquifer media and Impact of the vadose zone. These two factors have been each assigned a typical rating and a variable rating. The variable rating gives more flexibility to the user in case of specific knowledge. The minimum value that the empirical DRASTIC index can take is 23 and the maximum value is 226. The literature suggests that such extreme values are very rare, the most common values being within the range 50 to 200. The equation determining the vulnerability index is given in equation (1).

$$\text{Vulnerability Index} = D_rD_w + R_rR_w + A_rA_w + S_rS_w + T_rT_w + I_rI_w + C_rC_w + N_rN_w \quad (1)$$

where, D, R, A, S, T, I, C, and N are the parameters, subscript r is the rating value and subscript w is the weight associated to each parameter. Hence, the final equation after introducing the weights is as shown in equation (2).

$$\text{DRASTIC Index} = 5D + 4R + 3A + 2S + 1T + 5I + 3C + 5N \quad (2)$$

The areas with higher index value have greater susceptibility with respect to lower index value areas. DRASTIC was developed using four assumptions:

1. The introduction of the contaminant is at the ground surface.
2. The flushing of the contaminant into groundwater is through precipitation.
3. The mobility of the contaminant is similar to that of water.
4. DRASTIC can only be used for areas 100 acres or larger.

Table 1. Assigned weights for DRASTIC features

Feature	**Weight**
Depth to the water table	5
Net Recharge	4
Aquifer material	3
Soil type	2
Topography	1
Impact of the vadose	5
Hydraulic Conductivity	3
Nitrogen Loading	5

<u>Ordinal logistic regression</u>

The technique of binary logistic regression, commonly known as logistic regression (LR), is a statistical method used to estimate the aquifer vulnerability. LR models were used by Nolan, Hitt, and Ruddy (2002) to estimate aquifer vulnerability to NO_3^- contamination in the US. The ordinal logistic regression method considers more than one threshold value to obtain aquifer vulnerability; this is considered to be an improvement over LR methods (McCullagh, 1980). The background concentration and the MCL can both be used to assess the probabilities of occurrence of NO_3^- in groundwater (Twarakavi and Kaluarachchi, 2005). Epidemiological studies have seen extensive application of binary LR and more recently its applications extend to environmental research (Hosmer and Lemeshow, 1989).

The probability of response to be less than a threshold value is related to a set of influencing variables in LR, whereas, in classical linear regression the influencing variables are related to the response variable (Afifi and Clark, 1984; Hosmer and Lemeshow, 1989; Helsel and Hirsch, 1992; Kleinbaum, 1994). For instance, the probability of NO_3^- being less than the MCL for N loading, soil classes, slope, etc is considered in LR. The odds ratio, *O*, is given as in equation (3)

$$O = \frac{p}{1-p} \tag{3}$$

where, p is the probability of the response to be less than a given threshold value.

The natural logarithm of the *odds ratio* for the probability of the response to be less than the threshold value, and influencing variables is a linear regression in the LR model. The natural logarithm of the odds ratio, or *logit*, is linearly related to the influencing variables in binary LR and is written as shown in Equation (4)

$$\log(O) = \text{logit}(p) = a + \mathbf{bx} \qquad (4)$$

where, a is a constant, $\boldsymbol{b}$ is a vector of slope coefficients, and $\boldsymbol{x}$ is the vector of influencing variables.

The proportionality-odds model, commonly known as ordinal LR expands this concept to more than one threshold value (McCullagh, 1980). For example, if two thresholds (i =1, 2) are considered to categorize the response variable, ordinal LR relates the corresponding *logits* as follows (equation 5);

$$\log(O_i) = \text{logit}(p_i) = a_i + \mathbf{bx} \qquad i = 1, 2 \qquad (5)$$

where, O_i is the odds ratio for probability of response to be less than the i^{th} threshold, p_i is the probability of response to be less than the i^{th} threshold, and a_i is the constant for the i^{th} threshold, $\boldsymbol{b}$ is a vector of slope coefficients, and $\boldsymbol{x}$ is the vector of influencing variables.

The same slope coefficient, “b” is assumed to relate the probabilities of occurrence of response to the influencing variables, with respect to all the thresholds. Using binary LR for more than one threshold would result in the use of different slope coefficients, thereby resulting in a loss of physical significance (McCullagh, 1980; Helsel and Hirsch, 1992). The benefits of application of Ordinal LR models over LR models for ordinal nature of data are explained in the literature (McCullagh, 1980). Ordinal LR model is fit to the observed responses using the maximum likelihood approach. Unknown parameters are determined using maximum likelihood approach that best match the predicted and observed probability values. The application of maximum likelihood theory to ordinal LR models is explained in detail by Hosmer and Lemeshow (1989) and McCullagh (1980).

In this study, the ordinal LR is used to relate the probability of NO_3^- concentration with respect to background concentration and MCL, to the significant influencing variables like N loading, slope, soil hydrologic class, etc. The approach for implementing the ordinal LR model for a successful analysis of aquifer vulnerability to NO_3^- contamination includes a number of key steps as outlined in Twarakavi and Kaluarachchi (2005), which are: (a) categorizing of response values (concentration) based on *n* threshold values such as, MCL and background concentration into *(n+1)* discrete response categories, (b) identifying all possible influencing variables, discrete and continuous, of the physical system, (c) performing univariate ordinal LR between the response and each influencing variable and selecting the significant influencing variables using the Wald statistic and chi-square test, (d) performing multivariate ordinal LR between the probability of occurrence of response with respect to the threshold values, and the significant influencing variables and checking again for the significance of the influencing variables, (e) then repeating step (d) until only the significant influencing variables are included in the model, and (f) finally checking for the goodness-of-fit of the model results. In this study, an ordinal logistic regression model will be used to relate the probability of NO_3^- concentration to occur with respect to a concentration of 2 and 10 mg/l of NO_3^- - N, to the significant influencing variables.

Nitrate in Groundwater

This section presents a brief review of NO_3^- in groundwater, relevant to the present study, rather than a comprehensive review of the extensive literature available on NO_3^- in groundwater. NO_3^- is the most widespread contaminant among all inorganic constituents of health significance. The typical concentration of NO_3^- - N in natural groundwater is 2 mg/L or less (Mueller and Helsel, 1996). Considering many factors such as, occurrence of the contaminant in the environment, human exposure and associated health risks, economy, and impacts of regulation on water systems, etc, the US EPA established a drinking-water standard of 10 mg/L NO_3^- - N (U.S. Environmental Protection Agency, 1995).

NO_3^- toxicity and health effects are well documented in the literature. Methemoglobinemia results from NO_3^-, which is converted to nitrite ion in the oral cavity and the stomach. This is absorbed from the gastrointestinal tract into the blood (Shuval and Gruner, 1972). The ferrous iron (Fe^{+2}) present in the heme group is oxidized to ferric iron (Fe^{3+}), which bonds to NO_3^-, preventing the transport of oxygen by the blood (Jaffe, 1981). Infants are highly susceptible, and in certain cases it is fatal (Super et al., 1981; Keeney, 1986; Duijvenbooden, Van, and Matthijsen, 1987). It is also suspected that NO_3^- is a carcinogen (Van Duijvenbooden, and Matthijsen, 1987). Based on the correlation between stomach cancer mortality rates and previously published data on daily NO_3^- intake, it is suggested that there could be an association between nitrate intake and stomach cancer (Fine, 1982). Because of diseases like methemoglobinemia, cancer, and possibly other illnesses linked to NO_3^-, its concentration in public water supplies is monitored and regulated by federal law (Cast, 1985; Keeney, 1986).

Natural occurrence of NO_3^- can be predominantly classified into three categories, namely, geologic N, forests, and forage and pastoral agriculture (Keeney, 1989). Substantial quantities of NO_3^- were found in never fertilized rangeland of semiarid and western central Nebraska. This was attributed to leaching of NO_3^- from Pleistocene age deposits with the development of irrigation (Boyce et al., 1976). In the alluvium beneath the San Joaquin valley, California high levels of NO_3^- exist, and as in Nebraska has leached into the groundwater with the advent of irrigation (Strathouse et al., 1980). The NO_3^- in groundwater of Runnels County, Texas, is associated with natural soil (Kreitler and Jones, 1975). Forests also contribute large quantities of nitrogen usually in the form of NO_3^- to groundwater (Keeney, 1980). N losses and contamination of groundwater with NO_3^- were observed in grazed pastures in New Zealand (Ball et al., 1979).

There are several salts of nitrates such as sodium nitrate, potassium nitrate, and calcium nitrate etc, but the concern in water is simply nitrate. For example, when potassium nitrate dissolves in water it dissociates into potassium and nitrate to become independent quantities by a process called dissociation. There is no way of knowing whether a particular nitrate is from potassium nitrate or from calcium nitrate if both of them are dissolved in water. Certain organic chemicals are also nitrate, but they have very different properties, are very toxic and are not of concern in this study (Addiscott et al., 2005). In groundwater the cations are mainly of calcium, magnesium, potassium, sodium, iron and aluminum, and the salts that they form with nitrate are highly soluble.

In the developed world, most agricultural soils are maintained at a pH of 5.5 to 8.0 with the application of lime, thereby the soils being slightly acidic to slightly alkaline. The anions are repelled as the clays carry a negative charge at these pH values. Hence, it is advisable to assume that sorption does nothing to prevent NO_3^- from being transported to groundwater in the absence of clear evidence (Wong, Wild, and Juo, 1987; Duwig et al., 2003).

Major transformations in the N cycle are summarized as below (Madison and Brunett, 1985). 1. Absorption of inorganic forms of N (ammonia and NO_3^-) by plants and microorganisms. 2. Heterotrophic conversion of organic N from one organism to another organism. 3. Ammonification of organic N to ammonia during the decomposition of organic matter. 4. Nitrification of ammonia to NO_3^- and nitrite by the chemical process of oxidation. 5. Denitrification (bacterial reduction) of NO_3^- to nitrous oxide (N_2O) and molecular N (N_2) under anoxic conditions. 6. Fixation of N (reduction of N gas to ammonia and organic N) by microorganisms. The N cycle is shown in Figure 1 below.

By the process of nitrification, soil microbes readily convert ammonium (NH_4^+) to NO_3^-. Since, NH_4^+ is a cation, it is strongly attracted by clays and NO_3^-, an anion is not attracted. The form in which N is available to the crops is either in the form of NO_3^- or NH_4^+. Soils carrying any more of any of these two ions will usually result in the washing

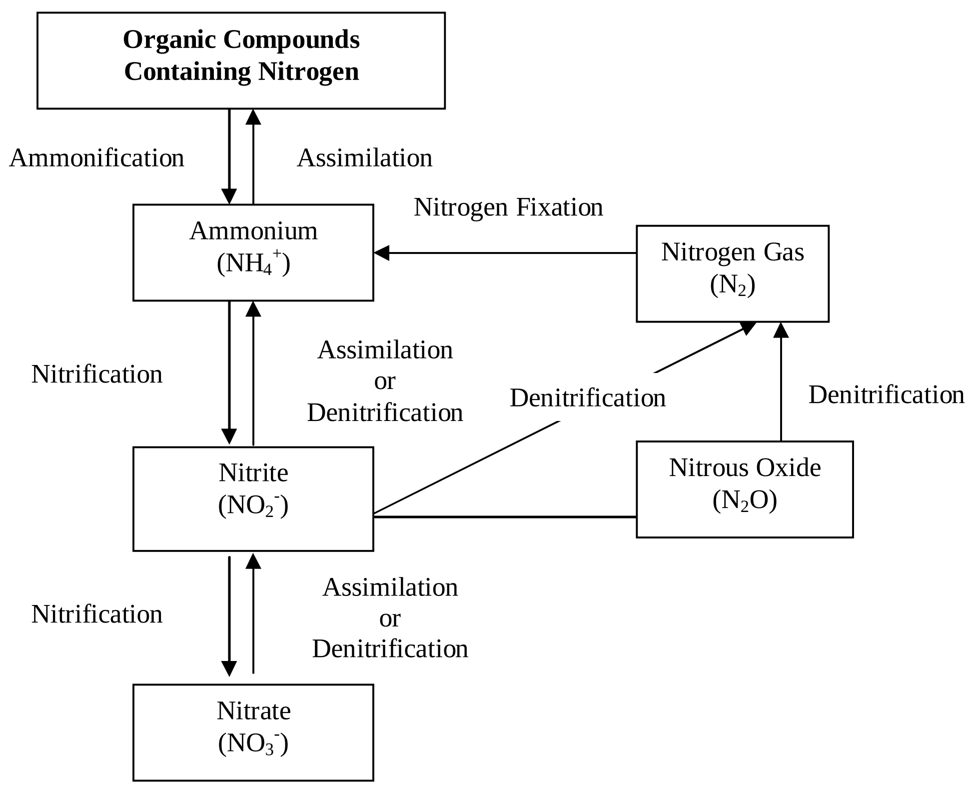

Figure 1. Simplified biological N cycle, Madison and Brunett, 1985.

away of NO_3^-. The N present in mineral form in the soil as NO_3^- and NH_4^+ constitute 1-2% of total soil N. This causes most of the environmental problems and also is most available to plants. Even though the quantity of N in humus is 50- 100 times more the quantity of mineral N, nothing happens to it rapidly and hence, is not an immediate problem.

The breakdown of organic matter by soil microbial activity, releasing CO_2, NH_4^+ and NO_3^- etc is known as mineralization. This process occurs in two stages called ammonification and nitrification in the case of N. Ammonification involves conversion of readily available N compounds to NH_4^+. The reaction is shown in the equation below:

$$R - NH_2 \rightarrow NH_4^+ + OH^- \tag{6}$$

NH_4^+ is converted to NO_3^- in two stages as shown in the equations below,

$$2NH_4^+ + 3O_2 \rightarrow 2NO_2^- + 4H^+ + 2H_2O + energy \tag{7}$$

$$2NO_2^- + O_2 \rightarrow 2NO_3^- \tag{8}$$

Some of the converted NO_3^- and NH_4^+ are simultaneously converted to various organic forms of N by a variety of soil organisms. Another process called denitification occurs in which some bacteria convert NO_3^- to N_2 or to N_2O. Production of N_2 is not a problem other than losing it to the atmosphere, whereas, partial denitrification resulting in N_2O is an environmental problem (Addiscott et al., 2005). The *Rhizobium* microbes in the root nodules of leguminous crops produce an enzyme called *nitrogenase*, which catalyses the N triple bond making N available to the plants.

According to Madison and Brunett (1985) the following are the major anthropogenic sources of NO_3^-: "fertilizers, septic tank drainage, feedlots, dairy and poultry farming, land disposal of municipal and industrial wastes, dry cultivation of mineralized soils, and the leaching of soil as a result of the application of irrigation water." The natural sources of NO_3^- are: "soil N, N-rich geologic deposits and atmospheric deposition." Hem (1989) suggests that N occurs in water as NO_3^- or NO_2^- anions, as ammonium cations, and in a range of organic compounds. In aerated water nitrite and organic species are unstable. Adsorption of ammonium cations to mineral surfaces is very strong, but the anionic species are readily transported in water and are stable over a wide range of conditions. As nitrate is the end product of reactions converting other forms of N in the soil, it is stable unless it is removed by plant uptake or denitrification. Given the wide range of NO_3^- sources associated with agriculture, its chemical stability in groundwater, high mobility and the frequency with which it has been measured in water; NO_3^- is a natural choice as an indicator for vulnerability of groundwater to contamination to non-point sources, other than the health concerns associated with it. Explanations of lower NO_3^- content in shallow groundwaters of the Southeast of the United States include dilution, denitrification, and uptake by plants (Hubbard and Sheridan, 1989).

CHAPTER III

NO_3^- CONTAMINATION IN THE UNITED STATES GROUNDWATER

Introduction

The distribution of NO_3^- in groundwater across the US is presented in this chapter. The variation of NO_3^- with respect to different parameters is assessed. The concentration of NO_3^- in ground water generally increases with higher N input and higher aquifer vulnerability (Nolan, Hitt, and Ruddy , 2002). The STATSGO database consists of the soil hydrologic group attribute, which has four major categories ranging from well-drained soils, soil hydrologic group A and B, to poorly drained soils, C and D (Service, 1994). Even in areas with high N input, poorly drained soils can reduce the risk of ground-water contamination (Mueller and Helsel, 1996). Additionally, water as runoff is carried away by drains and ditches off to streams rather than letting it seep to groundwater at the point of N input. The likelihood of groundwater contamination, even in areas with high N input and, in some cases, well-drained soils, can decrease with large amounts of woodland interspersed among cropland (Nolan, Hitt, and Ruddy, 2002). Explanations of lower NO_3^- content in shallow groundwaters of the southeast of the United States include dilution, denitrification, and uptake by plants (Hubbard and Sheridan, 1989).

Figure 2 shows the increasing N consumption in the United States traces an increasing trend. This trend is a disadvantage from groundwater protection perspective, as it is very clear from literature that increasing input leads to increasing concentration in the groundwater.

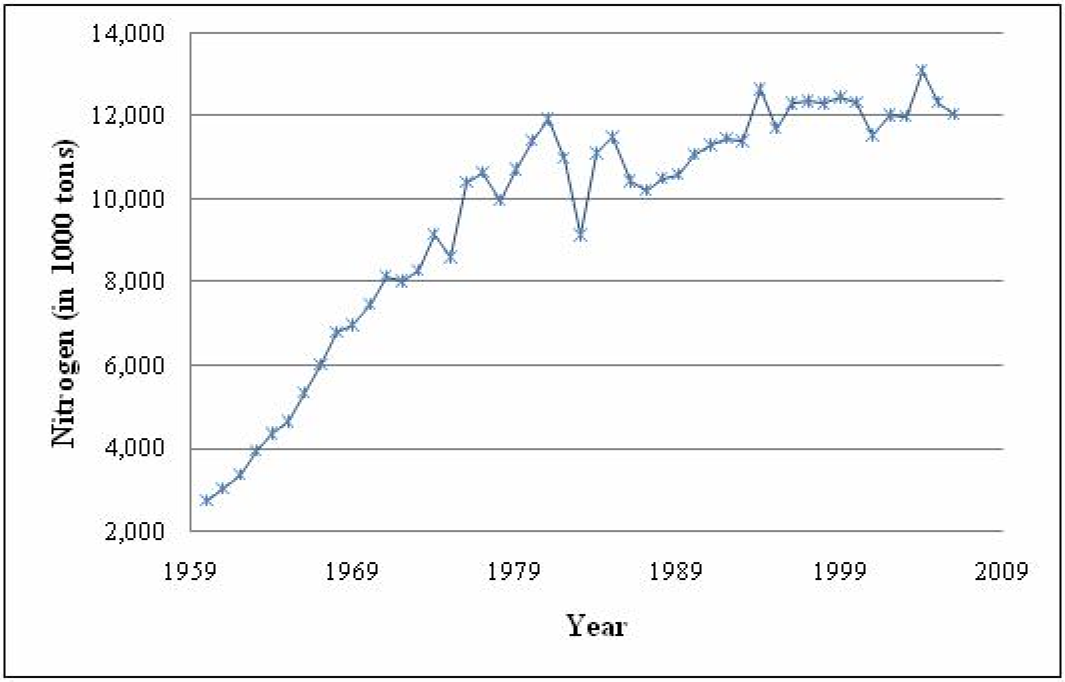

Figure 2. U.S. consumption of plant nutrients (N).

