Chapter 2: Background and Literature Review

The purpose of this chapter is to set the present study in the context of other studies of groundwater vulnerability. Since this study employs a statistical approach to vulnerability assessment, the literature review emphasizes those studies that have applied statistical methods to this problem. In addition, the use of nitrate as an indicator of vulnerability to contamination by agricultural chemicals is discussed.

This chapter addresses the following questions:

- What uses are there for groundwater vulnerability analysis?
- What methods are used for groundwater vulnerability analysis?
- Why use a statistical approach?
- How have statistical methods been applied to groundwater vulnerability analysis?
- What does the occurrence of nitrate indicate about agricultural contaminants?
- How does the method used in the present study differ from previous statistical approaches?

2.1 Uses for Groundwater Vulnerability Assessment

A groundwater vulnerability analysis identifies regions where groundwater is likely to become contaminated as a result of human activities. The objective of vulnerability analyses is to direct regulatory, monitoring, educational, and policy development efforts to those areas where they are most needed for the protection of groundwater quality. Fundamentally, this is an economic goal, rather than a scientific one. Vulnerability analysis should provide an answer to the question "Where should groundwater protection efforts be directed to return the most environmental and public health benefits for the least cost?"

In its 1991 final report, EPA's Ground-Water Task Force states as part of its "Ground Water Protection Principals" that "Efforts to protect ground water must also consider the use, value, and *vulnerability* of the resource, as well as social and economic values." (USEPA, 1991, emphasis added). The report goes on to list consideration of groundwater resource vulnerability as part of a "mature" method for setting priorities for groundwater protection. As an example of State efforts EPA regional offices should use as indicators while evaluating progress in the implementation of State Ground Water Protection Plans, the report cites development of

> a comprehensive State vulnerability assessment effort that can assist in developing State Pesticide Management Plans; targeting mitigation measures under State Nonpoint Source Management Plans; and prioritizing ground-water areas for geographically-targeted education; permitting; enforcement and clean up efforts across all ground-water related programs.

Two specific examples of EPA's intended use of groundwater vulnerability analysis are the existing regulations defining National Primary Drinking Water Standards, and the proposed differential protection strategy for imposing more restrictions on pesticide use where groundwater is vulnerable.

The first example was discussed in Chapter 1. The second example, EPA's proposed differential protection strategy for pesticides, is summarized as follows

> Under the new strategy of differential protection, if EPA determines that a pesticide poses a significant human health or environmental risk (because it may leach to groundwater) and the risk cannot be dealt with by labeling or national restricted use provisions, a state management plan (SMP) will be required for the sale and use of the pesticide in a state. The plan must describe how the risks will be addressed. As part of these plans, states will target specific areas, distinguishing those locales that warrant enhanced protection from those that merit less attention because of the lower value of the groundwater and/or their lower vulnerability to groundwater contamination. (GAO, 1992)

The National Research Council (NRC, 1993) has identified four general categories for the use of groundwater vulnerability analysis. These are: policy analysis and development, program management, to inform land use decisions, and to improve general education and awareness of a regions hydrologic resources. Judging by EPA's regulatory actions and stated groundwater protection strategy, by the publication of the NRC report, and by the results of a General Accounting Office survey (GAO 1992) stating that 42 of 45 responding states had conducted some form of groundwater vulnerability analysis, it is reasonable to conclude that groundwater vulnerability analyses are going to play some role in public policy on groundwater quality, and that methods for improving them should be studied.

2.2 GROUNDWATER VULNERABILITY ASSESSMENT METHODS

Comprehensive reviews of groundwater vulnerability assessment methods are presented in reports by the General Accounting Office (GAO, 1992) and the National Research Council (NRC, 1993). Both reports divide groundwater vulnerability assessment methods into three categories: (1) overlay and index methods, (2) methods employing process-based simulation models, and (3) statistical models. The same categories will be applied here.

Overlay and Index Methods. Overlay and index methods (the GAO report calls these "parameter weighting" methods), combine maps of parameters considered to be influential in contaminant transport. Each parameter has a range of possible values, indicating the degree to which that parameter protects or leaves vulnerable the groundwater in a region. Depth to the groundwater, for example, appears in many such systems, with shallow water considered more vulnerable than deep.

