ground water

Predicting Ground Water Nitrate Concentration from Land Use

by Kristin K. Gardner[1] and Richard M. Vogel[2]

Abstract

Ground water nitrate concentrations on Nantucket Island, Massachusetts, were analyzed to assess the effects of land use on ground water quality. Exploratory data analysis was applied to historic ground water nitrate concentrations to determine spatial and temporal trends. Maximum likelihood Tobit and logistic regression analyses of explanatory variables that characterize land use within a 1000-foot radius of each well were used to develop predictive equations for nitrate concentration at 69 wells. The results demonstrate that historic nitrate concentrations downgradient from agricultural land are significantly higher than nitrate concentrations elsewhere. Tobit regression results demonstrate that the number of septic tanks and the percentages of forest, undeveloped, and high-density residential land within a 1000-foot radius of a well are reliable predictors of nitrate concentration in ground water. Similarly, logistic regression revealed that the percentages of forest, undeveloped, and low-density residential land are good indicators of ground water nitrate concentration >2 mg/L. The methodology and results outlined here provide a useful tool for land managers in communities with shallow water tables overlain with highly permeable materials to evaluate potential effects of development on ground water quality.

Introduction

Over the past decade, Nantucket Island, located 30 miles off the south shore of Cape Cod, Massachusetts (Figure 1), was the fastest growing county in the state of Massachusetts. Most growth occurred as sprawl outside the two town centers, serviced by municipal sewers and municipal water. Now, population is scattered in varying densities across the island, serviced by on-site septic systems and private wells. The sole source of potable water for the nearly 10,000 year-round island residents and 50,000 summer residents is the underlying aquifer. This important resource is vulnerable to contamination by chemicals resulting from human activities.

Nitrate is a ground water contaminant of particular concern because of its high leachability in soils and correlation with development; the primary sources of ground water nitrate are domestic on-site sewage disposal and fertilizer (Puckett 1994). Because elevated concentrations of nitrate in drinking water can cause low oxygen levels in the blood of infants, known as methemoglobinemia, a potentially fatal condition (Coinly 1945), and have been associated with the occurrence of non-Hodgkin's lymphoma (Weisenberger 1990), the U.S. EPA (1995) has established a maximum contaminant level of 10 mg/L nitrate. Increased nitrogen levels also detrimentally affect coastal waters by expediting eutrophication, which impedes growth of submerged aquatic vegetation, necessary for nursery, spawning, and feeding for many species and by decreasing water's dissolved oxygen (D.O.) level, which is required by fish and shellfish. Determining where ground water is at risk of nitrate contamination can help land managers build aquifer protection strategies, which protect the water resources and health of humans.

In this study, historical ground water nitrate data, 798 samples collected island-wide from August 1985 to September 2000, combined with nitrate concentrations in 69 wells sampled for nitrate in August 2001 were used to determine a relationship between ground water quality and land use. Land use affects the quality of water in aquifers overlain by highly permeable material because land use determines the types and amounts of chemicals introduced at the land surface. Nitrate concentrations

[1]Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59715; (406) 539-2350; gardner@montana.edu

[2]WaterSHED Center, Department of Civil and Environmental Engineering, Tufts University, Medford, MA; (617) 627-4260; fax: (617) 627-3994; richard.vogel@tufts.edu

Received October 8, 2002; accepted May 7, 2004.

Published in 2005 by the National Ground Water Association.

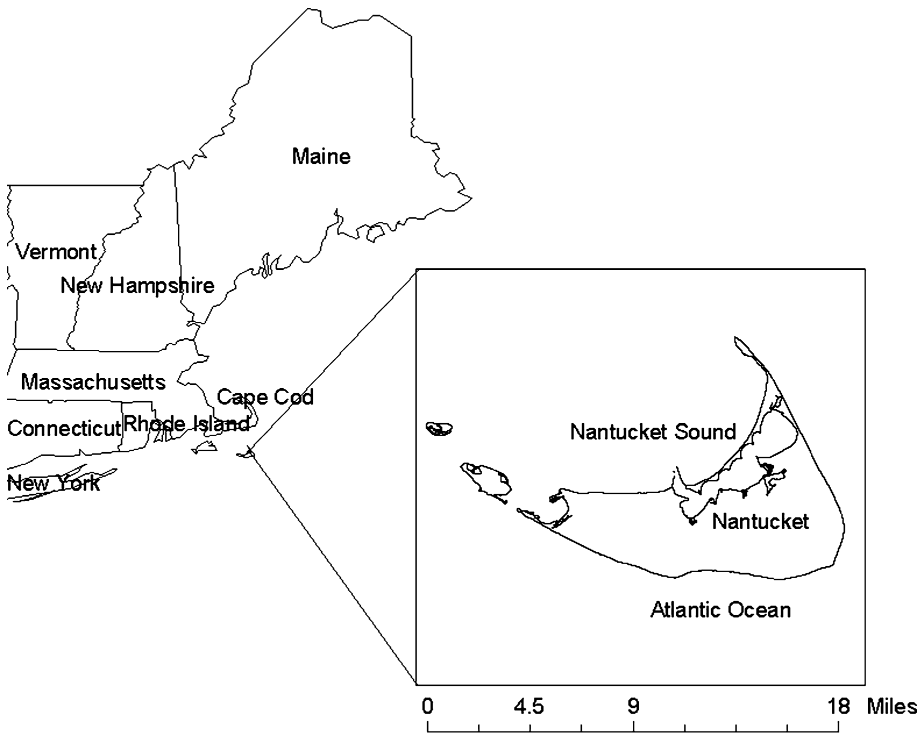

Figure 1. Study location.

measured at 20% of the wells were "censored," meaning the concentrations were reported as below the analytical laboratory reporting limit, or censoring threshold. For censored data, standard regression analysis, such as ordinary least squares, should not be used because it produces biased and inconsistent model parameter estimates (Kroll and Stedinger 1999). Alternatively, regression methods for censored data, Tobit and logistic models, were used (Helsel 1990; Helsel and Hirsch 1992). For both of these models, we employed maximum likelihood estimators of the slope and intercept. Tobit regression predicts a continuous dependent variable, whereas logistic regression predicts a discrete response. In this study, logistic regression predicted the likelihood of ground water nitrate exceeding 2 mg/L, while Tobit regression predicted the actual nitrate concentration at each well. The concepts and methods implemented in this investigation can be used by land managers and extrapolated to other communities with shallow water tables overlain with highly permeable material.