Data Synthesis and Analysis

Two datasets were used for the analysis of the behavior of NO_3^- with respect to groundwater vulnerability. First, the NO_3^- concentration values from the retrospective database, compiled by Hamilton (1994) from data and information provided by US Geological Survey's National Water-Quality Assessment (NAWQA) study units that began in 1991 is used. This dataset also consists of other parameters such as depth of the well, land use, type of well and nitrogen input in various forms. This database was used for analysis of NO_3^- variation with respect to other parameters, such as, depth to water table, N input and land use.

The second dataset is from NAWQA program's NO_3^- plus NO_2^- concentration values, which was used for the comparison of two methods considered. As the concentration of NO_2^- in groundwater is insignificant in comparison to NO_3^-

concentration (Hem, 1989), and also because this combination provides more wells to compare the result of the two methods, the data used is NO_3^- plus NO_2^- in mg/l of N. The NAWQA Program began in 1991 to describe the quality of the Nation's water resources, using nationally consistent methods. Hence, this data consisted of data only from the year 1991 (Koterba,Wilde, and Lapham, 1995).

The value used for comparison is the median of the concentration data for each well. The median is found to be more resistant to outliers typical of skewed data sets (Nolan and Stoner, 2000). The distribution of wells where NO_3^- concentration was measured is shown in Figure 3. The areas that showed pronounced problems from this dataset were: 1. northeastern USA, 2. intensely farmed area of the central USA grain belt, 3. irrigated agricultural regions of California and Idaho.

High NO_3^- concentration can be observed from southwestern and western central Nebraska. This observation is in agreement with Boyce et al. (1976), who found substantial quantities of NO_3^- under never-fertilized rangeland in this region. With the advent of irrigation the NO_3^- from the Pleistocene age loess were being leached into the groundwater. Literature also suggests that high levels of geologic NO_3^- exist in the alluvium beneath the San Joaquin valley in California and as in Nebraska some of this NO_3^- leached into the groundwater after the introduction of irrigation (Strathouse et al., 1980).

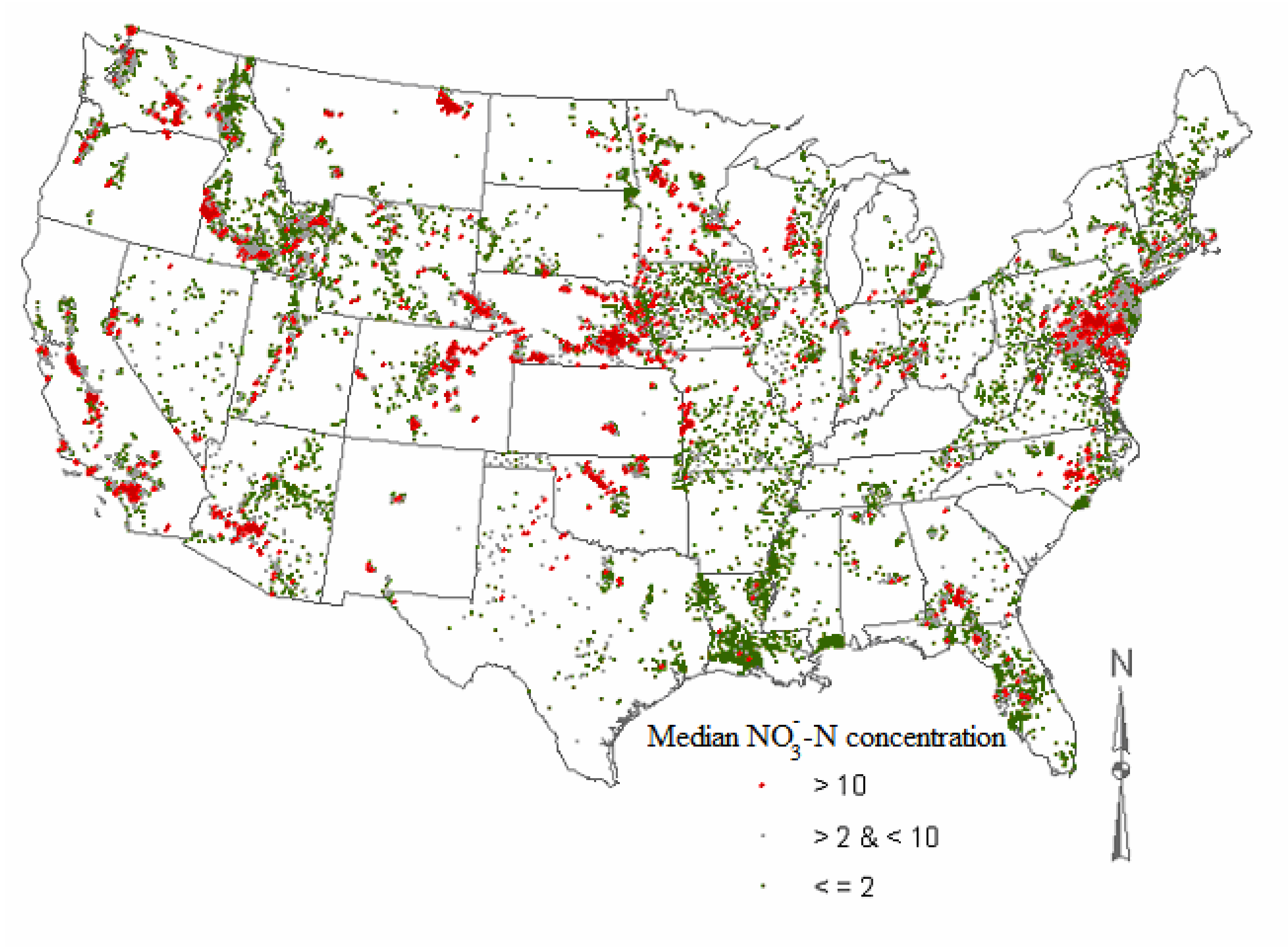

Figure 3. Map showing the distribution of the wells where NO_3^- concentration was measured in the U.S. from NAWQA program (1991-2006).

Figure 4 shows that 7.7 % of the wells where NO_3^- plus NO_2^- concentration was measured in the United States have median NO_3^- concentration values higher than the MCL of 10 mg/l of NO_3^--N, 27% of the wells have concentration between background and the MCL, and 65.1 % of the wells have a background concentration of 2 mg/l or less of NO_3^--N.

Out of the 7.7 % of the wells that have a NO_3^- concentration more than the MCL, 11.7 and 9.2% of the wells are in Nebraska and California respectively. Percentage distribution for some of the states with median nitrate concentration values greater than MCL is shown in Figure 5.

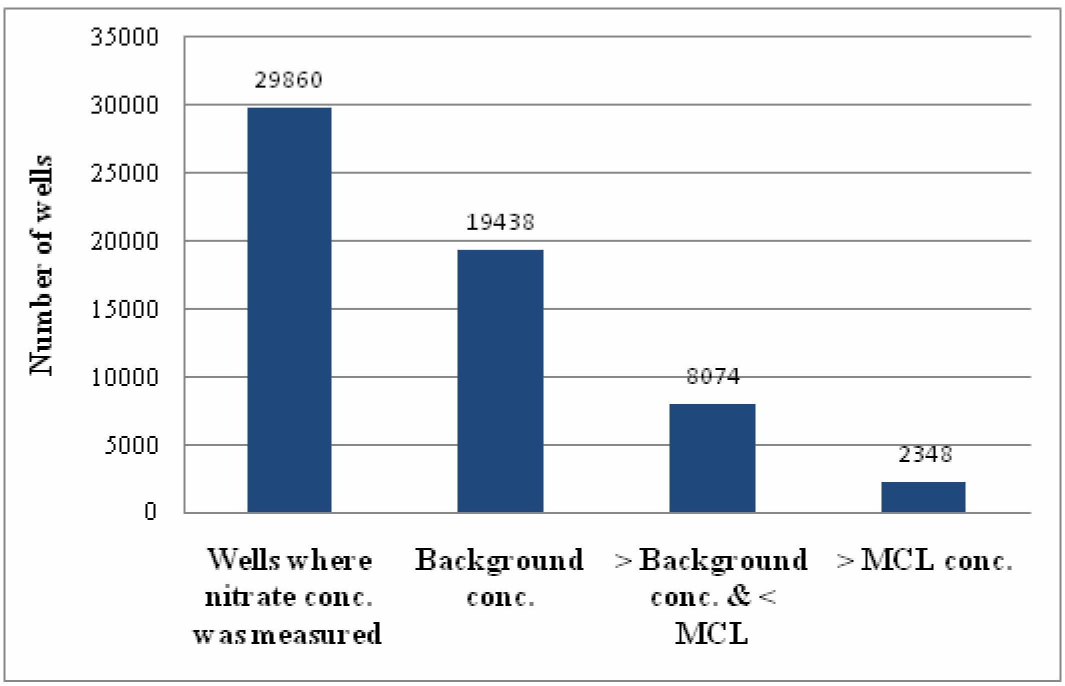

Figure 4. Median nitrate concentration levels in the U.S. (1991-2006).

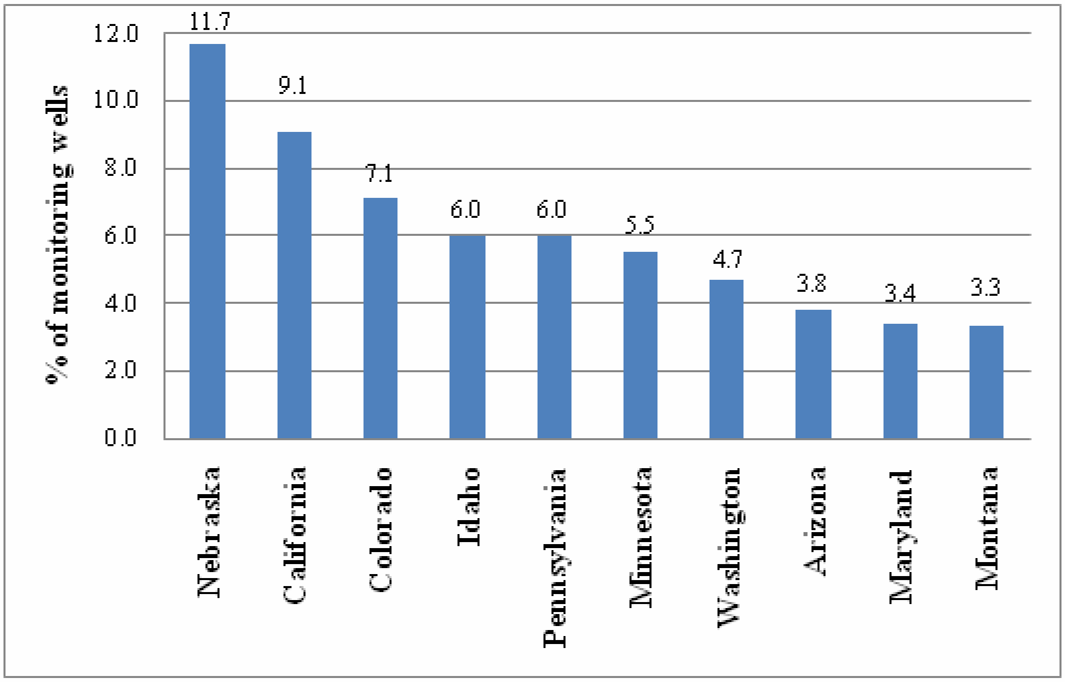

Figure 5. Percentage of wells greater than median MCL in top 10 states (1991-2006).

The four classifications considered in the trend analysis are, wells with an increasing and decreasing trend, no change, and in wells where only one concentration value was measured. The number of wells where NO_3^- plus NO_2^- concentration was measured, and used in this analysis, is 30,818. Around 60 % of the wells had only one concentration value measured, 7 % of the wells showed no trend, as indicated in Figure 7. 16.6 % of the wells showed an increase in the concentration of NO_3^--N. Figure 6 shows wells with increased NO_3^--N concentration are not very widely distributed in the United States. The states that show predominantly increasing trends are Iowa, Idaho, California, Nebraska, and Arizona.

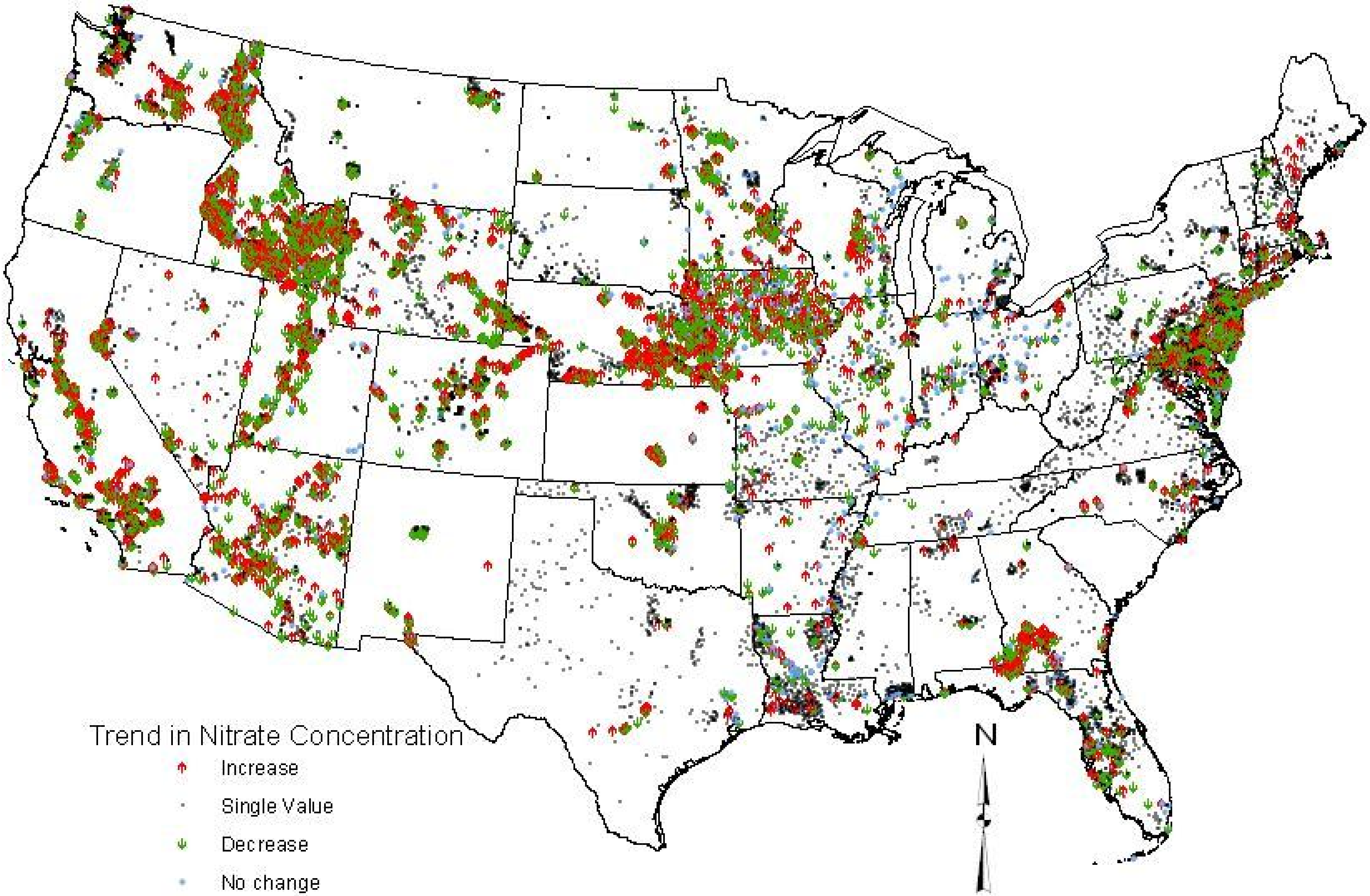

Figure 6. Trend in the median NO_3^- plus NO_2^- concentration in the U.S. groundwater (1991-2006).

Table 2. NO_3^- - N concentration data summarized by type of well. Data from 5361 wells across the U.S. from the retrospective database compiled by Hamilton, (1994) from the data and information provided by NAWQA study units that began in 1991

Type of well/ Number of wells	Maximum Value (mg/ l) NO_3^--N	Minimum Value (mg/ l) NO_3^--N	Median Concentration (mg/ l) NO_3^--N	Median Population Density (Number of people per km^2)
Domestic/ 3226	84.3 (1985)	0	1.2	15.2
Irrigation/ 838	52 (1987)	0	2.3	2
Public/ 1088	36 (1991)	0	0.2	45.5
Livestock/ 209	63 (1988)	0.01	2.9	6.2

It can be observed from Table 2 that highest value of median NO_3^- concentration is in livestock wells. The higher median value of the NO_3^- concentration is just a little over the background concentration, whereas the maximum value in the livestock wells may be explained based on high input of nutrients used. Similarly, irrigation wells have the second highest median NO_3^- concentration, but again are only slightly over the background value. The maximum value in this category may be explained from the leaching of the unused NO_3^- from fertilizers. Though the median value of the population density is higher in the public well category as compared to other wells, the median and maximum NO_3^- concentration values are low. This may be explained based on the extensive measures taken to protect the Public wells. Another observation from Table 2 is that though the maximum value concentration is very high in all well categories, the median value is well below the MCL. It is also to be noted that the high values were measured only once in these wells. The year of sampling is also given in the table along with the concentration values.

