Predicting the Probability of Elevated Nitrate Concentrations in the Puget Sound Basin: Implications for Aquifer Susceptibility and Vulnerability

by Anthony J. Tesoriero and Frank D. Voss[a]

Abstract

The occurrence and distribution of elevated nitrate concentrations ($\geq$ 3 mg/l) in ground water in the Puget Sound Basin, Washington, were determined by examining existing data from more than 3000 wells. Models that estimate the probability that a well has an elevated nitrate concentration were constructed by relating the occurrence of elevated nitrate concentrations to both natural and anthropogenic variables using logistic regression. The variables that best explain the occurrence of elevated nitrate concentrations were well depth, surficial geology, and the percentage of urban and agricultural land within a radius of 3.2 kilometers of the well. From these relations, logistic regression models were developed to assess aquifer susceptibility (relative ease with which contaminants will reach aquifer) and ground-water vulnerability (relative ease with which contaminants will reach aquifer for a given set of land-use practices). Both models performed well at predicting the probability of elevated nitrate concentrations in an independent data set. This approach to assessing aquifer susceptibility and ground-water vulnerability has the advantages of having both model variables and coefficient values determined on the basis of existing water quality information and does not depend on the assignment of variables and weighting factors based on qualitative criteria.

Introduction

Regional assessments of ground-water quality are complicated by the fact that constituent concentrations are often highly variable spatially. As such, these assessments require either (1) sampling a large number of wells that are randomly distributed throughout the regional aquifer system, or (2) understanding the factors that influence ground-water quality so that sampling can be targeted to areas that are most vulnerable to contamination. The U.S. Geological Survey's National Water Quality Assessment program is conducting both random and targeted ground-water sampling in 60 study units throughout the United States (Gilliom et al., 1995). The Puget Sound Basin, located in the northwestern United States, is one of these study units. As a part of the Puget Sound Basin investigation, an analysis of existing nitrate data was carried out to determine natural and anthropogenic factors that best explain the occurrence and distribution of elevated nitrate concentrations in ground water and subsequently to develop models that delineate areas more likely to be (or become) contaminated by nitrate and possibly other contaminants. The results of this analysis provide insights into the susceptibility and vulnerability of aquifers within the Puget Sound Basin as well as providing a methodology for studies in other areas.

Nitrate (NO_3^-) data were selected for evaluation because elevated concentrations of this constituent are typically caused by anthropogenic activities (e.g., crop fertilization, domestic on-site sewage disposal) and are relatively common compared to the frequency of detection of other constituents, such as pesticides or volatile organic compounds. In fact, it has been suggested that nitrate may be the most ubiquitous contaminant of ground water in the world (Spalding and Exner, 1993). As a widespread contaminant, nitrate may be an important indicator of environments that are susceptible to contamination. Aquifer susceptibility (equivalent to aquifer sensitivity in U.S. Environmental Protection Agency, 1993) refers to the relative ease with which a contaminant applied on or near a land surface can migrate to the aquifer of interest; ground-water vulnerability refers to the relative ease with which a contaminant applied at or near the land surface can migrate to the aquifer of interest, for a given set of land-use practices. The difficulty and high cost of remediating ground water (e.g. McKay and Cherry, 1989; Mercer et al., 1990) have increased the focus of regulatory agencies on the prevention of ground-water contamination. Therefore, the identifiction of areas that are susceptible to contamination has become an important tool for land-use planners and environmental regulators; additional precautions can be taken in these ares to minimize the risk of ground-water contamination. Ground-water vulnerability assessments are useful for water suppliers to help identify high-risk areas. In these areas, development of new supplies may be curtailed, and more frequent sampling of existing supplies may be required.

Aquifer susceptibility and ground-water vulnerability determinations have typically been made using a scoring system or a hydrogeological classification method based on estimates of the hydrogeologic factors that affect contaminant migration in the subsurface (e.g., Lemme et al., 1990; Hearne et al., 1992). Although the factors used in these assessments likely influence contaminant transport, their importance relative to each other for a given region is often not well-known. Weights for each factor are typically assigned on the basis of semiquantitative or qualitative information; existing water quality data typically are not used in this determination. The DRASTIC model (Aller et al., 1987) employs a scoring system and is probably the most widely known method for aquifer susceptibility assessment. DRASTIC has been tested against water-quality data with mixed results (Koterba et al., 1993; Kalinski et al., 1994).

[a]U.S. Geological Survey, Water Resources Division, 1201 Pacific Avenue, Suite 600, Tacoma, Washington 98402.

Received July 1996, revised January 1997, accepted February 1997.

In this analysis, susceptibility and vulnerability are assessed by using logistic regression to relate the occurrence of elevated nitrate concentrations to natural and anthropogenic variables. This method has the distinct advantage of selecting the significant variables and their relative importance based on the occurrence of contaminants in ground water in a given area.

Regional Hydrogeology

The Puget Sound Basin encompasses the 35,000 km^2 land area where runoff ultimately drains into the Puget Sound (Figure 1). The Puget Sound Basin aquifer system is composed primarily of unconsolidated sediments, which can be locally more than 900 meters thick (Jones, 1997). Sediment deposition was primarily the result of a series of regional glacial advances and retreats. A typical glacial sequence would consist of advance outwash, till, and recessional outwash. Interglacial deposits generally occurred in bays and lakes, with lesser amounts in rivers. Since the last glacial period, extensive alluvial materials have been deposited in the major river valleys. To construct a regional surficial geology map (Figure 1), geologic units were grouped according to their hydrologic properties (Jones, 1997). The term "fine-grained glacial deposits" in this report refers to the till and interglacial deposits that form the semiconfining and confining units in the basin. The term "coarse-grained glacial deposits" as used here refers to the recessional and advance outwash deposits that form the major aquifers in the basin. The alluvial deposits also form aquifers in this region, but differ from the coarse-grained glacial deposits in that they are often major discharge areas. The occurrence of the alluvial and glacial outwash aquifers generally coincides with the surficial deposits of alluvial and glacial (both fine- and coarse-grained) origin, respectively (Figure 1). The glacial outwash aquifer was divided into two groups for this analysis: (1) those areas with fine-grained surficial deposits (semiconfining unit) at the surface, and (2) those with coarse-grained surficial deposits. Further divisions of the outwash aquifer based on the presence of additional confining units at depth would be appropriate; however, this information was not available for many of the wells used in this study. Lastly, bedrock in this area consists of consolidated rocks of a wide variety of types (e.g., sandstone, limestone, and basalt); with a few exceptions, little or no ground-water development occurs in this unit (Molenaar et al., 1980).

Methods

Subsets of two large data sets were analyzed to estimate aquifer susceptibility and vulnerability: (1) the public water-supply data base maintained by the Washington State Department of Health (WDOH) and (2) the National Water Information System (NWIS) maintained by the U.S. Geological Survey (USGS). The WDOH data set contains water-quality data for samples collected from 1,967 wells from 1986 to 1994. The NWIS data base contains nitrate data for samples collected from 1,729 wells used for ground-water quality investigations conducted by the USGS from 1971 to present. Table 1 provides a summary of the two data sets. Because of a better spatial distribution and shorter sample collection period, the WDOH data set was used to determine relations between elevated concentrations of nitrate and potential explanatory variables; these relations were then verified using the NWIS data set. The minimum reporting level for nitrate plus nitrite as nitrogen (hereafter referred to as nitrate because nitrite concentrations are expected to be low relative to nitrate) was typically 0.2 mg/l for the WDOH data set and either 0.05 or 0.1 for the NWIS data set. Nearly 50 percent of the samples had concentrations of nitrate that were below the minimum reporting level.