Previous Studies

Many previous studies have correlated land use with water quality. Persky (1986) used ordinary least squares regression to show that the concentration of nitrate-nitrogen in ground water was positively correlated to the density of houses on Cape Cod. For censored data, several studies have used Tobit regression or logistic regression to determine significant sources of ground water nitrate. Yen et al. (1996) used an inverse hyperbolic sine Tobit model to determine the primary factors that affect nitrate concentrations in near-surface aquifers, based on nationwide data from the USGS collected in 1991. Factors found to have significant impacts on nitrate concentrations were well screen interval, depth to top of aquifer; percentages of urban-residential, forest, and pasture land within 3.2 km; D.O. concentration level; and the presence of a chemical facility or an animal feedlot. Estimated coefficients for one of the variables, animal feed lots, used in the regression analysis did not have an expected positive relationship with nitrate. Lichtenberg and Shapiro (1997) used Tobit regression to construct statistical relationships between land use and well water quality in Maryland community water system wells. Well depth was inversely related to nitrate levels, while unconfined formation, limestone, number of septic systems, and number of chicken farms directly contributed to nitrate concentration. Again, some of the explanatory variables did not relate as expected to ground water nitrate, and others were not significant.

Tesoriero and Voss (1997) used logistic regression to assess aquifer susceptibility and ground water vulnerability to nitrate concentrations >3 mg/L by prediction of the fraction of events occurring in Puget Sound Basin in northwestern Washington. Significant explanatory variables were well depth, surficial geology, and the percentage of urban and agricultural land within a 3.2-km radius of a well. Nolan (2001) developed a multivariate logistic regression model, which predicts the probability of exceeding 4 mg/L of nitrate in nationwide ground water. Significant explanatory variables were (1) nitrogen fertilizer loading; (2) percent cropland-pasture; (3) population density; (4) percent well-drained soils; (5) depth of the seasonally high water table; and (6) presence or absence of a bedrock fracture zone within an aquifer. In Long Island, New York, Eckhardt and Stackelberg (1995) used logistic regression to predict the probability of exceeding

3 mg/L of nitrate in ground water wells beneath agricultural, suburban, and undeveloped areas within a half-mile radius of the wells. The logistic regression models developed for nitrate had rank correlation coefficients of 0.87 to 0.88 and indicated that nitrate concentration generally increased with population density and amount of residential and agricultural land use and decreased as the depth to the water table increased.

Our study employed both logistic and Tobit regression methods to determine the relationship between ground water nitrate and land use. Both models improve upon those of previous studies by providing (1) more explanatory power and (2) relationships between explanatory variables and ground water nitrate, which are in accord with our physical knowledge of the problem. The results of both statistical methods indicate that land with limited human impact is negatively correlated with ground water nitrate. Therefore, efforts to preserve ground water quality by land managers should concentrate on land conservation techniques.

Multivariate Regression Models

Multivariate regression methods are employed here because there was no clear bivariate relationship between ground water nitrate concentration and any of the explanatory variables. Twenty percent of the nitrate concentrations were censored; therefore, Tobit and logistic models were developed with separate objectives. The Tobit model was used to predict a continuous variable, while the logistic model predicted a binary response defined by the occurrence of ground water nitrate concentration in excess of 2 mg/L.

The Tobit Model

Censored response data, represented as a range from zero to the measurement threshold, are integrated with uncensored observations in a procedure known as Tobit regression. The Tobit model (Tobin 1958) has been widely used by economists in modeling economic relations of censored observations. Kroll and Stedinger (1999) compared the performance of Tobit and ordinary least squares regression in a regional regression model of censored streamflow data using Monte Carlo simulation. Kroll and Stedinger concluded that ordinary least squares techniques performed poorly when compared to the Tobit model. The Tobit model is written in terms of the underlying latent variable, unobservable unless the nitrate concentration is above the censoring threshold

$$y_i^* = \ln(N_i)^* = \alpha + \beta \ln(\boldsymbol{x}_i) + \epsilon_i$$

$$y_i = y^* \text{ if } N_i > c$$

$$y_i = c \text{ otherwise} \tag{1}$$

where y_i^* is the latent dependent variable, N_i is the observed nitrate concentration, y_i is observed dependent variable, α is a constant, c is the measurement threshold, β is a vector of parameter slope estimates, $\ln(\mathbf{x})$ is a vector of independent explanatory variables, and ϵ_i is an error term assumed to be independently and normally distributed with a mean of zero and variance σ^2. If the error terms in the Tobit model are assumed to be homoscedastic and independent, the model parameters may be efficiently estimated by the method of maximum likelihood (Amemiya 1985).

In this study, model variables were chosen by Type 3 analysis in SAS statistical software, a process similar to stepwise selection in ordinary least squares regression except a likelihood ratio test statistic is used to select each set of independent variables. In this procedure, variables are added one at a time as long as they significantly contribute to the fit. The Wald chi-square statistic determines the significance of each model parameter in the resulting model. The maximum likelihood estimators for the parameters were derived from the LIFEREG procedure in SAS, by using a Newton-Raphson algorithm. The LIFEREG procedure fits a parametric model to right, left, or interval censored data. The model of the response variable consists of a linear effect composed of the covariates and a random disturbance term. The distribution of the random disturbance can be taken from a class of distributions that includes the extreme value, normal, logistic, and, by using a log transformation, the exponential, Weibull, lognormal, loglogistic, and gamma distributions (SAS Institute Inc. 1999).