The simplest overlay systems identify areas where parameters indicating vulnerability coincide, e.g. shallow groundwater and sandy soils. More sophisticated systems assign numerical scores based on several parameters. The most popular of theses methods, DRASTIC (Aller, et al. 1987) uses a scoring system based on seven hydrogeologic characteristics of a region.

The acronym DRASTIC stands for the parameters included in the method: Depth to groundwater, Recharge rate, Aquifer media, Soil media, Impact of vadose zone media, and hydraulic Conductivity of the aquifer. DRASTIC is applied by identifying mappable units, called hydrogeologic settings, in which all seven parameters have nearly constant values. Each parameter in a hydrogeologic setting is assigned a numerical rating from 0–10 (0 meaning low risk; 10 meaning high risk) which is multiplied by a weighting factor varying from 1–5. Two sets of weights, one for general vulnerability, another for vulnerability to pesticides can be used. A score for the setting is calculated as the sum of the seven products. DRASTIC scores are roughly analogous to the likelihood that contaminants released in a region will reach ground water, higher scores implying higher likelihood of contamination. DRASTIC is used to produce maps of large regions showing their relative vulnerability. Its authors recommend that it be applied on no region smaller than 100 acres.

Several other overlay and index systems for groundwater vulnerability assessment exist; the NRC report lists seven, including DRASTIC. Typically, such systems include variables related to ground water recharge rate, depth to the water table, and soil and aquifer properties. The relative importance of the variables and the methods for combining them vary from one method to another, but all share some common traits. In general, overlay and index methods rely on simple mathematical representations of expert opinion, and not on process representation or empirical data.

Mathematical Models. Process-based mathematical models such as PRZM, GLEAMS, and LEACHM can predict the fate and transport of contaminants from known sources with remarkable accuracy in a localized area by applying fundamental physical principals to predict the flow of water in porous media and the behavior of chemical constituents carried by that water. In the hands of knowledgeable analysts with the appropriate site-specific information, such models allow threats to the safety of ground water supplies to be recognized and can play an important role in planning remediation efforts. Unlike other groundwater quality prediction methods, mathematical models predict variations of water quality both in space and in time.

Although process models offer the most sophisticated, and potentially most accurate predictions of water quality, they are not widely used for regional groundwater vulnerability analysis. Reporting on the vulnerability assessment methods used by state agencies, the GAO found that none used mathematical process models (GAO, 1992).

The Federal Republic of Germany, however, has sponsored a modeling project to identify the regions most susceptible nitrate contamination of groundwater (Wendland et al. 1993). The data and model are based on a grid of the nation consisting of nearly 40,000 3 x 3 km cells. The data include five hydrologic themes, seven soil themes, three hydrogeologic themes, six themes describing regional groundwater flow, and five themes contributing to the nitrogen cycle. From this data, the model produces a map of "Denitrification Conditions" and three maps of potential nitrate concentrations under different flow assumptions. The quantity of data required for this study, both in terms of characteristics mapped and detail of mapping, requires greater resources than any State in the U.S. has presently devoted to groundwater vulnerability analysis.

Statistical Methods. Empirical or statistical methods are the least common vulnerability assessment methods in the literature. Although statistical studies are used as tests for other methods, and geostatistical methods such as kriging are frequently used to describe the distribution of water quality parameters, very few vulnerability assessment methods are directly based on statistical methods. The

GAO report identifies one statistically based method, and the NRC report adds one more. These will be discussed in the following section. In addition, the GAO reports that although twelve states use empirical methods for assessing the vulnerability of groundwater to pesticide contamination, their methods are not published, and have not been verified.

Checklists. A fourth category, not included in the GAO or NRC reports, encompasses the methods used by Texas and several other states for their Primary Drinking Water Standards enforcement. These methods provide a checklist or decision tree, based on well construction, geologic and soil factors, and the presence of chemical sources in the vicinity of the well. The vulnerability assessment method used by the Texas Natural Resource Conservation Commission (Blodgett 1993) is a representative example.

