

# Modeling the Probability of Arsenic in Groundwater in New England as a Tool for Exposure Assessment

JOSEPH D. AYOTTE,<sup>\*,†</sup>  
 BERNARD T. NOLAN,<sup>‡</sup>  
 JOHN R. NUCKOLS,<sup>§</sup>  
 KENNETH P. CANTOR,<sup>⊥</sup>  
 GILPIN R. ROBINSON, JR.,<sup>‡</sup>  
 DALSU BARIS,<sup>⊥</sup> LAURA HAYES,<sup>†</sup>  
 MARGARET KARAGAS,<sup>||</sup> WILLIAM BRESS,<sup>#</sup>  
 DEBRA T. SILVERMAN,<sup>⊥</sup> AND  
 JAY H. LUBIN<sup>⊥</sup>

*U.S. Geological Survey NH-VT Water Science Center, 361 Commerce Way, Pembroke, New Hampshire, 03275, U.S. Geological Survey, National Center, Reston, Virginia, 20192, Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, Colorado, 80523, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, 20892, Section of Biostatistics and Epidemiology, Department of Community and Family Medicine and the Norris Cotton Cancer Center, Dartmouth Medical School, Lebanon, New Hampshire, 03756, and Division of Health Protection, Vermont Department of Health, Burlington, Vermont, 05402*

We developed a process-based model to predict the probability of arsenic exceeding 5  $\mu\text{g/L}$  in drinking water wells in New England bedrock aquifers. The model is being used for exposure assessment in an epidemiologic study of bladder cancer. One important study hypothesis that may explain increased bladder cancer risk is elevated concentrations of inorganic arsenic in drinking water. In eastern New England, 20–30% of private wells exceed the arsenic drinking water standard of 10 micrograms per liter. Our predictive model significantly improves the understanding of factors associated with arsenic contamination in New England. Specific rock types, high arsenic concentrations in stream sediments, geochemical factors related to areas of Pleistocene marine inundation and proximity to intrusive granitic plutons, and hydrologic and landscape variables relating to groundwater residence time increase the probability of arsenic occurrence in groundwater. Previous studies suggest that arsenic in bedrock groundwater may be partly from past arsenical pesticide use. Variables representing historic agricultural inputs do not improve the model, indicating that this source does not significantly contribute to current arsenic concentrations. Due to the complexity of the fractured

bedrock aquifers in the region, well depth and related variables also are not significant predictors.

## Introduction

For the period 1950–1994, the northeastern part of the United States (U.S.) experienced higher than average mortality rates for bladder cancer in both males and females (1). Variations in bladder cancer mortality rates across the U.S. and the world suggest the role of environmental determinants (2). One possible explanation for the elevated rates is the consumption of water containing inorganic arsenic, a known bladder carcinogen at concentrations greater than 150 micrograms per liter ( $\mu\text{g/L}$ ) (3–6). Surveys of drinking water wells in northeastern New England have found arsenic concentrations exceeding 10  $\mu\text{g/L}$  (the U.S. Environmental Protection Agency's (USEPA) maximum contaminant level) in 20–30% of private wells (7–10). Private bedrock wells supply approximately 40% of the population in Maine, New Hampshire, and Vermont (7, 11).

Bedrock aquifers in New England consist primarily of fractured crystalline rocks of Pre-Cambrian to Mesozoic age and include metamorphic rock and intrusive granites. Metamorphic rocks include gneiss, marble, phyllite, slate, and schist that are found mostly in northeast trending belts. Aquifers in carbonate rocks are common in western New England. Aquifers in fault-bounded basins occur in south-eastern and south-central New England (12). Maps of New England showing geologic provinces and generalized lithology are in Appendix 1, Supporting Information.

Bedrock aquifers in crystalline rocks store and transmit water through intersecting fractures formed by various processes (13). The orientation, density, and hydraulic properties of fractures vary by rock type and structural setting and can affect groundwater flow, and ultimately water chemistry in the aquifer (14, 15). Crystalline rock aquifers have complex fracture characteristics that lead to mixing of waters from multiple flow paths (7, 12, 16, 17). Most bedrock groundwater supply wells in the region are less than 200 m deep. The aquifers in the fault-bounded basins and the carbonate rocks may be dominated by flow along lithologic contacts and connecting fractures. In addition, geochemical conditions in bedrock aquifers are complex, causing arsenic concentrations to vary markedly over short distances. These conditions make the prediction of arsenic levels in well water problematic (7–9, 16).