Logistic Model

Logistic regression has been used extensively in the health sciences since the late 1960s to predict a binary response from explanatory variables (Lemeshow et al. 1988) and more recently in the environmental sciences to identify variables that significantly affect ground water quality. Binary logistic regression is used to predict a binary response, such as the absence of a specific contaminant above a given concentration threshold, from independent explanatory variables. The probability, p, of being in one of the response categories is modeled. The predicted probability, p, is thus the predicted probability of the response being above the concentration threshold (response equal 1), while $1 - p$ is the predicted probability of the response being a 0. The odds ratio is based on the probability of exceeding a given concentration threshold value:

$$\text{odds ratio} = \left(\frac{p}{1-p}\right) \tag{2}$$

The main assumption of logistic regression is that the natural logarithm of the odds ratio or probability of being in a response category is linearly related to the explanatory variables. Regression coefficients are estimated using the method of maximum likelihood. The log of the odds ratio, the logit, transforms a variable constrained between 0 and 1 into a continuous unbounded variable. The logit is modeled as a linear function of the explanatory variables resulting in:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = b_o + \mathbf{bx} + \epsilon_i \tag{3}$$

where b_o is a constant and **bx** is a vector of slope coefficients and explanatory variables. The probability of contaminant exceeding a threshold is found by solving Equation 3 for p

$$p = \frac{e^{(b_o + \mathbf{bx})}}{1 + e^{(b_o + \mathbf{bx})}} \tag{4}$$

The overall likelihood ratio statistic (G) tests the significance of the explanatory variables in the model

$$G = -2(L_{\text{int}} - L_{\text{model}}) \tag{5}$$

where L_{int} is the log likelihood of intercept-only model and L_{model} is the log likelihood of model with one or more explanatory variables. The G statistic follows a chi-square distribution, with the number of degrees of freedom equal to the number of slope parameters, under the null hypothesis that slope coefficients for the explanatory variables in the model equal 0. For nested logistic models, the G statistic determines the significance of adding one or more new explanatory variables to the model. The number of degrees of freedom equals the number of additional variables in the more complex model. For comparison of non-nested logistic regression models, the partial likelihood ratios are not appropriate and, therefore, the Akaike's information criteria (AIC) is used

$$\text{AIC} = -L + k \tag{6}$$

where L is the log likelihood and k is the number of explanatory variables, adding a penalty for the addition of extra variables. Better models are those with a small AIC (Helsel and Hirsch 1992). A classification table summarizes the accuracy of the model by comparing the predicted and observed probabilities.

Study Area

Hydrogeology?

The island of Nantucket is ~16 miles long by 4 miles wide, overlying one major aquifer. Surficial geologic deposits on Nantucket Island are composed of Wisconsin age glacial moraine up to 300 feet thick. The northern portion of the island includes moraine and ice-contact deposits consisting of a mixture of bouldery sands, clayey sands, silts, and clays that form highlands up to 110 feet above present-day sea level. To the south and west are layered gravel and glacial outwash sand deposits. Relatively low-permeability clay lenses have been found along the island's northern shoreline but are laterally discontinuous (Oldale 1985). Topography is relatively flat, elevation ranges from 108 to 0 feet above sea level.

Water table elevations vary from ~12 feet above sea level in the widest part of the island, declining toward the coast to sea level (Masterson and Barlow 1994; Horsley, Witten, and Hegeman Inc. (1990); Knott and Olimpio 1986) (Figure 2). Water table elevations across the island are sustained by recharge derived from precipitation, which averages 46.8 inches a year (Walker 1980). Evapotranspiration calculation using the Thornthwaite model (Dingman 1994) and tritium data (Knott and Olimpio 1986) indicates ~42% of the annual precipitation recharges the fresh water aquifer. The distributions of the unsaturated thickness and well screen depths below the water table of wells sampled in August 2001 are illustrated using boxplots in Figure 3. The median screen depth below the water table is 28 feet, while the median unsaturated thickness at the wells is 24 feet. The horizontal component ground water flow at the wells is ~1 to 2 ft/d (Horsley 2000).

Land Use

Land use was determined from a 1999 MassGIS statewide land-use map interpreted from 1:25,000 aerial color infrared photography. The University of Massachusetts, Department of Forestry Resource Mapping Project, followed by a MassGIS quality assurance/quality checking routine, completed photo interpretation and digitizing. Because data were not available to establish the upgradient direction of each well sampled, land use within a 1000-foot radius of each well was identified through Geographic Information Systems (GIS) analysis (Figure 4). Barringer et al. (1990) found that the use of circular areas around water table wells is a simple and effective method for correlating land use and water quality. Because land-use patterns around the study area are fairly uniform and ground water in the aquifer flows at an average of 1 to 2 ft/d and moves <1000 feet in 2.5 years, a radius of 1000 feet was used. The ground water quality observed at wells sampled in August 2001 should still reflect the effects of 1999 land use. Although other well radii were examined, the 1000-foot radius performed best in the regression analyses.

Methods

Two ground water nitrate data sets were used in separate analyses: historical nitrate concentrations sampled between 1985 and 2000 obtained from Nantucket Health Department well records and 69 wells sampled for nitrate in August 2001. All wells sampled were analyzed by the Barnstable County Health Lab, a Massachusetts Department of Environmental Protection–certified lab, for nitrate by U.S. EPA Method 300, ion-exchange chromatography, with a detection limit of 0.1 mg/L and a measurement standard deviation of 0.018 mg/L.