Logistic regression was chosen for this analysis because (1) of its ability to treat the large number of censored values (i.e., concentrations below the minimum reporting level), and (2) it is well-suited to determine the variables and their coefficient values that best identify the wells that have elevated nitrate concentrations. Other regression technqiues (e.g., multiple linear regression) develop regression lines that are, for this data set, largely influenced by the vast majority of the data that are near background levels.

Logistic regression has been used extensively in the health sciences since the late 1960s to predict a binary response from explanatory variables (e.g., Truett et al., 1967; Lemeshow et al., 1988), and more recently in the environmental sciences to assess multiple variables that may explain the occurrence of contamination in ground water (e.g., Eckhardt and Stackelberg, 1995). To convert nitrate concentrations from a continuous variable to a binary variable, a concentration level must be established to separate events (concentrations greater than or equal to this level) from nonevents. In order to assess aquifer susceptibility or vulnerability, this level should represent a concentration that was the result of anthropogenic activities. Madison and Brunett (1985) have suggested a level of 3 mg/l for nitrate. Although historical data in the Puget Sound Basin indicate that background nitrate concentrations are much less than 3 mg/l (e.g., Brown and Caldwell, 1985), this value was selected as a conservative estimate to represent anthropogenic effects. The probability of a ground-water sample having a nitrate concentration ≥ 3 mg/l (hereafter referred to as an event) can be estimated for different explanatory variables using logistic regression. A brief description of logistic regression is provided below; detailed discussions can be found elsewhere (Hosmer and Lemeshow, 1989; Helsel and Hirsch, 1992). The form of the logistic regression model is:

$$P = \frac{e^{(b_0 + \mathbf{bX})}}{1 + e^{(b_0 + \mathbf{bX})}} \quad (1)$$

where P is the probability that nitrate is present at a concentration ≥ 3 mg/l, $\mathbf{X}$ is a vector of n explanatory variable values, b_0 is a scalar intercept parameter, and $\mathbf{b}$ is a vector of slope coefficient values, so that $\mathbf{bX} = b_1X_1 + b_2X_2 + \ldots\ldots b_nX_n$. The motivation behind selecting this function is described in Cox and Snell (1989). A transformation of P, called the logit transformation, is then performed to yield a linear function:

$$\ln \frac{P}{1 - P} = b_0 + \mathbf{bX} \quad (2)$$

Using an iterative procedure, the SAS statistical analysis computer software package (SAS Institute, 1990) calculates values for b_0 and $\mathbf{b}$ that maximize the ln likelihood function, L

$$L = \sum_{i=1}^{m} [(y_i \cdot \ln P_i) + (1 - y_i) \cdot \ln(1 - P_i)] \quad (3)$$

where m is the number of observations in the data set, and y_i is the outcome variable that is set to 1 when nitrate concentrations are ≥ 3 mg/l; otherwise, a value of zero is assigned. The model

Table 1. Number of Events (Nitrate Concentration ≥ 3 mg/l) and the Number of Total Wells for Each Data Set by Surficial Geology Type

	WDOH data set			*NWIS data set*		
Surficial geology	*Number of events*	*Number of wells*	*% of events*	*Number of events*	*Number of wells*	*% of events*
Coarse-grained	86	742	12	116	655	18
Fine-grained	40	912	4	38	793	5
Alluvium	6	197	3	12	227	5
Bedrock	2	50	4	1	53	2

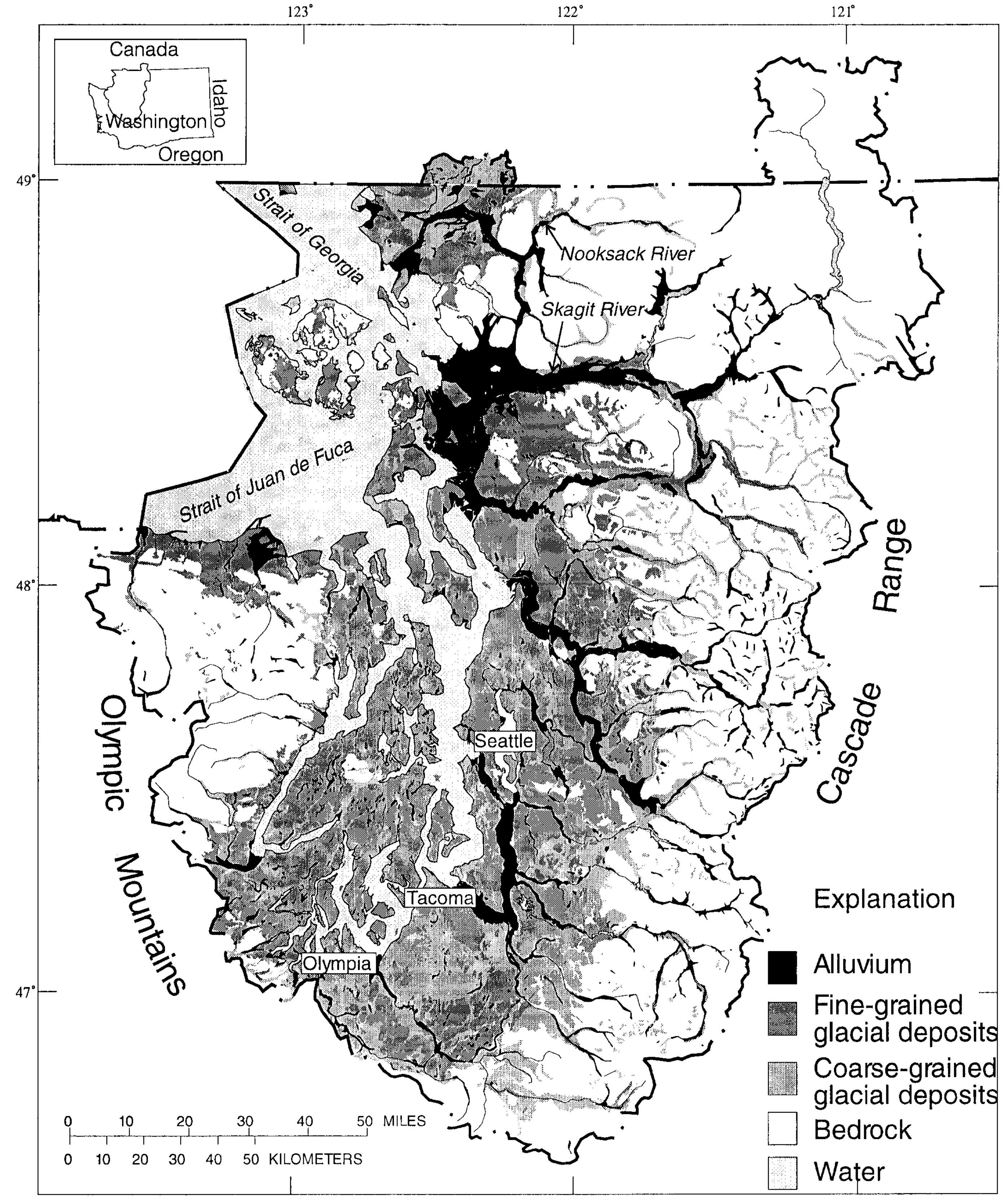

Fig. 1. Generalized surficial geology of the Puget Sound Basin (adapted from Jones, 1997).