The assessment consists of the following steps:

1. Determining the location of the water supply well.
2. Acquisition of well construction and material setting descriptions, and driller's logs for the well.
3. Verification of proper well construction, and identification of a *vulnerability point*—typically the bottom of a cemented well casing, the top of a gravel pack, or the top of the well's shallowest open interval. A well lacking cemented casing, or otherwise improperly constructed is considered susceptible to contamination.
4. Examination of driller's logs to determine geologic susceptibility. The thickness of aquitards (materials with low vertical conductivity) above

the vulnerability point is tabulated. If the vulnerability point lies below a single aquitard layer thirty feet thick (forty feet if the aquitard is exposed at the surface) or below multiple aquitard layers with a total thickness of 100 feet or more, the well is considered protected (not susceptible). A different method is used for wells in fractured rock or carbonate aquifers.

5. Delineation of a zone of contribution for susceptible wells. The limits of the zone for a forty-year time-of-travel are calculated with a semi-analytical computer model, WHPA, also used for wellhead protection programs in Texas and other states.
6. Review of contaminant use in the zone of contribution. A variety of databases with spatial coordinates are used for this purpose.
7. Waiver determination. Using the results of the preceding steps, a list of contaminants to be tested for is generated. Three- to nine-year waivers are given for contaminants not requiring monitoring.

The above procedure, and a similar vulnerability assessment method for Wisconsin (Wisconsin Bureau of Water Supply 1992), rely on a process similar to the overlay and index methods described earlier. Like those methods, the checklist applies expert knowledge and opinion systematically to the problem of vulnerability assessment, but does not employ a specific process model (except in an ancillary role) or an empirical/statistical basis for its recommendations.

2.3 STATISTICAL GROUNDWATER VULNERABILITY ASSESSMENT

Between them, the GAO and NRC reports on vulnerability assessment methods found only two published methods for statistical groundwater vulnerability analysis. Although a number of studies have applied statistical methods to verifying other methods, or have sought to prove or disprove a correlation between single environmental parameters (land use/land cover, for example) and groundwater quality, attempts to produce a predictive method for groundwater quality from empirical data are uncommon. A literature search revealed only six studies (including the two listed in the GAO and NRC reports) that attempt to identify and rate the importance of multiple indicators of groundwater vulnerability or groundwater quality. None of these studies used geostatistical methods.

Teso et al. (1988) used discriminant analysis—a statistical method for assigning objects to categories based on their location in a multi-dimensional data space—to identify sections (one mile squares) in Fresno County, California as susceptible (or not) to contamination by 1,2-dibromochloropropane (DBCP). They compiled both groundwater DBCP measurements and soil taxonomic groups for 835 sections. Based on the DBCP measurements they sorted the section into categories of "contaminated," meaning that DBCP had been detected in a well located in that section or "not contaminated," meaning that no wells in the section had detectable levels of DBCP. 511 of the 835 sections were classified as contaminated. In addition, the presence or absence of soils belonging to 228 taxonomic groups was encoded in a 228-dimensional binary vector for each section. A 1 in the n^{th} dimension of a section's soil vector indicates the presence of soil type n; a 0 in the same place indicates its absence. The 835 sections were used to calibrate a discriminant function that identifies any point in the 228-dimensional soil data space as "contaminated" or "not contaminated." A similar analysis with a smaller number of higher-order soil classifications (the 228 taxonomic groups were reduced to only six soil series) yielded a discriminant function based on the presence or absence of only six soil series in a section. This reduced discriminant function yielded a 0.776 success rate for classification of sections in Fresno County. When tested on an independent data set from nearby Merced County, the same function yielded a success rate of 0.573.

Chen and Druliner (1986) applied multiple linear regression to measurements of nitrate and herbicide concentrations in 82 wells tapping the High Plains Aquifer in Nebraska. They used the regression method to identify those environmental factors most strongly related to the concentration of nitrate and triazine herbicides (a class of herbicides that includes atrazine, cyanazine, and others). They found that three variables (well depth, irrigation-well density, and nitrogen-fertilizer use) explain 51% of the variation in nitrogen concentrations, and that two variables (specific discharge and well depth) explain 60% of the variation in triazine herbicide concentrations. Using nitrate concentration in combination with specific discharge explains 84% of the variation in triazine herbicide concentrations.