Because accurate historical land use and well characteristic information does not exist, exploratory data analysis was performed on the historical nitrate concentrations to determine temporal and spatial trends. The nitrate concentrations measured in wells sampled in August 2001 (Figure 5) were used in regression analysis to determine potential land use and well characteristics that significantly contribute to ground water nitrate. Descriptive statistics of potential explanatory variables are listed in Table 1. Potential explanatory land-use variables explored were percentages of undeveloped land (UND), forest (FRST), paved surface (PAVE), wetland (WET), high-density development (HD) (<1/4 acre lots), medium-density development (MD) (>1/4, <1 acre lots), low-density development (LD)

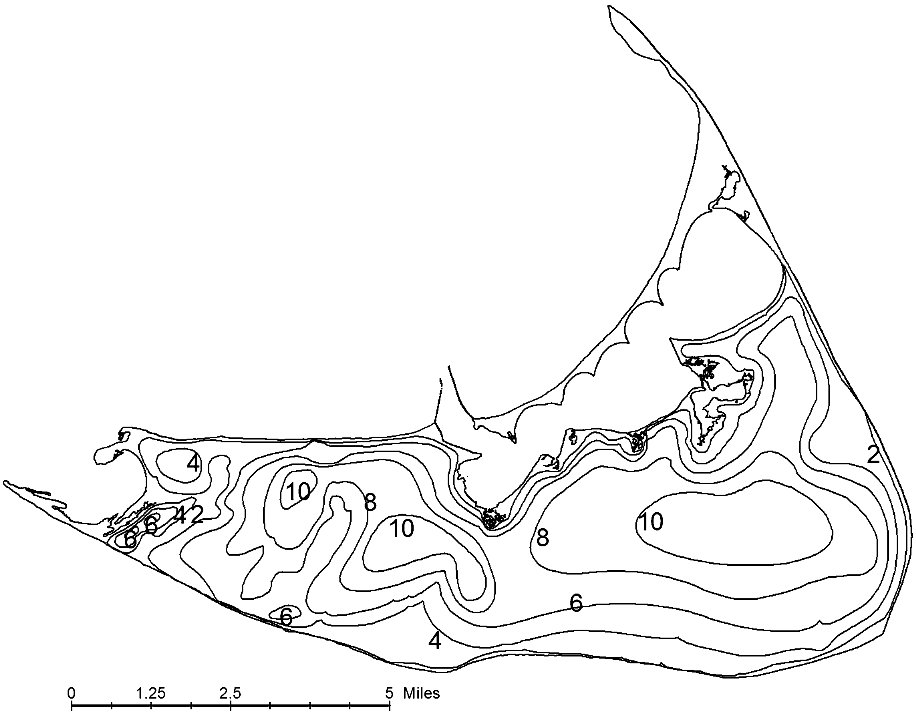

Figure 2. Water table elevations (feet above sea level).

(>1 acre lots), golf course (GOLF), and agriculture (AG), and number of septic tanks (SPTC) within the 1000-foot radius.

Well characteristics considered for regression analysis were the well screen depth below water table and unsaturated thickness. Well screen depth and water table depth, given as depth from surface, were found in well log reports available in town department records. Surface elevations and location at each well site were determined by a Trimble Pro XR GPS unit, with an estimated error of measurement of ±1.5 feet laterally and ±3.0 feet vertically. High nitrate concentrations at or near the water table declining with depth have been observed in many instances and interpreted as an indication of denitrification (Foster and Bath 1983; Andersen and Kristiansen 1984). Depth to the water table is an important factor because it determines the amount of subsurface material nitrate must travel through before reaching the aquifer. Nitrogen transformations are more likely with a larger vadose zone (Canter 1996).

Results

Historic Nitrate Concentrations

Exploratory data analyses indicate that a local farm has influence on ground water nitrate concentrations. Figure 6 uses boxplots to compare ground water nitrate concentrations directly downgradient from a commercial Nantucket farm (noted Agriculture) and ground water nitrate from all other island locations (noted Non-Agriculture). Both the boxplots and summary statistics

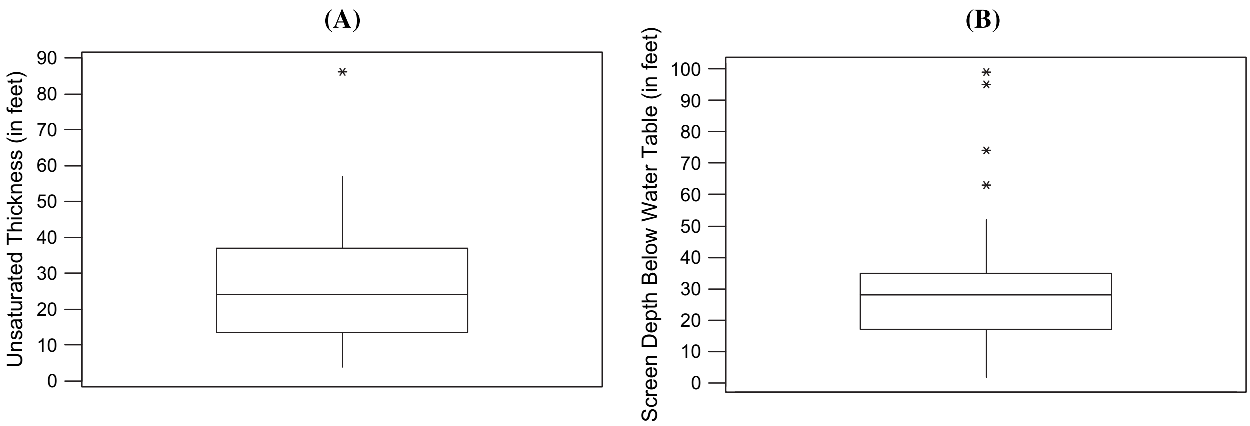

Figure 3. Statistical distributions of (A) unsaturated thickness and (B) screen depth below the water table for wells sampled in August 2001.

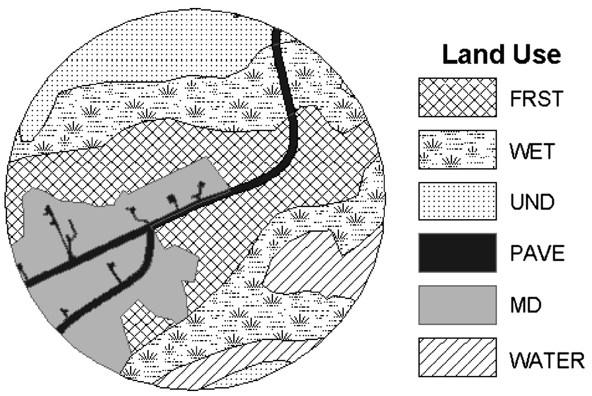

Figure 4. Land use determined within a 1000-foot radius of a sampled well: undeveloped land (UND), forest (FRST), paved surface (PAVE), wetland (WET), and medium-density development (MD).