Table 2. Regression Coefficients and Summary Statistics for Single Variable Models Developed Using WDOH Data

Model variable	Sample size[a]	b_0	b_1	G	p
Well depth (m)	1,967	-1.22	-0.030	89.7	<0.001
Recharge (cm)	1,818	-2.30	-0.036	3.1	0.079
Soil hydrologic group	1,967	-2.25	-0.124	1.1	0.284
Surficial Geology Types					
Coarse-grained glacial deposits	1,901	-3.14	1.110	37.2	<0.001
Fine-grained glacial deposits	1,901	-2.25	-0.884	19.6	<0.001
Alluvium	1,901	-2.51	-0.950	6.6	0.010
Bedrock	1,901	-2.57	-0.611	0.9	0.356
Land-Use Types & Population Density					
% Agriculture within 3.2 km of well	1,921	-3.08	0.034	62.7	<0.001
% Forest within 3.2 km	1,921	-1.72	-0.018	23.6	<0.001
% Urban within 3.2 km of well	1,921	-2.91	0.014	10.0	0.002
Population density, 1990 census data in persons/km^2	1,796	-2.53	-0.0001	0.2	0.688

[a]Sample size varies, depending on the availability of data for each variable.

[b_0 is the intercept parameter and b_1 is the regression coefficient as described in equation (1). G is the $-2 \ln L$ ratio, and p is the significance level for testing the hypothesis that the regression coefficient is zero. See text for details.]

can then be evaluated by calculating the -2 ln likelihood ratio, G, as:

$$G = -2(L_{int} - L_{model}) \quad (4)$$

where L_{int} and L_{model} are the ln likelihood values for an intercept-only model and the model being tested, respectively. G has a chi-square distribution under the null hypothesis that all regression coefficients are zero. G values and associated probability (p) values are provided in the data tables to follow. A significant p-value for this statistic provides evidence that at least one regression coefficient is nonzero. The G value is also a measure of the goodness-of-fit of the model; higher values indicate a better fit.

In the next section, figures are provided which show not only the logistic regression fits for estimating the probability of an event versus an explanatory variable but also the observed data for each decile of risk. Deciles of risk are determined by forming 10 equal groups based on the probability of an event (e.g., the 10 percent of the wells with the highest predicted probabilities of an event form one decile of risk). All logistic regression fits were determined by fitting individual observations, *not* the grouped data.

Data Analysis and Discussion

A number of factors can influence the concentration of nitrate in ground water. These factors are generally associated with the anthropogenic sources of nitrate, the nature of recharge to the ground-water flow system and the locations in the flow system where samples are taken. In this analysis, because the actual loadings of nitrate at or near the land surface are not available, they are estimated by land use and population density variables. Recharge is addressed directly using an estimate of the amount of recharge received in the vicinity of the well, and indirectly using variables describing the hydrologic characteristics of surficial deposits. The relative position of a sample point (well screen) in the flow system is approximated by a combination of variables describing well depth and surficial geology.

Single Variable Models

In order to develop multiple regression models to assess aquifer susceptibility and ground-water vulnerability, the relation between each variable and event occurrence was evaluated to determine which variables are individually significant predictors of elevated nitrate concentrations. The following variables were considered: well depth, surficial geology, soil hydrologic group (describes runoff potential of soil, Soil Conservation Service, 1993), recharge, land use, and population density (from 1990 census). Regression coefficients and statistics for models developed using these variables are provided in Tables 2 and 3. Well depth, surficial geology, and land use were found to be significant ($p \leq 0.05$) and are discussed below.

Well Depth

As expected, an inverse relation was found between well depth and the estimated probability of an event (Figure 2, Table 2). Similar relations have been observed in studies worldwide (e.g., Singh and Sekhon, 1978). This is a typical profile for surficially derived contaminants because ground water withdrawn from deeper wells generally has longer travel times than water from shallow wells because of the combined effects of travel through both the unsaturated and saturated zones. Nitrate

Table 3. Regression Coefficients and Summary Statistics for Aquifer Susceptibility and Ground-Water Vulnerability Models Developed from the WDOH Data

Model	Sample size[a]	b_0	b_1, Well depth (m)	b_2, % agriculture land use within 3.2 km of well	b_3, % urban land use within 3.2 km of well	G	p
Aquifer Susceptibility Models							
Coarse-grained	742	-0.46	-0.039	—	—	66.7	<0.001
Fine-grained	912	-2.18	-0.015	—	—	10.5	0.001
Alluvium	197	-0.04	-0.207	—	—	14.9	<0.001
Ground-Water Vulnerability Models							
Coarse-grained	719	-2.07	-0.028	0.034	0.029	108.8	<0.001
Fine-grained	899	-2.66	-0.016	—	0.023	15.9	<0.001
Alluvium	196	-1.95	-0.268	0.059	0.049	25.6	<0.001

[a]Sample size varies between susceptibility and vulnerability models due to availability of land-use data.

[b_0 is the intercept parameter, and b_1 through b_3 are the regression coefficient values corresponding to equation (1). G is the $-2 \ln L$ ratio, and p is the significance level for testing the hypothesis that the regression coefficients are zero. See text for details.]

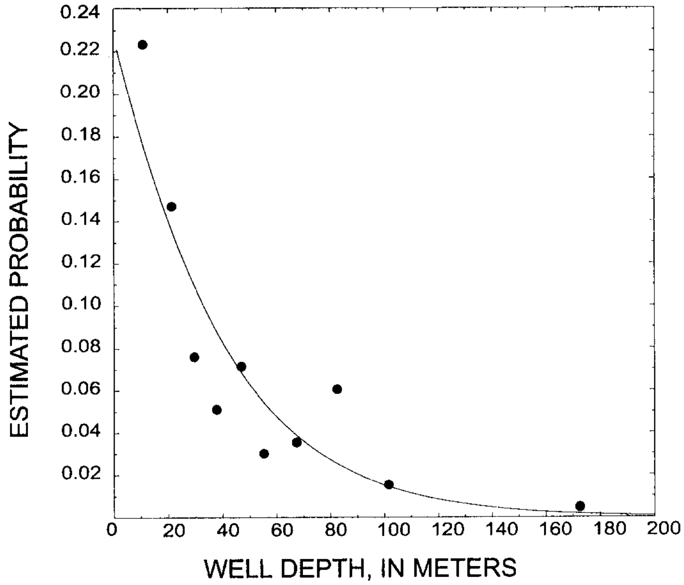

Fig. 2. The line depicts the estimated probability that nitrate concentrations in ground water are ≥ 3 mg/l as a function of well depth for the WDOH data set. Data points show the fraction of wells which have nitrate concentrations ≥ 3 mg/l vs. mean well depth for each decile of risk.

reduction may also be responsible for some of the decrease in events with well depth.

Surficial Geology

The regional surficial geology consists of four types of deposits: alluvium, coarse-grained glacial, fine-grained glacial, and bedrock (Figure 1, Table 1). Because surficial geology is neither a continuous variable nor an interval-scaled variable, its relation to events is described by the presence or absence of each type of deposit. The presence of coarse-grained glacial deposits at the surface is positively correlated with events; fine-grained glacial and alluvial deposits at the surface are inversely correlated with events (Table 2).

Land Use

The relation between the dominant land uses near a well and the occurrence of contaminants in ground water can often provide insights into possible sources of contamination. In fact, several investigators have related the occurrence of contaminants to surrounding land uses (e.g., Eckhardt and Stackelberg, 1995; Kolpin et al., 1994). Land-use classifications in the Puget Sound Basin were determined with USGS land-use and land-cover digital data from 1973-1980 (Fegeas et al., 1983). Areas were classified into several land-use categories: urban or built-up, agriculture, rangeland, forest, water, wetland, barren, tundra, and perennial ice and snow (level 1 scheme, Anderson et al., 1976). Forest (78 percent), urban (8 percent), and agriculture (6 percent) cover the largest land areas in the region. Forest land was not further classified according to the degree of timber harvesting. Most of the urban areas are centered on the major cities in the region: Seattle, Tacoma, and Olympia. Intensive agricultural activities, which include dairy farms, berries, wheat, and corn, occur along the Skagit and Nooksack River Valleys.