Statistical Studies of Groundwater Quality Indicators. In addition to the studies identified by the GAO and NRC reports, other research has used statistical methods to identify relationships between small numbers of indicators and measured water quality parameters, although not directed toward producing a vulnerability assessment method.

Burkart and Kolpin (1993a) examined the influence of a variety of hydrogeologic and land-use factors on the concentrations of nitrate and atrazine in shallow aquifers over an area encompassing portions of twelve States in the midwestern U.S. They sought to identify correlations between individual factors, such as aquifer type or depth to groundwater, and the concentrations of the constituents. Using non-parametric methods, including the Mann-Whitney rank sum test and contingency tables, they found significant differences in nitrate and herbicide concentrations when wells are grouped by aquifer class (bedrock or unconsolidated) and by depth of unconsolidated material over the aquifer.

Nightingale and Bianchi (1980) used linear correlation coefficients and multiple linear regression to examine the relationship between soil and aquifer permeabilities and measurements of conductivity, anion, and cation concentrations. Like the work of Teso et al., this study was based on historical measurements grouped by the sections from which they were taken. They found that salinity was correlated to soil and aquifer permeability, but that nitrate levels correlated only with the estimated specific yield of the aquifer system.

Helgesen et al. (1992), seeking a connection between land use and water quality, delineated discrete regions of uniform land use over a portion of the High Plains aquifer in southern Kansas. They selected one well at random from each region and tested a water sample for a variety of agricultural and petroleum related chemicals. Non-parametric hypothesis tests showed significantly higher mineral concentrations under irrigated croplands and petroleum-producing areas than under undeveloped range land.

Baker et al. (1994) used an approach similar to that of Burkart and Kolpin (1993), but applied it to a larger body of samples, collected through a voluntary well testing program. Samples of water from rural wells submitted by more that 43,000 participants in twelve states were analyzed for nitrate and herbicide concentrations. Non-parametric statistical methods were applied to compare the analysis results with descriptions of the wells and their surroundings submitted by the participants with the water samples. They found that the age of the well, its depth, and its proximity to feedlots or barnyards significantly influence the likelihood of finding elevated nitrate concentrations in the samples. Likelihhoods increased dramatically when two "risk factors" were combined. They also found that factors influencing nitrate exerted similar influences on herbicide concentrations.

2.4 CHOICE OF METHOD

A statistical approach was selected for this study for two reasons. The first is dissatisfaction with index/overlay methods and process-based models. The second is the appropriateness of this approach to GIS-based analysis.

Although they represent informed opinion, and apply consistent standards to all regions, overlay and index methods lack a sound methodological foundation, being based neither on direct observation nor first principles. "These methods are driven largely by data availability and expert judgment, with less emphasis on processes controlling ground water contamination. One can argue whether the factors included in the methods are the relevant ones for vulnerability assessment and whether the factor ratings are appropriate" (NRC, 1993). These doubts are supported by studies carried out to test DRASTIC. The GAO report observes that "…tests of DRASTIC generally indicated a poor relationship between model predictions (that is, relative groundwater vulnerability), and monitoring results (that is, where pesticides are found)" (GAO 1992).

Overlay and index methods are also difficult to interpret quantitatively and provide no estimates of uncertainty. Is a region with a DRASTIC score of 200 twice as vulnerable to contamination as one with a score of 100? Does a DRASTIC score of 150 mean "between 140 and 160" or "between 100 and 200?" DRASTIC's authors do not provide answers to these questions and caution against any absolute interpretation of the index. This places serious limitations on the value of DRASTIC as a guide to forming policy. Since DRASTIC is the most thoroughly studied of the index/overlay systems, others should be viewed with less confidence.

Mathematical models of groundwater processes have the great advantage of being based on sound principles, rather than opinion, but this does little to enhance their value for policy guidance at a state or regional level. The models require more expertise and (as illustrated by the German example) more detailed data than state agencies can provide on a regional scale. The NRC report offers the following view of process models.