The division of the states into various geographic regions is shown in the Figure 7. This division is according to Spalding and Exner (1993). As observed in Table 3, though the total N input is the highest in the Corn-belt states, the median NO_3^-

concentration is not the highest. It is, in fact, the lowest value of the maximum NO_3^- concentration. There may be two important reasons for this observation: (1) the presence of poorly drained soils in the Midwest region (Keeney, 1986) and (2) The regions are predominantly agricultural and the crop grown is corn. The N requirement of corn is the highest, which implies that most of the N applied is absorbed by the crops. The higher total input of N in these regions could be explained based on the high N requirement of corn as shown in Figure 8. Northeastern states have the highest median value of 1.8mg/ l NO_3-N, this corresponding to a median population density of 51.1 people per km^2. The maximum value is very high in this region. Higher numbers of home sewage disposal systems are present in these regions, and relatively high rainfall rates and low evapotranspiration leads to higher leaching rates. It can again be observed that the median nitrate concentration is well below the MCL, though there are certain wells with a very high value.

Agricultural land has highest value of median NO_3^- concentration. It can be concluded that agriculture is the single major contributor of NO_3^- to the groundwater. It can be seen from Table 4 that 11.2 % of the samples have exceeded the MCL of 10 mg/l of NO_3^--N. Though, NO_3^- leaching from the forest is a potential threat, its contribution is much less as compared to agricultural contribution. In a survey of eastern watersheds the total N levels were five times greater in streams draining from agricultural watersheds than from forested watersheds (Omernik, 1976).

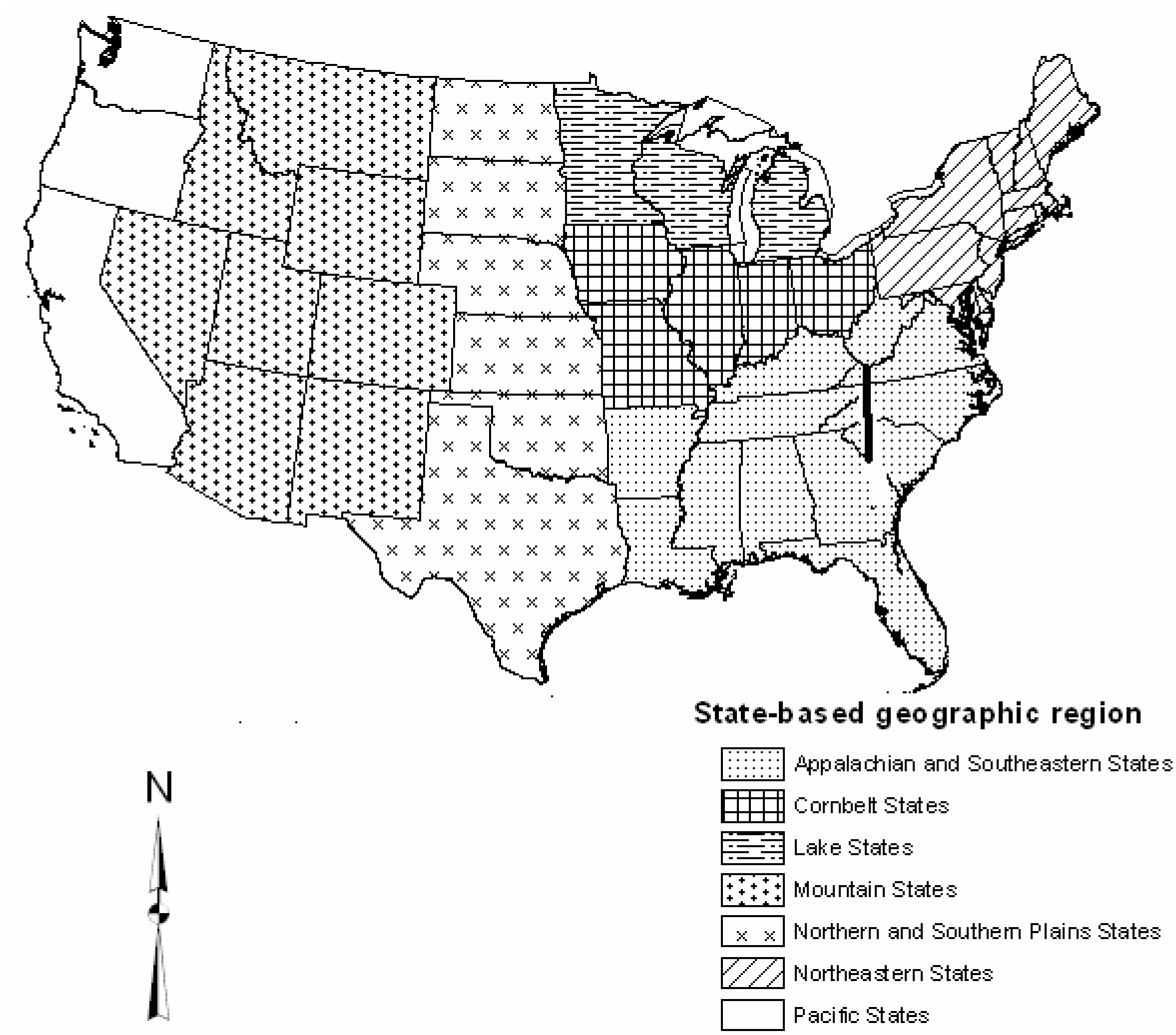

Figure 7. State-based geographic regions as defined by Spalding et al. (1991).

Table 3. NO_3^- -N concentration data, total N input, and population density summarized based on state-based geographic region. Data across the US from retrospective database compiled by Hamilton (1994) from the data and information provided by NAWQA study units that began in 1991

State based geographic region	Maximum Value (mg/ l) NO_3^-- N	Minimum Value (mg/ l) NO_3^-- N	Median Concentration (mg/ l) NO_3^-- N	Median Sum of N input from fertilizer, manure, & atmospheric sources (tons per mile^2)	Median Population Density (No. of people per km^2)
Corn-belt States	36 (1991)	0	0.2	16.1	12.9
Lake States	59 (1973)	0.01	0.06	14.4	18.3
Mountain States	46 (1981)	0	0.5	3.5	3.5
Northeastern States	70 (1989)	0.01	1.8	10	15.1
Northern and Southern Plains States	125.6 (1982)	0	0.9	12.4	4.2
Appalachian and Southeastern States	52.6 (1987)	0	0.05	5.9	13.7
Pacific States	83	0	1.1	14.8	7.3

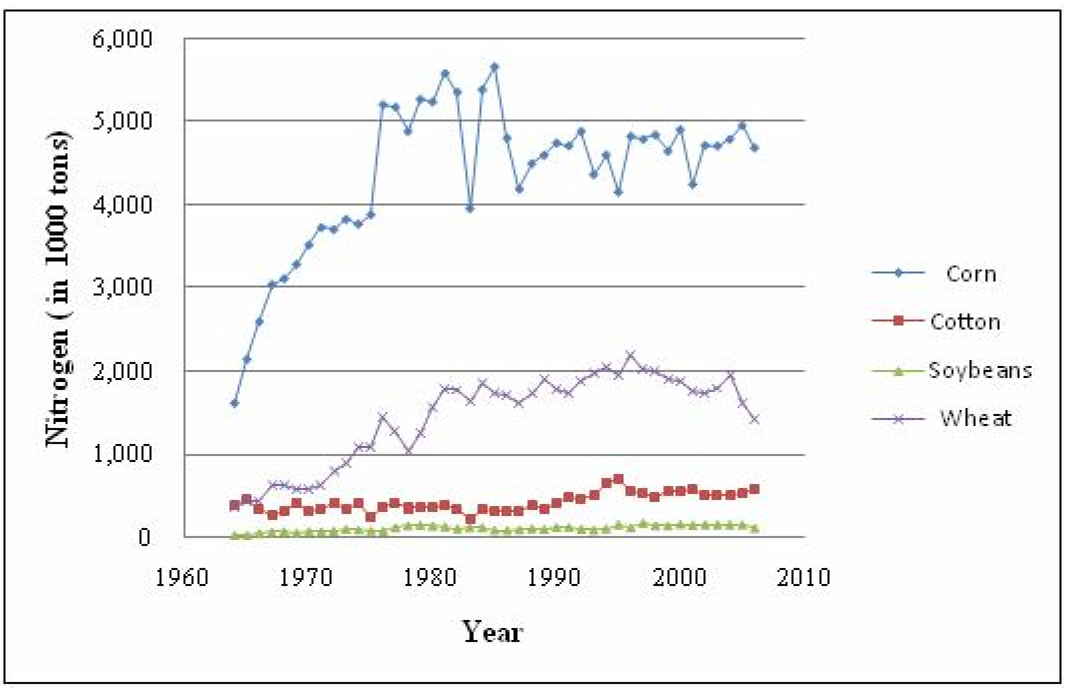

Figure 8. Total N applied for different crops in the U.S.

Table 4. Nitrate-N concentration data, total N input, and population density summarized based on Anderson level I land use category. Data across the US from the retrospective database complied by Hamilton (1994) from data and information provided by NAWQA study units that began in 1991

Anderson level I land use category	**Maximum Value (mg/ l) NO_3^--N**	**Minimum Value (mg/ l) NO_3^--N**	**Median Concentration (mg/ l) NO_3^--N**	**% samples exceeding MCL**	**Median Sum of N input from fertilizer, manure, & atmospheric sources (tons per mile^2)**	**Median population density (No. of people per km^2)**
Agricultural land	125.6 (1982)	0	1.4	11.2	13.7	12.2
Other land use, such as wetland	33 (1981)	0	0.1	2	5.7	18
Range land	84.3 (1985)	0	0.6	3.3	2.6	1.1
Urban or built-up land	31 (1984)	0	0.3	3.4	4.8	163.5
Forest land	24 (1973, 86, 90)	0	0.1	1.5	5.4	12.8

It can be observed from the Figure 9 that most of the points are within the first 300 feet of well depth below the land surface. Seventy-seven percent of the points are within 100 meters depth, thereby indicating that the problem is mainly a result of anthropogenic causes. This figure also exhibits a significant decline in the NO_3-N concentrations with the increasing depth below the land surface. All values corresponding to zero concentration were neglected; as such values could not be plotted on the logarithmic scale.

Figure 10 shows the distribution of NO_3^- wells with concentrations greater than the MCL of 10-mg/ l of NO_3^--N against the annual average precipitation map. It can be observed that the wells with pronounced NO_3^- contamination problems are not very dense in the regions where the precipitation is high; instead its distribution does not trace a

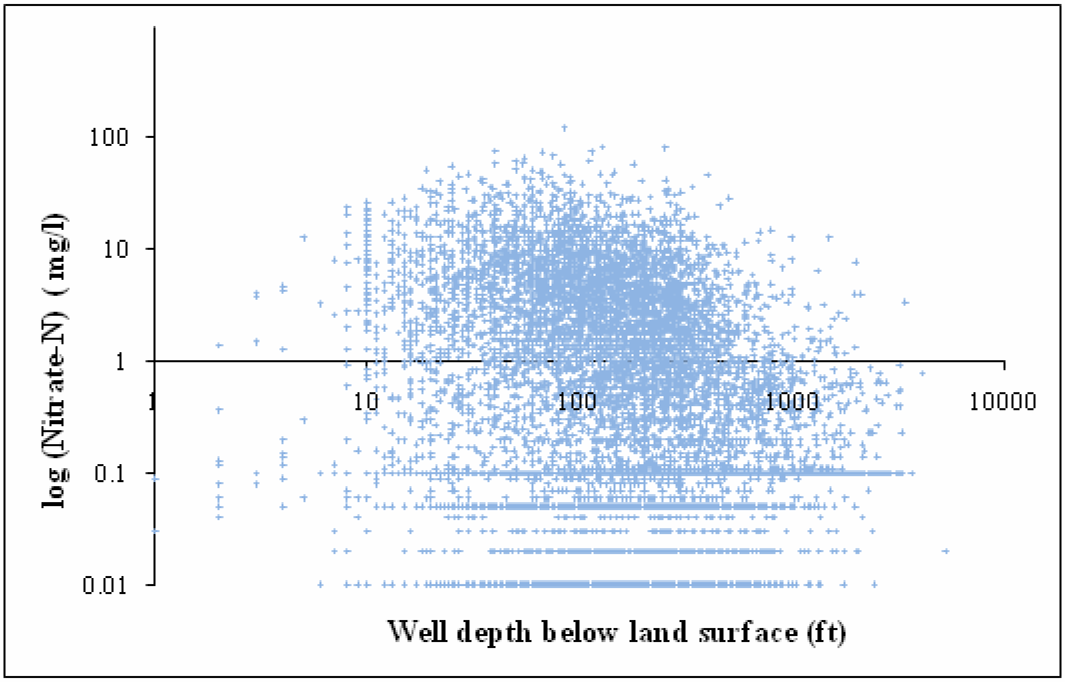

Figure 9. Nitrate-N concentrations in groundwater vs. well depth.

trend with the precipitation. Though, there is no trend that can be observed it can be stated that the precipitation is one of the important means by which NO_3^- travels to the groundwater. This implies that precipitation is not the only factor, but along with other parameters has an effect on the NO_3^- concentration in the groundwater.

To understand the behavior of NO_3^- at a smaller scale, the Central Nebraska basin with some similar attributes were examined. Similarity of some of the parameters provides an opportunity to compare the wells with different concentration values. The variation of NO_3^- concentration in the wells located at the Central Nebraska Basin at a well depth of 100 feet (this well depth was chosen as the number of values was enough to indicate a trend) is analyzed here. Domestic wells category was chosen for the similar reason of comparison. The lithologic description of aquifer is also the same for all the above wells as unconsolidated sand and gravel. The land use is predominantly agricultural land, based on the Anderson level I land use category. Four out of the 23 wells were in the rangeland category. The Soil hydrologic group (STATSGO) varies between A and B indicating well drained to moderately well drained types of soil for all the wells used in this comparison. The variation of NO_3^- concentration in the groundwater is plotted against the sum of N input from fertilizer, manure, and atmospheric sources in tons per square mile. This is shown in Figure 11.

Though there is scattering due to random variation, the points trace an increasing trend as they move to the right. The increasing trend indicates that as the N input increases the concentration of NO_3^- in the groundwater also increases. The relationship is

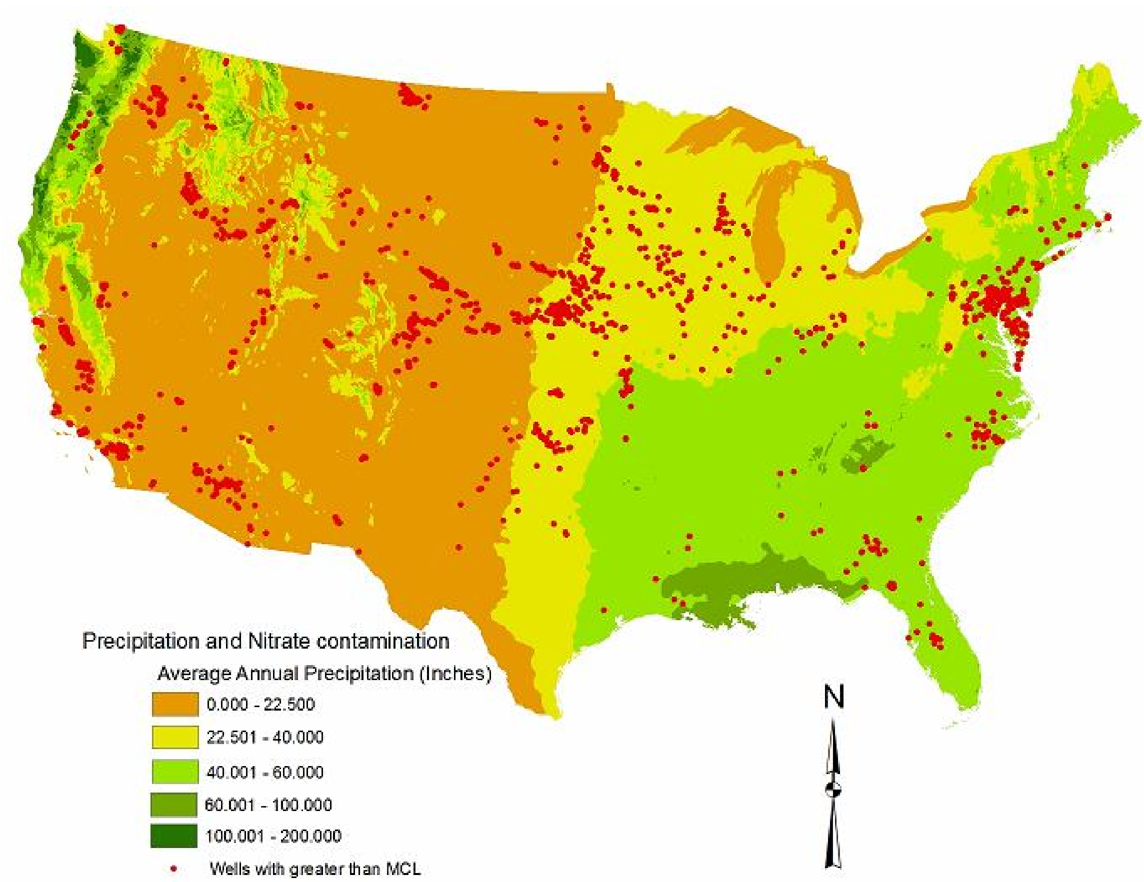

Figure 10. Distribution of NO_3^- wells with greater than MCL against average annual precipitation.

not perfect, because of the heterogeneity of the medium through which the contaminant travels, as is expressed by some random scattering in the distribution of the points. As the concentration of NO_3^- is dependent on various parameters, even considering certain parameters similarly, as is the case here, does not ensure a perfect increasing trend.

Summary

Conversion of N to NO_3^- in the aerobic natural environment is inevitable and contribution from anthropogenic activities is of great concern. On a national scale agriculture is recognized as the major contributor of NO_3^- to groundwater. The leaching of the NO_3^- depends on the type of soil, with poorly drained soils allowing little or no leaching. Various factors affect the concentration of NO_3^- in groundwater. Some of the factors are land use, depth to the water table, precipitation and evapotranspiration. Some of the regions with pronounced problems are the Midwest, Northeast, and the well irrigated regions of California. To protect groundwater from NO_3^- contamination, fertilizer use must be decreased, as there is a clear correlation between the two. The fertilizer consumption statistics, however, show an increasing trend in the United States. Emphasizes is on production of crops with greater efficiency with respect to N input. All states exhibit some degree of groundwater/wells contamination due to NO_3^-. In view of the health concerns associated with NO_3^-, the regions with higher aquifer vulnerability must be protected.