Because data were not available to establish the upgradient direction at each of the wells used in this study, land use within a circle of a specified radius centered on the well was used to determine the area of influence. The optimal land-use radius has been estimated by examining how models employing different radii fit the data (e.g., Vowinkel and Battaglin, 1988). Logistic regression analyses were performed for each of the three dominant land uses with eight different radii ranging from 0.8 to almost 13 kilometers. The vast majority of events were for shallow wells; therefore, to minimize computational effort, only wells less than 45 meters deep were used in these analyses. The G values for the three land uses were summed to calculate a total G value for each radius (Figure 3). The total G value is maximized at a radius of 3.2 km, suggesting that this is the optimal radius to evaluate the effect of land use on the occurrence of elevated nitrate concentrations in ground water (Figure 3).

The percent of each land-use type within 3.2 km of all wells was then related to events (Figure 4, Table 2). The probability of an event markedly increases as the percentage of agricultural land increases. This result is not surprising, as there are several potential sources of nitrate in agricultural areas, and is consistent with an analysis of nitrate data from across the United States (Mueller et al., 1995). The positive coefficient for the percentage of urban land suggests that urban sources (such as effluent from septic systems or fertilizers applied to lawns) also cause elevated nitrate concentrations in ground water. Percentage of forest land was inversely correlated with events, likely reflecting the relative absence of nitrate sources in forested areas; a similar observation has also been made by other investigators (e.g., Johnson, 1992; Mueller et al., 1995).

Multiple Variable Models

By incorporating the effects of multiple variables, the probability of an event for a more specific environment is obtained (e.g., urbanized area over coarse-grained deposits). The following models were developed using stepwise logistic regression to assess aquifer susceptibility and ground-water vulnerability. The most significant variables are entered first but only if the significance level of 0.1 is met. This process is continued until none of the remaining variables meet this significance level. A detailed

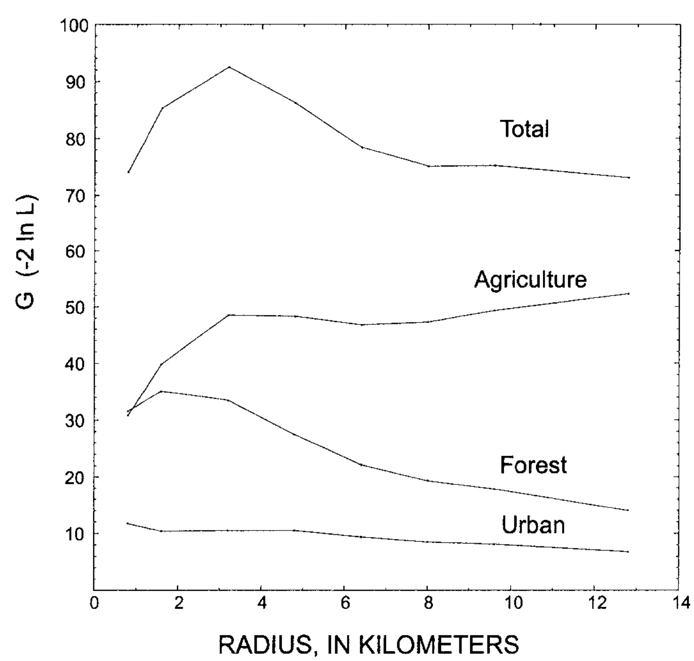

Fig. 3. G values vs. radius surrounding the WDOH wells used to calculate percent land use. Total was calculated by summing the three land-use types.

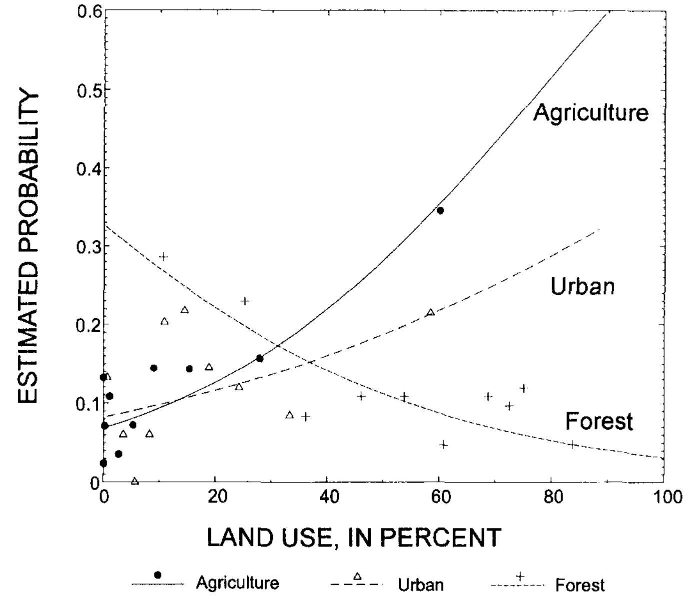

Fig. 4. Lines depict the estimated probability of encountering nitrate concentrations ≥ 3 mg/l as a function of percentage land use within a 3.2-kilometer radius of the well. Data points represent the fraction of WDOH wells that have nitrate concentrations ≥ 3 mg/l vs. mean land use for each decile of risk. Pluses, triangles, and circles are the data for forest, urban, and agricultural land uses, respectively.

description of this method is provided by Hosmer and Lemeshow (1989).

Aquifer Susceptibility

Assuming that nitrate sources are generally equal in the three surficial geology types, aquifer susceptibility to nitrate can be assessed by predicting the probability of an event using a model based solely on significant variables which influence contaminant transport (i.e., well depth and surficial geology). Unlike the vulnerability model to be discussed later, the probabilities calculated in this model do not indicate the risk that a well will have elevated nitrate concentrations as this will be dependent on land-use activities in the area. Rather, these probabilities are used as a relative measure of the likelihood that a well in this environment will have elevated nitrate concentrations, *if* a nitrate source was present.

The combined effects of well depth and surficial geology on estimated probabilities of an event were evaluated by developing separate models for each type of surficial geology (Figure 5A, Table 3). This approach is preferable to a single model for all wells because it allows for the possibility that different relations may occur between events and well depth for each surficial geology type. To provide a spatial representation of aquifer susceptibility, a probability map for each well depth of interest can be constructed using the relations shown in Figure 5A and the surficial geology map in Figure 1.

Prior to using the results of the logistic regression fit to assess aquifer susceptibility, the predicted probabilities for each of the surficial geology models were combined into a single data set to allow for an overall assessment of model fit. The overall model fit is shown in Figure 6a, which plots the average predicted probability of an event with the observed percentage of events for each decile of risk. A model that exactly predicts the percentage of events would have all points falling along a line with a slope of 1 and an intercept of zero; this "exact model" line is also shown in Figure 6a. The degree to which data fall along this line is a measure of how well the model estimates the probabilities of events. The coefficient of determination (r^2) for the "exact model" fit to the data was used to quantify goodness-of-fit. All r^2 values were determined using the mean-corrected sum of squares. Predicted probabilities calculated for the WDOH wells resulted in an r^2 value of 0.95. For comparison, an r^2 value of 0.76 is calculated when the single variable model for well depth is used to predict the probability of an event. Although this r^2 value indicates that the model provides a good fit of the WDOH data, fitting the verification data set (NWIS) provided a more rigorous test. The probability of an event for each NWIS well was calculated with surficial geology and well depth information and the regression coefficients estimated from fitting the WDOH data. Predicted probabilities calculated for the NWIS wells resulted in an r^2 value of 0.66 (Figure 6a).