> It must be recognized that sophisticated models may not necessarily provide more reliable outputs, especially for regional-

> scale, and even for field-scale applications. Since data for many of the required input parameters for sophisticated models are not always available, their values have to be estimated by indirect means using surrogate parameters or extrapolated from data collected at other locations. Errors and uncertainties associated with such estimates or extrapolations can be large and may negate the advantages gained from a more rigorous process description in the simulation model. (NRC 1993)

Given the state of available data, such models are not well suited to the task of regional assessment of groundwater vulnerability.

Statistical approaches offer the possibility of a method that is as easy to apply as an index/overlay method, but with a more defensible foundation. The weighted-sum approach of DRASTIC looks like the product of a multiple linear regression, and the NRC report observes that "Vulnerability assessment methods that use overlay/indexing techniques are an eyeballed form of multivariate discriminant analyses that lack probability estimates" (NRC 1993). Since overlay methods look like the results of statistical analysis, why not develop one that *is* what it looks like? Although it is risky to apply empirical methods outside the range of conditions over which it was calibrated, such methods are at least based on real measurements, not just a set of opinions.

Data Requirements. Statistical methods require data, the more data and the higher the quality, the better. Collection of groundwater quality data is expensive and time-consuming, driving up the cost of statistical investigations. Burkart and Kolpin orchestrated the collection of samples from 303 wells throughout the midwest during the spring and summer of 1991. This was a substantial undertaking with very careful quality control, and it produced roughly 600 measurements of herbicide, nitrate, and ammonium concentration. Given the size of the region under study, this is a small number of measurements on which to base broad conclusions of cause and effect. Anyone attempting a regional-scale study of water quality faces a very substantial problem in gathering sufficient data.

At the time this study was begun, the existing body of pesticide data in Texas was not sufficient to form the basis of a statistical study. EPA's *Pesticides in Groundwater Database* (EPA, 1992), which compiles monitoring study results over the period 1971–1991, contains only 511 pesticide measurements in Texas. The Texas Department of Agriculture (Aurelius, 1989) carried out a pilot study in 1897 and 1988 to estimate the extent to which rural domestic wells are contaminated by pesticides from nonpoint agricultural sources. 175 wells were tested for nine pesticides, arsenic, and nitrate. The study was confines to high-risk areas and cannot be considered as representative of the State as a whole.

Since pesticide measurements in groundwater were not adequate to support the development of a statistical method for groundwater vulnerability analysis, another constituent—nitrate, which has been extensively measured in groundwater—was chosen.

2.5 NITRATE IN GROUNDWATER

This section presents a brief review of nitrate in groundwater, relevant to the present study, rather than a comprehensive review of the extensive literature on nitrate in groundwater. In particular, the nitrate cycle is discussed, and important concentration values are identified.

High concentrations of nitrate (NO_3^-)in drinking water may cause the disease methemoglobinemea in small children (Hem 1989). Because of this and other diseases linked to nitrate (and possibly because it is inexpensive to measure), its concentration in public water supplies is monitored and regulated by federal law. The National Primary Drinking Water Standards (40 CFR 141) set the maximum contaminant level (MCL) for nitrate at 10 mg/l (measured as nitrogen). Groundwater systems must monitor for compliance with the MCL annually. If nitrate in excess of 5 mg/l is detected, the system must increase its monitoring to quarterly for at least one year.

Nitrate occurs naturally from mineral sources and animal wastes, and anthropogenically as a byproduct of agriculture and from human wastes. Nitrate is the most highly oxidized form of nitrogen in the nitrogen cycle, which includes activities in the atmosphere, hydrosphere, and biosphere. Figure 2.1 shows the following major transformations from the nitrogen cycle (Madison and Brunett, 1985)

Assimilation of inorganic forms of nitrogen (ammonia and nitrate) by plants and microorganisms.

Heterotrophic conversion of organic nitrogen from one organism to another.

Ammonification of organic nitrogen to produce ammonia during the decomposition of organic matter.

Nitrification of ammonia to nitrate and nitrite by the chemical process of oxidation.