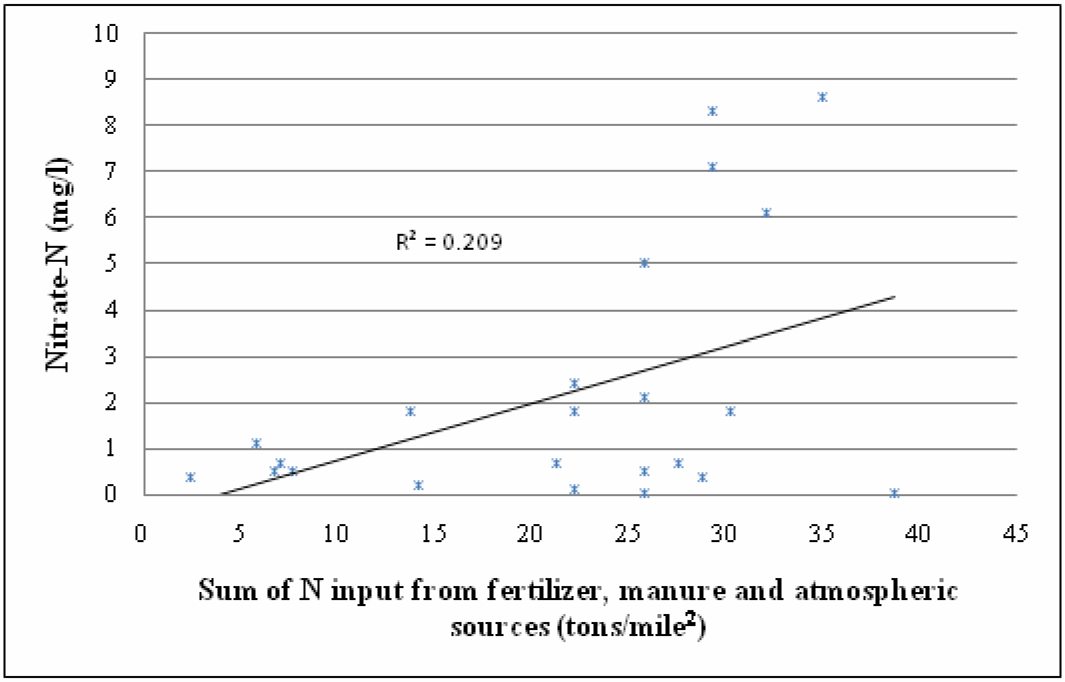

Figure 11. NO_3^- variations vs. the sum of N inputs from fertilizer, manure, and atmospheric sources.

CHAPTER IV

VULNERABILITY OF SHALLOW GROUNDWATER TO NO_3^- USING MODIFIED DRASTIC AND ORDINAL LOGISTIC REGRESSION

Introduction

This chapter discusses the approach and the development of modified DRASTIC and ordinal logistic regression models to compute the NO_3^- contamination vulnerability of shallow aquifers (less than 50 feet deep in this study) across the conterminous United States. The preparation of data along with the development of vulnerability map is discussed in detail. National scale assessment of the DRASTIC at 1:250,000 scales have so far not been done. This study uses modified DRASTIC approach, which included the contaminant loading parameter. Ordinal logistic regression methodology was used to obtain the probability map for heavy metal contamination only. This effort is to further its application for the case of nitrate contamination. This chapter consists of two sections, with the first section addressing the modified DRASTIC model and the second ordinal logistic regression.

Modified DRASTIC Approach

The focus of this section is the development of an aquifer vulnerability map across the conterminous United States using the Modified DRASTIC model. DRASTIC's methodology permits systematic evaluation of the groundwater pollution potential anywhere in the United States. Its methodology is designed such that only the hydro-geological factors are taken into consideration for the computation of the vulnerability of the groundwater. This model was developed to assist planners, managers, and administrators in the task of evaluating the groundwater vulnerability to various pollution sources. The intention is to help direct resources and land-use activities to the appropriate areas. According to the authors the model cannot replace any onsite inspections, nor can it be used to quote any type of facility or practice on any site. Rather, the purpose is to provide a preliminary procedure to evaluate the pollution potential of groundwater (Aller et al., 1987).

Description

This section is divided into five parts. Part one gives an overview of the model. Part two describes the development of the system, the description of the processes with respect to developing the methodology, assumptions, uses of the system and its limitations. Part three provides the description of factors and data sources. Part four describes the grounds in using the data and the development of the vulnerability map. Part five provides the results and conclusion.

DRASTIC is an empirical model developed by the National Water Well Association in conjunction with the US EPA to determine aquifer vulnerability on a regional basis. Although DRASTIC is physically based the final DRASTIC index is just a numerical index. This method was created to evaluate the aquifer vulnerability of any area in the United States and can only be used for an area larger than 100 acres. Due to wide variability of the pollutants DRASTIC assumes a generic pollutant. It is assumed that the pollutant has the mobility of water (Aller et al., 1987). NO_3^- is prone to leaching through soil with infiltrating water due to its solubility and mobility (Nolan, Hitt, and Ruddy, 2002). The solubility and mobility makes NO_3^- an appropriate choice for performance indication of the model. The breadth of NO_3^- concentration data in groundwater again allows for a reliable comparison of the performance of the model.

The DRASTIC model does not readily assess the condition of leaky aquifers or confined aquifers. The system encompasses two portions, namely, the hydrogeologic settings and the relative ranking of the hydrogeologic parameters. In the hydrogeologic setting are the physical characteristics, which affect the pollution potential of the groundwater. Since DRASTIC does not take into account the specifics of a particular contaminant, its result can be used only to compare contaminants which have the mobility of water, such as NO_3^-. Hence, modified DRASTIC approach is an improvement over the DRASTIC model.

Data sources

The Modified DRASTIC index is the outcome of seven hydrogeologic parameters and the N loading, namely, the Depth to water, Net Recharge, Aquifer Media, Soil Media, Topography (Slope), Impact of the vadose zone Media, Conductivity (Hydraulic) of the Aquifer and the N loading . The data for depth to water was obtained from the STATSGO database developed by United States Department of Agriculture. The field in the STASGO attribute table for shallow water table depth is "*wtdeph,*" which is the maximum value for the range in depth to the seasonally high water table during the months specified. This field is found in the "comp" table of the STATSGO dataset. The STATSGO data provides a national coverage, at a scale corresponding to 1:250,000; except for Alaska, where the scale corresponds to 1:20,000,000 (Service, 1994). The net recharge data used was obtained from the data "Estimated mean annual natural ground-water recharge in the conterminous United States" from USGS (Wolock, 2003). The data for the aquifer media was obtained from the information compiled by the U.S. Geological Survey. This dataset contains Principal Aquifers of the 48 Conterminous United States, Hawaii, Puerto Rico, and the U.S. Virgin Islands (USGS, 2003). This data generally contains information regarding the uppermost principal aquifer. The soil data was obtained from the STATSGO dataset. This database consists of Soil hydrologic group, which was used as a surrogate for the actual soil texture classification data described in the DRASTIC approach. The topography information was obtained again from the STATSGO dataset as well. The attributes which contained this information are “slopeh” and “slopel,” which are abbreviations for the maximum and minimum value for the range of slope of a soil component within a map unit, respectively. The aquifer media data was used as a surrogate for the attribute Impact of the Vadose Zone. The conductivity (Hydraulic) of the aquifer was obtained by assigning values of hydraulic conductivity obtained from Freeze and Cherry (1979) to the principal aquifers of the conterminous United States, Hawaii, Puerto Rico, and the U.S. Virgin Islands data. Finally the data for N loading was obtained from the dataset “Estimates of N-fertilizer sales for the conterminous United States in 1990,” (Battaglin and Goolsby, 1994) and the N deposition from the atmosphere obtained from national atmospheric deposition program website.

Methodology

The rating of each of the factors considered in the evaluation of modified DRASTIC was done using the procedure outlined in the DRASTIC manual and Almasari et al. (2005). The cell size used in the computation of the final DRASTIC index is of 1 km resolution. This was used due to the limitation of the availability of all the data at a resolution finer than 1 km.

Depth to water: The shape files containing the depth to the seasonally high water table from STATSGO dataset were converted into a raster (a grid of rows and columns of cells). The cell size of the raster grid is 1 km. This raster file is rated for different ranges according to the procedures outlined in the DRASTIC manual and shown in Table 5 below. The final depth to water rating raster was obtained by adding the individual rated files. Figure 12 below is the map showing the depth to water rating computed according to the DRASTIC approach.

Table 5. Rating for depth to water table

Depth to water table (feet)	
Range	**Rating**
0-5	10
5-15	9
15-30	7
30-50	5
50-75	3
75-100	2
>100	1

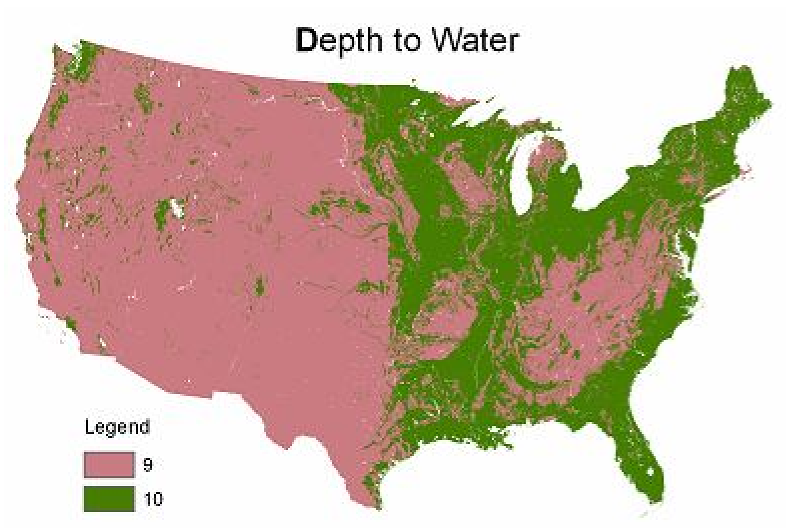

Figure 12. Depth to water rating.

Net recharge: The net recharge data was obtained from the raster dataset of mean annual natural ground-water recharge developed by the USGS. According to the authors, the grid of base-flow index values and the grid of mean annual runoff values derived from a 1951-80 mean annual runoff contour map were multiplied to obtain the ground water recharge values (Wolock, 2003). This data are then rated according to Table 6 given below. The final net recharge rating raster was obtained by adding the individually rated raster files. Figure 13 below is the map showing the net recharge rating computed according to the DRASTIC approach.

Table 6. DRASTIC rating for net recharge

Recharge (Inches)	
Range	**Rating**
0-2	1
2-4	3
4-7	6
7-10	8
>10	9

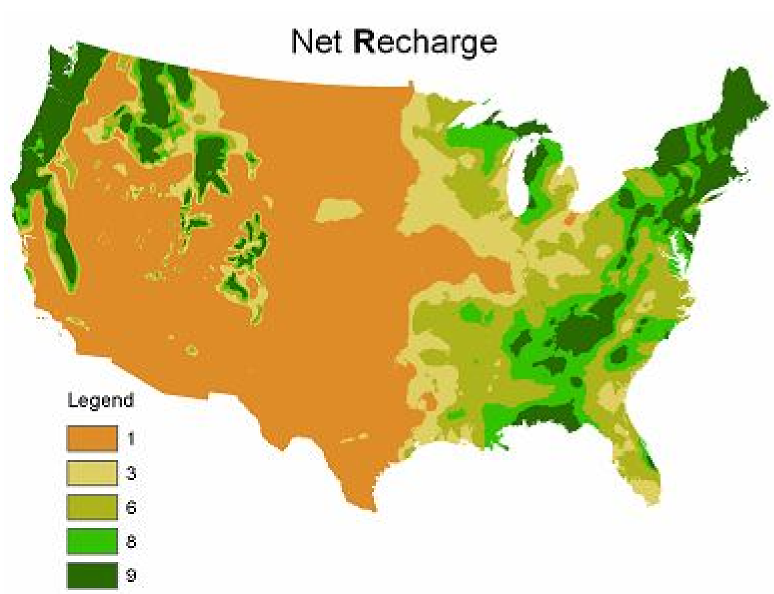

Figure 13. Ratings of net recharge.

Aquifer media: The shape file for the shallow aquifers for the conterminous US is rated according to the ratings provided by the DRASTIC manual and the rated shape file was converted to a raster. The ratings used are given in Table 7 below. The final rated map is shown in Figure 14 below.

Soil media: The raster layer for the Soil Media was prepared from the Soil Hydrologic group, which was used as a surrogate for the individual soil texture. This approach was adopted as the number of soil texture classification was very large and the classification given in the DRASTIC manual did not include all soil types. The general soil description of the soil rating system used in DRASTIC was limited to a few soil types, thereby introducing subjectivity in the choice of rating. The following are the ratings for the soil media: **A:** 8, **B:** 5, **C:** 4, **D:** 3, **A/D:** 6, **B/D:** 4, **C/D:** 4, which are rated based on the permeability of the soil group. The map generated from this rating is shown in Figure 15.

Table 7. Rating used for the aquifer media

Aquifer Media	**Rating**
Other Rocks	3
Carbonate-rock aquifers	8
Igneous and metamorphic rock aquifers	3
Sand stone and carbonate rock aquifers	6
Sandstone aquifers	6
Semi consolidated sand aquifers	4
Unconsolidated sand and gravel aquifers	8

Topography: The fields from the STATSGO attribute data for the conterminous US, namely, Slopel and Slopeh (the minimum and maximum value for the range of slope of a soil component, respectively) are used to calculate the final slope value. The average of the above two fields is used to determine the final slope value which is used to determine the ratings for this factor, based on Table 8 shown below. The map produced using this approach is shown in Figure 16.

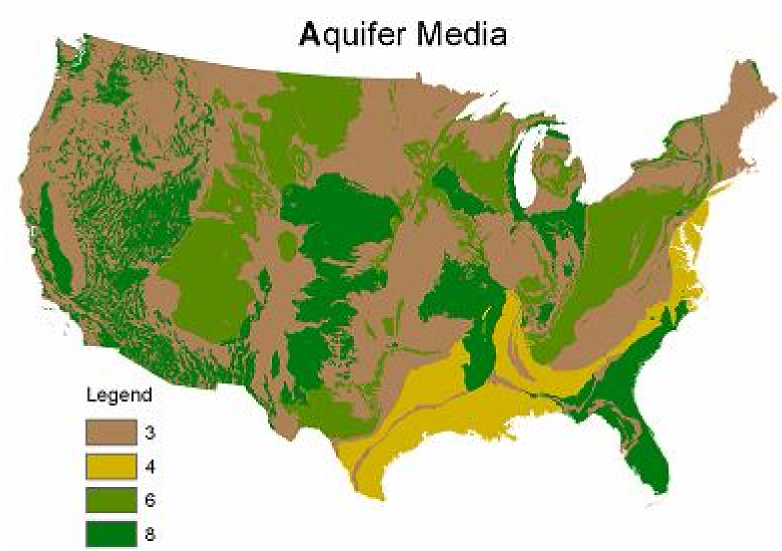

Figure 14. Ratings for the aquifer media.

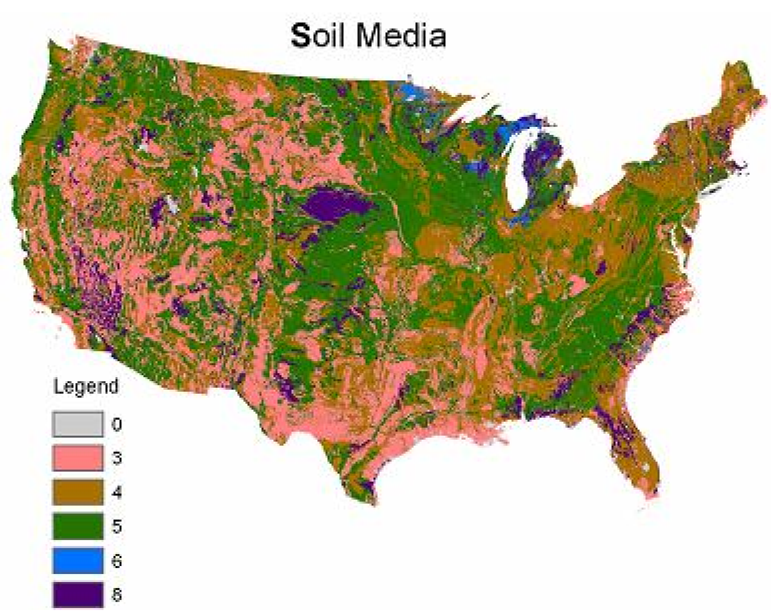

Figure 15. Ratings for the soil media.

Impact of the vadose zone media: The shape file used for the calculation of aquifer media is used with the assumption that the geology present just above the water table will be the similar to the geology below the water table. With this assumption the ratings map is prepared, with the procedure outlined in the DRASTIC manual. The ratings are shown in Table 9. The map generated using the above mentioned approach is shown in Figure 17.

Table 8. DRASTIC rating for topography (slope)

Topography	
Range	**Rating**
0-2	10
2-6	9
6-12	5
12-18	3
>18	1

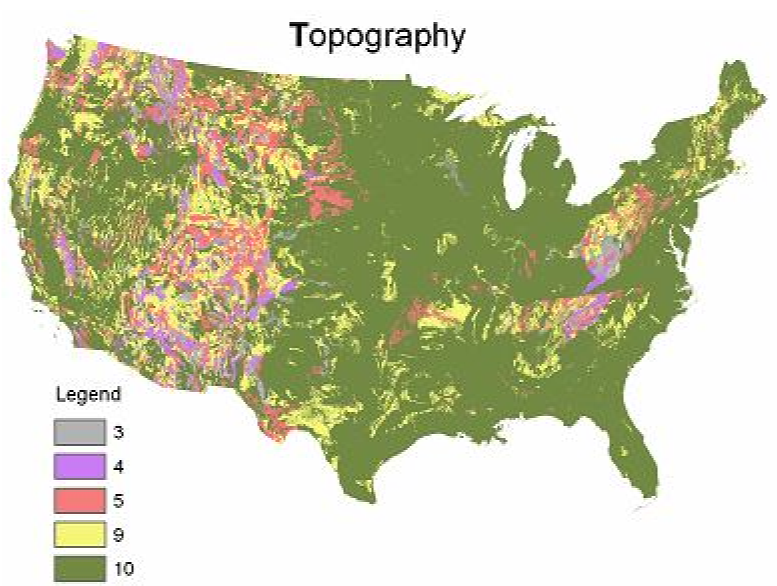

Figure 16. Ratings for the topography.