(Table 2) of historic nitrate concentrations indicate that agricultural land has influenced ground water nitrate concentrations. A 5% level rank sum test, using a large sample approximation, was used to determine if nitrate concentrations are greater in wells near the farm compared to wells located elsewhere on Nantucket. The large sample approximation rank sum test is a nonparametric test chosen for the following reasons: (1) the data appear nonnormal; (2) 20% of nitrate data are censored, <0.1 mg/L; and (3) the sample size is >10 (Helsel and Hirsh 1992).

The hypothesis test is as follows:

H_o: median concentration (agriculture) = median concentration (nonagriculture)

H_a: median concentration (agriculture) > median concentration (nonagriculture)

The test statistic is

$$Z_{rs} = \frac{W_{rs} - d/2 - \mu_w}{\sigma_w} \quad (7)$$

where W_{rs} is the sum of the ranks of the smaller sample group, d is the minimum difference between possible values of the test statistic, $\sigma_W = \sqrt{nm(N+1)/12}$, n is the size of smaller sample, m is the size of larger sample, and $N = n + m$. The subscript w is used to denote that these are the moments of the test statistic W. The null hypothesis, H_o, is rejected if $Z_{rs} > Z_{\alpha=0.05} = 1.645$ from standard a standard normal table. The resulting test statistic, $Z_{rs} = 14.44$, is substantially greater than $Z_\alpha = 1.645$, so the null hypothesis is easily rejected, and the conclusion is that nitrate concentrations in wells downgradient from the farm are significantly different from those in wells located elsewhere on the island. Therefore, historic Nantucket ground water nitrate data support the premise that agricultural land contributes significantly to ground water nitrate concentration.

Present Nitrate Concentrations

Results of the ground water nitrate monitored in August 2001 are displayed in Figure 7 with a lognormal probability plot, ignoring the 16 observations below the detection limit of 0.1 mg/L. A probability plot correlation coefficient of 0.991 implies that one can accept the lognormal hypothesis at the 5% significance level. Therefore, the logarithms of nitrate concentrations were plotted against each potential explanatory variable. The median nitrate concentration was 0.9 mg/L, the maximum concentration was 20.1 mg/L, the minimum concentration was <0.1 mg/L, and the interquartile range was 0.1 to 2.1 mg/L.

Tobit Regression

Explanatory variables (Table 1) were selected using a stepwise process with log ratio statistics. All variables found insignificant at the 5% level were dropped. The

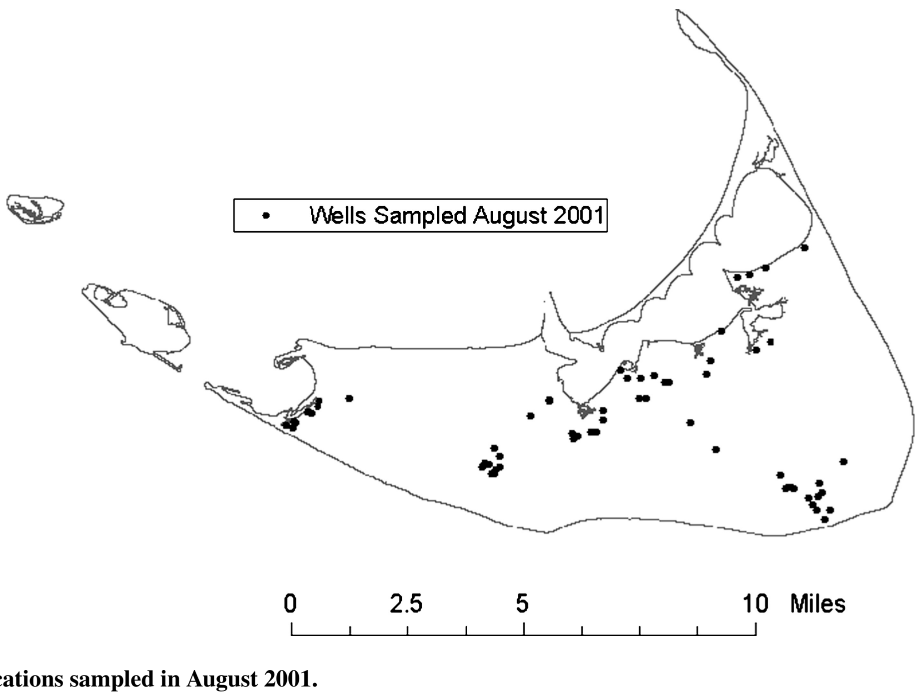

Figure 5. Well locations sampled in August 2001.

Table 1
Land-Use Explanatory Variables and Descriptive Statistics

Variable	Percent of 1000-Foot Buffer Minimum	Median	Maximum	Interquartile Range
Land use				
Undeveloped land	0.00	0.40	0.97	0.14 to 0.62
Forest	0.00	0.00	0.47	0.00 to 0.03
Paved surface	0.00	0.06	0.46	0.03 to 0.08
Wetland	0.00	0.00	0.33	0.00 to 0.03
High-density development	0.00	0.00	0.71	0.00 to 0.00
Medium-density development	0.00	0.06	0.69	0.00 to 0.21
Low-density development	0.00	0.13	0.77	0.03 to 0.26
Golf course	0.00	0.00	0.39	0.00 to 0.00
Agriculture	0.00	0.00	0.56	0.00 to 0.00
Number of septic tanks	0.00	7.00	48.00	1.75 to 23.25
Well characteristics				
Unsaturated thickness	4.00	24.00	86.00	14.00 to 36.00
Screen depth below water table	1.80	28.00	99.00	17.00 to 35.00

estimates of slope and intercept by Tobit regression are based on two important assumptions: (1) homoscedasticity of the error terms and (2) normal distribution of the error terms. The performance of the Tobit model was evaluated based on the error assumptions, the distribution of the Wald chi-square statistics and their associated p values, and a plot of predicted vs. actual nitrate concentration.

