USGS
science for a changing world

Estimating the Probability of Elevated Nitrate (NO_2+NO_3-N) Concentrations in Ground Water in the Columbia Basin Ground Water Management Area, Washington

Prepared in cooperation with the
COLUMBIA BASIN GROUND WATER
MANAGEMENT AREA

Water-Resources Investigations Report 00-4110

U.S. Department of the Interior

Estimating the Probability of Elevated Nitrate (NO_2+NO_3-N) Concentrations in Ground Water in the Columbia Basin Ground Water Management Area, Washington

By Lonna M. Frans

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 00-4110

Prepared in cooperation with the
COLUMBIA BASIN GROUND WATER MANAGEMENT AREA

Tacoma, Washington
2000

U.S. DEPARTMENT OF THE INTERIOR
BRUCE BABBITT, *Secretary*

U.S. GEOLOGICAL SURVEY
Charles G. Groat, *Director*

For additional information write to:

Director, Washington Water Science Center
U.S. Geological Survey
1201 Pacific Avenue – Suite 600
Tacoma, Washington 98402
http://wa.water.usgs.gov

Copies of this report can be purchased from:

U.S. Geological Survey
Information Services
Building 810
Box 25286, Federal Center
Denver, CO 80225-0286

CONTENTS

FIGURES

TABLES

Estimating the Probability of Elevated Nitrate (NO_2+NO_3-N) Concentrations in Ground Water in the Columbia Basin Ground Water Management Area, Washington

By Lonna M. Frans

ABSTRACT

Logistic regression was used to relate anthropogenic (man-made) and natural factors to the occurrence of elevated concentrations of nitrite plus nitrate as nitrogen in ground water in the Columbia Basin Ground Water Management Area, eastern Washington. Variables that were analyzed included well depth, depth of well casing, ground-water recharge rates, presence of canals, fertilizer application amounts, soils, surficial geology, and land-use types. The variables that best explain the occurrence of nitrate concentrations above 3 milligrams per liter in wells were the amount of fertilizer applied annually within a 2-kilometer radius of a well and the depth of the well casing; the variables that best explain the occurrence of nitrate above 10 milligrams per liter included the amount of fertilizer applied annually within a 3-kilometer radius of a well, the depth of the well casing, and the mean soil hydrologic group, which is a measure of soil infiltration rate. Based on the relations between these variables and elevated nitrate concentrations, models were developed using logistic regression that predict the probability that ground water will exceed a nitrate concentration of either 3 milligrams per liter or 10 milligrams per liter. Maps were produced that illustrate the predicted probability that ground-water nitrate concentrations will exceed 3 milligrams per liter or 10 milligrams per liter for wells cased to 78 feet below land surface (median casing depth) and the predicted depth to which wells would need to be cased in order to have an 80-percent probability of drawing water with a nitrate concentration below either 3 milligrams per liter or 10 milligrams per liter. Maps showing the predicted probability for the occurrence of elevated nitrate concentrations indicate that the irrigated agricultural regions are most at risk. The predicted depths to which wells need to be cased in order to have an 80-percent chance of obtaining low nitrate ground water exceed 600 feet in the irrigated agricultural regions, whereas wells in dryland agricultural areas generally need a casing in excess of 400 feet. The predicted depth to which wells need to be cased to have at least an 80-percent chance to draw water with a nitrate concentration less than 10 milligrams per liter generally did not exceed 800 feet, with a 200-foot casing depth typical of the majority of the area.

INTRODUCTION

Over 80 percent of drinking water in the Columbia Basin of eastern Washington comes from ground water. Since the early 1950’s, nitrite plus nitrate as nitrogen concentrations (hereafter referred to in this report as nitrate) in ground water in some areas of the Columbia Basin have increased by as much as two orders of magnitude (Ebbert and others, 1995). Nationally, nitrate concentrations have increased from 3-fold to 60-fold as a result of agricultural practices (Alley, 1993). A recent study found that nitrate concentrations in over 23 percent of sampled wells in Adams, Grant, and Franklin Counties exceeded the U.S. Environmental Protection Agency maximum contaminant level (MCL) of 10 milligrams per liter (mg/L) as nitrogen, while an additional 37 percent had concentrations between 3 and 10 mg/L (Ryker and Frans, 2000).

In February 1998, the Washington State Department of Ecology (Ecology) approved the formation of the Columbia Basin Ground Water Management Area (GWMA) in Adams, Franklin, and Grant Counties, with an initial emphasis on the reduction of nitrate concentrations in ground water (Columbia Basin GWMA Steering Committee, written commun., October 31, 1997). Ground-water characterization and monitoring for the GWMA are being carried out by a partnership of the U.S. Geological Survey (USGS) and a private consultant, Kennedy/Jenks Consultants, Federal Way, Wash., with the common purpose of producing scientific products supporting the overall Columbia Basin GWMA effort to minimize nitrate concentrations in ground water (Columbia Basin GWMA Characterization and Monitoring Workgroup, written commun., April 23, 1998).

Toward the goal of reducing concentrations of nitrate in ground water, a detailed quantitative risk analysis, which correlates factors influencing nitrate concentrations, is needed to identify the major anthropogenic and natural variables that affect nitrate concentrations, and to map out regions with the highest probabilities for the occurrence of elevated nitrate concentrations. By identifying the major anthropogenic factors affecting elevated nitrate concentrations, corrective actions can be undertaken through best management practices (BMP's) to potentially reduce the nitrate concentrations in ground water. Ground-water risk assessment helps identify higher-risk regions on which to focus BMP implementation.

Purpose and Scope

This report presents the results of the analysis relating the anthropogenic and natural factors that may influence the distribution and occurrence of elevated nitrate concentrations in ground water of Adams, Franklin, and Grant Counties. The probability that elevated nitrate concentrations will occur is predicted using a logistic regression model. The logistic regression model was entered into a geographic information system (GIS) to produce maps that display the probability of elevated nitrate occurrence.

Description of the Study Area

The Columbia Basin GWMA (fig. 1) includes Adams, Franklin, and Grant Counties in eastern Washington, totalling an area of 5,985 square miles (15,501 square kilometers), with a population of approximately 150,000. The climate is arid to semiarid, with average annual precipitation ranging from 6 to 13 inches per year. Franklin County is bordered by the Columbia River to the west and by the Snake River to the south and east; Grant County is bordered by the Columbia River to the west. The basin is one of the Nation's top two producers of potatoes and wheat and is a significant producer of apples and many specialty crops. Much of the southwestern part of the area is intensively irrigated with Columbia River water, while other parts of the basin are dominated by dryland or ground-water-irrigated farming and rangeland grazing (fig. 2). The geology of the GWMA area is dominated by sediments of varying thickness that overlie basalt and some sandstone (fig. 3). Since the start of surface-water irrigation in the 1950's, the water table in some parts of the GWMA area has risen by 50 to 500 feet (Jones and Wagner, 1995).

Acknowledgments

The author thanks Jim Tesoriero, USGS, for his technical advice regarding logistic regression, Frank Voss, USGS, for his help with the GIS methods used in the analysis, and Sarah Ryker, USGS, who developed the original project design, and who helped in compiling the Columbia Basin GWMA data set on which this report is in part based.

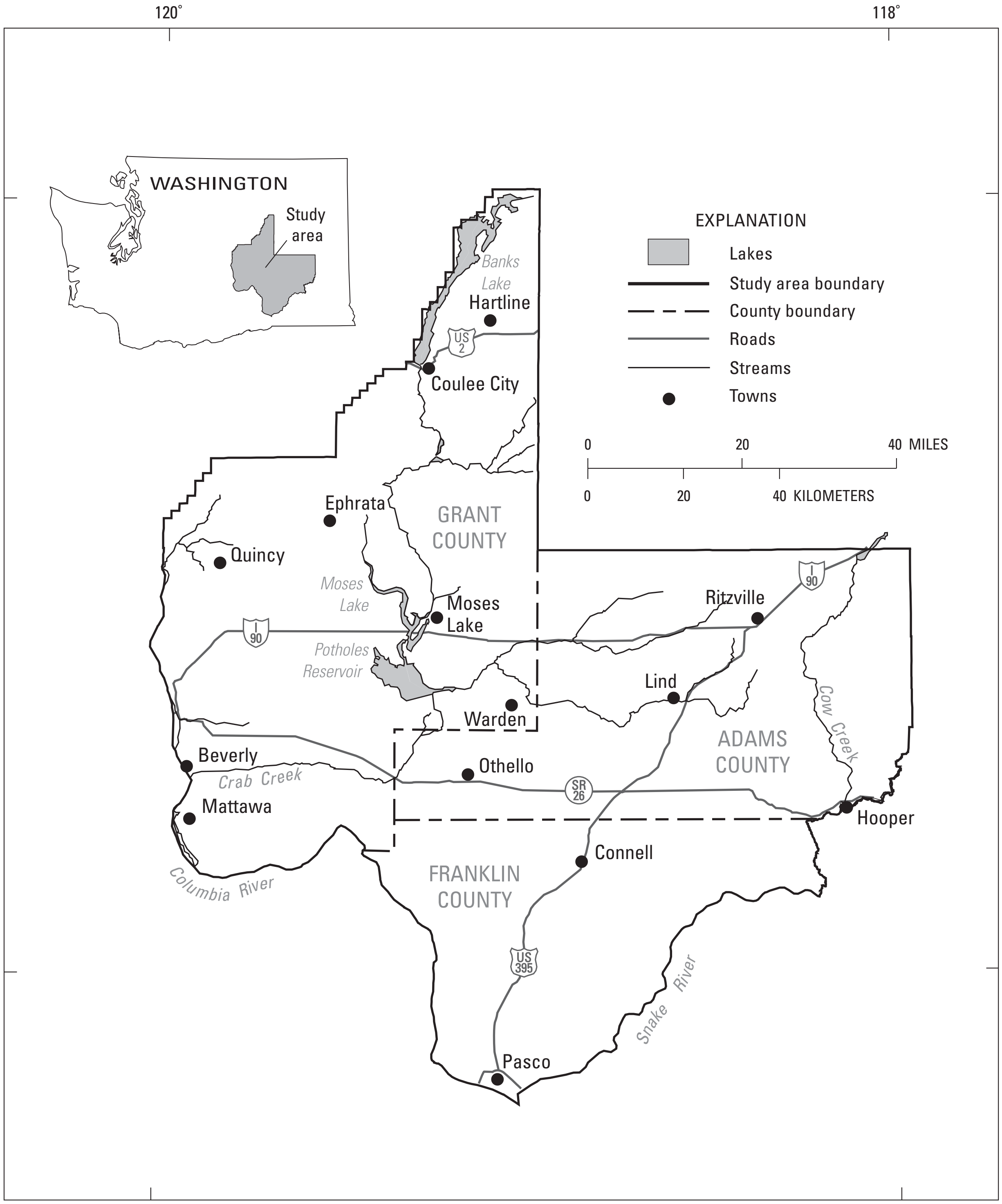

Figure 1. Location of the Columbia Basin Ground Water Management Area, eastern Washington.