The relation between well depth, surficial geology, and elevated nitrate concentrations shown in Figure 5A can be

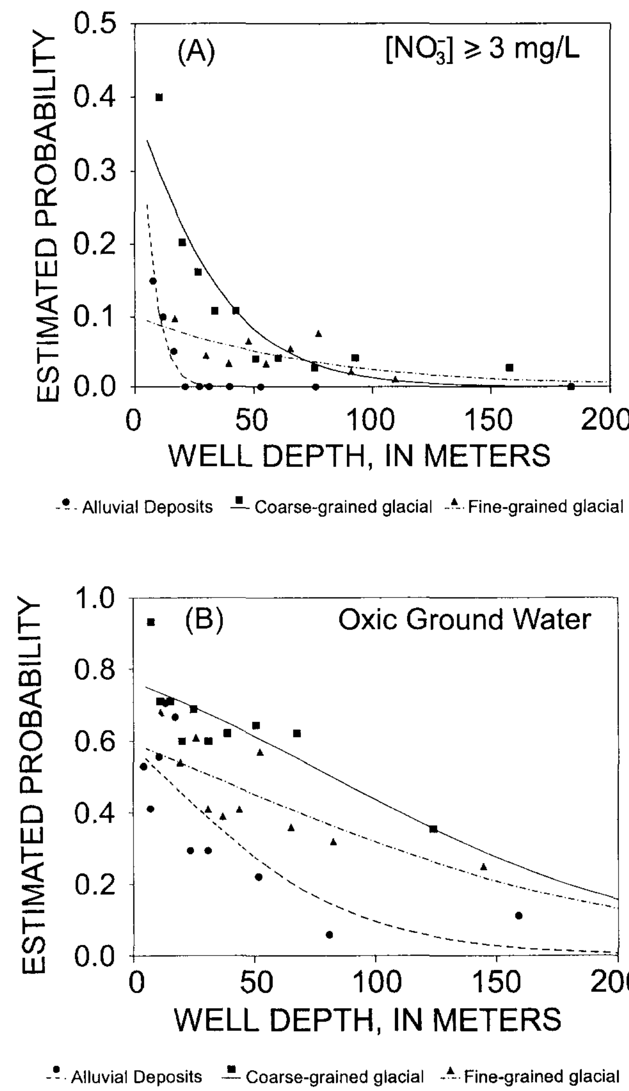

Fig. 5. (A) Lines depict the estimated probability that nitrate concentrations are ≥ 3 mg/l vs. well depth for each surficial geology type. Data points show the fraction of WDOH wells that have nitrate ≥ 3 mg/l vs. mean well depth for each decile of risk. (B) Lines depict estimated probability that dissolved oxygen concentrations are 1 mg/l in the NWIS data set. Data points represent the fraction of NWIS wells that have dissolved oxygen concentrations ≥ 1 mg/l vs. mean well depth for each decile of risk.

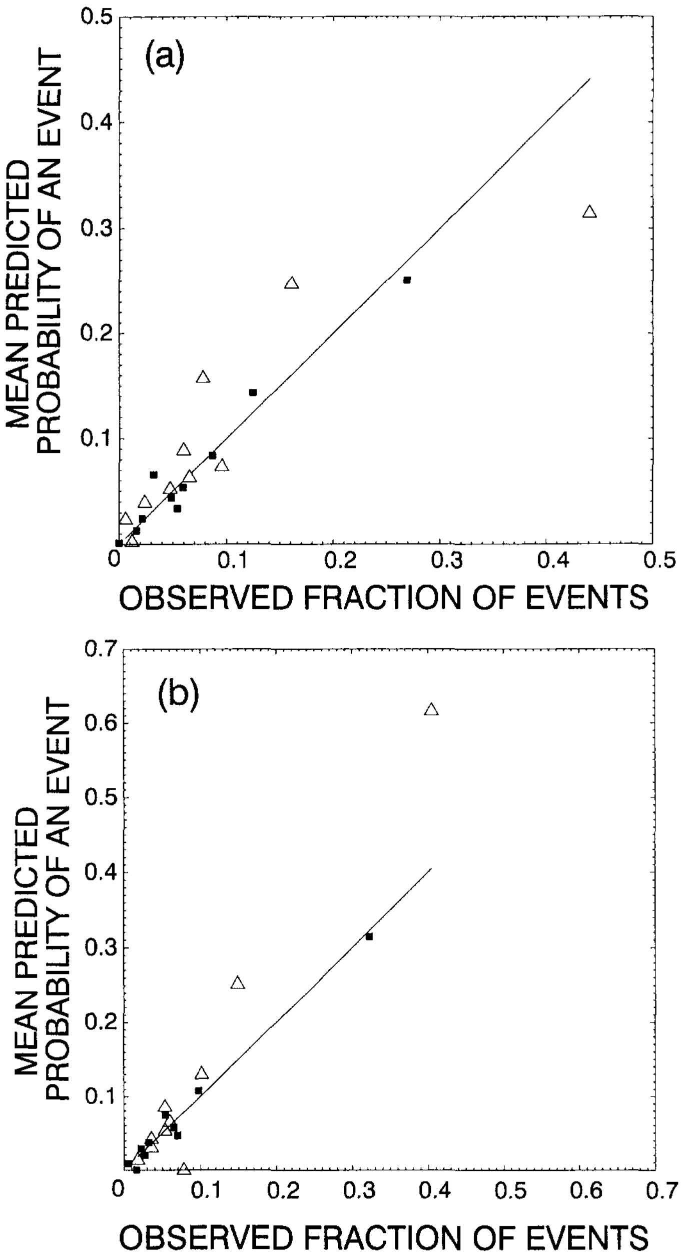

Fig. 6. The average predicted fraction of wells with nitrate concentrations ≥ 3 mg/l vs. observed for each decile of risk: aquifer susceptibility model (6a) and ground-water vulnerability model (6b). Boxes and trianges display WDOH and NWIS data respectively. Predicted = observed lines are also shown.

largely explained by the predominant flow paths in the region; in some cases nitrate reduction may also play a role. Figure 7 presents a generalized cross section and flow paths for an alluvial valley in the region (adapted from Vaccaro et al., 1997). Wells in areas with coarse-grained deposits at the surface (Point A in Figure 7) have the highest probabilities of elevated nitrate as they are more likely to receive recharge from the vicinity of the well than are wells with fine-grained deposits at the surface. Wells installed in areas with fine-grained deposits at the surface are apt to be screened in the underlying coarse-grained deposits and may represent an intermediate part of the flow system (Point B). Very shallow alluvial wells are also likely to tap into recharge from local flow systems and therefore have higher probabilities. However, as well depths increase, wells in the alluvial deposits increasingly intercept ground water from regional flow paths which are away from their recharge area (Point C), resulting in low probabilities of events in these wells. For a given surficial geology type, deeper wells intercept longer flow paths (Points D, E, and F).