Hydraulic conductivity of the aquifer: The shape file for the shallow aquifers for conterminous United States was used along with the hydraulic conductivity values for the corresponding aquifers. The aquifers hydraulic conductivity was determined using the information from Freeze and Cheery (1979). This is rated with the ranges given in Table 10 and the rated shape file is converted to a raster with a cell size of 1 km. The map generated from the above mentioned procedure is shown in Figure 18.

Table 9. DRASTIC rating for impact of the vadose zone media

Impact of Vadose Zone Media	
Material	**Rating**
Other rocks	1
Igneous and Metamorphic rocks	4
Semi-Consolidated sand aquifers	5
Sandstone aquifers	6
Unconsolidated sand and gravel aquifers	8

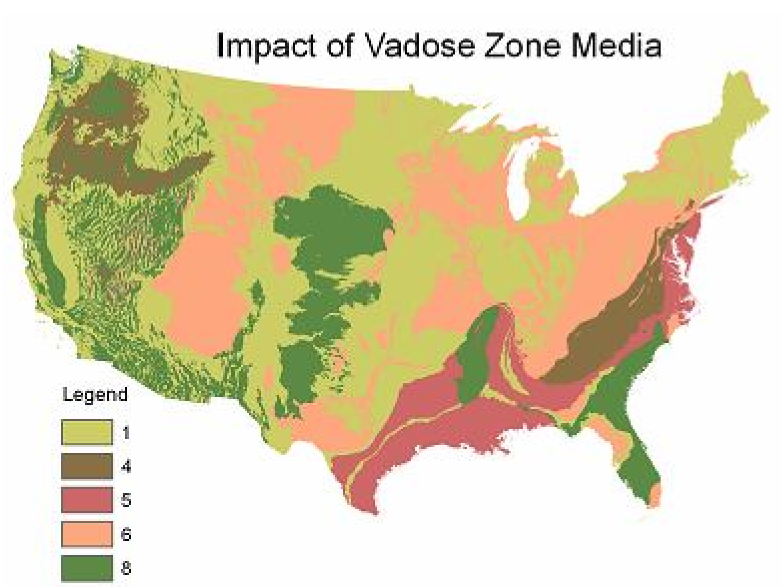

Figure 17. Ratings for the Impact of the vadose zone media.

Table 10. DRASTIC rating for conductivity (hydraulic) of the aquifer

Conductivity (hydraulic) of the aquifer	
Material	**Rating**
Sandstone aquifers	1
Sandstone and Carbonate rock aquifers	1
Igneous and metamorphic rock aquifers	1
Semi consolidated sand aquifers	8
Unconsolidated sand and gravel aquifers	10

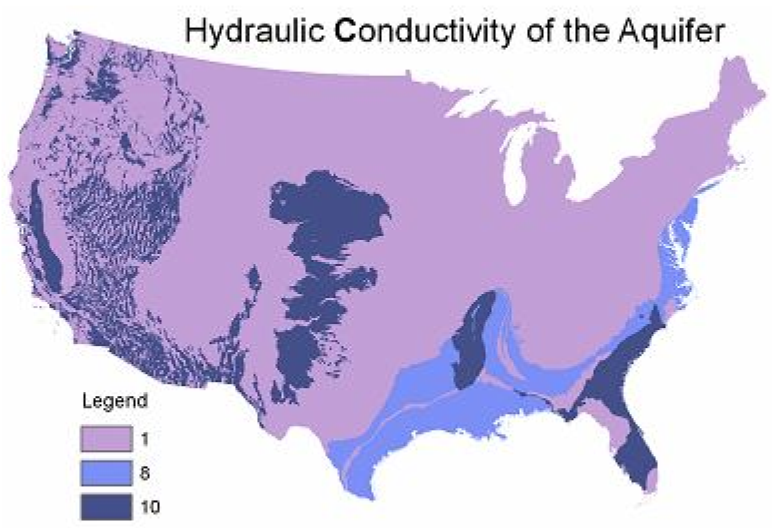

Figure 18. Ratings for the hydraulic conductivity of the aquifers.

Nitrogen loading: The total N loading considered here consists of N loadings from two major sources, namely farm fertilizer and confined animal manure. The farm fertilizer N loading was compiled at the county level from national databases of fertilizer sales (Ruddy, Lorenz, and Mueller, 2006). The approach used to estimate the farm fertilizer loading is based on the procedure described by Nolan and Hitt (2006). The county level N loading was allocated to the land use categories comprising, orchards/vineyards/other, pasture/hay, row crops, small grains, and fallow land. The loadings were determined based on a weighting factor obtained by dividing the area of the above mentioned lands in the particular county (Nolan and Hitt, 2006). The base land use data called National Land Cover Data (NLCD) at 30-m resolution was used (Vogelmann et al., 2001). The final N fertilizer application was based on the enhanced version of the land use data designated as “NLCDe”. NLCDe reclassifies the misclassified NLCD data with the aid of 1970s-1980s aerial photography data (USGS, 1990). The misclassified data pertaining to N fertilizer application resulted from orchards and residential areas with tree canopy being classified as forest. This error occurred as these are difficult to distinguish with satellite imagery (Nakagaki and Wolock, 2005). The annual estimates of the farm fertilizer loading in kilograms per hectare applied to agricultural lands were averaged for the years 1992-2001. This approach was adopted based on the data availability at the time of the study.

An approach similar to the farm fertilizer loading was adopted to determine the N loading from confined manure. Annual N input in kilograms per hectare from confined animal manure was averaged for the years from 1992-1997. Confined manure was applied to pasture/hay, row crops, small grains, and fallow land use categories from the NLCD (Ruddy, Lorenz, and Mueller, 2006). Confined manure estimates for other years were not available at the time of this study (Nolan and Hitt, 2006).

The total N loading used here is the sum of the farm and confined animal manure N loading sources as mentioned above. The atmospheric deposition of N was very insignificant as compared with the other two sources and hence was not considered in this study. The data was compiled at a one km by one km resolution for the conterminous

United States. The use of only certain years of data in the study is based significantly on the availability of the data at the time of the study. The ratings map for N loading shown in Figure 19 was generated from the procedures outlined in the modified DRASTIC procedure, with a weight of five (Secunda, Collin, and Melloul, 1998; Almasari et al., 2005). The ratings were determined by dividing the N loading into 10 different categories and then rating the highest category with a rate of 10 and so on.

DRASTIC index: The computation of the modified DRASTIC index is done using the empirical equation from the DRASTIC manual and the guidelines set forth in Almasari et al. (2005), which are given below. The final DRASTIC index map is shown in Figure 20 below.

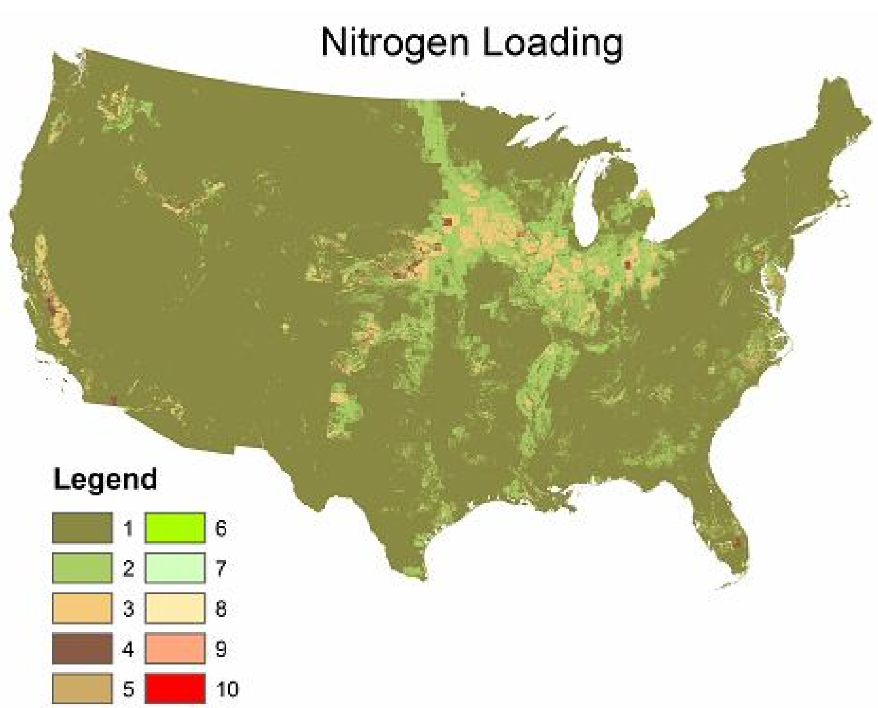

Figure 19. Ratings for N loading.

DRASTIC index = 5D + 4R + 3A + 2S + 1T + 5I + 3C + 5N

where

D= Depth to water; R= Net Recharge; A = Aquifer Media; S= Soil Media; T= Topography; I= Impact of the vadose zone media; C= Conductivity (Hydraulic) of the Aquifer; N= Nitrogen Loading.

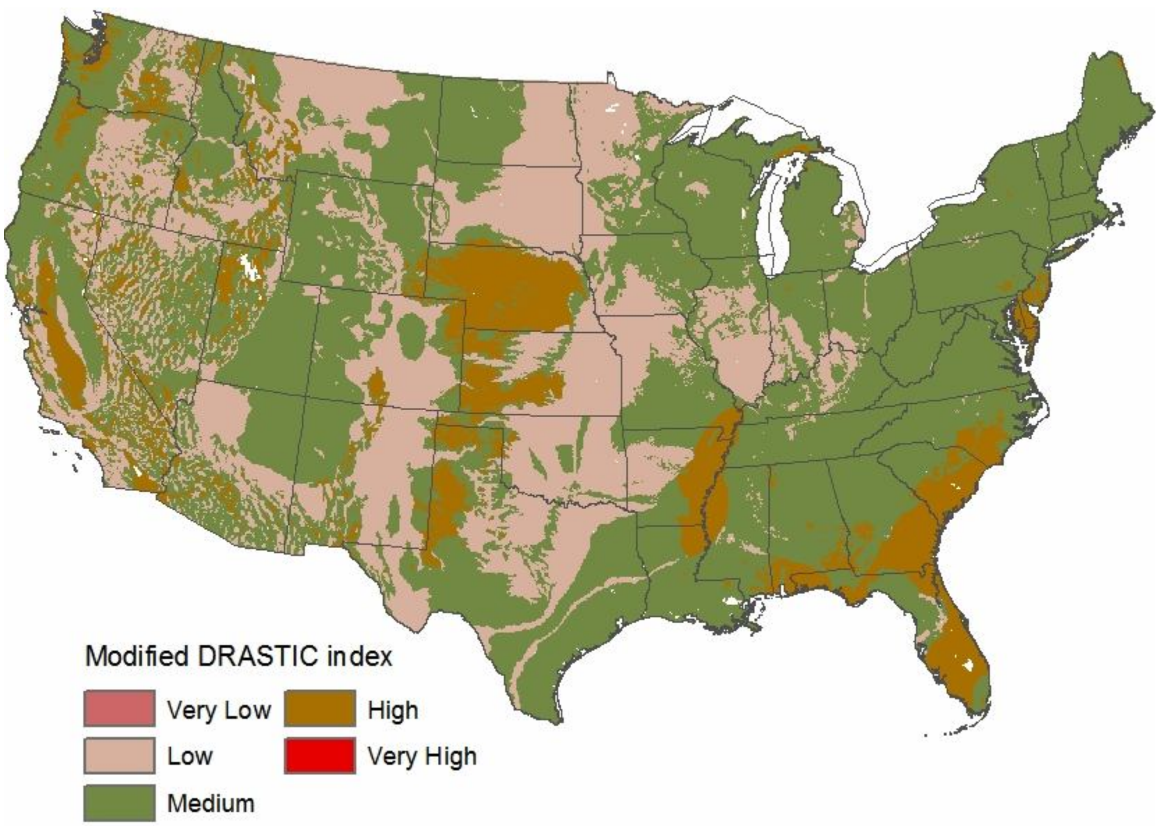

Figure 20. Final modified DRASTIC index.

Ordinal Logistic Regression Approach

Introduction

Binary logistic regression, commonly known as logistic regression (LR) is a statistical method used to estimate aquifer vulnerability. LR models were used by Nolan, Hitt, and Ruddy (2002) to estimate aquifer vulnerability to NO_3^- contamination in the U S. Ordinal logistic regression, henceforth referred to as OLR considers more than one threshold value to obtain aquifer vulnerability; this is considered to be an improvement over binary LR methods (McCullagh, 1980). The background concentration and the MCL can both be used to assess the probabilities of occurrence of NO_3^- in groundwater (Twarakavi and Kaluarachchi, 2005).

Description

The probability of the response being less than a threshold value is related to a set of influencing variables in LR, whereas, in classical linear regression the influencing variables are related to the response variable (Afifi and Clark, 1984; Hosmer and Lemeshow, 1989; Helsel and Hirsch, 1992; Kleinbaum, 1994). For instance, the probability of NO_3^- being less than the MCL for N loading, recharge and groundwater withdrawal etc, is considered in LR. The natural logarithm of the *odds ratio* for the probability of response to be less than the threshold value, and influencing variables is a linear regression in the LR model.

The odds ratio, *O*, is given as in equation (9)

$$O = \frac{p}{1-p} \tag{9}$$

where p is the probability of the response to be less than a given threshold value.

The natural logarithm of the odds ratio, or *logit*, is linearly related to the influencing variables in binary LR and is written as shown in Equation (10)

$$\log(O) = \text{logit}(p) = a + bx \qquad (10)$$

where a is a constant, b is a vector of slope coefficients, and x is the vector of influencing variables.

The proportionality-odds model, commonly known as OLR expands this concept to more than one threshold value (McCullagh, 1980). For example, if two thresholds (i =1, 2) are considered to categorize the response variable, OLR relates the corresponding *logits* as follows (Equation 11);

$$\log(O_i) = \text{logit}\,(p_i) = a_i + bx \qquad i=1,\,2 \qquad (11)$$

where, O_i is the odds ratio for probability of response to be less than the i^{th} threshold, p_i is the probability of response to be less than the i^{th} threshold, and a_i is the constant for the i^{th} threshold, b is a vector of slope coefficients, and $\boldsymbol{x}$ is the vector of influencing variables.

The same slope coefficient, “b” is assumed to relate the probabilities of occurrence of response to the influencing variables, with respect to all the thresholds. Using binary LR for more than one threshold would result in the use of different slope coefficients, thereby resulting in a loss of physical significance (McCullagh, 1980; Helsel and Hirsch, 1992). The benefits of application of OLR models over LR models for ordinal nature of data are explained in the literature (McCullagh, 1980). OLR model is fitted to the observed responses using the maximum likelihood approach. Unknown parameters are determined using the maximum likelihood approach that best match the predicted and observed probability values. The application of maximum likelihood theory to OLR models is explained in detail by Hosmer and Lemeshow (1989) and McCullagh (1980).

Data sources

The NO_3^- data used for the analysis is from NAWQA program's land use type of groundwater studies which sample shallow groundwater. The data used in this study is taken during the 1991-2005 periods, thereby ensuring consistency in the collection procedures (Fishman, 1993; Koterba, Wilde, and Lapham, 1995). The data in the form of NO_2^- plus NO_3^- in mg/L as N was used and is henceforth referred to as nitrate as the concentration of NO_2^- in groundwater is negligible (Nolan and Stoner, 2000).

The data classification used in this study is based on the approach adopted by Nolan and Hitt (2006). This approach segregates the influencing parameters into three different categories based on N sources, factors influencing the transport and its attenuation in groundwater. The selections of the parameters were based, considering the various processes that influence the accumulation, transport and its attenuation.

The N loadings are represented by including the various sources considered by Nolan and Hitt (2006). The different sources are farm fertilizer, confined manure, orchards/vineyards, population density and cropland/pasture/fallow. The farm fertilizer N loading was compiled at the county level from national databases of fertilizer sales (Ruddy, Lorenz, and Mueller, 2006). The approach used to estimate the farm fertilizer loading is based on the procedure described by Nolan and Hitt (2006). The county level N loading was allocated to the land use categories comprising, orchards/vineyards/other, pasture/hay, row crops, small grains and fallow land. The loadings were determined based on a weighting factor obtained by dividing the area of the above mentioned lands in the particular county (Nolan and Hitt, 2006). The base land use data called National Land Cover Data (NLCD) at 30-m resolution was used (Vogelmann et al., 2001). The final N fertilizer application is based on an enhanced version of the land use data designated as “NLCDe”. NLCDe reclassifies the misclassified NLCD data with the aid of 1970s-1980s aerial photography data (USGS, 1990). The misclassified data pertaining to N fertilizer application resulted from orchards and residential areas with tree canopy being classified as forest. This error occurred as these are difficult to distinguish with the satellite imagery (Nakagaki and Wolock, 2005). The annual estimates of the farm fertilizer loading in kilograms per hectare applied to agricultural lands were averaged for the years 1992-2001. This approach was adopted based on the data availability at the time of the study. The use of one application rate for a county is reasonable as range of crops grown in a county is fairly limited (Nolan, Hitt, and Ruddy, 2002).

An approach similar to the farm fertilizer loading was adopted to determine the N loading from confined manure. Annual N input in kilograms per hectare from confined animal manure was averaged for the years from 1992-1997. Confined manure was applied to pasture/hay, row crops, small grains and fallow land use categories from the NLCD (Ruddy, Lorenz, and Mueller, 2006). Confined manure estimates for other years were not available at the time of this study (Nolan and Hitt, 2006). The confined manure estimates were obtained from Ruddy, Lorenz, and Mueller (2006).