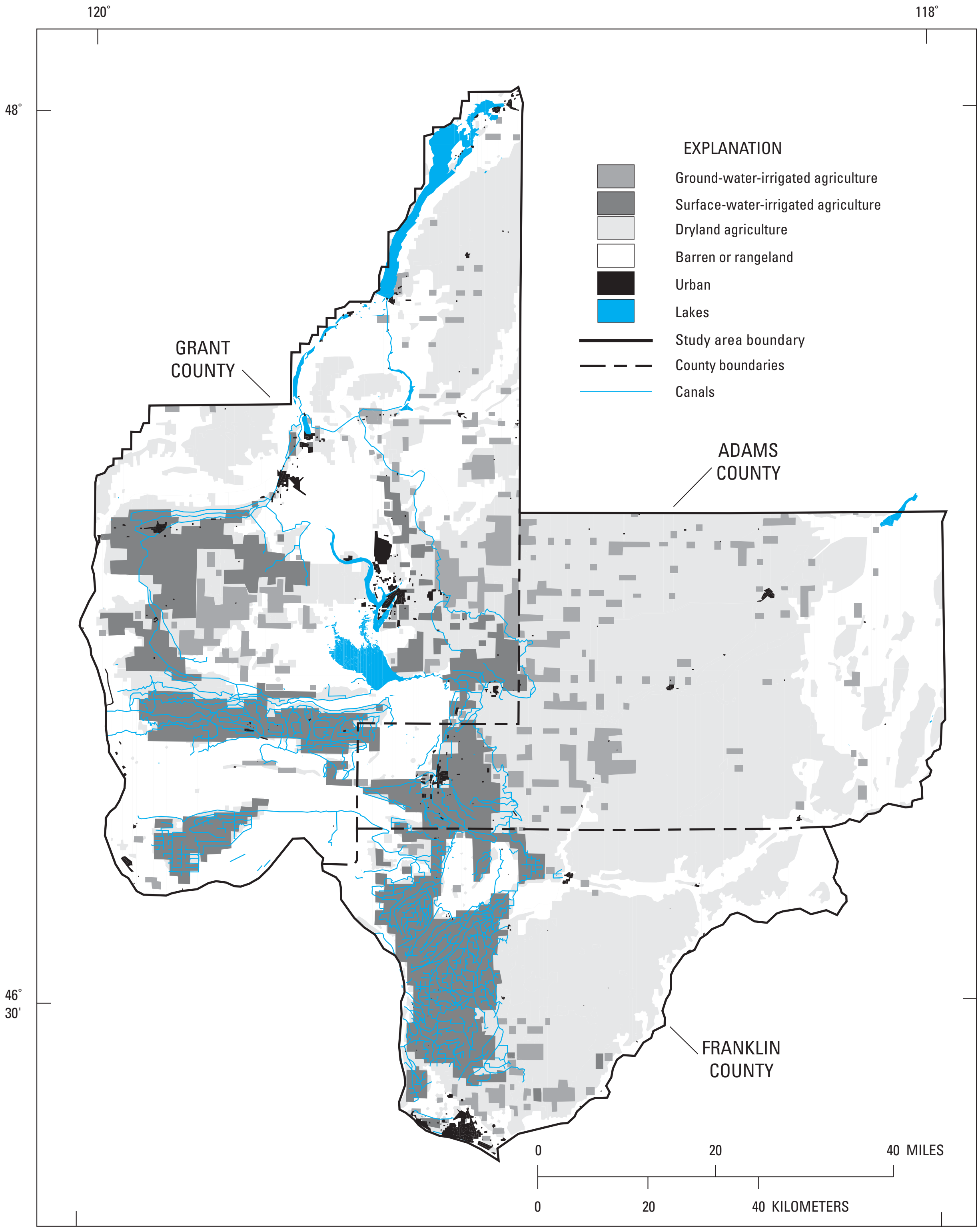

Figure 2. Land-use types in the Columbia Basin Ground Water Management Area, eastern Washington.

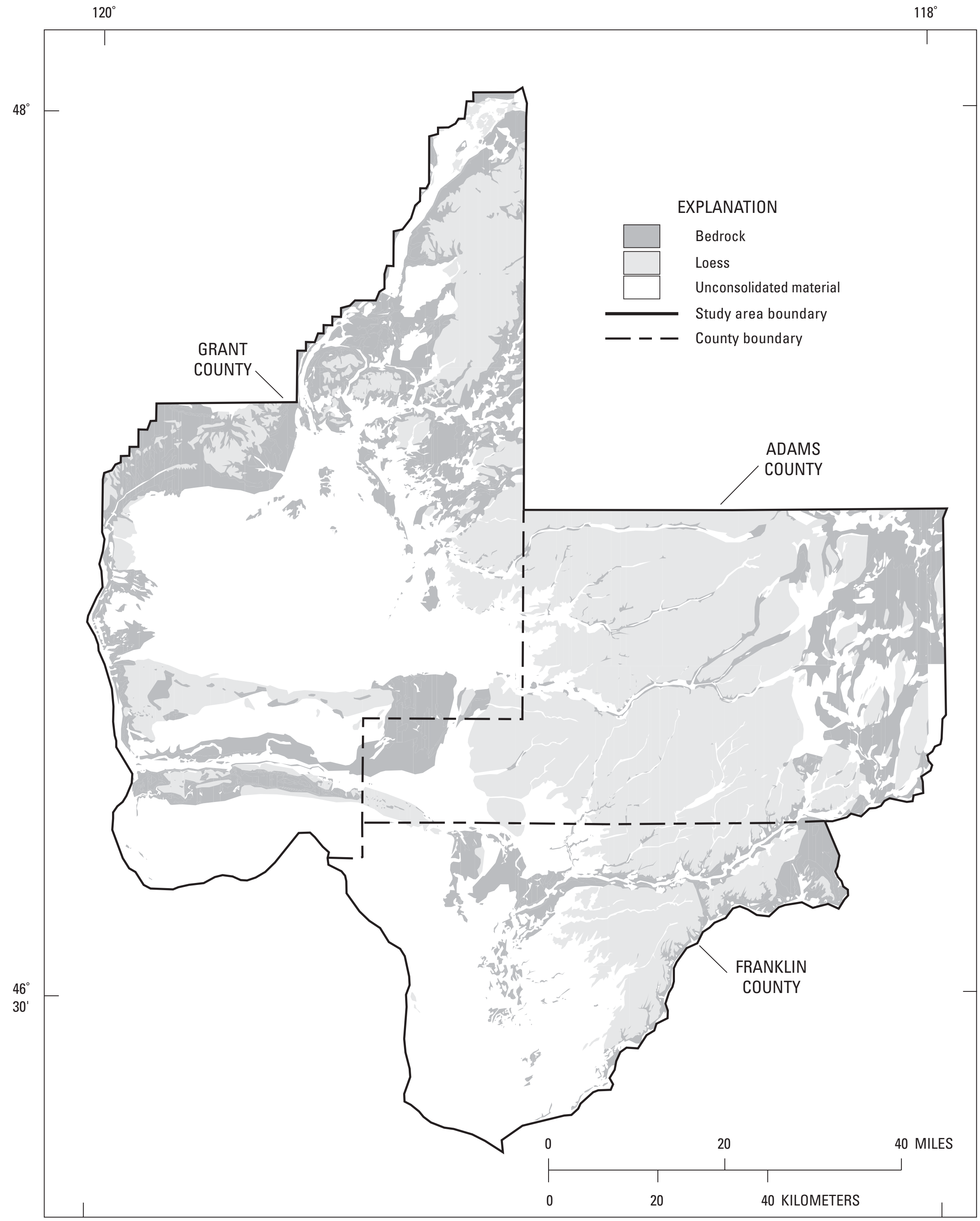

Figure 3. Surficial geology of the Columbia Basin Ground Water Management Area, eastern Washington.

MODEL DEVELOPMENT

Logistic regression was selected for this analysis because it quantifies the relation between a variable of interest (response variable) and one or more variables which affect the variable of interest (explanatory variables). This is conceptually similar to multiple linear regression. However, in logistic regression the response variable is transformed into a binary response variable (yes or no variable). This makes it an excellent tool to use for modelling aquifer vulnerability to nitrate because it can quantify the probability that nitrate will exceed a certain level. According to Tesoriero and others (1998), “Logistic regression may identify relations between the occurrence of a constituent and explanatory variables when other methods do not because logistic regression answers a simpler question—whether a well is expected to have water with a concentration greater than a specified value. In contrast, other methods (e.g., linear regression) try to estimate the absolute concentration of a constituent. Given the many factors related to the source and transport of a constituent in ground water, it is often impossible to predict its concentration; however, it may be possible to predict the probability that the concentration is above a specified level.”

In this study, two logistic regression models are used to determine the probability of exceedance of two different nitrate concentrations in ground water (3 mg/L and 10 mg/L). For each model, a binary response variable was defined by dividing the nitrate concentrations into two groups. For the 3-mg/L model, the groups are those nitrate concentrations that were greater than or equal to 3 mg/L (hereafter known as exceedances) and those that were less than 3 mg/L (known as nonexceedances). For the 10-mg/L model, nitrate concentrations were divided into those that were greater than or equal to 10 mg/L (exceedances) and those that were less than 10 mg/L (nonexceedances). The threshold of 3 mg/L was chosen because nitrate concentrations in excess of that level are generally the result of anthropogenic effects (Madison and Brunett, 1985), and 10 mg/L was chosen because it is the MCL for nitrate.

The logistic regression models take the form of

$$p = \frac{e^{(b_0 + b\boldsymbol{X})}}{1 + e^{(b_0 + b\boldsymbol{X})}}, \qquad (1)$$

where p is the probability of an exceedance, b_0 is the intercept, $\boldsymbol{X}$ represents a set of explanatory variables such as land use, soils, or well depth, and b represents the slope for each of the explanatory variables so that $b\boldsymbol{X}$=[b_1(land use)+b_2(soils)+b_3(well depth)+.....] (Hosmer and Lemeshow, 1989; Helsel and Hirsch, 1992). When the probability of an exceedance is plotted versus an explanatory variable, the result is an S-shaped curve with the probability being bounded by 0 on the lower end and 1 on the upper end. The SAS system statistical software was used to determine values of b_0 and b that best fit the data using an iteratively reweighted least-squares algorithm (SAS Institute, 1990).

Model performance was evaluated using several statistical measures for which a brief description is given here. Individual regression coefficients were considered to be statistically significant if the p-value of the Wald statistic was less than or equal to 0.05 (95-percent confidence or greater). The Wald statistic follows a chi-squared distribution and is used to indicate whether the model coefficients are significantly different from zero. The Hosmer Lemeshow goodness-of-fit statistic, HL, was used to evaluate how well the model fit the data; a smaller value indicates a better fit of the data. The HL statistic is computed by grouping wells based on their predicted probabilities and comparing the frequency of actual exceedances to the predicted frequency of exceedances. The groups are determined by listing the wells in the order of their predicted probabilities and dividing the data into 10 groups of roughly equal number. Since each group contains approximately 10 percent of the wells, each group is termed a decile of risk. The p-value for the HL statistic indicates how well the model fits the data. The model performs best as the p-value of the HL statistic approaches one. The HL statistic cannot be computed for categorical explanatory variables that indicate the presence or absence of a given feature, such as geology type. The Akaike Information Criterion (AIC) was also used as a measure of goodness-of-fit. The smaller the AIC value, the better the fit. Additional information regarding these statistical measures can be found in Hosmer and Lemeshow (1989) and Helsel and Hirsch (1992).

For most explanatory variables, the p-values for both the Wald statistic and the HL statistic are reported. The p-value for the Wald statistic can be thought of as similar to the p-value for the slope of a line in linear regression. In linear regression, the slope of the line is considered to be statistically different from a horizontal line (with a slope of zero) if the p-value is less than 0.05. Similarly, the regression coefficients are considered to be statistically different from zero if their p-values are less than 0.05. Additionally, the p-value for the HL statistic can be thought of as similar to an r-squared value in linear regression. An r-squared value is an indication of how well the linear regression line fits the data points, and the HL statistic p-value is an indication of how well the logistic regression line fits the data points. Both a low r-squared value and a low HL statistic p-value indicate that a predicted line does not fit the data well.