The Tobit model yielded the following significant variables: high-density residential (HIGH), number of septic tanks (SPTC), percent of forest land (FRST), percent of undeveloped land (UND), and percent of agricultural land (AG).

$$\mathrm{lN}(N) = 2.74\,\mathrm{HIGH} + 0.042\,\mathrm{SPTC} - 11.26\,\mathrm{FRST} - 3.02\,\mathrm{UND} + 6.36\,\mathrm{AG} \qquad (8)$$

Figure 8 illustrates the spatial distribution of land use in Equation 8. The model parameter estimates along with their standard errors, confidence intervals, Wald chi-square statistics, and associated p values are presented in Table 3. Wald chi-square statistics indicate all variables significant at the 5% level. The value of the Wald chi-square statistic gives a measure of importance of each explanatory variable in the model; therefore, undeveloped land has the most explanatory power in the model, since the Wald chi-square statistic is larger than that of other predictor variables. The relationship between predictors and nitrate concentrations is as expected: nitrate concentrations increase with number of septic tanks and percentage of high-density residential and agricultural land and decrease with percentage of forest and undeveloped land. The required assumptions of the Tobit model are validated: (1) we could not reject the normality hypothesis of residuals at the 5% significance level since the probability plot correlation coefficient $R = 0.993$ is greater than critical value of $R = 0.98$ and (2) the residuals appear homoscedastic. Figure 9 compares the actual vs. predicted nitrate values, which had a correlation of 0.82. We conclude that Equation 8 adequately estimates nitrate concentrations in ground water as a function of five land-use characteristics, with the most accurate predictions between 0.1 and 5.0 mg/L nitrate.

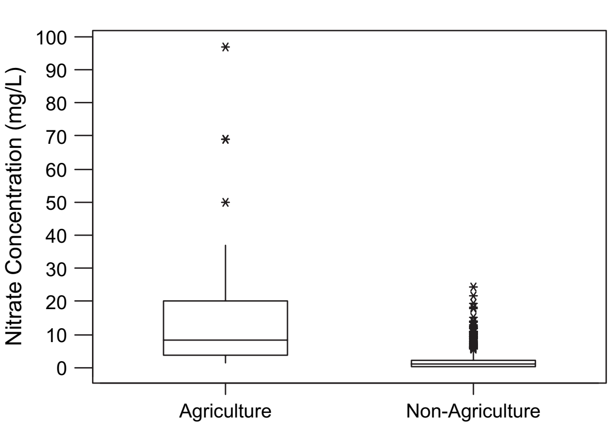

Figure 6. A boxplot comparison of historic ground water nitrate concentrations.

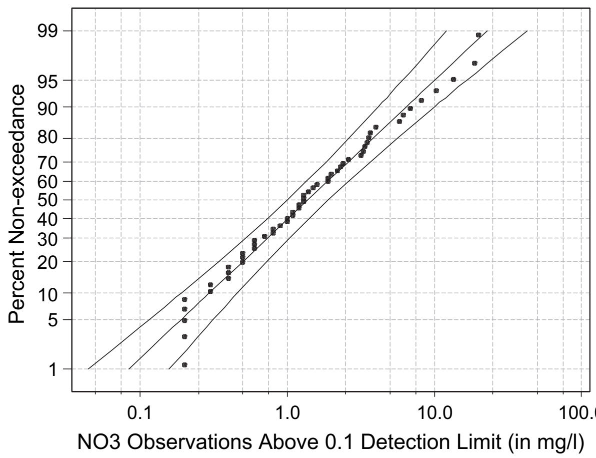

Figure 7. Probability plot of ground water nitrate data sampled in August 2001.

Table 2
Summary Statistics of Historic Ground Water Nitrate Concentrations

	Median	Minimum	Maximum	Lower Quantile	Upper Quantile
Agriculture	8.3	1.4	96.9	3.7	19.2
Nonagriculture	0.80	0.05	24.1	0.2	2.3

Logistic Regression

Binary logistic regression predicts a binary event from continuous explanatory variables. To convert nitrate concentration from a continuous variable to a discrete variable, a concentration threshold must be established to separate events (concentration level greater than or equal to the threshold) from nonevents. In order to assess aquifer susceptibility or vulnerability, this level should represent a concentration that is the result of anthropogenic activities. Meuller and Helsel (1996) have suggested a nitrate level of 2 mg/L as a conservative estimate to represent anthropogenic effects. Therefore, a response is considered a nitrate concentration >2 mg/L, while a nonresponse is a nitrate concentration $\leq$2 mg/L.

A model containing all explanatory variables from Table 1 was developed initially. A stepwise process dropped variables if their p values were >0.05. Nonnested models were compared using AIC and a classification table comparing observed vs. predicted probabilities. The resulting logistic model

$$\mathrm{Prob}[N > 2\,\mathrm{mg/L}] = \frac{\exp[6.33 - 6.61\,\mathrm{FRST} - 15.12\,\mathrm{UND} - 9.92\,\mathrm{LD}]}{1 + \exp[6.33 - 6.61\,\mathrm{FRST} - 15.12\,\mathrm{UND} - 9.92\,\mathrm{LD}]} \quad (9)$$

includes percentage of forest (FRST), undeveloped (UND), and low-density residential (LD) land as predictors of the probability of nitrate concentration >2 mg/L. Equation 9 demonstrates that the direction of relationships between nitrate concentration and the explanatory variables are as expected. The probability of detection of nitrate >2 mg/L decreases with percentage of forest, undeveloped, and low-density residential land. Table 4 describes the results of the logistic model parameter estimates.