The other three variables considered in the initial model building process are believed to be surrogates for additional sources of N. Percent orchards/vineyards, population density and percent cropland/pasture/fallow are believed to be surrogates to N loading. Orchards/vineyards used in this study are the percent of orchards/vineyard land cover in the conterminous United States. This data was developed by computing the percentage of the area pertaining to the orchards/vineyards in that particular 1km resolution national grid cell. The population density data was obtained from the initial dataset originating from Hitt (2007) at a resolution of 100 m for the 1990 population density. The data was resampled at a 1 km resolution and by multiplying the grid values by 0.1. The resulting dataset represents the 1990 block group population density of people per square km for the conterminous United States (Nolan and Hitt, 2006). Percent cropland/pasture/fallow was again derived using the similar procedure as orchards/vineyards.

The data for transport to the aquifer was represented by the following variables: (1) water input in km^2/cm, (2) presence or absence of carbonate rocks, (3) presence or absence of basalt and volcanic rocks, (4) drainage ditch in km^2, (5) percent slope, presence or absence of glacial till, (6) depth to water, and (7) percent clay sediment. Water input here is used in the same meaning as that of Nolan (1998). It is defined as the ratio of the total area of irrigated land to precipitation in square km per cm for the conterminous United States. The national precipitation grid was obtained from DAYMET (Thornton and Running, 1999). The presence or absence of carbonate rocks and basalt and volcanic rocks were derived by coding 1 for presence and 0 for absence. This data was developed from the principal aquifers in the National Atlas of the United States.

The data representing the drainage ditch in km^2 was developed from the National Resources Inventory surface drainage, and field ditch conservation practice in the conterminous United States. The land cover classification of the NLCDe dataset where this was applied were orchards/vineyards/other, LULC orchards/vineyards/other, pasture/hay, row crops, small grains and fallow. Each grid cell consists of the percentage of the above mentioned land cover classes. The source of the percent slope (topography) data is from STATSGO (State Soil Geographic). The average slope was computed by determining the average of the weighted value of high and low value for the range in slope expressed in percent. The attributes representing these two values are SLOPEL and SLOPEH for the low and the high value of the weighted average for the range in slope respectively. The grid was of the resolution of 1km containing these slope values. The data for the presence or absence of poorly sorted glacial till east of the Rocky Mountains in the conterminous United States was developed from the dataset "Digital representation of a map showing the thickness and character of Quaternary sediments in the glaciated United States east of the Rocky Mountains: surficial Quaternary sediments." The presences of the glacial till were coded as 1 and the absence as 0. The depth to water data was obtained from the NAWQA data warehouse along with the nitrate concentration data. The data for the percent clay sediment originated from the STATSGO dataset. A detailed description of the procedure is given in Nolan and Hitt (2006). The procedure to develop all the data is described in detail in Nolan and Hitt (2006).

The attenuation in groundwater is represented using similar parameters as mentioned in Nolan and Hitt (2006). The parameters that represent attenuation are (1) fresh surface water withdrawals, (2) areas with irrigation tail water recovery, (3) percent histosol soil types and wetlands. The data for fresh surface water withdrawal for irrigation in mega liters per day was developed from a national grid consisting of 1995 fresh surface water withdrawal for irrigation (Solley, Pierce, and Perlman, 1998). The county level data was applied to agricultural land within a county. The NLCDe 92 at 1-

km resolution was used, with land cover classification consisting of agricultural lands such as, orchards/vineyards/other, LULC orchards/vineyards/other, pasture/hay, and row crops and small grains. The area with the irrigation tail water recovery data was compiled and weighted in a similar manner as the 1992 NRI data. The factors representing the fresh surface water withdrawal for irrigation and areas with tail water recovery both represent dilution of nitrate.

The percent histosol soil type and the percentage of woody wetlands and emergent herbaceous wetlands cover in the conterminous United States both represent denitrification (Nolan and Hitt, 2006). The data on histosols was obtained from the STATSGO dataset, i.e. soils containing high organic matter content. Detailed explanation of developing this data is provided in Nolan and Hitt (2006). The percent of wetlands was defined as the sum of the percentages of the woody wetlands and emergent herbaceous wetlands from the NLCDe 92 dataset at a 1 km resolution.

Methodology

In this study, the OLR model is used to relate the probability of NO_3^- concentration with respect to background concentration and MCL, to the significant influencing variables, such as N loading, clay sediments, presence or absence of carbonate rocks, drainage ditch, and glacial till. The approach for implementing the OLR model for a successful analysis of aquifer vulnerability to NO_3^- contamination includes a number of key steps as outlined in Twarakavi and Kaluarachchi (2005), which are: (a) categorizing of response values (concentration) based on *n* threshold values such as, MCL and background concentration into *(n+1)* discrete response categories, (b) identifying all possible influencing variables, discrete and continuous, of the physical system, (c) performing univariate ordinal LR between the response and each influencing variable and selecting the significant influencing variables using the Wald statistic and chi-square test, (d) performing multivariate ordinal LR between the probability of occurrence of response with respect to the threshold values, and the significant influencing variables and checking again for the significance of the influencing variables, (e) then repeating step (d) until only the significant influencing variables are included in the model, and (f) finally checking for the goodness-of-fit of the model results. Detailed explanations of Steps (a) through (f) are discussed in further detail in the next sections.

The preparation of response data in step (a) involves assigning the NO_3^- concentration to one of the *(n+1)* response categories formed by *n* thresholds, namely background concentration and MCL. The background concentration is 2 mg/L of NO_3^- as N, and the MCL is 10 mg/L of NO_3^- as N. In this study, OLR model will be used to relate the probability of nitrate concentration to occur with respect to a concentration of 2 mg/l of NO_3^- - N, to the significant influencing variables. Based on their magnitude relative to background concentration and MCL the NO_3^- concentrations are grouped into three categories. A concentration less than or equal to the background concentration is listed under the response category 1, all concentrations between the background and MCL under the response category 2, and all concentrations greater than or equal to the MCL under the response category 3.

The distribution of the wells where NO_3^- was measured is shown in Figure 21. These wells were used in the development of the model. There are 3,770 wells that were sampled mainly during the first decade of the NAWQA program with depth to water less than or equal to 50 feet, to satisfy the condition of shallow wells. As DRASTIC model does not readily consider the confined aquifer condition lower depths would mean the unconfined condition and also a better comparison of DRASTIC and OLR. A simple regression analysis was done between the latitude and longitude to show a uniform scatter of wells. A regression coefficient of 0.04 indicated a good scatter of wells across the CONUS. Table 11 shows the descriptive statistics of NO_3^- concentration with respect to various influencing variables.

All the influencing variables that may influence the occurrence of NO_3^- in ground water are identified in step (b). A variety of influencing variables were considered in the OLR model that would influence the concentration of NO_3^-. The knowledge gathered during literature review guided in the process of determining the influencing variables. Some of the possible influencing variables that were considered are (a) depth to water table, (b) fresh surface water withdrawal, (c) histosol soil type, (d) confined manure and (e) N farm fertilizer loading, etc.

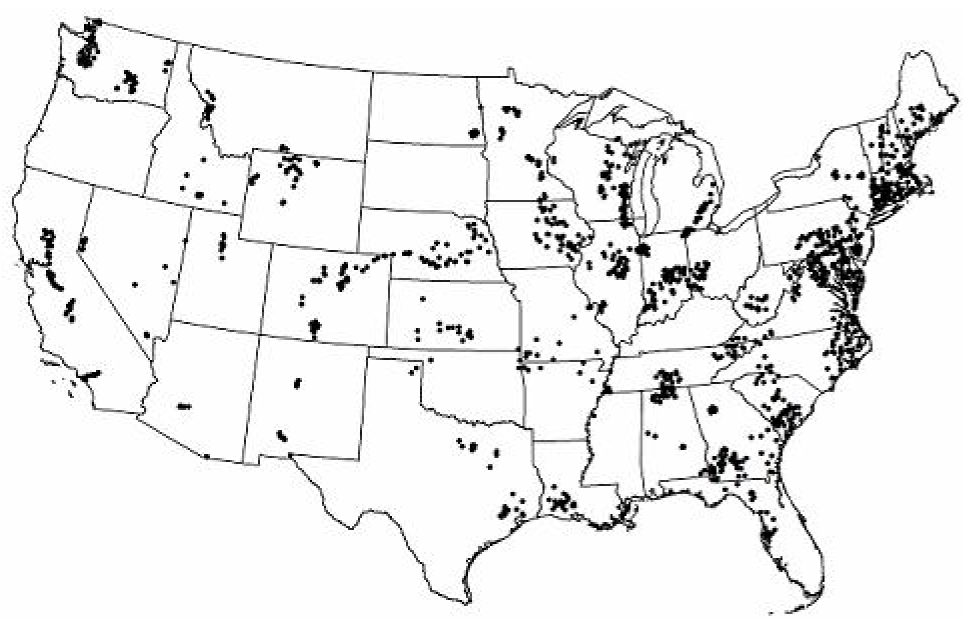

Figure 21. A map showing the distribution of shallow wells sampled during the NAWQA program used in the OLR model development.

Table 11. Influencing variables considered in the OLR analysis (approach similar to Nolan and Hitt (2006))

Variable	**10^{th} percentile**	**Median**	**90^{th} percentile**
N Sources			
Farm Fertilizer (kg/ ha)	0	18.2	81.44
Confined manure (kg/ ha)	0	3.01	24.77
Orchards/vineyards (percent)	0	0	0
Population density (people/km^2)	3.5	26.3	1002.8
Cropland/pasture/fallow (percent)	0	33	92
Transport parameters			
Water input (km^2/cm)	0	4.51E-5	9.66E-3
Carbonate rocks (binary indicator)	NA	NA	NA
Basalt and volcanic rocks (binary indicator)	NA	NA	NA
Drainage ditch (km^2)	0	0	0.052
Slope (percent)	1.0	3.18	12.21
Glacial till (binary indicator)	NA	NA	NA
Clay sediment (percent)	3.6	17.43	34.55
Depth to water (feet)	4.1	13.47	33.8
Attenuation Parameter			
Fresh surface water withdrawal (MLD)	0	0.0007	0.879
Irrigation tail water recovery (km^2)	0	0	0
Histosol soil type (percent)	0	0	4
Wetlands(percent)	0	0	15

Equations (12) to (14) show the generic equation to predict the probability of occurrence of NO_3^- concentration.

$$\begin{aligned}\log it(\Pr ob.\ \ of\ \ \ k=1\ \ or\ \ \ C \le background) = {} & \beta_1 + \beta_{FF}FF + \beta_{CM}CM + \beta_{OV}OV + \beta_{PD}D + \beta_{CPF}CPF + \\ & \beta_{WI}WI + \beta_{CR}CR + \beta_{BVR}BVR + \beta_{DD}DD + \beta_S S + \beta_{GT}GT + \beta_{CS}CS + \beta_{FSW}FSW + \\ & \beta_{ITW}ITW + \beta_{HST}HST + \beta_W W\end{aligned} \tag{12}$$

$$\begin{aligned}\log it(\Pr ob.\ \ of\ \ \ k \le 2\ \ or\ \ \ C < MCL) = {} & \beta_2 + \beta_{FF}FF + \beta_{CM}CM + \beta_{OV}OV + \beta_{PD}D + \beta_{CPF}CPF + \\ & \beta_{WI}WI + \beta_{CR}CR + \beta_{BVR}BVR + \beta_{DD}DD + \beta_S S + \beta_{GT}GT + \beta_{CS}CS + \beta_{FSW}FSW + \\ & \beta_{ITW}ITW + \beta_{HST}HST + \beta_W W\end{aligned} \tag{13}$$

$$\log it(\Pr ob.\ \ of\ \ \ k=3\ \ or\ \ \ C \ge MCL) = 1-(\Pr ob.\ \ of\ \ \ k \le 2) \tag{14}$$

where, C is the nitrate concentration in ground water; k is the response category; β_1 and β_2 are constants; β_{FF}, β_{CM}, $\beta_{OV,}$ $\beta_{PD,}$ $\beta_{CPF,}$ $\beta_{WI,}$ $\beta_{CR,}$ $\beta_{BVR,}$ $\beta_{DD,}$ $\beta_{S,}$ $\beta_{GT,}$ $\beta_{CS,}$ $\beta_{FSW,}$ $\beta_{ITW,}$ $\beta_{HST,}$ and β_W are slope coefficients representing farm fertilizer, FF, confined manure, CM, orchards/vineyards, OV, population density, PD, cropland/pasture/fallow, CPF, water input, WI, carbonate rocks, CR, basalt ad volcanic rocks, BVR, drainage ditch, DD, slope, S, glacial till, GT, clay sediment, CS, fresh surface water withdrawal, FSW, irrigation tail water recovery, ITW, histosol soil type, HST and wetlands, W, respectively. It should be noted that the OLR model represented from Equations (12) through (14) is without consideration of significance of the influencing variables. As explained earlier, the influencing variables were broadly classified into three groups, namely, sources of N, transport parameters, and attenuation parameters.

The next important step in OLR model building process is the selection of the significant influencing variables as not all variables affect the response. In Step (c) univariate OLR analysis is performed to eliminate the non-significant influencing variables. The p-value of the chi-square (χ^2) test and Wald statistic, W, is used to estimate the significance of the influencing variables of the system. The expected value of the parameter is divided by its standard error to give the Wald statistic. A p-value <0.25 and an absolute Wald statistic exceeding 2 from the univariate test is a significant influencing variable and is a candidate for the model (Hosmer and Lemeshow, 1989). The results of the univariate OLR are shown in Table 12 below.

Table 12. Statistics indicating the relative significance of influencing variables

Variable	W	p-value
N Sources		
Farm Fertilizer (kg/ ha)	6.49	0.000
Confined manure (kg/ ha)	6.18	0.000
Orchards/vineyards (percent)	3.62	0.000
Population density (people/km^2)	0.52	0.606
Cropland/pasture/fallow (percent)	4.36	0.000
Transport parameters		
Water input (km^2/cm)	5.83	0.000
Carbonate rocks (binary indicator)	2.65	0.008
Basalt and volcanic rocks (binary indicator)	0.77	0.440
Drainage ditch (km^2)	2.46	0.014
Slope (percent)	1.88	0.060
Glacial till (binary indicator)	3.94	0.000
Clay sediment (percent)	3.11	0.002
Depth	2.44	0.015
Attenuation Parameter		
Fresh surface water withdrawal (MLD)	3.03	0.002
Irrigation tail water recovery (km^2)	2.44	0.015
Histosol soil type (percent)	2.19	0.029
Wetlands(percent)	1.76	0.078

Maximum likelihood estimation approach is used to fit the multivariate ordinal LR model using the significant influencing variable. The importance of each influencing variable should again be verified by estimating the Wald statistic and comparing the estimated slope coefficients of the influencing variable from the multivariate and univariate ordinal LR analysis performed in step (c) (Hosmer and Lemeshow, 1989). In step (e) insignificant variables are eliminated and only the significant influencing variables are fitted in the new model. The significant influencing variable after the multivariate OLR analysis is shown in Table 13.

The primary source of nitrate to the groundwater is from farm fertilizer and confined manure. Its significance is demonstrated by the Wald statistic. It is then followed by confined manure. Confined manure is not as much of a concern in comparison to farm fertilizer as a contributor of nitrate to groundwater. Clay sediments can form a barrier obstructing the passage of nitrate. Carbonate rocks are very porous and consists of large cracks or spaces in between them. This forms a easy pathway for the contaminant to be transported to the groundwater. The variable drainage ditch indicates percent area of ditches in 1 km square area. Higher percentage of drainage ditches would indicate that most of the surface water is transported elsewhere. This would imply transport of nitrate to a different location.

The final ordinal LR model relates the probability of occurrence of NO_3^- concentration with respect to the MCL and background. The parameter estimates obtained from this analysis are shown in Table 14.

Table 13. Significance of major influencing variables after multivariate OLR

Variable	**Wald Statistic**
Farm fertilizer	7.05
Confined manure	4.10
Clay sediments	5.41
Carbonate rocks	2.47
Drainage ditch	3.93
Glacial till	4.20

Table 14. Excepted values of slope coefficients of influencing variables in the ordinal logistic regression analysis

Description	Coefficient
β_0, constant (probability less than 2)	0.11
β_1, constant (probability less than 10)	2.02
Farm fertilizer	-0.012
Confined manure	-0.012
Clay sediments	2.82E-5
Carbonate rocks	-1.046
Drainage ditch	2.659
Glacial till	0.687

The *goodness-of-fit* is checked in step (f) to determine the accuracy of the final model as compared to the observed data. The *measures of association approach* is used to perform the goodness-of-fit test (Hosmer and Lemeshow, 1989). A table of the number and the percentage of concordant, discordant, and tied pairs of observed and predicted probabilities is obtained from the measures of association approach. A concordant pair is formed if the predicted and corresponding observed probabilities are similar, a discordant pair if they are not similar, and a tie if it is difficult to relate them. In other words, the accuracy of the model is greater if the percentage of concordant pairs is higher. In order to check the extent of similarity between the observed and predicted probabilities graphical procedures may also be used wherever possible. This analysis is performed in the next chapter, where a comparison of both the models is done. Figure 22 shows the spatial distribution of the probability of nitrate greater than or equal to MCL. In Figures 23 and 24 the spatial distributions of the probability of nitrate less than or equal to background and less than MCL are respectively shown.