Nitrate is generally very mobile in oxic ground water (refers to ground water with ≥ 1 mg/l of dissolved oxygen) but under anoxic conditions may be reduced. Thus the presence of oxic ground water can be used as indicator of environments which are more susceptible to elevated nitrate concentrations. Unfortunately, dissolved oxygen data were not available for the WDOH wells, so adding it as a variable to the nitrate model was not possible. Logistic regression plots for dissolved oxygen data from the NWIS data set (Figure 5B) provides some clues as to the role that nitrate reduction plays in the relations between elevated nitrate, surficial geology, and well depth (Figure 5A). The probability of an elevated nitrate concentration decreases much more quickly with depth in the coarse-grained glacial deposits than does the probability of oxic water, suggesting that nitrate reduction is not a major factor limiting the transport of nitrate in this setting. In the alluvial deposits, dissolved oxygen probabilities decrease quickly with depth suggesting that some of the rapid decrease in nitrate probabilities with depth may be due to nitrate reduction.

In summary, shallow wells with coarse-grained glacial surficial deposits are the most susceptible to elevated nitrate concentration both because they tend to receive water with short flow paths, and these parts of the aquifer system are more likely to have oxic water. Conversely, wells with alluvial and fine-grained glacial surficial deposits likely represent longer flow paths and are more likely to have waters which favor nitrate reduction.

The methodology described above can also be used to assess aquifer susceptibility to other constituents (e.g., volatile organic compounds and pesticides) by using the detection of these constituents as the outcome variable. When this is not possible, using a model based on the occurrence of elevated nitrate to evaluate aquifer susceptibility to other constituents would require the assumption that the constituent of interest behaves similarly to nitrate. For most constituents this assumption will not be valid. In this case, the nitrate model can only be used to make a qualitative assessment of aquifer susceptibility, based on using nitrate as a conservative tracer. This approach has two major limitations: (1) retardation of the constituent of interest is not included in the assessment, leading to overestimates of susceptibility for areas (e.g., surficial geology types) which retard the constituent more strongly than other areas in the basin, and (2) areas where nitrate reduction has occurred may have aquifer susceptibility estimates for the constituent of interest which are too low.

Ground-Water Vulnerability

To determine the likelihood that ground water from a well is contaminated by nitrate, both natural and anthropogenic factors are considered. The percentages of urban, agricultural, and forest land within the optimal radius of a well and well depth formed the list of possible variables, with models developed for each surficial geology type (Table 3). Both urban and agricultural land uses were added to all models except the model for wells in the fine-grained deposits, for which only urban land use was found to be significant. Soils derived from the fine-grained glacial deposits in the Puget Sound Basin tend to be poor for agricultural purposes. Accordingly, most of the intensive crop and dairy farms in this region are on either the alluvial or coarse-grained deposits. Percent forest land use did not meet the

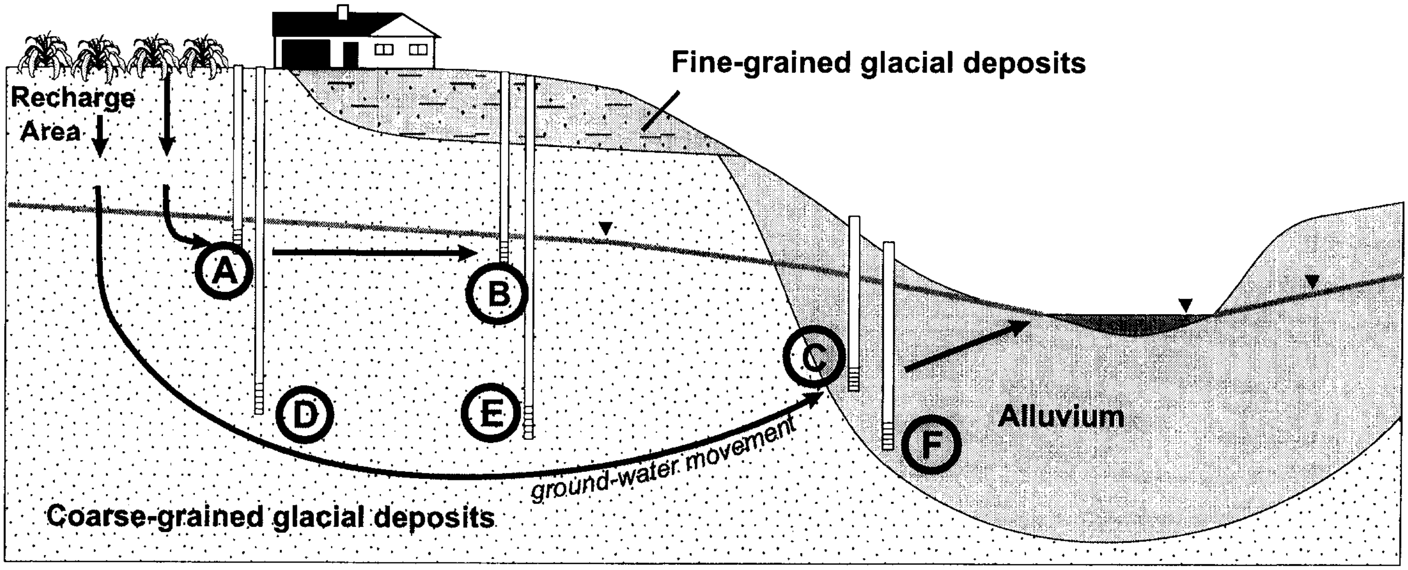

Fig. 7. Generalized cross section for an alluvial valley in the Puget Sound Basin (adapted from Vaccaro et al., 1997). Letters are points along flow paths as described in the text.

significance level for addition to any of the models. This is likely because there are only three major land-use types in this basin. As a result, percent forest land can be largely described by the variables for percent urban and agriculture. The vulnerability model is an improvement over the susceptibility model with regard to its ability to predict the percentage of events in both the WDOH data set ($r^2 = 0.98$, Figure 6b) and the verification data set ($r^2 = 0.79$, Figure 6b).

Maps displaying the probability of elevated nitrate concentrations for ground water withdrawn from 15- and 70-meter deep wells are shown in Figures 8 and 9, respectively. These maps were generated by first determining the percentage of urban and agricultural land use within a 3.2 km radius from the center of each 200 m by 200 m cell. The probability of an event for each cell was then calculated using the regression coefficients in Table 3 and equation (1). Maps of this type should help both water suppliers and water resource managers delineate areas, both spatially and with depth, which are at risk to nitrate contamination. For example, it is apparent that the lower valleys of the Skagit and Nooksack Rivers, which are intensive agricultural areas, are quite vulnerable ($P > 0.4$) to nitrate contamination. Similarly, urban areas over coarse-grained deposits (e.g., south of Tacoma) are also vulnerable to contamination. Vulnerability decreases in all areas with well depth (Figure 9). This model (and the maps) can be updated periodically as new nitrate data become available to monitor changes in ground-water vulnerability over time.

By using the detection of other constituents of interest such as pesticides and volatile organic compounds as the outcome variable, ground-water vulnerability to these constituents can be assessed. When this is not possible, using the nitrate vulnerability model to provide a relative measure of ground-water vulnerability to other constituents would require that (1) the constituent behaves similarly to nitrate, and (2) the sources of nitrate and the constituent of interest are similar. Limitations regarding the first assumption are the same as discussed in the aquifer susceptibility section. With regard to similar sources, pesticides may have the most in common with nitrate as both are often associated with crop cultivation. Unfortunately, nitrate has generally been a poor predictor of pesticide occurrence (see Barbash and Resek, 1996). As such, using a nitrate model to predict pesticide occurrence should be done with caution and not until it is established that the occurrence of nitrate and pesticides in ground water are related in the region of interest.