Denitrification (bacterial reduction) of nitrate to nitrous oxide (N_2O) and molecular nitrogen (N_2) under anoxic conditions.

Fixation of nitrogen (reduction of nitrogen gas to ammonia and organic nitrogen) by microorganisms.

Madison and Brunett (1985) list the following as major anthropogenic sources of nitrate: "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 the result of the application of irrigation water." Natural sources include: "soil nitrogen, nitrogen-rich geologic deposits, and atmospheric deposition."

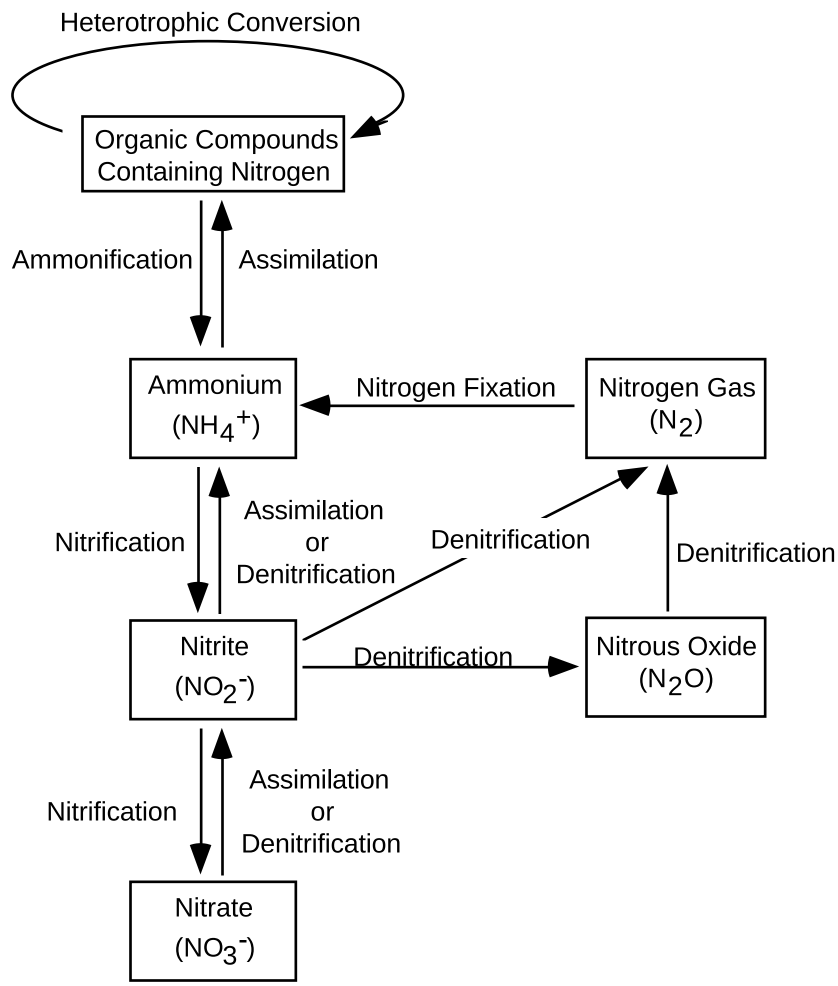

Figure 2.1 Simplified Biological Nitrogen Cycle
[after Madison and Brunett (1985)]

According to Hem (1989), nitrogen occurs in water as nitrate or nitrite anions, as ammonium cations, and in a variety of organic compounds. Nitrite and the organic species are unstable in aerated water. Ammonium cations are strongly adsorbed on mineral surfaces, but the anionic species are readily transported in water and are stable over a wide range of conditions.

Given the wide range of nitrate sources associated with agriculture, its chemical stability in water, and its high mobility—to say nothing of the frequency with which it has been measured in water—nitrate is a natural choice as an indicator for vulnerability of groundwater to contamination to nonpoint agricultural sources. This use has been suggested by Cohen et al. (1984), and has been tested by a number of investigators. Domagalski and Dubrovsky (1992) found no significant difference in nitrate concentrations between wells with and without triazine herbicide residues in the San Joaquin valley of California. An examination of the report by Burkart and Kolpin (1993a) shows that the geological factors associated with high frequencies of herbicide contamination are also associated with high frequencies of excess nitrate detection. Baker et al. (1994) found a similar correspondence between nitrate and pesticide vulnerability in samples collected from rural wells in 17 States.

