Chapter 1: Introduction

The objective of the study presented in this dissertation is to develop a methodology for assessing the vulnerability of groundwater to contamination by agricultural chemicals. Federal water quality regulations have created a need for such assessments, but do not specify the methods to be used, or rigorously define groundwater vulnerability. The present work advocates a statistical approach to vulnerability assessment, and, in keeping with that approach, suggests that *probability of contamination*, a quantity that can be expressed numerically, be used as a surrogate for *vulnerability*, which remains a rather nebulous and unquantifiable commodity.

In this work, the words *susceptibility*, *vulnerability*, and *risk* represent related, but distinct, ideas. A groundwater supply is said to be *susceptible* to contamination if it is possible for a contaminant to reach it, even if no source exists for that contaminant. The supply is *vulnerable* to a particular contaminant if it is susceptible and a source of the contaminant is present. The *risk* of contamination is the likelihood or probability that the contaminant is actually present in the groundwater. Risk, unlike susceptibility and vulnerability can be described by a number. In other words, risk is quantifiable, while susceptibility and vulnerability are not.

In addition to risk of contamination, there are other risks associated with groundwater quality: risk of human exposure to the contaminant, risk of adverse public health effects, and so on. Although these risks are important to the formation of public policy, they lie beyond the scope of this study, which is concerned solely with the likelihood that a contaminant is present in a groundwater supply.

Although risk of contamination is quantifiable, it is not measurable. Water quality measurements describe the degree to which chemical constituents are present in water—their concentration—not risk or probability. How, then, is it possible to conduct a *statistical* investigation of groundwater susceptibility or vulnerability, which cannot be quantified, or of risk of groundwater contamination, which cannot be measured?

This work proposes that an answer to this question lies in the following postulate: For any body of groundwater and any chemical constituent, there exists a probability distribution function, $P(C_t)$, describing the likelihood that a sample, chosen at random from that body, will contain a concentration of the constituent greater than a threshold concentration, C_t. While this concentration probability distribution is not identical to risk of contamination, susceptibility, or vulnerability, it is closely related to all three, and is both quantifiable and measurable, to the extent that its parameters can be estimated from measurements of concentration taken from the groundwater body.

A body of groundwater contains an infinite number of potential water samples—a *population*, in statistical argot—the concentration probability distribution $P(C_t)$, describes that population. Actual measurements of constituent concentrations in this body of groundwater make up a *sample* of that population. Properties (called *parameters*) of the concentration probability distribution can be estimated by calculations performed on the sample. These estimates are called *statistics*. The methodology advocated here uses statistics calculated from groups of groundwater quality (i.e. constituent concentration) measurements as surrogates for risk of contamination, which cannot be measured, and for susceptibility and vulnerability, which cannot be quantified.

The particular results presented here form a spatial and statistical study of the presence of nitrate in groundwater in Texas. This work analyses nitrate measurements collected throughout the State from 1962 to 1993 and recorded in the Texas Water Development Board's Ground Water Data System (Nordstrom and Quincy, 1992), and uses statistical methods in conjunction with a geographic information system (GIS) and a relational database management system to organize the data and form conclusions.

Although the present work was directed toward the vulnerability of groundwater to agricultural chemicals, of which nitrate is a widely measured representative, the methods developed in the course of this study are not specific to agricultural chemicals or to groundwater. The same approach could easily be applied to industrial contamination of air, or any number of other forms of pollution.

1.1 MOTIVATION

This impetus for this study comes from the National Primary Drinking Water Regulations (40 CFR 141), which took effect in January 1993. These regulations implement provisions of the revised Safe Drinking Water Act by listing 60 maximum contaminant levels (MCLs) for constituents that must be monitored by operators of public water supplies, and imposing schedules for monitoring those constituents. Earlier regulations listed only 34 MCLs, so the costs of monitoring have increased significantly, especially since most of the additional MCLs are for organic chemicals such as industrial solvents, like toluene and trichloroethylene, and pesticides, like atrazine and alachlor, which require more expensive analytical methods than inorganic or nutrient constituents. To reduce the financial burden on regulators and water systems, the regulations allow the State agencies responsible for enforcement of the Safe Drinking Water Act to waive some monitoring requirements for a number constituents, including several agricultural chemicals, in water systems that have been shown, over several monitoring cycles, to be free from contamination from those constituents.