The overall likelihood ratio statistic for the model in Equation 9 is 54.70 with 2 degrees of freedom resulting in a p value of 0.0001. This result confirmed that the logistic model is significantly better than simply estimating the proportion of data above the response level without considering these three explanatory variables. Finally, the reliability of the model is evaluated by comparing the predicted and observed responses in Table 5. When nitrate concentrations are <2 mg/L, the logistic model is 97.8% correct, while when nitrate concentrations are >2 mg/L, the logistic model is 94.1% correct. Apparently the logistic model in Equation 9 is quite accurate for predicting the probability that ground water nitrate concentrations exceed 2 mg/L, which implies that unaltered land is the most important predictor of ground water nitrate concentrations among the variables considered.

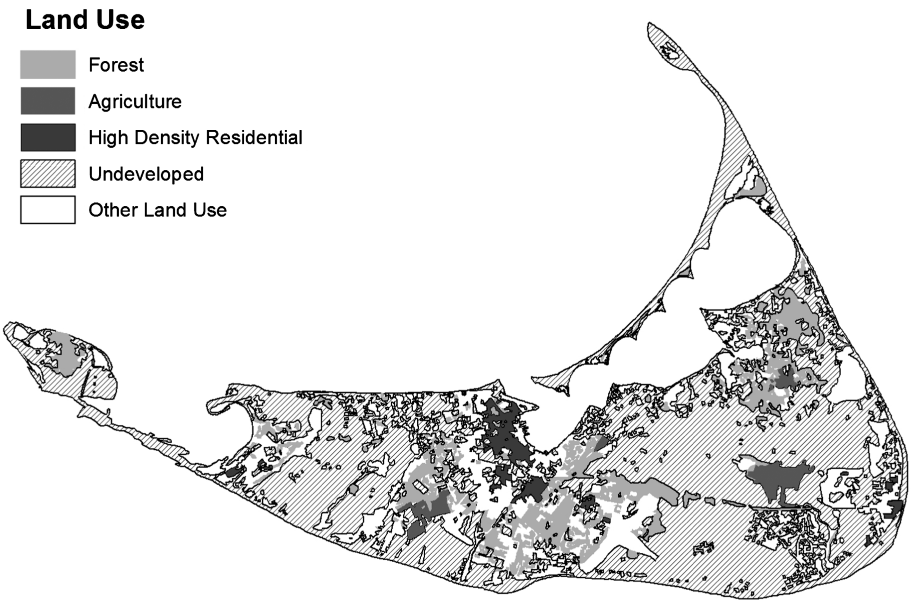

Figure 8. Distribution of significant land-use variables from the Tobit model—Equation 8 (septic is not visible due to large scale).

Parameter	Model Parameter Estimate	Standard Error	Model Parameter 95% Confidence Limits		Wald Chi-Square	Pr > Chi-Square
UND	−3.02	0.30	−3.60	−2.44	103.92	<0.0001
SPTC	0.042	0.0067	0.029	0.056	39.16	<0.0001
IAG	6.36	1.073	3.25	8.46	35.06	<0.0001
HIGH	2.74	0.55	1.66	3.82	24.69	<0.0001
FRST	−11.26	2.88	−16.91	−5.61	15.26	<0.0001

Table 3
Model Parameter Estimates for the Tobit Model

Conclusions

Concern in Nantucket, Massachusetts, has been rising over nitrate leaching into ground water from fertilizers and sewage. In this study, both historic and present ground water nitrate concentrations in Nantucket were correlated to land uses that reflect potential sources of nitrate. Historic nitrate concentrations downgradient from agricultural land had significantly higher median nitrate concentrations than concentrations collected elsewhere on island, 8.25 vs. 0.80 mg/L.

Multivariate statistical analyses indicate that the presence of nitrate in ground water is directly correlated with explanatory factors that describe land use. Tobit regression results demonstrated that the number of septic tanks and the percentages of forest, undeveloped, and high-density residential land within a 1000-foot radius of a well are reliable predictors of nitrate concentration. Logistic regression analyses revealed that the percentages of forest, undeveloped, and low-density residential land areas are excellent predictors of ground water nitrate concentrations in excess of 2 mg/L. The strength of the multivariate correlations indicates that land use is an excellent predictor of ground water quality on Nantucket. Both models introduced here indicate that percent of undeveloped land is inversely correlated with nitrate concentration. These results may be used by the Nantucket planning department to develop land management strategies to attempt to minimize future nitrate pollution in ground water.

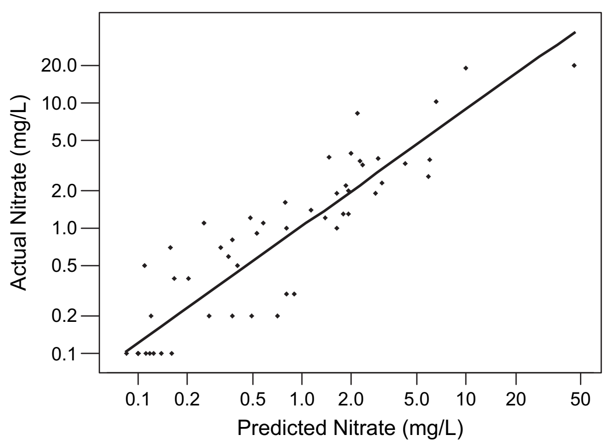

Figure 9. Measured nitrate concentration vs. predicted nitrate concentration based on Tobit regression (Equation 8).