Nitrate concentrations are usually reported in units of miligrams per liter (mg/l) with the mass representing either the total mass of nitrate ion in the water (nitrate-NO_3), or as the mass of only the nitrogen (nitrate-N). The molecular weight of nitrate is 62; the molecular weight of nitrogen 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 of 10 mg/l nitrate-N is equivalent to 44.3 mg/l nitrate-NO_3.

In their nationwide study of nitrate in the groundwater of the U.S., Madison and Brunett assigned the following interpretations to ranges of nitrate concentrations (in nitrate-N)

- Less than 0.2 mg/l—Assumed to represent natural background concentrations.

- 0.21 to 3.0 mg/l—Transitional; concentrations that may or may not represent human influence.

- 3.1 to 10 mg/l—May indicate elevated concentrations resulting from human activities

- More than 10 mg/l—Exceeds maximum concentration for National Interim Primary Drinking-Water Regulations.

Their selection of 3.0 mg/l as a threshold to indicate human influence has been followed by many investigators, including Burkart and Kolpin, and Baker et al. The use of individual concentration levels in this study is discussed further in Section 4.1.

2.6 OUTLINE OF PROPOSED VULNERABILITY ANALYSIS METHOD

The general form of the approach to statistical groundwater vulnerability analysis advanced in this work can be summarized in six steps. These are:

1. Select a constituent or set of constituents, whose presence will indicate the degree of vulnerability of a groundwater source.
2. Identify a set of distinct mappable regions of the surface or subsurface.
3. Assemble a body of measurements of the constituent identified in step 1 that can be linked with the regions identified in step 2.
4. Calculate descriptive statistics for the body of measurements linked with each region.
5. Map the variation of the descriptive statistics from region to region.
6. Relate the variation of the descriptive statistics to the variation of indicator parameters by forming a mathematical expression that mimics

the relationship between the descriptive statistics and indicator values mapped over the same set of regions.

The results of these steps include maps and numerical values associated with the regions, indicating their vulnerability to contamination as represented by the descriptive statistics, and a mathematical model that permits those results to be extended to areas where water quality data have not been collected, but values of the indicator parameters are known.

2.6.1 Comparison of Method with Previous Studies

The six steps are proposed as a synthesis of the approaches taken in the statistical studies cited in Section 2.2. The work of Teso et al. (1988), and Nightingale and Bianchi (1980) follows steps 1 through 4 by dividing the study area into regions by square-mile section, forming groups of water quality measurements from historic data based on the location of sampling sites in the sections, and forming summary statistics for each section—binary classifications in based on the presence or absence of DBCP in any well in the section in Teso et al, arithmetic averages of nitrate concentrations and electrical conductivity for all measurements from the section in Nightingale and Bianchi. Similarly, Burkart and Kolpin (1993b) grouped the measurements collected in their reconnaissance of agricultural contaminants in the mid-continental U.S. by their location in major land resource areas (MLRAs) and calculated a third type of summary statistic—the frequency with which threshold concentrations of nitrate and herbicides were exceeded in measurements collected in the MLRAs.

Burkart and Kolpin, Baker et al. and Teso et al. mapped their results (step 4), but not Nightingale and Bianchi did not.

Comparison of summary statistics to indicator parameters and formation of a mathematical model (step 5) is carried out in all of the cited studies except for Burkart and Kolpin (1993b). Chen and Druliner (1986), and Helgesen et al. (1992) compared indicator parameters directly to concentrations reported in individual water samples rather than statistics calculated on groups of measurements, although Helgesen et al. intend each well to represent a region. Burkart and Kolpin (1993a) re-group their measurements for each indicator, rather than forming one set of groups and comparing their statistics to indicator variations over the same groups. Teso et. al and Nightingale and Bianchi base their results on region-based statistics and indicator values from the same regions.