Response Variables

The binary response variable of nitrate concentrations exceeding or not exceeding a selected concentration (exceedances and nonexceedances, respectively) was computed from two nitrate data sets. Data from 574 wells collected during the fall 1998 GWMA baseline sampling (Ryker and Frans, 2000) were combined with an additional 102 samples collected by the USGS between 1993 and 1995 as part of the Central Columbia Plateau National Water-Quality Assessment program (wells that were resampled as part of the GWMA fall sampling were excluded) (Williamson and others, 1998) (fig. 4). Both the GWMA wells and the USGS wells were originally selected for ground-water sampling using the USGS Stratified Random Site-Selection software program (Scott, 1990) to ensure an aerially random selection of wells. The GWMA data set includes both shallow and deep wells, while the USGS data set contained mostly shallow wells.

Twenty-eight wells were sampled during both the 1993–95 and 1998 time periods. In order to verify that no statistically significant changes in nitrate concentrations had occurred between the two time periods that would preclude the combination of the two data sets, a t-test was performed on the 28 wells that were sampled during both time periods. No statistically significant difference between the two data sets was found ($p=0.94$), indicating that the two data sets could be combine. In cases where a well had more than one nitrate sample, the most recent sample was used.

Explanatory Variables

Explanatory variables that were evaluated included well depth (depth to bottom of well), casing depth (depth to bottom of casing), well diameter, ground-water recharge rate, soil hydrologic group, soil drainage group, soil permeability rate, soil clay content, amount of organic matter in soil, land use, surficial geology, the presence of nearby irrigation, the presence of nearby canals, and nitrogen fertilizer application amounts (table 1). Explanatory variable data were obtained from several sources.

Well depth, well diameter, and the depth to the bottom of the well casing were obtained for each of the wells, if available, from the USGS National Water Information System (NWIS). The surficial geology data were obtained from digital maps generated by the Washington Department of Natural Resources (Eric Schuster, written commun., 1999), and were classified into three types: bedrock (includes basalt and sandstone), fine-grained loess deposits, and coarser-grained unconsolidated deposits (fig. 3). Land-use classifications were determined from National Land Cover Data (NLCD) (Vogelmann and others, 1998). Areas were classified as either urban, agricultural (includes both irrigated and dryland agriculture), or other (which includes rangeland, wetlands, and barren areas). The data delineating irrigated areas in Grant and Franklin Counties were obtained from the Franklin Conservation District (Pat Daly, written commun., 2000). Irrigated areas in Adams County were determined from the extent of the Bureau of Reclamation irrigation blocks and from Van Metre and Seevers (1991). All soil criteria (hydrologic group, drainage group, clay content, permeability rate, and organic matter content) were obtained from the STATSGO soil data base (U.S. Department of Agriculture, 1993). The soil hydrologic groups (A, B, C, or D) represent rates of infiltration and were assigned a value between 1 and 4, respectively, so that a mean value for each soil could be calculated (fig. 5).

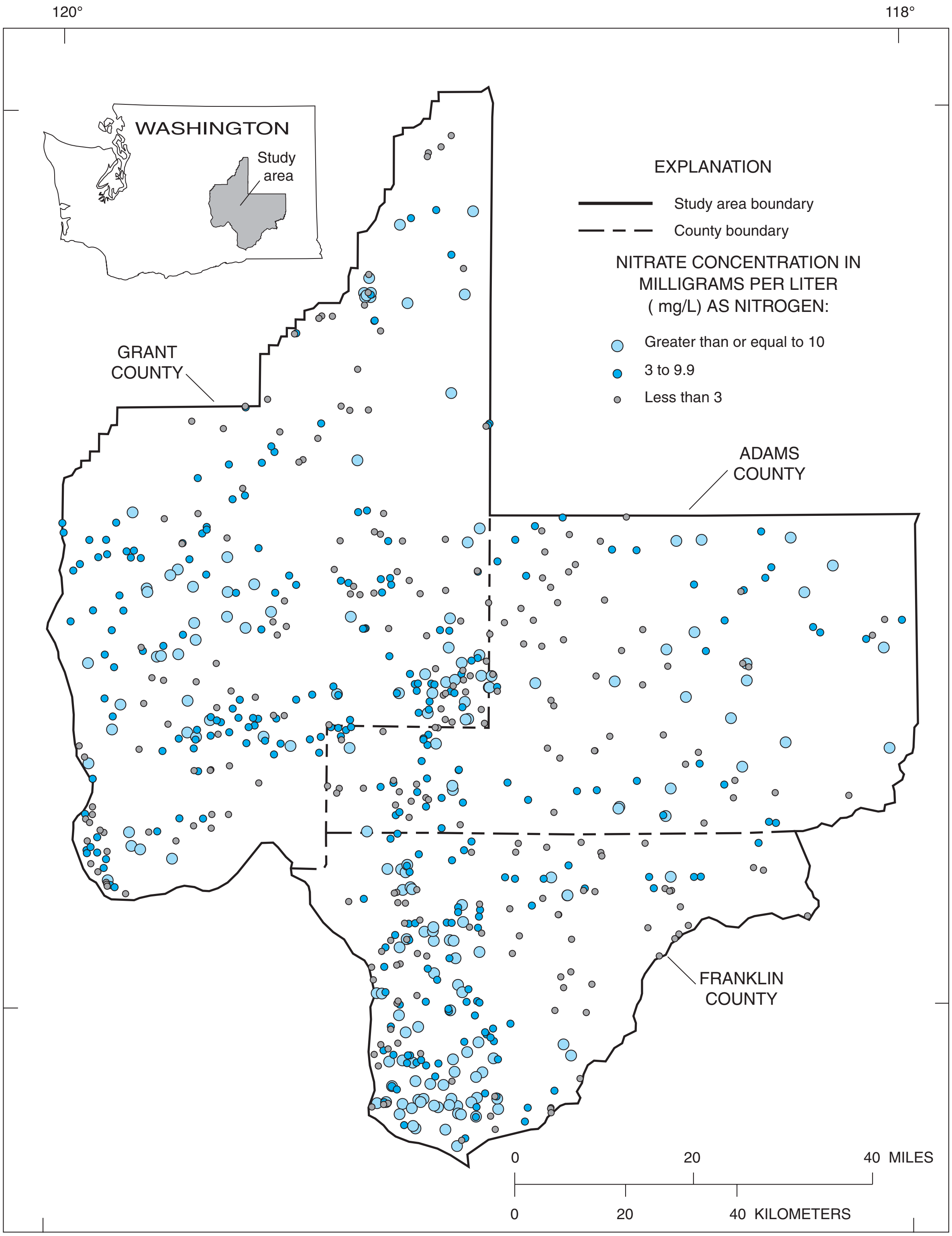

Figure 4. Locations and ranges of nitrate concentrations of the 676 wells included in the data set.

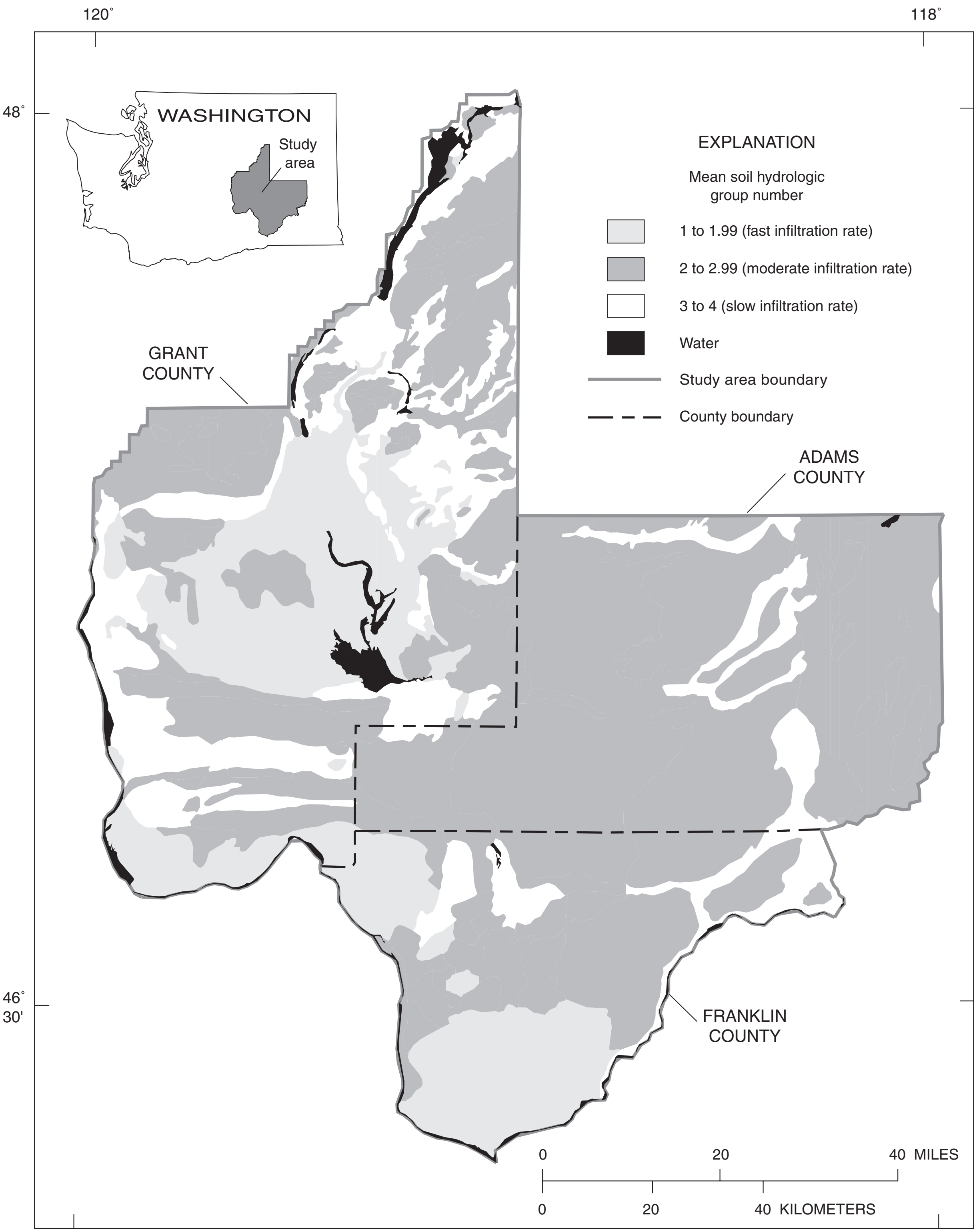

Figure 5. Mean soil hydrologic group numbers for the Columbia Basin Ground Water Management Area, eastern Washington.

Table 1. Explanatory variables used in the logistic regression models

[lb/yr, pounds per year; in/hr, inches per hour; --, not applicable;

Variable	Units	Minimum	Maximum
Well depth	feet	10	2,020
Casing depth	feet	2	1,209
Well diameter	inches	1.5	72
Nitrogen application within given radius of well	lb/yr	0	[1]3,180,000
Ground-water recharge	inches/year	0.02	46.35
Mean soil hydrologic group	unitless	1.32	3.74
Mean soil drainage group	unitless	1.28	3.53
Clay content	percent	2.99	19.34
Organic matter content	percent	0.31	1.10
Permeability rate	in/hr	0.99	12.7
Presence of irrigation within given radius of well[2]	--	--	--
Presence of canals within given radius of well[2]	--	--	--
Land use types within given radius of well			
Agricultural	percent	0	[3]100
Urban	percent	0	[3]100
Other	percent	0	[3]100
Surficial geology[2]			
Bedrock	--	--	--
Loess	--	--	--
Coarse unconsolidated material	--	--	--

[1]Maximum value is for a 5-kilometer radius around well. Maximum will vary because the value is related to the length of the chosen radius around each well. A smaller radius will decrease the maximum value.