Summary and Conclusions

Large existing data sets have been used to better understand the factors that influence regional ground-water quality. Logistic regression was used to determine relations between the occurrence of elevated nitrate concentrations in ground water and natural and anthropogenic variables. From these relations, models were developed to assess aquifer susceptibility and ground-water vulnerability to nitrate. This methodology can also be used to evaluate the susceptibility and vulnerability to other constituents by using the detection of these constituents as the outcome variable. Developing aquifer susceptibility and vulnerability models in the manner described here has several advantages:

1. Both the variables and their coefficient values in the model are determined statistically by relating the occurrence of the constituent of interest (in this case, elevated nitrate) to explanatory variables.

2. Aquifer susceptibility and ground-water vulnerability estimates are given as a probability that the constituent will be detected at a specified level, rather than providing relative risk (e.g., high/low). Ground-water vulnerability estimates can then be directly compared to other regions even if different variables were found to be significant. Also, the concentration level for the outcome variable can be changed depending on the objective of the study (e.g., determining areas where ground water exceeds drinking water standards).

3. Aquifer susceptibility and ground-water vulnerability estimates are provided both spatially and with depth. Changes in ground-water vulnerability over time can be evaluated by updating probability estimates by using newly collected water quality and land-use data.

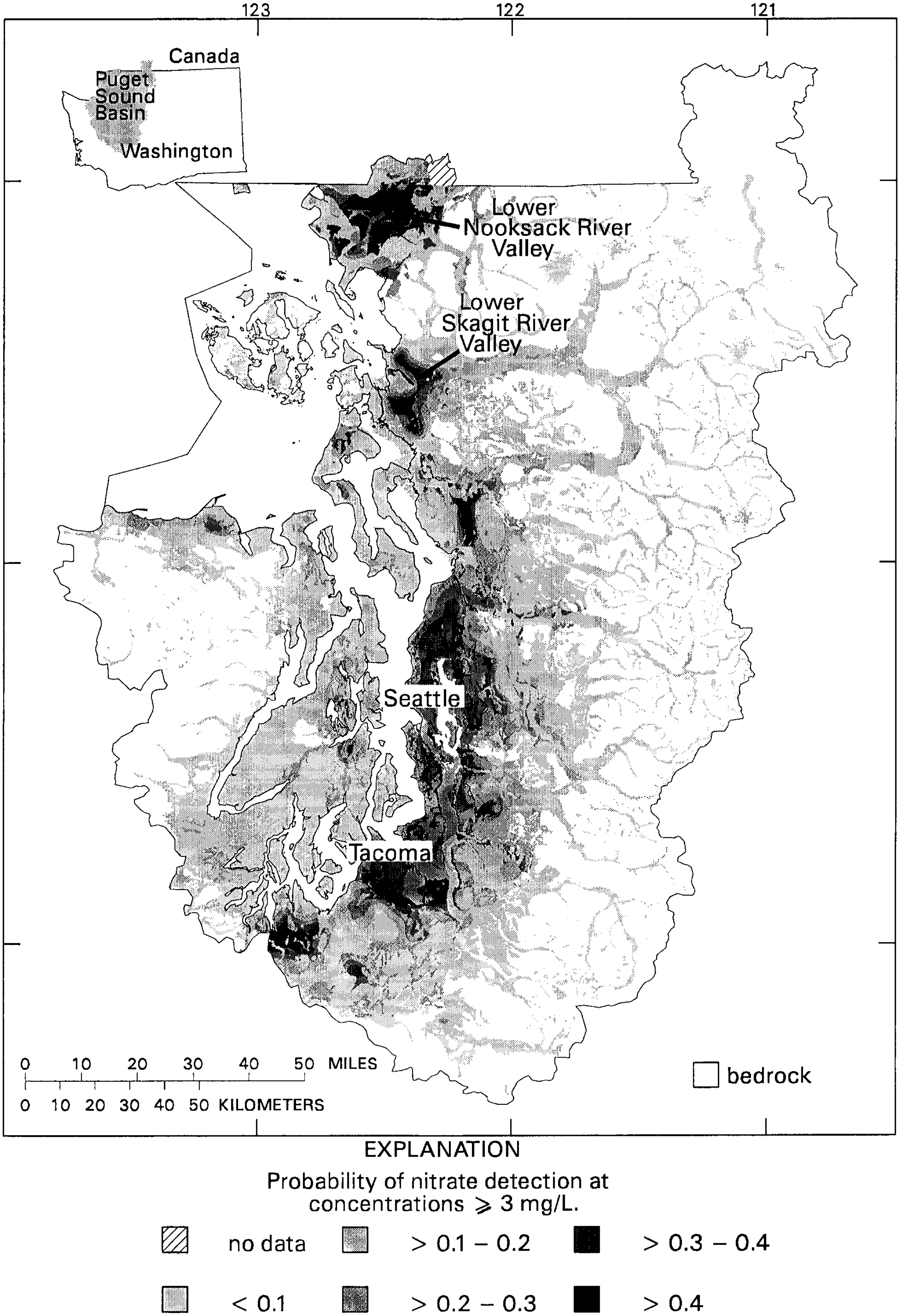

Fig. 8. Map depicts the probability that nitrate concentrations in ground water withdrawn from 15 meter deep wells are ≥ 3 mg/l.