The cited studies approach data compilation in one of two ways. These can be identified as the *well-oriented* approach and the *region-oriented* approach.

The *well-oriented* approach, taken by Burkart and Kolpin, by Chen and Druliner, and by Baker et al. is to select a relatively small number of wells to represent a each region or setting. Measured variations in constituent concentration from well to well are compared to variations in the characteristics of the wells and their surroundings. Barringer et al. (1990) point out that results from such studies can be biased due to spatial autocorrelation if the wells are too close together. A well-oriented study requires careful planning or data screening to assure that the selected wells are typical of the regions where they are located.

The *region-oriented* approach is to define a set of regions, calculate two sets of statistics on the regions—one of water quality and one of potential indicators—and study the relationships between the two sets of statistics. This is the method that Teso et al. and Nightingale and Bianchi applied in their studies California. In both studies the regions were surveying sections. In Teso et al., the water quality statistics were the binary classification of the sections by having or not having DBCP detections, the indicator statistics were the soil taxonomy vectors, and the relationship between the two was analyzed with discriminant analysis. In Nightingale and Bianchi, the water quality statistics were arithmetic averages of conductivity or cation and anion concentrations, the indicator statistics were averages of aquifer and soil permeability, and the relationships were examined with linear correlation coefficients for paired variables and multiple linear regression for multiple variables. Helgeson et al. identified regions by land use, and characterized each by a single randomly selected water sample. In another report on the results of their groundwater reconnaissance of the midwest, Burkart and Kolpin (1993b) used a GIS to identify regions—STATSGO soil polygons (see Chapter 3) or Major Land Resource Areas—as more or less vulnerable to contamination, based on the frequency that atrazine was detected in wells in those regions.

Region-oriented studies avoid some of the problems of well-oriented studies, but are subject to some limitations. Bias due to autocorrelation is reduced by aggregating samples, giving each region equal weight in evaluating the relationship between indicators and water quality. The potential for an atypical well to incorrectly characterize a region is reduced (if sufficient data is available) by the contributions of several wells to the description of water quality in the region. The regional orientation, however, precludes any study of the effects of well-specific characteristics such as pumping rates or construction characteristics. On balance, the regional approach was judged more suitable for the data available, and the objectives of the study.

2.6.2 Application in Present Work

In this study, the five steps were implemented as follows.

1. Use *nitrate* to represent the vulnerability of groundwater.
2. Divide Texas into a grid of *7.5′ quadrangles*, based on the well-numbering system used by the Texas Water Development Board (TWDB) in its Ground-Water Data System (Nordstrom and Quincy, 1992). The well-numbering system and the quadrangles are described in Section 4.2.
3. Form groups of groundwater nitrate measurements recorded in the TWDB Ground-Water Data System based on the location by quadrangle of the wells from which the water samples were collected.
4. Calculate statistical estimates of *exceedence probabilities*, the likelihood that nitrate concentrations measured in water samples collected in the quadrangles will exceed selected threshold values.
5. Prepare maps of the quadrangles showing the variation of the exceedence probabilities for the selected thresholds.
6. Prepare maps of four indicator parameters—average annual precipitation, average soil thickness, average soil organic matter content, and average

annual nitrogen fertilizer sales—and use *stepwise multiple linear regression* to construct a simple linear model of exceedence probabilities based on these indicators.

The italicized words in the list above indicate specific choices made in the course of this investigation that make it distinct from the general model described at the beginning of this section. All of these choices will be discussed in later sections of this chapter.

In addition to the 7.5' quadrangles, five aquifers—the Carrizo-Wilcox, the Edwards (Balcones Fault Zone), the Hueco-Mesilla Bolson, the Ogallala, and the Seymour—were used as an alternate set of regions to divide a subset of the TWDB data into groups for an analysis similar to that performed on the quadrangles. The variation of exceedence probabilities for this subset was compared from aquifer to aquifer as well as by the four parameters listed in step 5 above.

The choice of nitrate for study, the methods used to form the data into groups for analysis, the methods used to calculate the exceedence probabilities, and the use of stepwise multiple linear regression are described in Chapter 4. The data used in the analyses are described in Chapter 3.