[2]Variables were coded as either the presence or absence of each attribute. A 1 was assigned if attribute was present. A 0 was assigned if attribute was absent.

[3]Theoretical maximum is 100 percent. The actual maximum varies because the value is dependent on the length of the chosen radius around each well.

Each soil contains different percentages of the soil hydrologic groups so a weighted average was computed for the soil as a whole. Lower soil hydrologic group numbers indicate soils with faster infiltration rates. Soil drainage groups (ranging from excessively drained to very poorly drained) were assigned a value between 1 and 7, and a mean value was calculated for the soil in the same manner as for the hydrologic group. Ground-water recharge data for the GWMA were taken from Bauer and Vaccaro (1990) (fig. 6).

Estimates of agricultural nitrogen fertilizer application amounts were computed in this study as the product of the rate of nitrogen fertilizer application for a crop (in pounds per acre per year) and the total crop acreage within a specified radius of each well. Nitrogen application rates were estimated from nitrogen application guides published by the Washington State University Cooperative Extension and from the Washington Agricultural Statistics Service (table 2). The typical value was used in all calculations.

Crop acreages for the irrigated areas were obtained from two sources. Van Metre and Seevers (1991) used Landsat imagery to classify crops in areas irrigated with ground water in 1985 (fig. 2). For crops irrigated with surface water, an average of the crop acreages grown within irrigation blocks for the 5-year period 1987–91 were obtained from reports from the Bureau of Reclamation (Alan Hattrup, written commun., 1991). Crop acreages for dryland farmed areas were taken from the USGS digital land-use and land-cover data sets stored in the Geographic Information Retrieval and Analysis System (GIRAS) (U.S. Geological Survey, 1986).

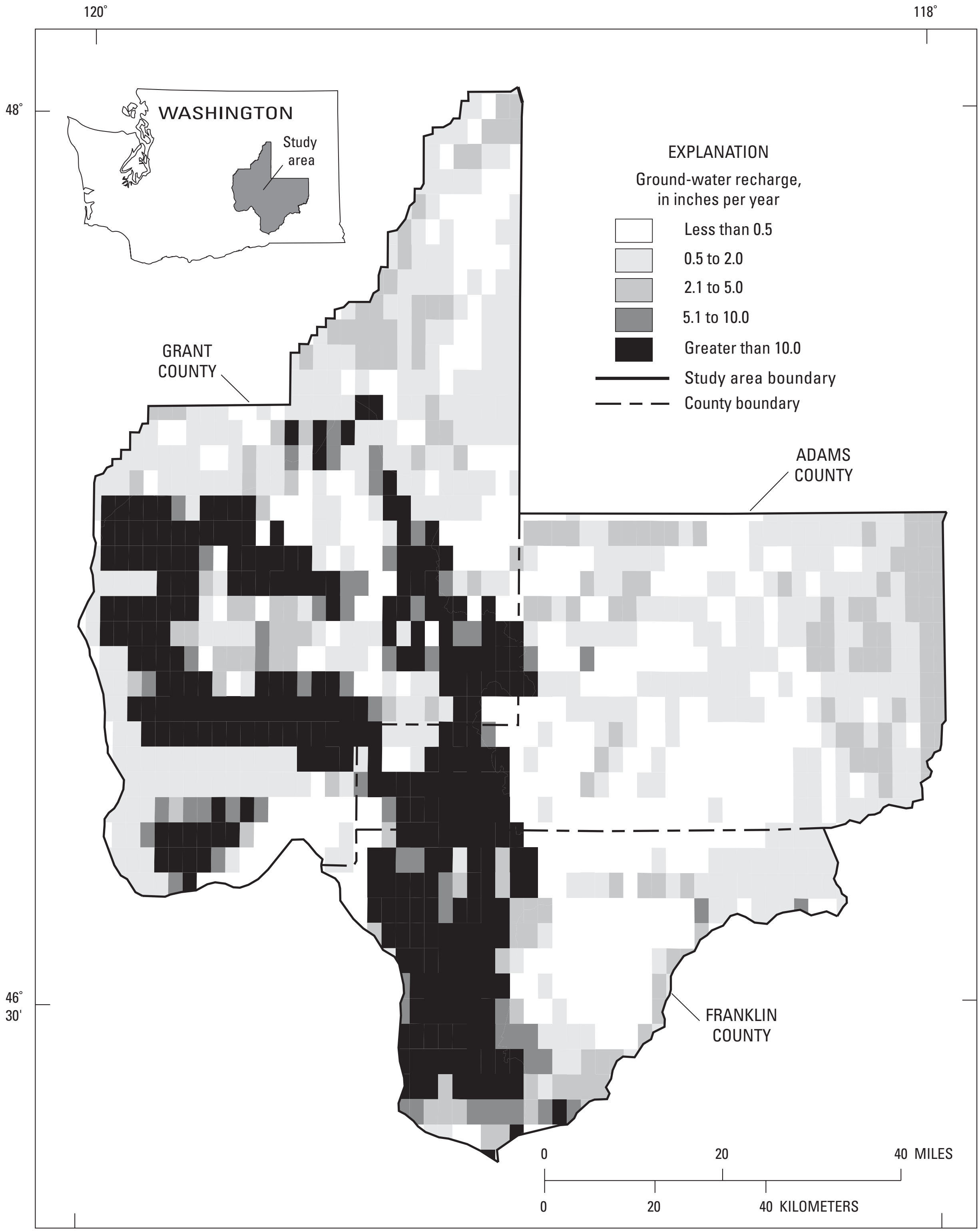

Figure 6. Mean annual regional ground-water recharge in the Columbia Basin Ground Water Management Area, Washington, 1956-77. (Modified from Bauer anc Vaccaro, 1990.)

Table 2. Estimates of nitrogen fertilizer application for crops grown in the Columbia Basin Ground Water Management Area, eastern Washington

Crop	Estimated application, in pounds of nitrogen per acre per year		
	Low	High	Typical[1]
Asparagus[2]	118	128	123
Alfalfa[3]	0	40	10
Corn[4]	0	280	275
Potatoes[2]	308	353	330.5
Beans[5]	0	120	60
Peas[2]	26	32	29
Apples[6,7]	59	72	65.5
Grapes[6]	57	66	61.5
Pasture[8]	0	200	100
Mint[9]	0	300	180
Wheat-irrigated[10]	0	240	160
Wheat-dryland[11]	30	70	50
Other	0	218	125

[1]Typical value is the average value for crops with data from the Washington Agricultural Statistics Service and is an estimated value for crops with data from the Cooperative Extension Fertilizer Guides.

2Washington Agricultural Statistics Service, 1995.

[3] Washington State University Cooperative Extension, 1976.

[4] Low and high values are from Washington State University Cooperative Extension, 1979, 1970c; Typical number is from John Holmes, oral commun., Feb. 1996.

[5] Washington State University Cooperative Extension, 1980.

[6] Washington Agricultural Statistics Service, 1992 and 1994.

[7] Rate was applied to all orchards.

[8] Washington State University Cooperative Extension, 1970b.

[9] Washington State University Cooperative Extension, 1970a.

[10] Washington State University Cooperative Extension, 1988.

[11] Washington State University Cooperative Extension, 1986.

All areas of cropland in the GIRAS that were outside of the ground- and surface-water irrigated areas were assumed to be dryland farmed. Except for the 5-year crop averages in the surface-water irrigated areas, yearly changes in crop distributions, which can affect the amount of fertilizer used, are not reflected in these estimates.

A GIS was used to determine the recharge amount, soil characteristics, and the surficial geology type at the location of each well, and to compute the percentages of each land-use type, the total amount of fertilizer applied, and whether a canal or irrigation was present within a specified radius of each well.

ESTIMATING THE PROBABILITY OF ELEVATED NITRATE CONCENTRATIONS IN GROUND WATER

Based on the relations between the explanatory variables and elevated nitrate concentrations, two models were developed using logistic regression that predict the probability that ground water will exceed a nitrate concentration of either 3 mg/L or 10 mg/L, respectively. Prior to the development of the logistic regression models based on multiple explanatory variables, the relation between each individual explanatory variable and the binary response variable was determined using logistic regression to identify which of the explanatory variables were statistically significant. The process of analyzing the relation between individual explanatory variables and the binary response variable is referred to as a univariate analysis.

Univariate Analysis of Each Explanatory Variable

Except for the percentage of urban land use near a well and the presence of coarser unconsolidated geologic units, all explanatory variables were found to be significant at a 95-percent confidence level (p_{b1}<0.05) for the 3-mg/L model (table 3). For the 10-mg/L model, none of the surficial geology types were significant at the 95-percent confidence level, nor was the presence of nearby canals or irrigation, nor was the rate of ground-water recharge (table 4). The depth to the bottom of the casing (p_{b1}=0.07) was technically not significant at the 95-percent confidence level, but it was included in the final multivariable calculations despite this because it had one of the larger HL statistics, indicating a good fit with the data. The lack of a significant relation of some of the variables with 10-mg/L exceedances compared to 3-mg/L exceedances could in part be due to the lower number of wells that exceeded the threshold level of 10 mg/L, as opposed to 3 mg/L.

Table 3. Logistic regression coefficients and summary statistics for single-variable models that predict the probability of nitrate concentrations greater than or equal to 3 milligrams per liter

[b_0, intercept parameter; b_1, regression coefficient; P_{b1}, p-value for the Wald statistic for b_1; HL, Hosmer Lemeshow goodness-of-fit statistic; P_{HL}, p-value for the HL statistic; lb/yr, pounds per year; in/hr, inches per hour; km, kilometer; <, less than; --, not available]

Explanatory variable	Sample size	b_0	b_1	[1]Pb_1	HL	P_{HL}[2]
Nitrogen applied within 2 km of well (lb/yr)	676	-0.419	4.1×10^{-6}	<0.001	5.5	0.70
Well casing depth (feet)	526	0.938	-0.00336	<0.001	7.7	0.47
Recharge (inches/year)	676	0.151	0.0313	<0.001	11.4	0.18
Well depth (feet)	665	1.09	-0.00199	<0.001	18.1	0.02
Soil drainage group (unitless)	672	1.42	-0.344	0.007	22.1	0.001
Organic matter (percent)	672	1.27	-1.53	0.001	26.4	<0.001
Permeability rate (in/hr)	672	0.240	0.0537	0.004	27.6	<0.001
Clay content (percent)	672	1.06	-0.0763	0.015	27.4	<0.001
Well diameter (inches)	590	0.744	-0.0273	0.02	18.9	<0.001
Soil hydrologic group (unitless)	672	1.08	-0.250	0.034	24.6	<0.001
Presence of irrigation within 0.5 km of well	676	-0.111	0.881	<0.001	--	--
Presence of canals within 2 km of well	676	0.310	0.432	0.007	--	--
Land-use types						
Percent agriculture within 2 km of well	676	-0.262	0.0129	<0.001	8.3	0.4
Percent other within 2 km of well	676	0.997	-0.0136	<0.001	8.7	0.36
Percent urban within 2 km of well	676	0.532	-0.00333	0.79	13.8	0.06
Surficial geology types						
Loess	676	0.624	-0.573	0.005	--	--
Bedrock	676	0.426	0.440	0.027	--	--
Coarse unconsolidated material	676	0.493	0.0454	0.78	--	--

[1]A p-value less than 0.05 shows there is a statistically significant correlation at a greater than 95-percent confidence level.

[2]As the p-value approaches 1, performance of the model improves.