The concepts and methods outlined in this study have broad application to any location where the quality of shallow water table aquifers overlain by highly permeable material is potentially affected by land use. This study determined significant relationships between land use and ground water nitrate concentrations and is potentially useful because it relies on common information and the regression methods are relatively simple, easy to apply, and capture most of the spatial variability of nitrate concentrations. However, there are significant limitations associated with this approach. The regression models express purely statistical associations, which have been observed on the basis of historical data. Since the relations expressed by our models are not scientifically process based, they may have to be refit if land-use conditions on Nantucket should change. Applicability to prediction in other types of ground water systems is uncertain. Whenever possible, knowledge of water table

Table 4
Results of Binary Logistic Regression

Predictor	Estimate	Standard Error	Z statistic	*p* value
Constant	6.33	1.95	3.26	0.001
FRT	−66.61	23.01	−2.89	0.004
UND	−15.12	4.67	−3.24	0.001
LD	−9.19	3.55	−2.59	0.010

Table 5
Observed Vs. Predicted Response for Logistic Model 1

Predicted Outcome Logistic Model 1	True State Of Nitrate Concentration $N \leq 2$ mg/L	True State Of Nitrate Concentration $N > 2$ mg/L
$N > 2$ mg/L	1	16
$N \leq 2$ mg/L	45	1
Model reliability (%)	97.8	94.1

and flowpaths should be included when determining land-use impacts to ground water quality.

Acknowledgments

We appreciate the support of the Nantucket Land Council staff: Linda Holland, Lynn Zimmerman, Cormac Collier, Kara Courtney, and Mary Kenelly, who fully funded and supported the project. We also thank Don Harleman and Lynn Gelharfor for technical advice, and Hilyard Wood, the town of Nantucket GIS coordinator, who provided the GIS data. Fieldwork and analysis were conducted with help from Mark Willet, of the Wannacomet Water Company. Reviews were conducted by Martin Minter, Keith Halford, and two anonymous reviewers.

References

Amemiya, T. 1985. *Advanced Econometrics*. Cambridge, Massachusetts: Harvard University Press.

Andersen, L.J., and H. Kristiansen. 1984. Nitrate in ground water and surface water related to land use in the Karup Basin, Denmark. *Environmental Geology* 5,no. 4: 207–212.

Barringer, T.H., D. Dunn, W.A. Battaglin, and E.F. Vowinkel. 1990. Problems and methods involved in relating land use to ground-water quality. *Water Resources Bulletin* 226, no. 1: 1–9.

Canter, L.W. 1996. *Nitrates in Ground Water*. New York: Lewis Publishers.

Coinly, H.H. 1945. Cyanosis in infants caused by nitrates in well water. *Journal of the American Medical Association* 129, no. 2: 112.

Dingman, S.L. 1994. *Physical Hydrology*. New York: Macmillan Publishing Company.

Eckhardt, D.A., and P.E. Stackelberg. 1995. Relation of ground-water quality to land use on long island, New York. *Ground Water* 33, no. 6: 1019–1033.

Foster, S.S.D., and A.H. Bath. 1983. Distribution of agricultural soil leachates in the unsaturated zone of the British Chalk. *Environmental Geology* 5, no. 2: 53–59.

Helsel, D.R. 1990. Less than obvious. *Environmental Science & Technology* 24, no. 12: 1767–1774.

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

Horsley, S. 2000. Interview, May 2000. Medford, Massachusetts.

Horsley, Witten, and Hegeman Inc. 1990. Final report: Nantucket water resources management plan, Nantucket, Massachusetts.

Knott, J.F., and J.C. Olimpio. 1986. Estimation of recharge rates to the sand and gravel aquifer using environmental tritium, Nantucket Island, Massachusetts. USGS Water-Supply Paper 2297.

Kroll, C.N., and J.R. Stedinger. 1999. Development of regional regression relationships with censored data. *Water Resources Research* 35, no. 3: 775–784.

Lemeshow, S., D. Teres, J. Avrunin, and H. Pastides. 1988. Predicting the outcome of intensive care unit patients. *Journal of the American Statistical Association* 83, no. 402: 348–356.

Lichtenberg, E., and L. Shapiro. 1997. Agriculture and nitrate concentrations in Maryland community water system wells. *Journal of Environmental Quality* 26, no. 1: 145–153.

Masterson, J.P., and P.M. Barlow. 1994. Effects of simulated ground-water pumping and recharge on ground-water flow in Cape Cod, Martha's Vineyard, and Nantucket Island Basins, Massachusetts. USGS Open-File Report 94–316.

Meuller, D.K., and D.R. Helsel. 1996. Nutrients in the nations waters—too much of a good thing? USGS Circular 1136.

Nolan, B. 2001. Relating nitrogen sources and aquifer susceptibility to nitrate in shallow ground waters of the United States. *Ground Water* 39, no. 2: 290–299.

Oldale, R.N. 1985. Seismic investigations on Cape Cod, Martha's Vineyard, and Nantucket Massachusetts, and a topographic map of the basement surface from Cape Cod Bay to the Islands. USGS Professional Paper 650-B.

Persky, J.H. 1986. The relation of ground-water quality to housing density, Cape Cod, Massachusetts. USGS Water Resources Investigation Report 86–4093.

Puckett, L. 1994. Nonpoint and point sources of nitrogen in major watersheds of the United States. USGS Water Resources Investigations Report 94–4001.

SAS Institute Inc. 1999. SAS/STAT User's Guide, Version 8, Volumes 1, 2, and 3.

Tesoriero, A., and F. Voss. 1997. Predicting the probability of elevated nitrate concentrations in the Puget Sound basin: Implications for aquifer susceptibility and vulnerability. *Ground Water* 35, no. 6: 1029–1039.

Tobin, J. 1958. Estimation of relationships for limited dependent variables. *Econometrica* 26, no. 1: 24–36.

U.S. EPA. 1995. *Drinking Water Regulations and Health Advisories*. Washington, DC: Office of Water.

Walker, E.H. 1980. Water resources of Nantucket Island, Massachusetts. USGS Hydrological Investigations Atlas HA 615.

Weisenberger, D.D. 1990. Environmental epidemiology of non-Hodgkin's lymphoma in Eastern Nebraska *American Journal of Industrial Medicine* 18, no. 3: 303–305.

Yen, S.T., S. Liu, and D.W. Kolpin. 1996. Analysis of nitrate in near-surface aquifers in the Midcontinental United States: An application of the inverse hyperbolic sine Tobit model. *Water Resources Research* 32, no. 10: 3003–3011.