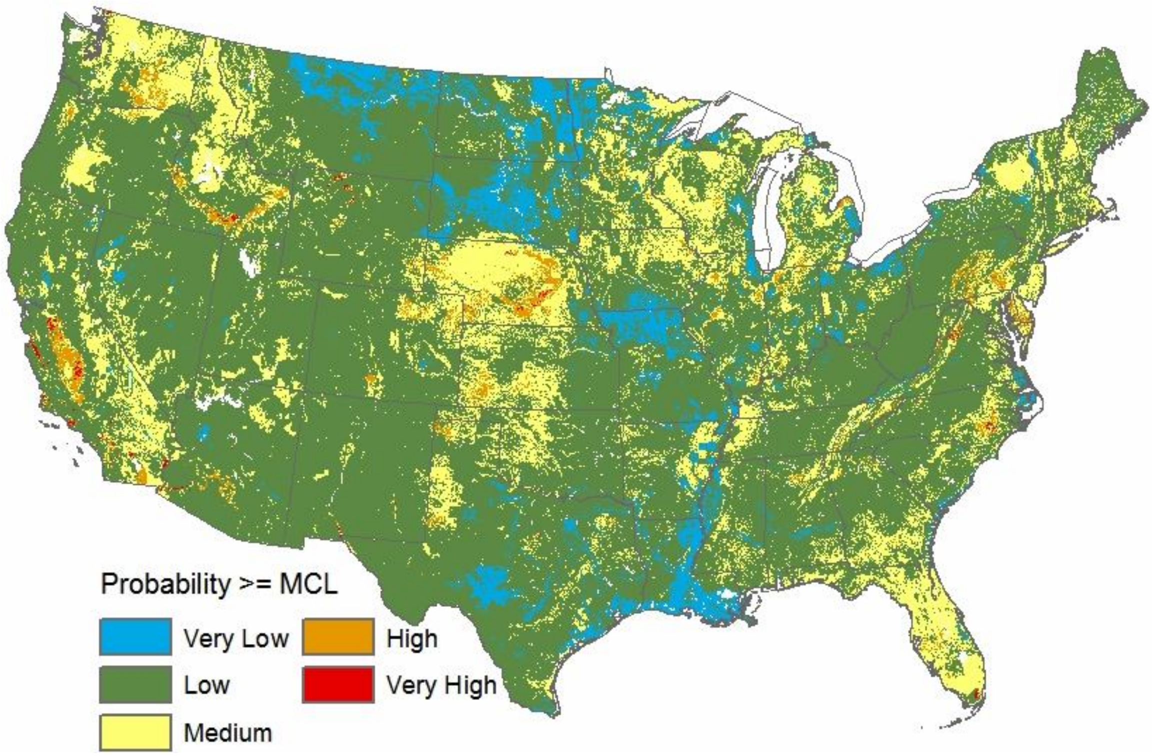

Figure 22. Spatial distribution of probability of occurrence of NO_3^- greater than or equal to MCL.

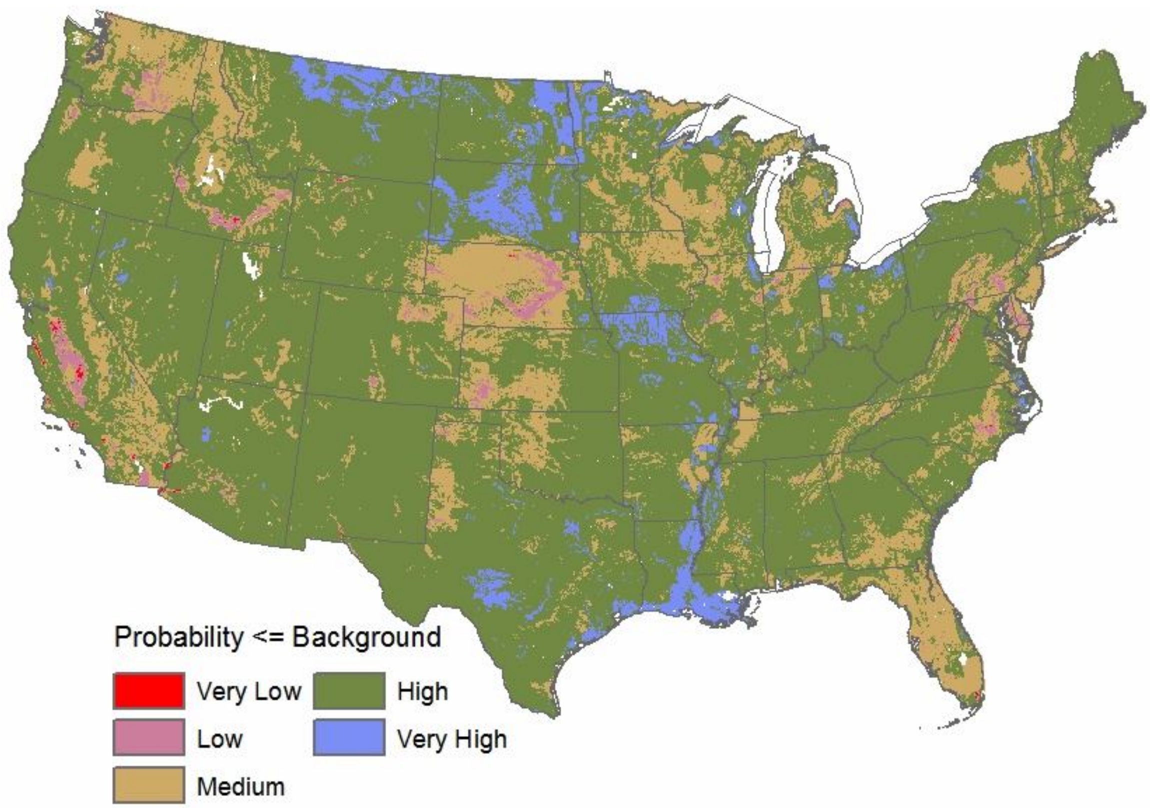

Figure 23. Spatial distribution of probability of occurrence of NO_3^- $\leq$ background.

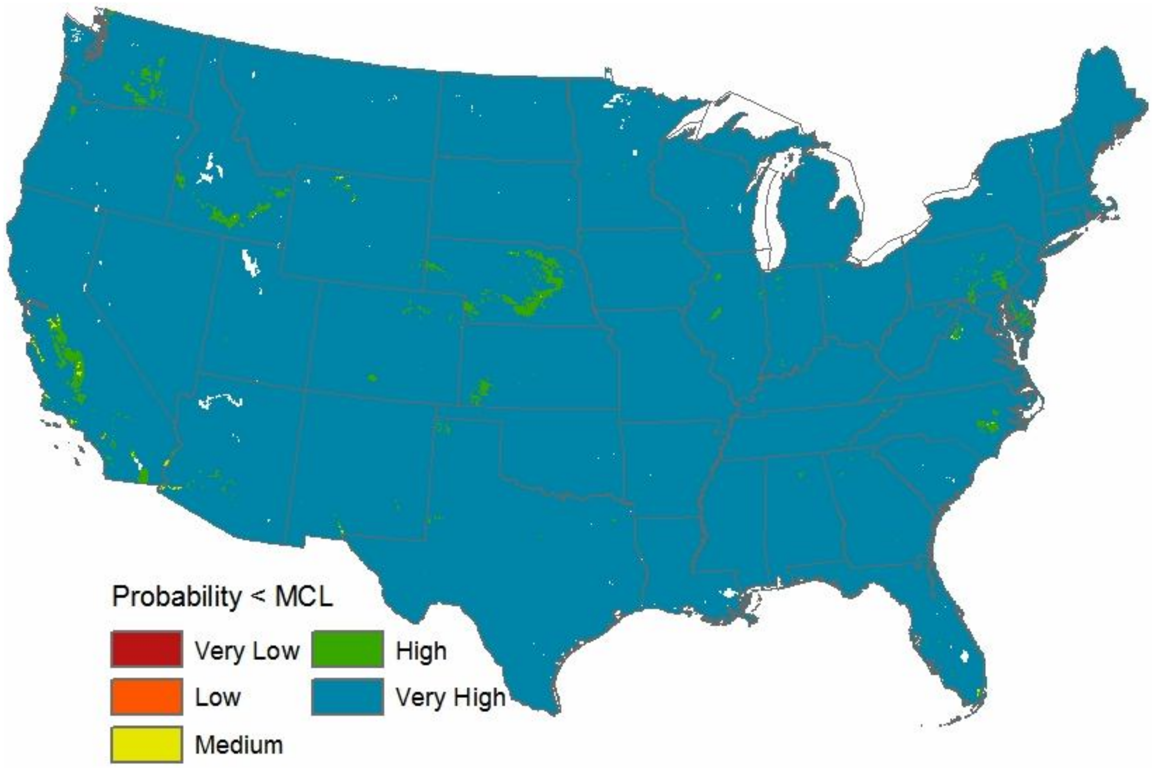

Figure 24. Spatial distribution of probability of occurrence of NO_3^- < MCL.

Results and Discussions

According to modified DRASTIC 18.2 % of the area of conterminous US lies in the greatest groundwater vulnerability region. This was based on a modified DRASTIC score of greater than or equal to 165. 50.8% in the lower vulnerability regions, based on less than or equal to 125; and 31.0% lies in the moderate regions, based on greater than 125 and less than 165.

Seventeen variables which could have an influence on the nitrate concentration were initially considered. The univariate analysis indicated that only 13 variables were significant. Further multivariate analysis results indicated that only six variables were significant. Washington, California, Nebraska, parts of Idaho, Kansas, Pennsylvania, Delaware, Virginia, and Maryland indicated pronounced problems with the OLR analysis. These states had significant areas with high probability of nitrate contamination. Utah did not show any significant contamination problems with regards to nitrate in this analysis.

The results of the OLR model indicate that depth to water is not a very significant variable for shallow water depths. This is in direct contrast to the DRASTIC approach which assigns a high weight value to depth to water table.

CHAPTER V

COMPARISION

The focus of this chapter is to analyze modified DRASTIC and OLR results by comparing them with NO_3^- concentration data. A linear regression fit is done to verify modified DRASTIC index values with concentration data. In the case of OLR, a linear regression analysis is done with the observed and average predicted probabilities yielding an R^2 value that would indicate the goodness of fit.

The assumptions of the modified DRASTIC model were considered while developing the model for the CONUS, namely, (1) confined aquifers cannot be readily modeled using the DRASTIC approach. This criterion was satisfied by considering only shallow aquifers which are predominantly unconfined. (2) Generic contaminant has the mobility of water. NO_3^- satisfies the generic contaminant assumption of DRASTIC model. As DRASTIC considers the contaminant has the mobility of water, NO_3^- would satisfy this criteria very well. In the case of any other contaminant, this would have been a serious limitation. These two conditions were considered to account for DRASTIC's inadequacies, and hence provide a better condition for evaluating the results. In the case of modified DRASTIC, the concentration values at each of the well locations were analyzed with respect to their index scores at those locations. The plot between the index and NO_3^- concentration value at 3,770 well locations are shown in Figure 25. It can be observed from the scatter plot that there is no clear trend.

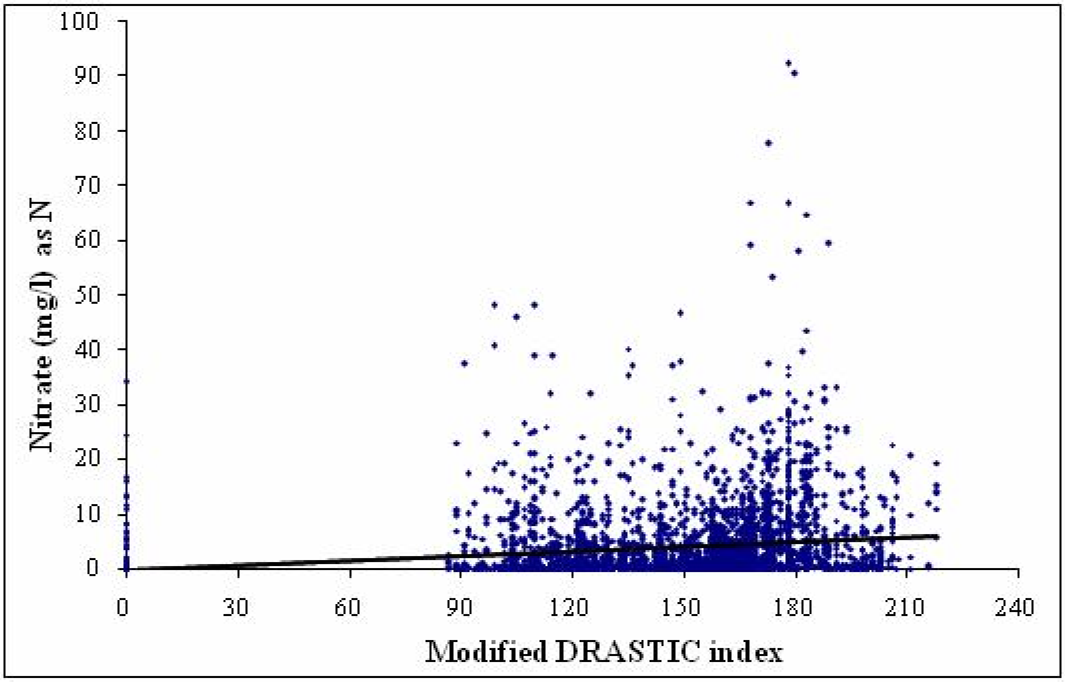

Figure 25. Plot between modified DRASTIC index and NO_3^- plus NO_2^- concentration for depth to water table up to 50 feet.

An R^2 value of 0.017 was computed for a linear regression fit. The R^2 value indicates a very poor correlation between the modified DRASTIC index and the concentration values. Several possibilities may have lead to this inadequacy in the predictability of the model. Some of the possibilities are:

(1) Fixed weights: The weights of the model do not change to accommodate any difference in the influencing parameters significance. This could be a serious limitation in attributing a higher weight to a less significant parameter or vice versa.

(2) Data resolution: The data resolution could be another significant parameter leading to less precise results. The hydraulic conductivity varies by orders of magnitude in a very small area, but in this analysis the values were averaged for larger areas. Other data was also averaged to cover larger areas.

(3) Inadequacy of DRASTIC to model confined aquifers: An important constraint with the DRASTIC model is its inefficiency to effectively address the condition of confined aquifer (Aller et al., 1987). This inadequacy to some extent can be resolved by the use of DRASTIC to model shallow depths only, which are predominantly unconfined. The data sets used to develop both the models contained shallow aquifers, thereby overcoming this limitation. Deeper groundwater is older and chances of any contamination by anthropogenic sources could be minimal. Another important reason for considering shallow groundwater is that the likelihood of encountering a less permeable layer increases with greater depth (Nolan and Hitt, 2006). The N loading factor may not be very significant influencing variable at greater depths. This may also limit the models predictability as there is no way to accommodate this parameter in the modified DRASTIC model.

A linear regression fit of observed and predicted probabilities was determined to evaluate the result of the OLR model. To perform this analysis probability values from the calibration data set were compared with probability values from the validation data set. The plot between observed and predicted probability values is shown in Figure 27. The validation dataset had 1885 wells with NO_3^- concentration values measured at depth to water values less than or equal to 50 feet. The plot showed a close match between the two sets indicating the validity of the OLR method to analyze the NO_3^- occurrence. The regression coefficient for this analysis was determined to be 0.63.

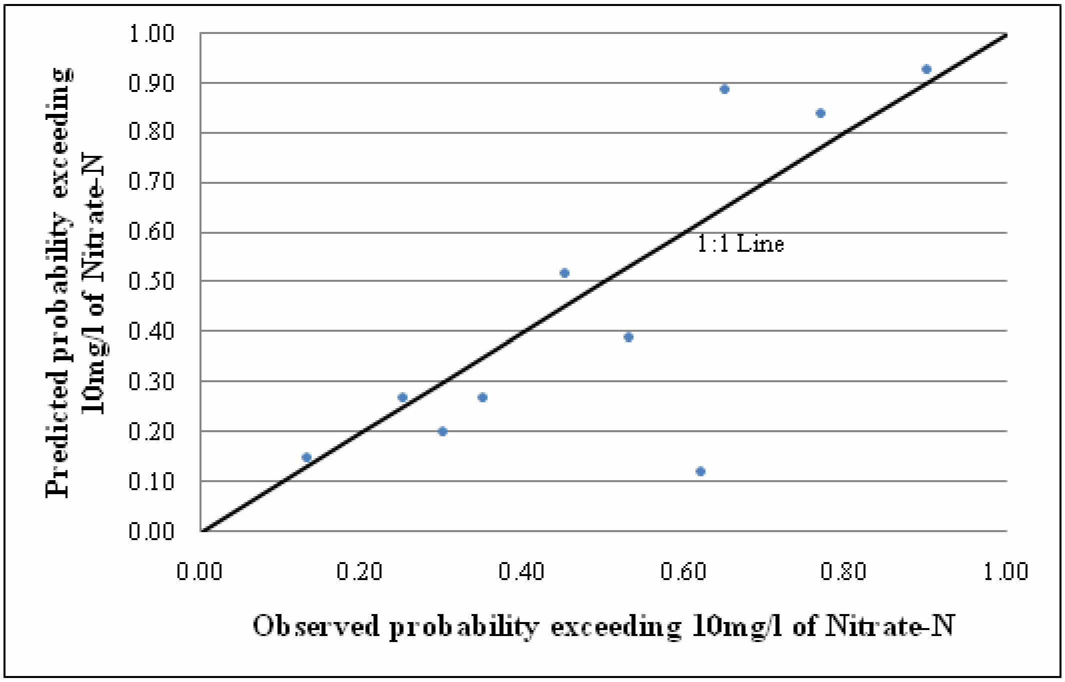

Figure 26. Observed and predicted probabilities for NO_3^- contamination in the conterminous U.S.

The use of modified DRASTIC requires considerable experience with the application of the model. There is subjectivity in the choice of the ratings which can result in different scores from user to user. With regards to the OLR model, its data requirement is flexible and can be analyzed with the available data. The best use of the DRASTIC model is when available data is in the form required by the model, and the user has good background knowledge in hydrogeology. The OLR model offers the flexibility of data and only requires the skill set needed for performing statistical analysis. The OLR analysis is scientifically more defensible in comparison with the DRASTIC approach.

Table 15 shows an area of 11.41% of high and very high vulnerability from modified DRASTIC. In contrast the OLR approach has only 1.99% in these categories. The division of different vulnerability classes was done based on equal intervals and was similar for both the approaches. A R^2 value of 0.71 between the areas under each vulnerability class by both the approaches indicates a good trend in the overall prediction by both the models. The plot is shown in Figure 27.

The performances of each of the models were analyzed at randomly selected 1000 similar well locations. The values were extracted at each of the raster cells based on the nearest value approach. The R^2 value for the OLR model was 0.10 and that of modified DRASTIC was 0.03. In general, though both R^2 values are on the lower side, the prediction capability of OLR model is much better than modified DRASTIC. The plot between groundwater nitrate concentration and modified DRASTIC and OLR models is shown in Figure 26.