Well depth and casing depth both showed an inverse relation with the probability of elevated nitrate occurrence (fig. 7). This was not unexpected, as previous studies have shown that nitrate concentrations in the study area generally tend to decrease with depth below land surface (Jones and Wagner, 1995; Ryker and Frans, 2000). The casing-depth models fit the data much better than the well-depth models based on the HL statistic. The well-depth model probably does not fit the data as well due to the complications arising from the very long open intervals present in many of the wells. A deep well with a short casing may draw water from a wide range of depths below land surface. Thus, well depth is an imprecise indicator of nitrate concentrations at depth. Because nitrate concentrations tend to increase with shallower depth, the upper limit of an open interval (casing depth) is a better indicator of the relation between nitrate and depth than the lower limit of the open interval (well depth). Therefore, due to these complications, casing depth was used in the multivariable analysis instead of well depth.

Well diameter showed a significant inverse relation with both the 3-mg/L and 10-mg/L models. Typically, monitoring wells, which have the smallest diameter, are installed at the shallowest depths and draw water with the highest nitrate concentrations. Wells with a large diameter, such as public supply wells, tend to be installed to greater depths and draw water with lower nitrate concentrations.

Table 4. Logistic regression coefficients and summary statistics for single-variable models that predict the probability of nitrate concentrations greater than or equal to 10 milligrams per liter

[b_0, intercept parameter; b_1, regression coefficient; P_{b1}, p-value for the Wald statistic for b_1; HL, Hosmer Lemeshow goodness-of-fit statistic; P_{HL}, p-value for the HL statistic; lb/yr, pounds per year; in/hr, inches per hour; km, kilometer; m, meters; <, less than; --, not available]

Explanatory variable	Sample size	b_0	b_1	[1]Pb_1	HL	P_{HL}[2]
Well casing depth (feet)	526	-0.946	-0.00169	0.07	7.72	0.5
Well diameter (inches)	590	-0.817	-0.0461	0.03	6.48	0.2
Well depth (feet)	665	-0.808	-0.00139	0.001	12.34	0.1
Recharge (inches/year)	676	-1.13	-0.0017	0.8	11.95	0.1
Nitrogen applied within 3 km of well (lbs/year)	676	-1.94	1.44×10^{-6}	<0.001	17.86	0.02
Soil drainage group (unitless)	672	0.159	-0.520	<0.001	17.4	0.008
Soil hydrologic group (unitless)	672	0.181	-0.632	<0.001	28.3	<0.001
Permeability rate (in/hr)	672	-1.51	0.0617	0.003	28.2	<0.001
Organic matter (percent)	672	-0.505	-1.36	0.02	37.5	<0.001
Clay content (percent)	672	-0.579	-0.0837	0.03	46.8	<0.001
Presence of irrigation within 0.5 km of well	676	-1.42	0.355	0.09	--	--
Presence of canals within 500 m of well	676	-1.134	-0.064	0.7	--	--
Land use types						
Percent urban within 5 km of well	676	-1.27	0.0334	0.031	7.80	0.45
Percent agriculture within 5 km of well	676	-1.86	0.0118	0.003	7.89	0.44
Percent other within 5 km of well	676	-0.636	-0.0143	<0.001	8.27	0.41
Surficial geology types						
Bedrock	676	-1.15	-0.0199	0.927	--	--
Loess	676	-1.14	-0.0595	0.805	--	--
Coarse unconsolidated material	676	-1.18	0.0498	0.788	--	--

[1]A p-value less than 0.05 shows there is a statistically significant correlation it a greater than 95-percent confidence level.
[2]As the p-value approaches 1, performance of the model improves.

The rate of ground-water recharge showed a significant positive relation with 3-mg/L exceedances, indicating that nitrate concentrations increased where recharge was greater. Recharge was not significantly related to 10-mg/L exceedances. Recharge ranged from 0.02 inches per year to more than 45 inches per year and was highest in the irrigated areas. Higher rates of recharge mean more water is available to carry nitrate from the surface to the ground water.

All of the soil criteria were significantly related to 3-mg/L exceedances and 10-mg/L exceedances. The mean soil hydrologic group number and drainage group number showed significant inverse relations with both the 3-mg/L and 10-mg/L exceedances, indicating the probability of elevated nitrate concentration increases beneath soils that are well drained or that have high infiltration rates. Nitrate concentrations were also inversely correlated with an increase in clay and organic matter content and positively correlated with permeability rate. Coarse-textured soil tends to favor movement of water and nitrate to ground water, while finer-textured soils favor storage of water and nitrate in soils and runoff of nitrate to streams. Additionally, soils with slow infiltration rates and high organic matter content may have an increased likelihood of fostering denitrifying conditions, thereby decreasing the amount of nitrate available to leach into the ground water.

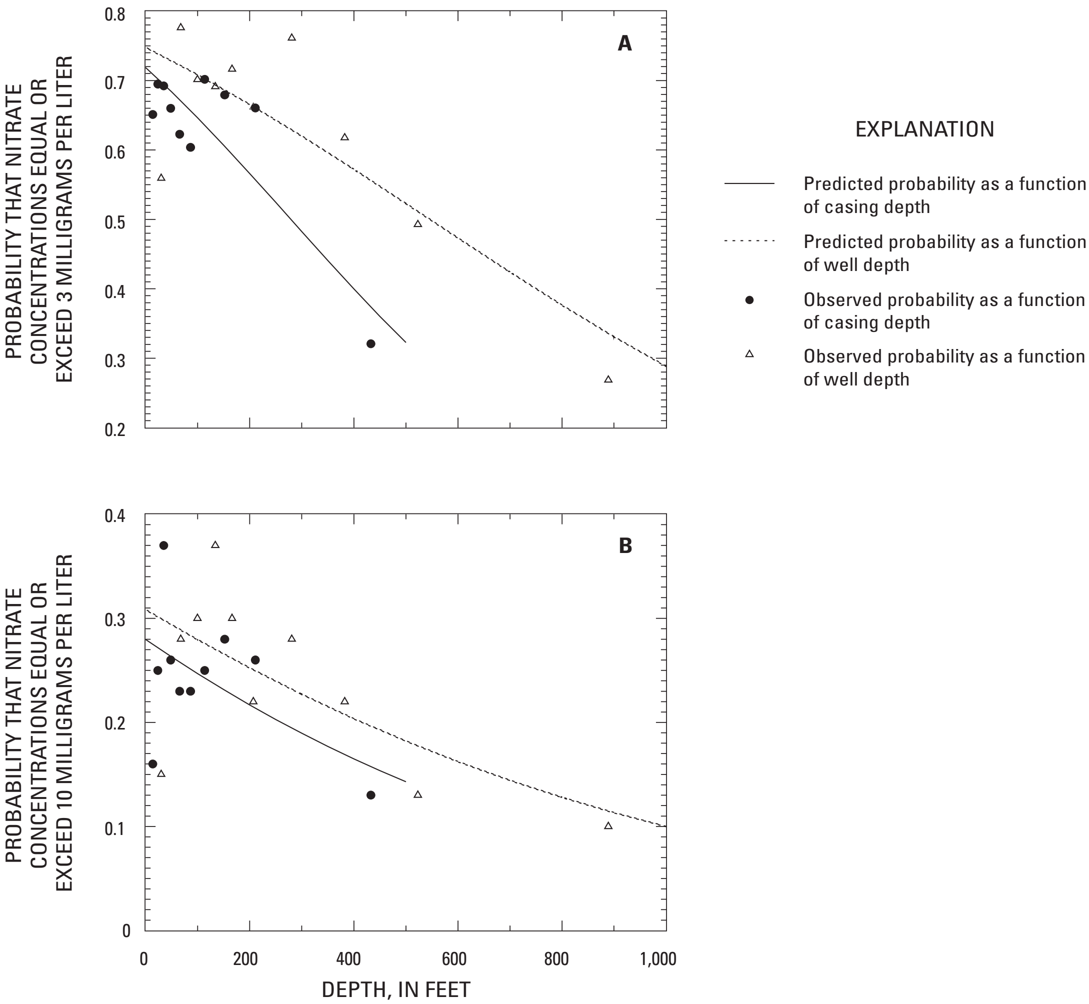

Figure 7. Predicted and observed probabililities that nitrate concentrations are (*A*) greater than or equal to 3 milligrams per liter or (*B*) greater than or equal to 10 mg/L as a function of casing or well depth below land surface.
(Each data point is the fraction of actual wells with a nitrate concentration equal to or exceeding either 3 or 10 mg/L.)

Surficial geology was related to exceedances based on the presence or absence of each geology type because it is not a continuous numeric variable. The presence of bedrock units was positively correlated with exceedances, while the presence of fine-grained loess deposits was inversely correlated with exceedances for the 3-mg/L model, and it was not significantly correlated with exceedances for the 10-mg/L model. The fine-grained nature of the loess deposits inhibits the movement of water and dissolved nitrate into the ground-water system. Coarser unconsolidated deposits were not significantly correlated with either 3- or 10-mg/L exceedances.

The percentage of each land-use type and the total amount of nitrogen fertilizer applied within a specified radius of each well were calculated and then related to nitrate concentrations. The optimal radius to use for the land-use and fertilizer application data was determined by performing logistic regression analyses at different radii around the wells. The radius that best fit the data was determined by comparing AIC values at the different radii (table 5). The AIC values for the three land-use types at each radius were summed to get the total AIC value. A radius of 2 km was selected for the 3-mg/L model because logistic regression analysis at that radius yielded the smallest AIC values (best fit) for both nitrogen application amounts and total land use. For the 10-mg/L model, a radius of 3 km was chosen for nitrogen application amounts, and 5 km was chosen for land use.

Table 5. Akaike Information Criterion (AIC) values for land-use types, fertilizer application amounts, and the presence of irrigation at various radii from wells

[Bold values indicate the radius with the best fit. km, kilometer; mg/L, milligrams per liter]

Radius (kilometers)	Land-use types				Nitrogen application	Presence of irrigation
	Agriculture	Urban	Other	Total		
			3 mg/L exceedances			
0.5	883.1	895.6	884.0	2,662.7	865.9	**872.2**
1	879.5	895.3	880.3	2,655.2	860.5	878.1
2	875.4	896.8	874.4	**2,646.6**	**851.7**	883.6
3	879.4	890.8	892.0	2,662.2	853.9	878.4
4	881.3	896.7	880.1	2,658.2	856.4	881.5
5	883.7	895.8	881.8	2,661.4	857.5	878.5
			10 mg/L exceedances			
0.5	744.6	748.2	744.8	2,237.7	735.3	**745.7**
1	743.4	748.5	743.1	2,235.0	735.1	746.2
2	739.5	748.3	738.0	2,225.8	729.5	748.2
3	739.5	745.3	745.8	2,230.6	**728.8**	747.9
4	739.0	746.7	736.6	2,222.4	729.4	748.1
5	739.3	744.0	735.6	**2,219.0**	729.5	747.9

The percentage of agricultural land near a well had a positive relation to exceedances. This was not surprising, given the abundance of nitrate sources available in agricultural areas. The percentage of other (scrub and grasslands, wetlands, barren areas) land-use types was inversely correlated with exceedances. This reflects the lack of large nitrate sources in these areas. The percentage of urban land use showed a significant positive correlation with the 10-mg/L exceedances, but no significant relation with the 3-mg/L exceedances.

The amount of nitrogen fertilizer applied per year within a 2-km radius of each well for the 3-mg/L model or within a 3-km radius for the 10-mg/L model was positively correlated with exceedances (fig. 8). The 3-mg/L model fit the data the best of all the single-variable models based on the HL statistic, but the 10-mg/L model did not have a good fit based on the HL statistic. The positive correlation was expected because as more fertilizer is applied, the likelihood is greater that some of the excess nitrate will be transported to the ground-water system.