Waivers may also be granted to systems that have been shown, through a vulnerability assessment, to be secure from contamination. The choice of vulnerability assessment method is left to the State, subject to approval by the Environmental Protection Agency (EPA), but must include either sufficient knowledge of previous use of the constituent in regions contributing to the water supply that the State can be sure that no source of the constituent is present, or evidence that the water supply is protected by soil or geological conditions, and the structure of the well.

The Water Utilities Division of the Texas Natural Resource Conservation Commission is responsible for enforcement of the National Primary Drinking Water Regulations in Texas. The Water Utilities Division is engaged in an ongoing effort to record the locations and descriptions of public water supply wells in Texas in a geographic information system (GIS), in part to facilitate the analysis of wells and their surroundings for the purpose of granting monitoring waivers.

The original purpose of this study was to devise an automated system for vulnerability analysis using the Water Utilities Division's GIS data.

It soon became apparent that the data that was available in Statewide GIS coverages and databases was not adequate to form the basis of a vulnerability assessment system. In particular, hydrogeologic information such as aquifer composition, degree of confinement, and groundwater flow direction do not exist in GIS form for the State as a whole. In the absence of such data, the study focused on evaluating the usefulness of the data that *is* available in GIS for predicting groundwater vulnerability, and developing a method for deriving a statistically based groundwater vulnerability assessment method from existing groundwater quality measurements.

Concentration Thresholds. Laws like the Safe Drinking Water Act and related regulations set thresholds to trigger regulatory action, so the likelihood of exceeding thresholds is of more practical value as an measure of vulnerability than other statistical measures such as average concentrations. This study explores the use of exceedence probability as a measure of vulnerability.

Databases and Geographic Information Systems. Data management technologies, such as GIS, will play an increasingly large role in forming environmental policy and EPA has identified GIS as an important technology for groundwater protection. The "Ground-Water Data Management Summary and Recommendations" chapter of the 1991 final report of EPA's Ground-Water Task Force states that

> GIS is an emerging tool for cross-media planning and integrated environmental management, and base program activities such as permitting, inspection, and enforcement. In addition, it is particularly useful in risk-based priority setting of Regional program commitments and resource requirements. GIS has been found to be increasingly useful in program planning and priority setting activities, once the investment in area-specific mapping has been accomplished. As EPA begins using GIS in its decision making, it is also important to begin promoting the use of GIS by the State's [*sic*] in their decision making process. (USEPA, 1991)

Data Stockpiles. Government agencies have collected and stored huge amounts of environmental data. GIS and database management systems offer a means for manipulating and analyzing this data *en masse*. This study attempts to address questions like "What benefits do this mass of data offer?" "What additional value does GIS give to that data?" "What are the shortcomings of publicly available data sets, and how can they be improved for future use?"

Spatial Patterns of Nitrate in Texas Groundwater. Figure 1.1 illustrates some of the essential points of the methods developed in this study. The figure shows three maps of the 254 counties of Texas. In each map, the counties are collected in groups containing one-fifth (20% or 51) of the counties, based on the level of a nitrate-related value defined for all counties. For , for example, a well is considered "vulnerable" if the Texas Water Development Board's groundwater database shows that a nitrate concentration in excess of the MCL of 10 mg/l has been detected there. The counties are ranked by the proportion of vulnerable wells to the total number of wells listed for that county in the database. The 50 counties with the highest proportion of vulnerable wells are shaded red. The next-highest 51 counties are shaded orange, and so on. The resulting ranking of the counties can be used as an estimate of the relative vulnerability of the groundwater supplies in those counties to contamination by nitrates.

This estimate of vulnerability can then be compared to a candidate indicator, such as nitrogen fertilizer sales, to test the value of that candidate for predicting groundwater vulnerability.

The figures rank the counties according to:

Figure 1a: The proportion of wells where nitrate concentrations above 10 mg/l (as Nitrogen) have been detected

Figure 1b: The proportion of wells where nitrate concentrations above 1 mg/l (as Nitrogen) have been detected

Figure 1c: Nitrogen fertilizer sales during the years 1986–1991

The data sources for the three maps are described in Chapter 3.

The figures show some clear patterns, some of which run counter to intuition. A striking contrast can be seen between the fertilizer sales and the appearance of nitrate in groundwater. The belt of high fertilizer sales in east Texas does not appear to create a corresponding high level of nitrate in groundwater. In fact, the region with the highest rate of nitrate concentrations in excess of 10 mg/l (the MCL for nitrate) lies northwest of Dallas, spatially separate from the regions of highest fertilizer use. Fertilizer sales figures appear to have relatively little value as an indicator of the likelihood of finding groundwater nitrate concentrations in excess of either of the two thresholds considered in Figure 1.1. The figure does, however, show large-scale regional variation in frequency of elevated nitrate concentrations and different patterns, which suggests that data with coarse spatial resolution can have some value as indicators.