Table 15. A comparison of areas representing different vulnerability classes from modified DRASTIC and Ordinal logistic regression (100% represents the whole study area)

Vulnerability Class	**Modified DRASTIC (% Area)**	**Ordinal logistic regression (% Area)**
Very low	0.00031	6.069
Low	69.15	72.46
Medium	57.73	19.48
High	11.41	1.84
Very High	0.00071	0.15

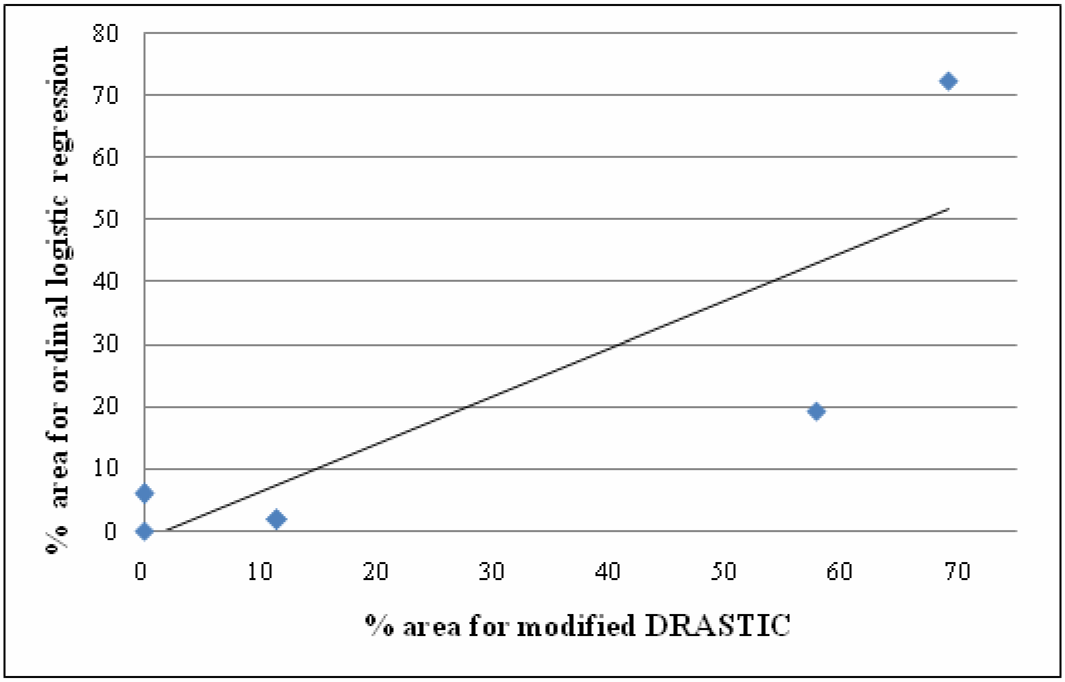

Figure 27. Percent area in each vulnerability class as predicted by both the models.

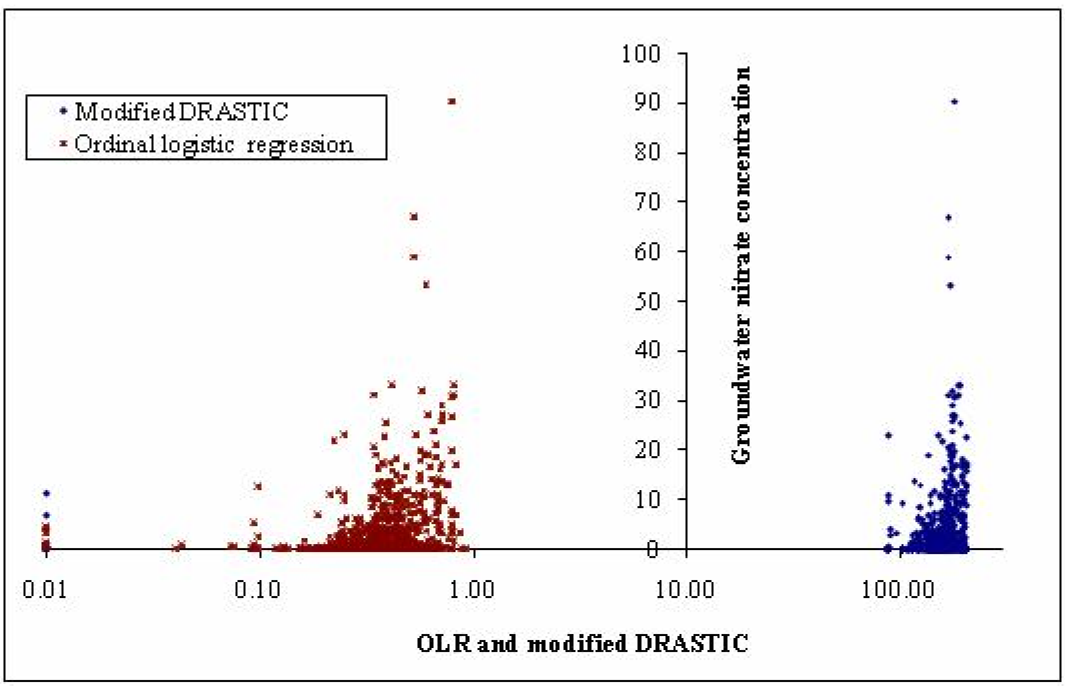

Figure 28. Groundwater nitrate concentration vs. modified DRASTIC and OLR.

CHAPTER VI

SUMMARY AND CONCLUSIONS

This research contains two parts: (1) applying two models, namely, modified DRASTIC and ordinal logistic regression across conterminous US (CONUS) to NO_3^- contamination, and (2) analyzing the results to look into the model performance with respect to various factors, such as, data requirement in a particular format, model inadequacies, level of skill required by user, interpretation of the output and use of the output to aid any policy making.

In the first part of the research, the modified DRASTIC model was applied to determine the shallow aquifer vulnerability across CONUS. The results were empirical values which did not clearly follow a particular trend, such as an increase in the NO_3^- value with respect to an increase in the score. The regression analysis yielded a very poor result. The performance of the OLR model was satisfactory as the observed and predicted probabilities were well correlated.

This study provides a comprehensive analysis of aquifer vulnerability of US to nitrate from two methods. Previous studies to evaluate the vulnerability using DRASTIC were done at scales coarser than 1:250,000. There is evidence in the literature to suggest improvement of the results with an improvement of scale. Ordinal logistic regression model was applied only to heavy metal contamination. This study extends its application to nitrate.

Benefits

Modified DRASTIC is a widely used method to determine the aquifer vulnerability of the US and around the world. The aim of this research was to compare the performance of this widely used method with a relatively new vulnerability assessment model, Ordinal logistic regression.

1. This study exposed the limitations of each of these models with respect to the other. A lower value of the regression coefficient between modified DRASTIC index and NO_3^- concentration indicates the models inability to predict accurately, and the observed and predicted probability values from the OLR analysis indicate its ability to predict better. This could possibly be due to some factors not considered in the modeling processes.
2. This study also examined the appropriate use of a model for a particular region in the US and removed the subjectivity in the choice of these two methods. For instance, there is considerable subjectivity in the ratings of the modified DRASTIC model, where data is a limitation. The ratings used in modified DRASTIC require the procedure be done according to the classification provided in the manual, thereby limiting the accuracy of results by introducing subjectivity in the rating process. In such cases, Logistic Regression model can be used.
3. The study also targets the optimal utilization of the available data for prediction purposes. Certain data might be redundant, and this study has addressed the issue. The parameter selection done using Ordinal LR allows selection of only those parameters which have a significant impact on the concentration values of nitrate and eliminates the rest.

Scope for Future Work

Data was one of the major constraints with regards to developing the modified DRASTIC model. Future work could involve including higher resolution data to verify the model's performance. Also, the application of the OLR model to other contaminants could be explored. Extending or modifying the theoretical framework of the methods, especially DRASTIC can be explored to better represent the weights.

REFERENCES

Addiscott, T. M., A. J. Gold, C. A. Oviatt, N. Benjamin, and K. E. Giller 2005. The chemistry and physics of nitrate. Nitrate, agriculture and the environment. CABI Publishing, Harpenden, U.K. 14-28.

Afifi, A. A., and V. Clark 1984. Logistic regression in computer-Aided multivariate analysis. Lifetime Learning Publications, Belmont, CA.

Aller, L., T. Bennet, J. H. Lehr, and R. J. Petty 1987. DRASTIC: A standardized system for evaluating ground water pollution potential using hydro geologic settings, U.S. EPA.

Almasari, M. N., J. J. Kaluarachchi, S. Ghabayen, A. Jarrar, M. McKee, A. Jayyousi, and A. A. 2005. Assessment of groundwater vulnerability to nitrate contamination in Gaza Strip, Palestine. EWRI Conference Alaska.

Ball, R., D. R. Keeney, P. W. Theobald and P. Nes. 1979. Nitrogen balance in urine-affected areas of a New Zealand pasture. Agronomy Journal 71: 309-314.

Battaglin, W. A., and D. A. Goolsby. 1994. Estimates of nitrogen-fertilizer sales for the conterminous United States in 1990. Lakewood, CO, USGS < http://water.usgs.gov/GIS/metadata/usgswrd/XML/nit90.xml>.

Bedient, P. B., H. S. Rifai, and C. J. Newell. 1999. Groundwater contamination-transport and remediation. Prentice Hall, Englewood Cliff, New Jersey. 604 p.

Boyce, J. S., J. Muir, A. P. Edwards, E. C. Seim, and R. A. Olson. 1976. Geologic nitrogen in Pleistocene loess of Nebraska. Journal of Environmental Quality 5: 93-96.

Cast 1985. Agriculture and groundwater quality. Council for Agricultural Science and Technology Report 103: 62.

Centers for Disease Control and Prevention. 1996. Morbidity and mortality weekly report 45: 569-572.

Duwig, C., T. Becquer, L. Charlet, and B. E. Clothier. 2003. Estimation of nitrate retention in a Ferrosol by a transient-flow method. Europeon Journal of Soil science 54: 505-515.

Fine, D. H. 1982. Endogenous synthesis of volatile nitrosamines: Model calculations and risk assessment. IARC Science Publication 41: 379-396.

Fishman, M. J. 1993. Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory: Determination of inorganic and organic constituents in water and fluvial sediments, U.S. Geological Survey Open-File Report 93-125.

Foster, S. S. D. 1987. Fundamental concepts in aquifer vulnerability, pollution risk and protection strategy. Proceedings of International Conference, Vulnerability of Soil and Groundwater to Pollutants, Noordwijk, The Netherlands.

Foster, S. S. D., and R. Hirata 1988. Groundwater pollution risk assessment: A methodology using available data. WHO-PAHO/HPE-CEPIS technical manual. Lima, Peru. 81 p.

Freeze, R. A., and J. A. Cherry 1979. Groundwater. Prentice Hall, Englewood Cliffs, New Jersey. 604 p.

Gogu, R. C., V. Hallet, and A. Dassargues 2003. Comparison of aquifer vulnerability assessment techniques. Application to the Neblon river basin (Belgium). Environmental Geology 44(8): 881-892.

Haertle, A. 1983. Method of working and employment of EDP during the preparation of groundwater vulnerability maps. IAHS Publication 142: 1073-1085.

Hamilton, P. A. 1994. NAWQA retrospective database for nutrients in ground water and surface water. U.S. Geological Survey. http://water.usgs.gov/nawqa/nutrients/datasets/retrodesc.html

Helsel, D. R., and R. M. Hirsch 1992. Statistical methods in water resources. Elsevier Publishers, New York, New York.529 p.

Hem, J. D. 1989. Study and interpretation of the chemical characteristics of natural water. Water-Supply Paper, U.S. Geological Survey: 2254.

Hitt, K. J. 2007. Vulnerability of shallow ground water and drinking-water wells to nitrate in the United States: Model of predicted nitrate concentration in shallow, recently recharged ground water -- Input data set for population density. U.S. Geological Survey. Reston, Virginia. http://water.usgs.gov/GIS/metadata/usgswrd/XML/gwava-dw_popd.xml

Hosmer, D. W., and S. Lemeshow 1989. Applied logistic regression. John Wiley and Sons. New York. 373 p.

Hubbard, R. K., and J. M. Sheridan 1989. Nitrate movement to groundwater in the southeastern Coastal plain. Journal of Soil and Water Conservation 44(1): 20-27.

Jaffe, E. R. 1981. Methaemoglobinemia. Clinical Haeatol 10: 99-122.

Keeney, D. R. 1980. Prediction of soil nitrogen availability in forest ecosystems: A literature review. Forest Science 26: 159-171.

Keeney, D. R. 1986. Sources of nitrate to groundwater. Critical Review in Environmental Control 16(3): 257-304.

Keeney, D. R. 1989. Sources of nitrate to groundwater 21: 23-34. In: R. F. Follett (ed.) Nitrogen management and ground water protection. Elsevier Science Publishers, New York.

Kleinbaum, D. G. 1994. Logistic regression - a self-learning text. Springer-Verlag, New York.

Koterba, M. T., F. D. Wilde, and W. W. Lapham 1995. Ground-water data collection protocols and procedures for the National Water- Quality Assessment Program: collection and documentation of water-quality samples and related data. U.S. Geological Survey**:** Open-File Report 95-399.

Kreitler, C. W. and D. C. Jones 1975. Natural soil nitrate: The cause of the nitrate contamination of groundwater in the Runnels County, Texas. Ground Water 15: 53-58.

Madison, R. J. and J. O. Brunett 1985. Overview of the occurrence of nitrate in groundwater of the United States. Water Supply Paper 2275. U.S. Geological Survey.

McCullagh, P. 1980. Regression models for ordinal data. Journal of the Royal Statistical Society Series B (Methodological) 42(2): 109-142.

Mueller, D. K., and D. R. Helsel 1996. Nutrients in the nation's waters too much of a good thing? U.S. Geological Survey Circular.

Nakagaki, N., and Wolock, D. M. 2005. Estimation of agricultural pesticide use in drainage basins using land cover maps and county pesticide data. U.S. Geological Survey Open-File Report 2005-1188.

Nolan, B. T. 1998. Modeling approaches for assessing risk of non-point contamination of ground water: U.S. Geological Survey Open-File Report 98-531, 22 p.

Nolan, B. T., and K. J. Hitt 2006. Vulnerability of shallow groundwater and drinking-water wells to nitrate in the United States. Environmental Science Technology 40(24): 7834 - 7840.

Nolan, B. T., K. J. Hitt, and B. C. Ruddy 2002. Probability of nitrate contamination of recently recharged ground waters in the conterminous United States. Environmental Science and Technology 36(10): 2138-2145.

Nolan, B. T., and J. D. Stoner. 2000. Nutrients in ground waters of the conterminous United States, 1992-1995. Environmental Science and Technology 34(7): 1156-1165.

N. R. C. 1993. Groundwater vulnerability assessment. Predicting relative contamination potential under conditions of uncertainty. National Academy Press, Washington D.C. 224 p.

Omernik, J. M. 1976. The influence of land use on stream nutrient levels. Corvallis, Oregon, Environmental Research Laboratory, U. S. Environmental Protection Agency**:** EPA-600/3-76-014.

Ruddy, B. C., D. L. Lorenz, and D. K. Mueller. 2006. County level estimates of nutrient inputs to the land surface of the conterminous United States, 1981-2001, U.S. Geological Survey Scientific Investigations Report 2006-5012.

Secunda, S., M. L. Collin, and A. J. Melloul. 1998. Groundwater vulnerability assessment using a composite model combining DRASTIC with extensive agricultural land use in Israel's Sharon region. Journal of Environmental Management 54(1): 39-57.

Service., S. C. 1994. State Soil Geographic (STATSGO) Data Base for the United States and Puerto Rico. Ft. Worth, Texas, U.S. Department of Agriculture, Soil Conservation Service.

Shuval, H. I., and N. Gruner. 1972. Epidemiology and toxicological aspects of nitrates and nitrites in the environment. American Journal of Public Health(62): 1045-1052.

Solley, W. B., R. R. Pierce, and H. A. Perlman. 1998. Estimated use of water in the United States in 1995. U. S. Geological Survey Circular 1200.

Spalding, R. F., and M. E. Exner. 1993. Occurrence of nitrate in groundwater-A review. Journal Environmental Quality 22(3): 392-402.

Strathouse, S. M., G. Sposito, P. J. Sullivan, and L. J. Lund. 1980. Geologic nitrogen: A potential geochemical hazard in the San Joaquin valley, California. Journal of Environmental Quality 9: 54-60.

Super, M., H. Heese, D. MacKenzie, and W. Dempster. 1981. An epidemiological study of well water nitrates in a group of south west African/Namibian infants. Water Resources 15: 1265-1270.

Thornton, P.E., and S.W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. Agriculture and Forest Meteorology 93:211-228.

Twarakavi, N. K. C., and J. J. Kaluarachchi 2005. Aquifer vulnerability assessment to heavy metals using ordinal logistic regression. Ground Water 43(2): 200-214.

U.S. Environmental Protection Agency. 1995. Drinking water regulations and health advisories. Office of Water. U.S. Government Printing Office, Washington, D.C. 18 p.

U.S. Environmental Protection Agency. 1990. Citizen's guide to ground-water protection. Office of water, U. S. EPA 440/6-90-004, 33 p.

U.S. Geological Survey. 1990. Land use land cover digital data from 1: 250,000- and 1: 100,000-scale maps: data users guide. U.S. Geological Survey, Reston, Virginia. 4, 33 p.

U.S. Geological Survey. 2003. Principal aquifers of the 48 conterminous United States, Hawaii, Puerto Rico, and the U.S. Virgin Islands. U.S. Geological Survey, Madison, Wisconsin. http://nationalatlas.gov/mld/aquifrp.html

Van Duijvenbooden, W., and A. J. C. M. Matthijsen. 1987. Basis document Nitraat. National Institute of Public Health and Environmental Hygiene. Bilthoven, The Netherlands. RIVM Report No. 758473012.

Vogelmann, J. E., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and N. VanDriel. 2001. Completion of the 1990's national land cover dataset for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. Photogram, Engineering Remote Sensing, 67: 650-662.

Ward, M. H., S. D. Mark, K. P. Cantor, D. D. Weisenburger, A. Correa-Villaseñor, and S. H. Zahm. 1996. Drinking water nitrate and the risk of non-Hodgkins lymphoma. Epidemiology 7: 465-471.

Wolock, D. M. 2003. Estimated mean annual natural ground-water recharge in the conterminous United States. U.S. Geological Survey Open-File Report, 03-311.

Wong, M. T. F., A. Wild, and A. S. R. Juo. 1987. Retarded leaching of nitrate measured in monolith lysimeters in southeast Nigeria. Journal of Soil Science 38(3): 511-518.