The presence of irrigation within 0.5 km was positively related to nitrate concentrations exceeding 3 mg/L, but no significant relation could be found with 10-mg/L exceedances. In both cases, the radius with the lowest AIC values was 0.5 km (table 5). The presence of elevated nitrate concentrations in locations with a close proximity to irrigation is likely due to the increased amount of recharge occurring in those areas. The extra water applied through irrigation means that more water is available to carry nitrate from the land surface to the ground water.

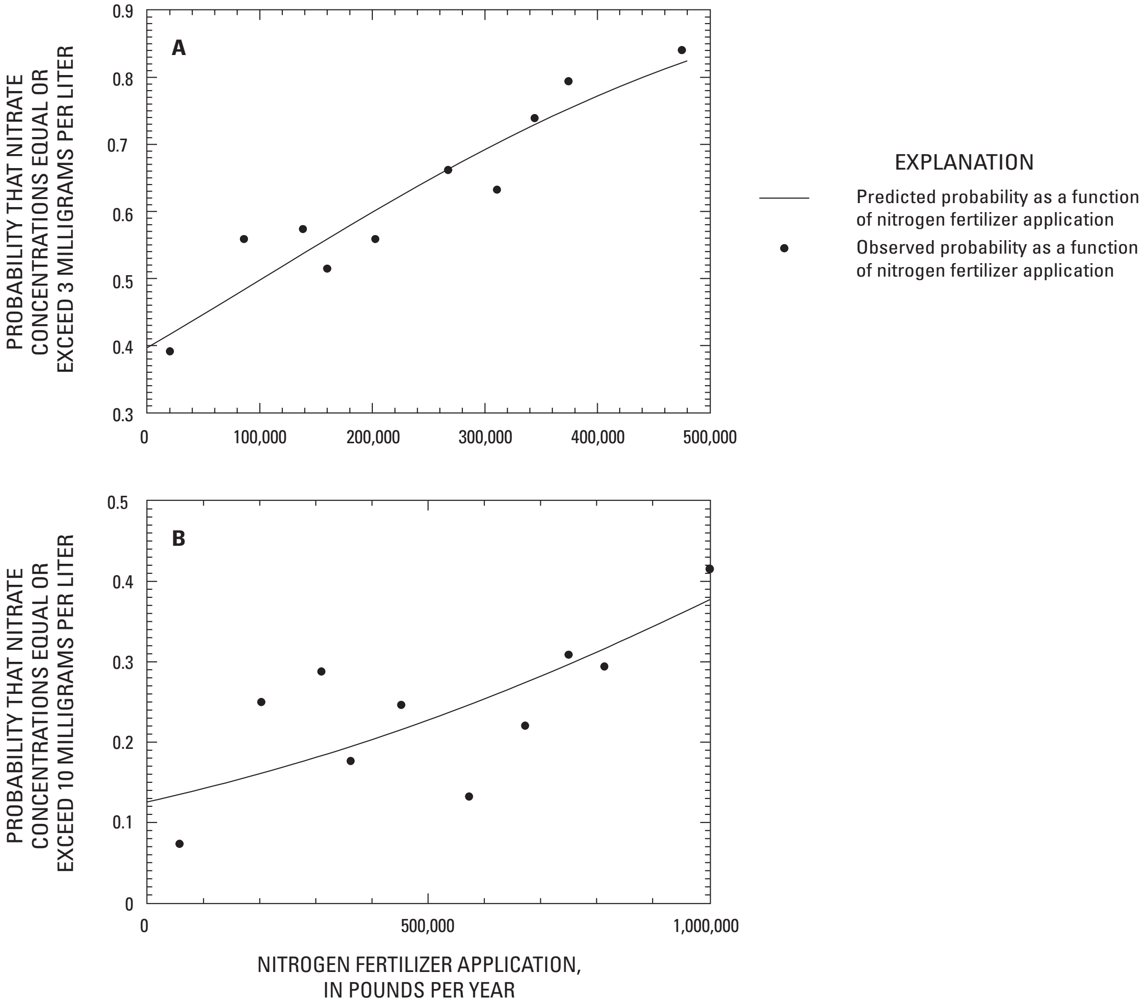

Figure 8. Predicted and observed probabilities that nitrate concentrations are (A) greater than or equal to 3 milligrams per liter or (B) greater than or equal to 10 mg/L as a function of the total amount of nitrogen fertilizer applied within (A) 2 kilometers of a well or (B) 3 km of a well.

(Each data point is the fraction of actual wells with a nitrate concentration exceeding either 3 or 10 mg/L.)

The presence of canals within a certain radius around the well was positively related to exceedances for the 3-mg/L model, but no significant relation was found for 10-mg/L exceedances. For the 3-mg/L model, the most significant radius was 2 km, while it was only 500 meters for the 10-mg/L model. The presence of canals in the 3-mg/L model is likely a surrogate for other explanatory variables such as the presence of irrigated agriculture; previous studies show the presence of canals should be negatively correlated with nitrate exceedances. Canal water has lower nitrate concentrations than leachate from applied irrigation water, and wells located in regions where canal leakage is a dominant source of recharge usually have lower nitrate concentrations (Ebbert and others, 1995). In addition to correlating the presence or absence of nearby canals with nitrate exceedances, an attempt was made to correlate the distance to the nearest canal with nitrate exceedances. None of these attempts produced any significant results.

Multivariate Analysis

All explanatory variables, except where previously noted in the univariate analysis, were considered as possible variables in the final multivariable analysis. In order to calculate coefficients for the two multivariable models, a stepwise selection procedure was used, and the p-values for entry and retention of variables in each model were set at 0.05. In stepwise selection, variables are entered one at a time, starting with the most significant, until no additional variables meet the entry criterion ($p<0.05$). The confidence level (and thus p-values) for variables in the multivariable models are generally different than those calculated for the single-variable models because logistic regression adjusts the confidence level based on the presence of other explanatory variables in the model. The final two multivariable models were selected based on how well they fit the data according to the Hosmer Lemeshow statistic.

The explanatory variables selected for the 3-mg/L model included fertilizer application amounts and casing depth, while those selected for the 10-mg/L model included fertilizer application amounts, casing depth, and the mean soil hydrologic group (table 6). A Hosmer Lemeshow p-value of 0.93 for the 3-mg/L model and 0.85 for the 10-mg/L model indicate that the models fit the data quite well. The predicted and observed number of exceedances and nonexceedances were very similar for the respective models, which indicates that the models do a good job of predicting the occurrence of elevated nitrate concentrations (table 7). Based on the 3-mg/L model, wells with short casings located in areas of heavy fertilizer application will have the highest probability of exceeding a nitrate concentration of 3 mg/L. For the 10-mg/L model, wells with short casings located in areas of heavy fertilization on soils with high infiltration rates will have the highest probability of exceeding a nitrate concentration of 10 mg/L.

Using the logistic regression models, maps showing regions of high vulnerability were generated. To generate vulnerability maps, a 100-meter-by-100-meter grid was constructed that covered the entire GWMA. It was assumed that a well was in the center of each grid cell and that the nitrate concentration in each well represents the nitrate concentrations in the ground water for the entire grid cell. For the 3-mg/L model, the amount of nitrogen fertilizer applied within a 2-km radius from the center of each cell was calculated, a casing depth of interest was selected and then, using equation (1) and the coefficients in table 6, probabilities of nitrate concentration exceedances were determined for each cell. The probabilities were then mapped using a GIS. The process for the 10-mg/L model was similar, with the inclusion of the mean soil hydrologic group number of each cell along with a casing depth of interest and the amount of nitrogen fertilizer applied within 3 km of each cell.

Predicted probabilities for wells cased to approximately 80 feet below land surface (the median casing depth for the data set was 78 feet) ranged from a 20-percent to greater than 90-percent chance of exceeding 3 mg/L (fig. 9). The regions with the highest probabilities were those areas of irrigated agriculture that use the most nitrogen fertilizer. Conversely, the eastern portions of Franklin and Adams Counties, which do not have large regions of irrigated agriculture, had lower probabilities of exceeding a nitrate concentration of 3 mg/L. The probability that nitrate concentrations will exceed 10 mg/L for wells cased to approximately 80 feet is much less than the probability that nitrate concentrations will exceed 3 mg/L (fig. 10). The probability that nitrate concentrations will exceed 10 mg/L is typically less than 40 percent, with only a few regions in Grant and Franklin Counties exceeding 40 percent. Maps such as these can be generated for any depth of interest, with vulnerability to elevated nitrate concentrations decreasing in all areas as casing depth increases.

Table 6. Regression model coefficients for predicting ground-water vulnerability to nitrate

[b_0, intercept parameter; b_1, b_2, and b_3, regression coefficients; HL, Hosmer-Lemeshow statistic; P_{HL}, p-value for the HL statistic; mg/L, milligrams per liter; lb/yr, pounds per year; --, not applicable]

b_0	b_1, Fertilizer application amount (lb/yr)	b_2, Well casing depth (feet)	b_3, Mean soil hydrologic group (unitless)	HL	[1]P_{HL}
			3-mg/L model		
-0.106	4.8×10^{-6}	-0.1000414	--	2.98	0.936
			10-mg/L model		
-0.342	1.45×10^{-6}	-0.00286	-0.636	4.06	0.852

[1]As the p-value approaches 1, performance of the model improves.

Table 7. Observed and predicted numbers of times that nitrate concentrations did or did not exceed the logistic regression model threshold

[mg/L, milligrams per liter]

Decile	Number of wells	Exceedances		Nonexceedances	
		Observed	Predicted	Observed	Predicted
			3-mg/L model		
1	52	15	15.0	37	37.0
2	52	23	24.0	29	28.0
3	52	27	27.5	25	24.5
4	52	33	30.2	19	21.8
5	52	37	32.7	15	19.3
6	52	33	35.0	19	17.0
7	52	36	37.6	16	14.4
8	52	39	39.6	13	12.4
9	52	41	41.8	11	10.2
10	54	46	46.5	8	7.5
			10-mg/L model		
1	52	3	3.5	49	48.5
2	52	7	5.8	45	46.2
3	52	9	7.9	43	44.1
4	52	13	9.8	39	42.2
5	52	12	11.4	40	40.6
6	52	9	13.1	43	38.9
7	52	13	15.0	39	37.0
8	53	17	17.5	36	35.5
9	52	20	19.4	32	32.6
10	53	25	24.5	28	28.5