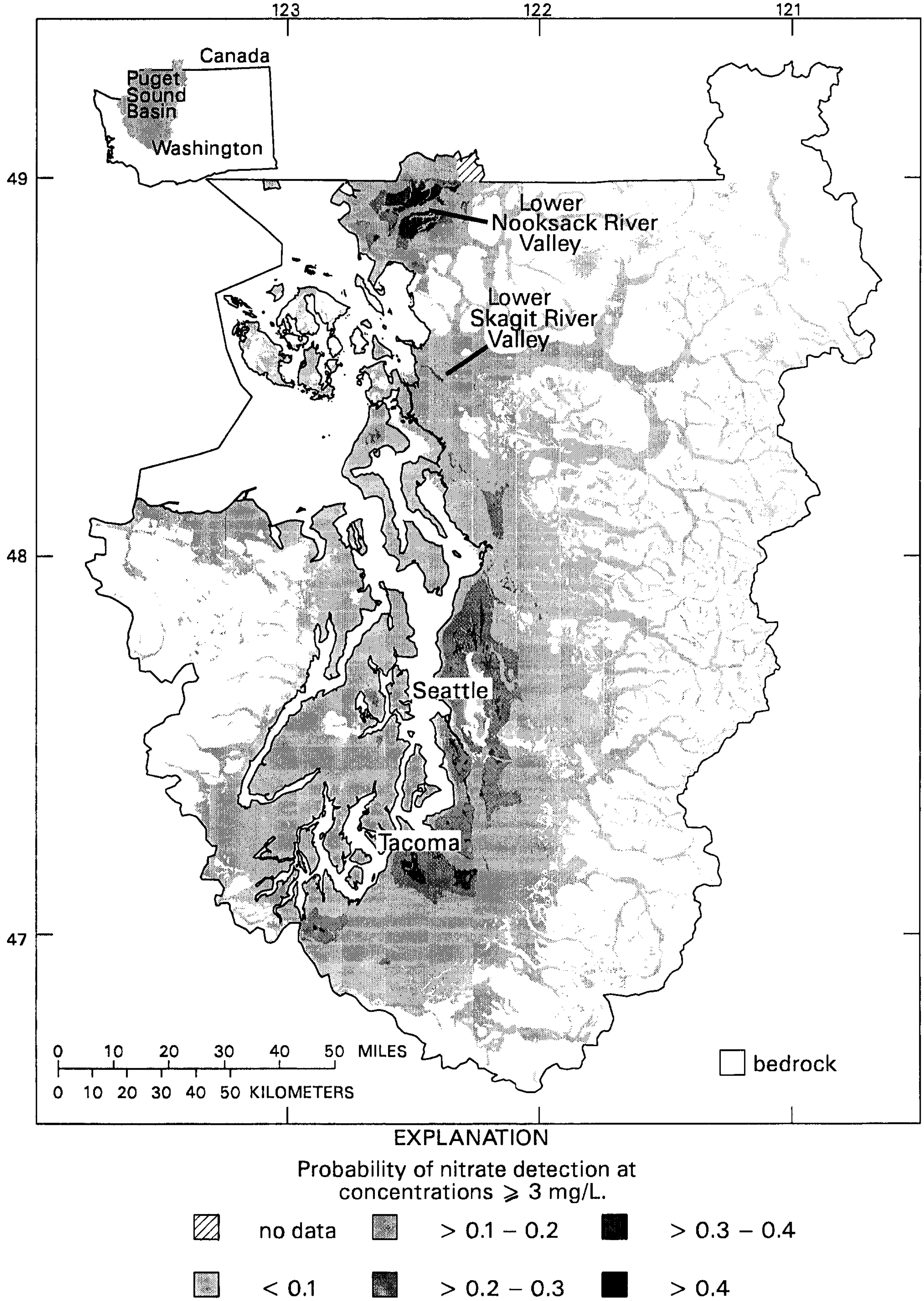

Fig. 9. Map depicts the probability that nitrate concentrations in ground water withdrawn from 70 meter deep wells are ≥ 3 mg/l.

Acknowledgments

The authors thank David K. Mueller of the U.S. Geological Survey for helpful discussions regarding logistic regression, and Ginny Stern of the Washington State Department of Health (WDOH) for her help with the WDOH data set.

References

Aller, L., T. Bennett, J. H. Lehr, R. J. Petty, and G. Hackett. 1987. DRASTIC: A standardized system for evaluating ground water pollution using hydrogeologic settings. U.S. Environmental Protection Agency, Ada, OK. EPA/600/2-87/035. 622 pp.

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer. 1976. A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Prof. Paper 964. 28 pp.

Barbash, J. E. and E. A. Resek. 1996. Pesticides in Ground Water. Ann Arbor Press, Chelsea, MI.

Brown and Caldwell. 1985. Clovers/Chambers Creek Geohydrologic Study. Prepared for the Tacoma-Pierce County Health Department.

Cox, D. R. and E. J. Snell. 1989. Analysis of Binary Data. Chapman Hall, New York.

Eckhardt, D.A.V. and P. E. Stackelberg. 1995. Relation of groundwater quality to land use on Long Island, New York. Ground Water. v. 33, no. 6, pp. 1019-1033.

Fegeas, R. G., R. W. Claire, S. C. Guptill, K. E. Anderson, and C. A. Hallam. 1983. Land use and land cover digital data. U.S. Geological Survey Circular 895-E.

Gilliom, R. J., W. M. Alley, and M. E. Gurtz. 1995. Design of the National Water-Quality Assessment program: Occurrence and distribution of water-quality conditions. U.S. Geological Survey Circular 1112.

Hearne, G., M. Wireman, A. Campbell, S. Turner, and G. P. Ingersall. 1992. Vulnerability of ground water in the greater Denver area. U.S. Geological Survey Water Resources Investigation Report 92-4143.

Helsel, D. R. and R. M. Hirsch. 1992. Statistical Methods in Water Resources. Elsevier, New York.

Hosmer, D. W. and S. Lemeshow. 1989. Applied Logistic Regression. John Wiley and Sons, New York.

Johnson, D. W. 1992. Nitrogen retention in forest soils. J. of Environmental Quality. v. 21, pp. 1-12.

Jones, M. A. 1997. Geologic framework for the Puget Sound aquifer system, Washington and British Columbia. U.S. Geological Survey Prof. Paper 1424-C (in press).

Kalinski, R. J., W. E. Kelly, I. Bogardi, R. L. Ehrman, and P. D. Yamamoto. 1994. Correlation between DRASTIC vulnerabilities and incidents of VOC contamination of municipal wells in Nebraska. Ground Water. v. 32, pp. 31-34.

Kolpin, D. W., M. R. Burkart, and E. M. Thurman. 1994. Herbicides and nitrate in near surface aquifers in the midcontinental United States. U.S. Geological Survey Water-Supply Paper 2413.

Koterba, M. T., W.S.L. Banks, and R. J. Shedlock. 1993. Pesticides in shallow groundwater in the Delmarva peninsula. J. Environ. Qual. v. 22, pp. 500-518.

Lemeshow, S., D. Teres, J. S. Avrunin, and H. Pastides. 1988. Predicting the outcome of intensive care unit patients. J. of the Amer. Statistical Assoc. v. 83, pp. 348-356.

Lemme, G., C. G. Carlson, R. Dean, and B. Khakural. 1990. Contamination vulnerability indices: A water quality planning tool. J. of Soil and Water Conserv. v. 2, pp. 349-351.

Madison, R. J. and J. O. Brunett. 1985. Overview of the occurrences of nitrate in groundwater of the United States. U.S. Geological Survey Water Supply Paper 2275. pp. 93-105.

McKay, D. M. and J. A. Cherry. 1989. Groundwater contamination: Pump and treat remediation. Environ. Sci. and Technol. v. 23, pp. 630-636.

Mercer, J. W., D. C. Skipp, and D. Giffin. 1990. Basics of pump-and-treat ground-water remediation technology. U.S. Environmental Protection Agency, Ada, OK. EPA/600/8-90/003. 31 pp.

Molenaar, D., P. Grimstad, and K. L. Walters. 1980. Principal aquifers and well yields in Washington. Washington Dept. of Ecology Geohydrologic Monograph 5.

Mueller, D. K., P. A. Hamilton, D. R. Helsel, K. J. Hitt, and B. C. Ruddy. 1995. Nutrients in ground water and surface water of the United States—An analysis of data through 1992. U.S. Geological Survey Water Resources Investigations Report 95-4031.

SAS Institute. 1990. SAS/STAT User's Guide. SAS Institute, Inc., Cary, NC.

Singh, B. and G. S. Sekhon. 1978. Nitrate pollution of groundwater from farm use of fertilizers—A review. Agriculture and the Environ. v. 4, pp. 207-225.

Soil Conservation Service. 1993. State soil geographic data base (STATSGO): Data users guide. U.S. Dept. of Agriculture, Soil Conservation Service, Miscellaneous Publication Number 1492.

Spalding, R. F. and M. E. Exner. 1993. Occurrence of nitrate in groundwater—A review. J. Environ. Qual. v. 22, pp. 392-402.

Truett, J., J. Cornfield, and W. Kannel. 1967. A multivariate analysis of the risk of coronary heart disease in Framingham. J. of Chronic Diseases. v. 20, pp. 511-524.

U.S. Environmental Protection Agency. 1993. A review of methods for assessing aquifer sensitivity and ground water vulnerability to pesticide contamination. EPA/813/R-93/002. 147 pp.

Vaccaro, J. J., A. J. Hansen, and M. A. Jones. 1997. Hydrogeologic framework of the Puget Sound aquifer system, Washington and Canada. U.S. Geological Survey Professional Paper 1424-D (in press).

Vowinkel, E. F. and W. A. Battaglin. 1988. Methods of evaluating the relation of ground-water quality to land use in a New Jersey coastal plain aquifer system. U.S. Geological Survey Water Resources Investigation Report 88-4220.