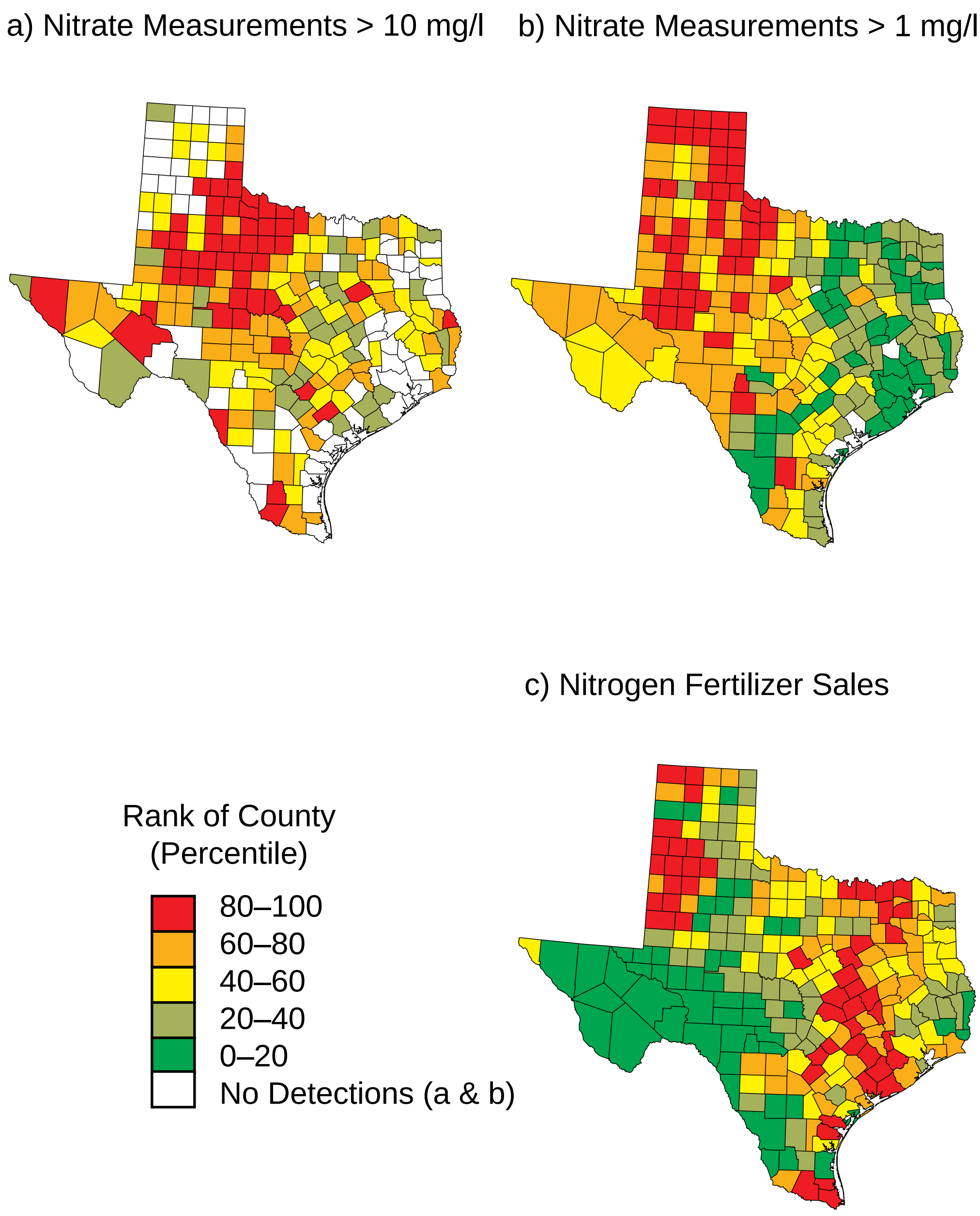

Figure 1.1 Nitrate-Related Ranking of Texas Counties

1.2 OBJECTIVES

At the time this research was proposed, the intended objective was to develop an automated system for granting vulnerability-based waivers for water quality monitoring under the National Primary Drinking Water Regulations. Because those waivers require a high degree of certainty in identifying regions that are *not* vulnerable, and because of a lack of statewide geologic data in GIS form, this goal was found to be impractical. The focus of the study shifted to improving vulnerability assessment methods using available data.