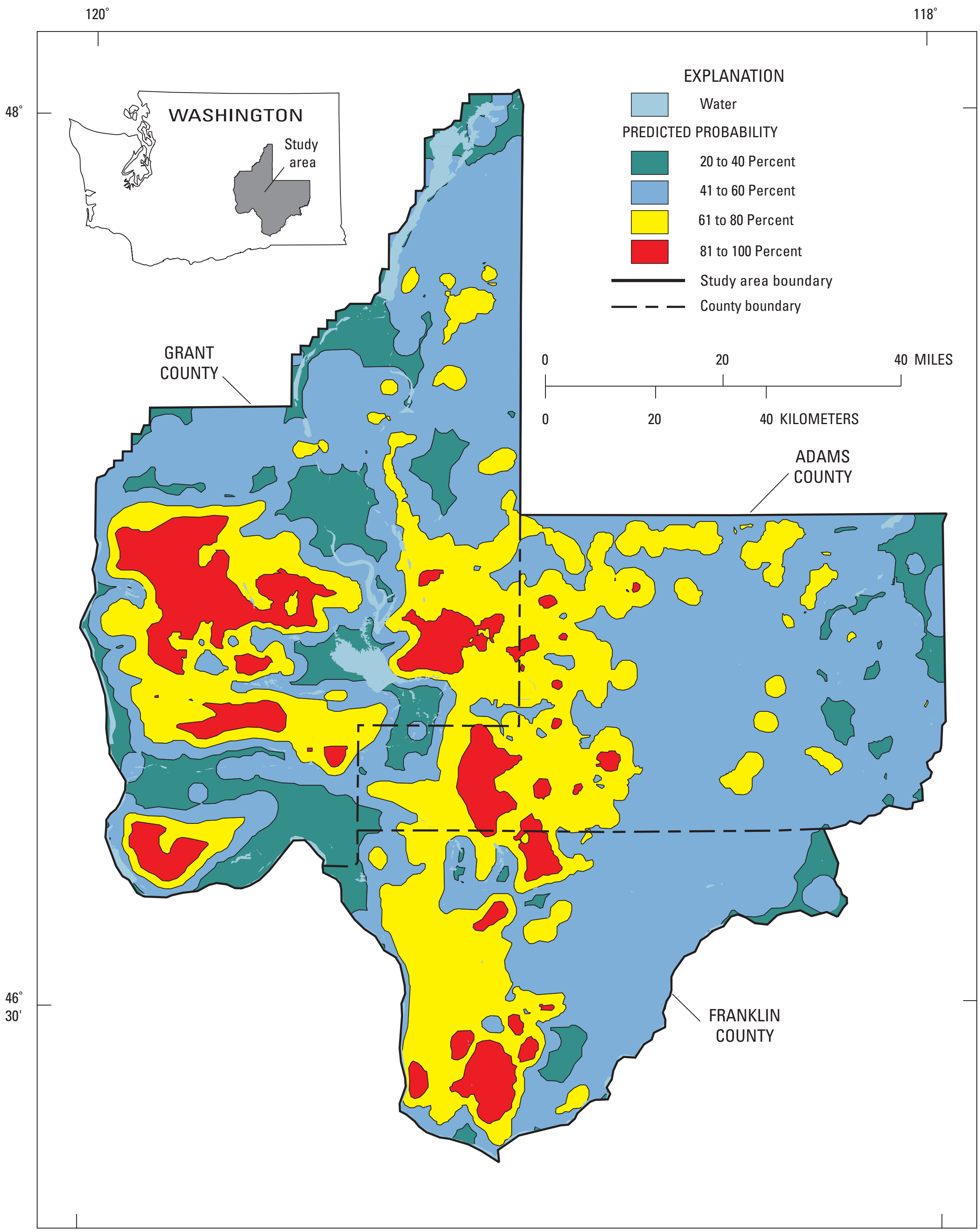

Figure 9. Predicted probability that nitrate concentrations will exceed 3 milligrams per liter for wells cased to a depth of approximately 80 feet below land surface.

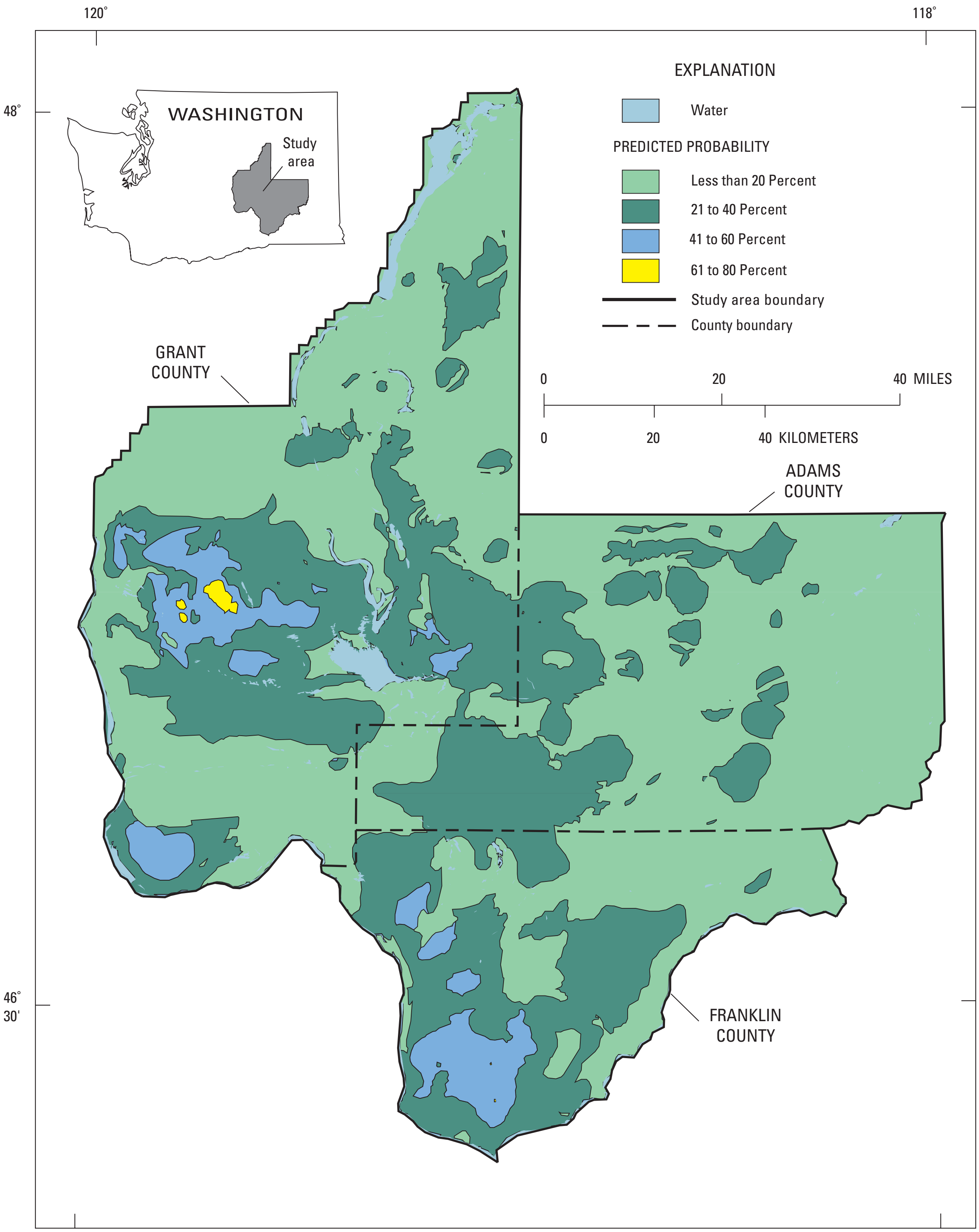

Figure 10. Predicted probability that nitrate concentrations will exceed 10 milligrams per liter for wells cased to a depth of approximately 80 feet below land surface.

In addition to generating maps displaying the probability of elevated nitrate concentrations, maps can also be generated to show the depth to which wells need to be cased to have a certain probability of obtaining water with a nitrate concentration below 3 mg/L or 10 mg/L. This was accomplished by fixing the probability at a certain level, using the coefficients in table 6, and solving equation (1) for the casing depth. For example, by fixing the probability at a 20 percent chance of exceeding 3 mg/L, it can be seen that in some regions, wells need to be cased to over 1,000 feet in order to have an 80-percent chance of obtaining low nitrate ground water (fig. 11). In general, wells installed in irrigated agricultural areas need a casing of at least 600 to 1,000 feet to have an 80-percent chance of obtaining low nitrate ground water, while wells in dryland agricultural areas need a casing of at least 400 feet and wells in barren or rangeland areas need a casing of at least 200 feet. Wells need to be cased to a much shallower depth in order to have at least an 80-percent chance of obtaining water with a nitrate concentration below 10 mg/L (fig. 12). The majority of wells in the GWMA need a casing less than 200 feet to obtain water with a nitrate concentration below the MCL, whereas only a few small areas in Franklin and Grant Counties need a well casing of at least 600 feet.

It should be noted that although fertilizer application, casing depth, and soil hydrologic group (10-mg/L model only) were the three most significant explanatory variables in the models, many other variables can also have an important effect on nitrate concentrations. The absence of a variable from the model does not mean that it does not have an important effect on nitrate concentrations. In many cases, if two or more explanatory variables are closely correlated with each other, only one of the variables will be incorporated into the model and it will account for the effects of the other variable. Additionally, there are many local or small-scale complexities in the ground-water system that affect the concentration of nitrate in the water that cannot be accounted for. Therefore, although a well may be installed in a region with a predicted high probability of elevated nitrate occurrence, it may in fact actually yield water with low nitrate concentrations due to complexities in the hydrogeologic system. Such small-scale effects must be dealt with on a small-scale basis rather than through regional-scale models such as those developed in this study.

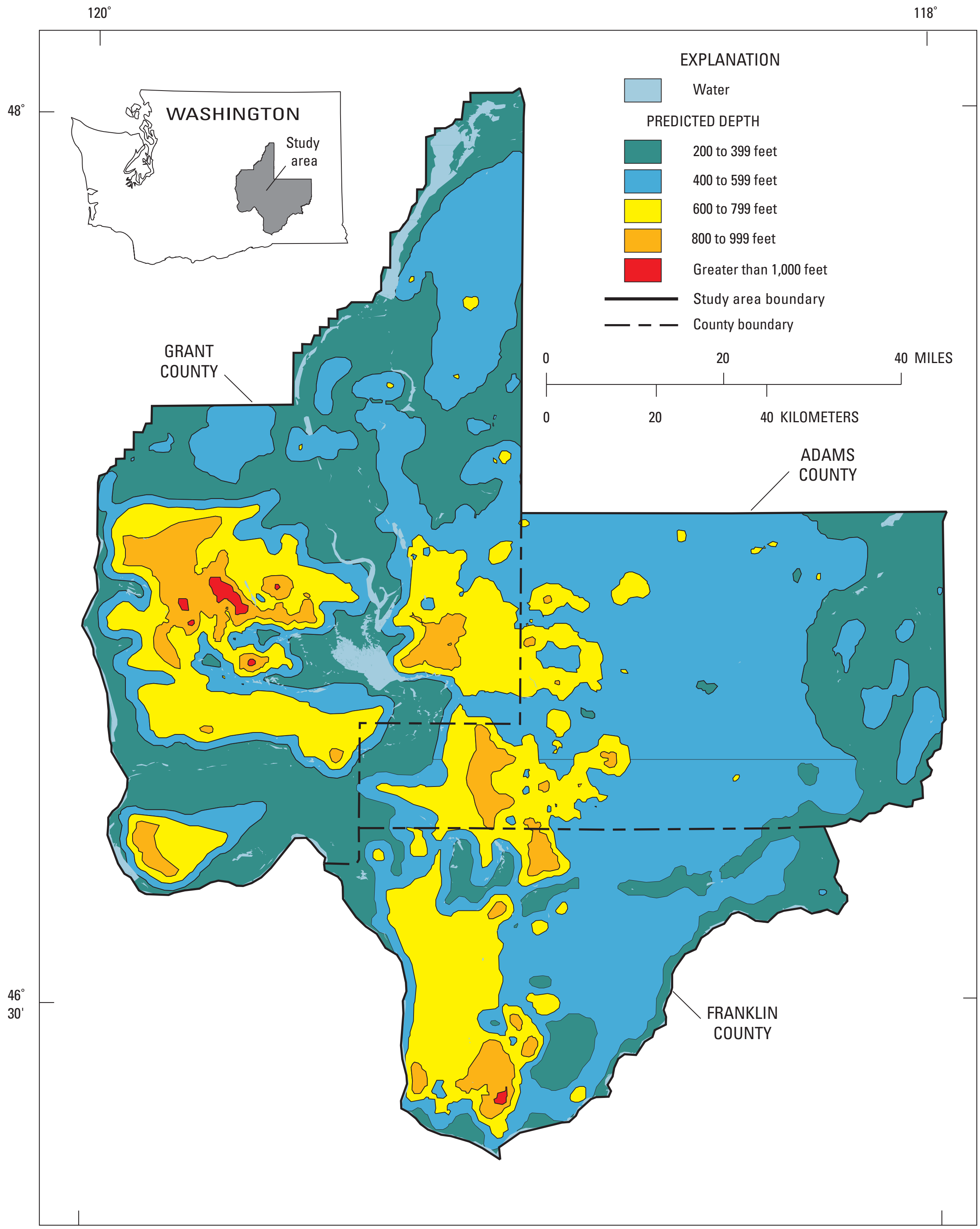

Figure 11. Predicted depth to which wells need to be cased to have at least an 80-percent chance of obtaining water with a nitrate concentration less than 3 milligrams per liter.

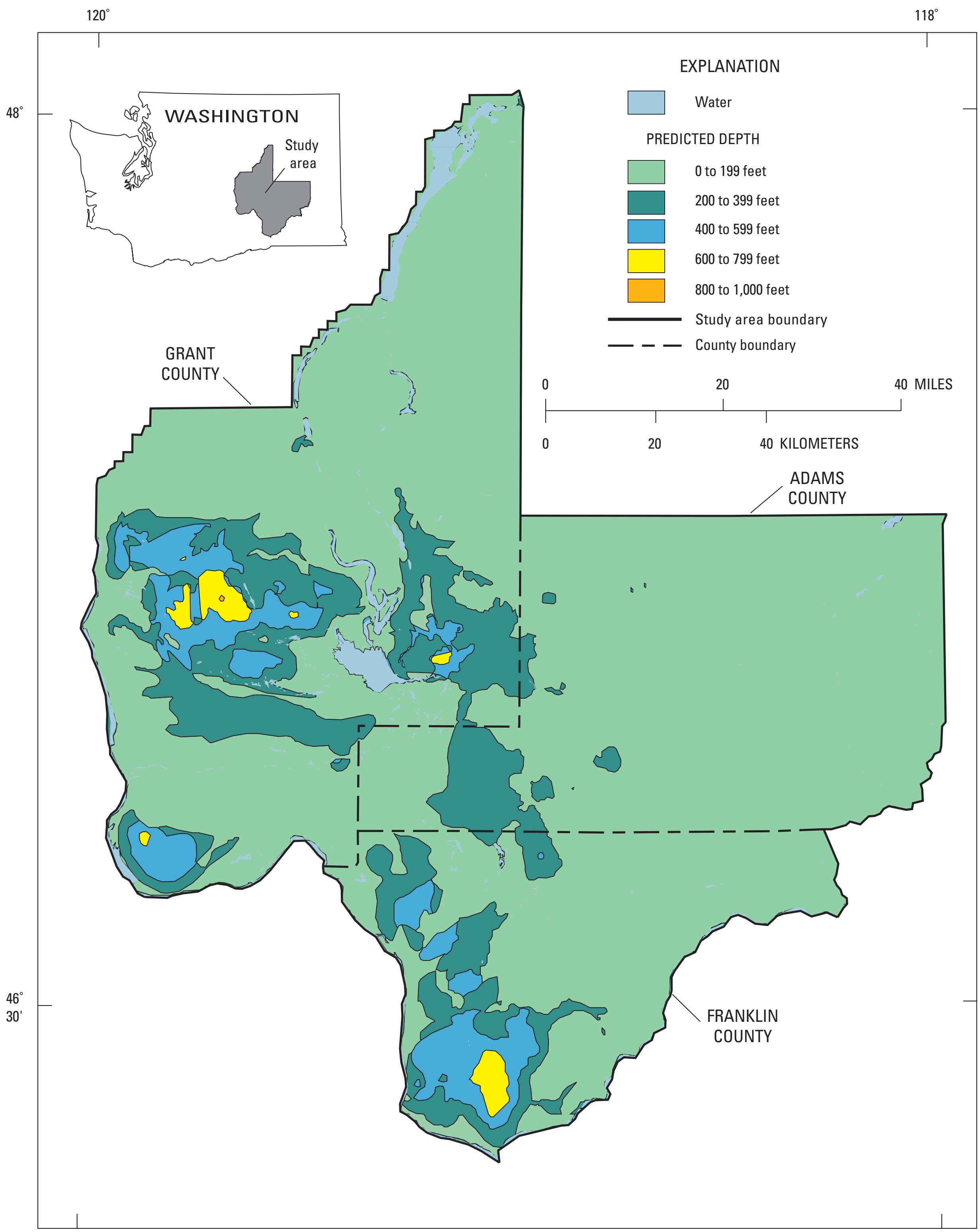

Figure 12. Predicted depth to which wells need to be cased to have at least an 80-percent chance of obtaining water with a nitrate concentration less than 10 milligrams per liter.

SUMMARY

Logistic regression was used to relate anthropogenic and natural factors to the occurrence of elevated nitrate concentrations in the Columbia Basin GWMA and develop models that predict the probability that nitrate concentrations in ground water exceed either 3 mg/L or 10 mg/L. Explanatory variables that were considered for inclusion in the models included surficial geology, soils, ground-water recharge rates, land use, presence of canals near the well, presence of irrigation near the well, well depth, well diameter, depth to the bottom of well casing, and the amount of fertilizer applied annually near the well. When each variable was individually related to elevated nitrate concentrations, most were significantly related to nitrate concentrations above 3 mg/L and about half were significantly related to nitrate concentrations above 10 mg/L. The variables that best predict the occurrence of elevated nitrate in a multivariable model were casing depth and the amount of fertilizer applied within 2 km of a well for the 3-mg/L model, while casing depth, soil hydrologic group, and the amount of fertilizer applied within 3 km of a well make up the variables in the 10-mg/L model. Maps showing the predicted probability for the occurrence of elevated nitrate concentrations indicate that the irrigated agricultural regions are most at risk. The predicted depths to which wells need to be cased in order to have an 80-percent chance of obtaining low nitrate ground water exceed 600 feet in the irrigated agricultural regions, while wells in dryland agricultural areas generally need a casing in excess of 400 feet. The predicted depth to which wells need to be cased to have at least an 80-percent chance to draw water with a nitrate concentration less than 10 mg/L generally did not exceed 800 feet, with a 200-foot casing depth being typical of the majority of the study area.

REFERENCES CITED

Alley, William, ed., 1993, Regional ground-water quality: New York, Van Nostrand Reinhold, 634 p.

Bauer, H.H., and Vaccaro, J.J., 1990, Estimates of ground-water recharge to the Columbia Plateau regional aquifer system, Washington, Oregon, and Idaho, for predevelopment and current land-use conditions: U.S. Geological Survey Water-Resources Investigations Report 88-4108, 37 p., 2 pl.

Ebbert, J.C., Cox, S.E., Drost, B.W., and Schurr, K.M., 1995, Distribution and sources of nitrate, and presence of fluoride and pesticides, in parts of the Pasco Basin, Washington, 1986-1988: U.S. Geological Survey Water-Resources Investigations Report 93-4197, 173 p.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: New York, Elsevier, 522 p.

Hosmer, D.W., and Lemeshow, S., 1989, Applied logistic regression: New York, John Wiley and Sons, 307 p.

Jones, J.L., and Wagner, R.W., 1995, Water-quality assessment of the Central Columbia Plateau in Washington and Idaho—analysis of available nutrient and pesticide data for ground water, 1942-92: U.S. Geological Survey Water-Resources Investigations Report 94-4258, 119 p.

Madison, R.J., and Brunett, J.O., 1985, Overview of the occurrence of nitrate in ground water of the United States: U.S. Geological Survey Water-Supply Paper 2275, p. 93-105.

Ryker, S.J., and Frans, L.M., 2000, Summary of nitrate concentrations in ground water of Adams, Franklin, and Grant Counties, Washington, Fall 1998—a baseline for future trend analysis: U.S. Geological Survey Water-Resources Investigations Report 99-4288, 39 p.

SAS Institute, 1990, SAS/STAT user's guide: Cary, NC., SAS Institute, Inc., 1,686 p.

Scott, J.C., 1990, Computerized stratified random site-selection approaches for design of ground-water-quality sampling networks: U.S. Geological Survey Water-Resources Investigations Report 90-4101, 109 p.

Tesoriero, A.J., Inkpen, E.L., and Voss, F.D., 1998, Assessing ground water vulnerability using logistic regression: Source Water Assessment and Protection 1998 Conference Proceedings, p. 157-165.

U.S. Department of Agriculture, 1993, State soil geographic data base (STATSGO)--data users guide: U.S. Department of Agriculture, Natural Resource Conservation Service, Miscellaneous Publication no. 1492, 88 p.

U.S. Geological Survey, 1986, Land use and land cover digital data from 1:250,000- and 1:100,000-scale maps: U.S. Geological Survey Data Users Guide 4, 33 p.

Van Metre, P., and Seevers, P., 1991, Use of Landsat imagery to estimate ground-water pumpage from irrigation on the Columbia Plateau in eastern Washington, 1985: U.S. Geological Survey Open-File Report 89-4157, 38 p.

Vogelmann, J.E., Sohl, T.L., Campbell, P.V., and Shaw, D.M., 1998, Regional land cover characterization using Landsat thematic mapper data and ancillary data sources: Environmental Monitoring and Assessment, v. 51, p. 415-428.

Washington Agricultural Statistics Service, 1992, Washington agricultural statistics 1991-1992: Olympia, Wash., Washington Agricultural Statistics Service, 137 p.

______1994, Washington agricultural statistics 1993-1994: Olympia, Wash., Washington Agricultural Statistics, 126 p.

______1995, Washington agricultural statistics 1994-1995: Olympia, Wash., Washington Agricultural Statistics, 123 p.

Washington State University Cooperative Extension, 1970a, Irrigated mint fertilizer guide for central Washington: Washington State University Cooperative Extension publication FG-8, 2 p.

———1970b, Irrigated pasture fertilizer guide for central Washington: Washington State University Cooperative Extension publication FG-4, 2 p.

———1970c, Irrigated sweet corn fertilizer guide for central Washington: Washington State University Cooperative Extension publication FG-35, 2 p.

———1976, Alfalfa fertilizer guide for central Washington Washington State University Cooperative Extension publication FG-3, 2 p.

———1979, Irrigated field corn for grain or silage fertilizer guide for central Washington: Washington State University Cooperative Extension publication FG-6, 2 p.

———1980, Irrigated field beans fertilizer guide for central Washington: Washington State University Cooperative Extension publication FG-0005, 2 p.

———1986, Winter wheat fertilizer guide for eastern Washington dryland area: Washington State University Cooperative Extension publication EB 1390, 4 p.

———1988, Winter wheat fertilizer guide for central Washington: Washington State University Cooperative Extension publication EB 1487, 4 p.

Williamson, A.K., Munn, M.D., Ryker, S.J., Wagner, R.J., Ebbert, J.C., and Vanderpool, A.M., 1998, Water quality in the Central Columbia Plateau, Washington and Idaho, 1992-1995 U.S. Geological Survey Circular 1144, 35 p.

USGS

L.M. Frans

ESTIMATING THE PROBABILITY OF ELEVATED NITRATE (NO_2+NO_3-N) CONCENTRATIONS IN GROUND WATER IN THE COLUMBIA BASIN GROUND WATER MANAGEMENT AREA, WASHINGTON

USGS WRIR 00-4110

Printed on recycled paper