The objectives of this study are:

1) To formulate a spatially variable statistical model capable of representing in a compact form the information contained in tens of thousands of water-quality measurements spread over an area the size of Texas (691,000 km^2).
2) To apply this model in identifying spatial patterns of nitrate detection in Texas as a whole and in five selected major aquifers.
3) To estimate the relative importance of a small number of indicators—soil conditions, precipitation rates, fertilizer sales—in predicting the likelihood of contamination of groundwater by agricultural chemicals.
4) To evaluate the usefulness of a geographic information system and a database management system in carrying out an empirical study based on historic data.
5) To evaluate the usefulness of publicly available, computerized environmental data for estimating the vulnerability of groundwater to contamination.

1.3 SCOPE OF STUDY

The following limitations define the scope of this study.

1) The analysis of nitrate concentrations is restricted to data in the Texas Water Development Board's Groundwater Data System. This limits the study area to Texas and provides a single, consistent source for well descriptions and nitrate concentration data.

2) All the data used in this study comes from databases maintained by government agencies and available on a Statewide basis. This excludes, for example, data collected for studies of single aquifers or groundwater systems, unless they have been incorporated into Statewide databases. For example, maps of dominant groundwater flow direction, which exist for some aquifers, are not used because this data is unavailable over most of the State.

1.4 PROJECT SUMMARY

The study can be divided into the following three major steps.

1) Define bodies of groundwater for this study and sort measurements of water quality by their association with these bodies. Two types of definition are used. The first, based purely on location, divides Texas into seven-and-a-half minute (7.5') quadrangles, and defines a distinct body of water for each quadrangle. A measurement is associated with a given quadrangle if the well from which it was collected is located in that quadrangle. The second set of groundwater bodies is composed of five aquifers selected from the Texas Water Development Board's map of major and minor

aquifers (Ashworth and Flores, 1991). A measurement is associated with a given aquifer if the well from which it was collected draws water from a hydrogeologic unit associated with that aquifer.

2) Calculate statistical estimates of nitrate concentration probability distributions associated with the bodies of groundwater. Both discrete probabilities (estimates of the probability that various nitrate levels will be exceeded) and continuous distributions (estimates of the parameters of a probability density function) are calculated for the groundwater bodies identified in step 1.

3) Relate the statistics calculated in step 2 to indicator variables. Potential indicators of water quality considered in this study are: average annual precipitation, average soil thickness, average soil organic content, and estimated nitrate fertilizer application rate. These indicators were chosen as candidates because they were readily available, and could be plausibly linked to the degree of vulnerability of the groundwater in a region. The variation in the chosen indicators will be compared with the variation in the statistics using stepwise linear regression.

1.5 CONTRIBUTIONS OF STUDY

The study makes the following contributions to knowledge and understanding of groundwater vulnerability analysis:

1) The formulation of a spatially varying statistical mode from which exceedence probabilities (estimates of the likelihood that a constituent

will be found in concentrations exceeding a selected threshold value) can be calculated as a quantifiable measure of groundwater vulnerability.

2) The development of a quantitative, statistical method for assessing the relative value of indicators of groundwater vulnerability, and a demonstration of this method with a small number of potential indicators.

3) Application of the above to a large body of data drawn from a diversity of hydrologic and geologic settings.

4) Insight into the variation of groundwater vulnerability in Texas, and the factors that influence that vulnerability.

1.6 OUTLINE OF DISSERTATION

This dissertation consists of seven chapters.

Chapter One, this chapter, sets out the motivation, goals, scope and plan of the research project.

Chapter Two, Literature Review, summarizes the existing state of knowledge about the problems of groundwater vulnerability analysis, with particular emphasis on statistical and empirical approaches.

Chapter Three, Data Sources and Description, describes the data that are analyzed in the research, where they came from and how they were manipulated to support the needs of the research.

Chapter Four, Methods, describes the mathematical models and methods that were employed in the research. The emphasis in this chapter falls on the theoretical and mathematical aspects of the research.

Chapter Five, Procedures, focuses on the details of carrying out the analyses. This "how to" chapter describes the computer programs that actually carried out the mathematics described in Chapter Four.

Chapter Six, Results, presents maps, tables, and summary statistics that describe the distribution of nitrate in Texas, its relation to indicator variables, and the relationship between nitrate distribution and the occurrence of pesticides in groundwater in the midwestern United States.

Chapter Seven, Conclusions, finishes the dissertation by offering a summary of the completed project and the meaning of the results presented in Chapter Six.