The source of arsenic in groundwater from New England bedrock wells is mainly geologic (9). Arsenic concentrations in well water have been linked to geologic variables, such as bedrock type or bedrock lithochemistry (7, 8, 18), although some studies have suggested the role of anthropogenic sources (19–23). Arsenical pesticides were commonly used in New England from the early 1900s through the late 1960s and may contribute to well water arsenic in some locations (21, 22).

Evaluation of source and mobility factors for arsenic in groundwater elsewhere in the world suggests the importance of strongly reducing conditions where arsenic often co-occurs with iron and manganese, indicative of desorption from metal oxides (10, 24–26). More recently, arsenic release through biogeochemical redox processes have been suggested (27–29). Geologic sources of arsenic in the U.S. include sulfide minerals, iron oxides, and mixing with geothermal waters, although the mechanism for mobility is most often release from iron and manganese oxides (9, 10). In New England,

\* Corresponding author phone: 603.226.7810; fax: 603.226.7894; e-mail jayotte@usgs.gov.

<sup>†</sup> U.S. Geological Survey NH-VT Water Science Center.

<sup>‡</sup> U.S. Geological Survey, National Center.

<sup>§</sup> Colorado State University, this work was conducted while on sabbatical at the Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, 20892, USA.

<sup>⊥</sup> National Cancer Institute.

<sup>||</sup> Dartmouth Medical School.

<sup>#</sup> Vermont Department of Health.

oxidation of sulfide minerals is a likely source (9). Redox and pH-driven release of arsenic from metal oxides in the aquifer may be important in the mildly alkaline waters of New England bedrock aquifers.

The National Cancer Institute (NCI), the Departments of Health in New Hampshire and Vermont, the Maine Bureau of Health, Dartmouth Medical School (DMS), and the U.S. Geological Survey (USGS) are collaborating in a population-based epidemiologic study of bladder cancer in northern New England. Estimates of lifetime exposure to inorganic arsenic in drinking water are required for approximately 2700 study participants with approximately 11 000 current and past homes in the 6-state region of New England (i.e., Connecticut, Massachusetts, Maine, New Hampshire, Rhode Island, and Vermont). We are measuring arsenic in water supplied to all current and some past homes of participants, as well as abstracting records for arsenic in public water supplies. A prediction model, however, is necessary to estimate arsenic levels of water from wells at past homes that cannot be directly measured. In a pilot study, 53% of past residences fell into this category (30).

For this purpose, we developed a model that identifies regional-scale geologic, hydrologic, geochemical, and land-use factors and improves our ability to predict locations where arsenic exceeds 5  $\mu\text{g/L}$  in bedrock wells.

## Methods

**Development of the Arsenic Regression Model.** We used logistic regression to model the probability that arsenic concentrations in bedrock wells are  $\geq 5 \mu\text{g/L}$ . This technique is well established and ideally suited for modeling censored dependent variable data, allowing the use of measurements below laboratory reporting levels (LRLs). (31–34). The model takes the following form:

$$P[y = 1|x] = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}} \quad (1)$$

where  $P$  is the probability of observing the event,  $y$  is an indicator variable ( $y = 1$  denoting an event or measurement  $\geq 5 \mu\text{g/L}$  and  $y = 0$  denoting a nonevent or measurement  $< 5 \mu\text{g/L}$ ),  $x_1, x_2, \dots, x_k$  are explanatory variables, and  $\beta_0, \beta_1, \dots, \beta_k$  are unknown parameters. The exponential of a parameter,  $\exp(\beta_i)$ , specifies the proportional increase in the odds of an arsenic concentration being  $\geq 5 \mu\text{g/L}$  per unit increase in the explanatory variable. An  $\exp(\beta)$  value greater than 1 represents an increasing effect and values less than 1 represent a decreasing effect of the factor. The value 5  $\mu\text{g/L}$  was used since it represents the maximum detection level among all measurements (77% of the 2470 observations used in the model were  $< 5 \mu\text{g/L}$ ) and yielded reasonable model sensitivity. Other thresholds, including 10  $\mu\text{g/L}$ , were tested and produced similar but less sensitive models. Models developed with larger thresholds are necessarily more variable since the probability of an “event” is smaller and the corresponding binomial variance is greater.

We used SAS System software (35) with backward selection, followed by selective evaluation of variables. The Hosmer–Lemeshow goodness-of-fit test (HL) was used to compare observed to fitted values for the model, and the Wald statistic was used to test individual model parameters, using a 0.10 significance level (31). Pearson residuals were used to identify poorly fitting predictions (36). Model discrimination is represented by the  $c$  statistic, which is the area under the Receiver Operating Statistics curve (31). We computed the overall rate of correct classification, sensitivity (the rate of predicting true positive measurements) and specificity (the rate of predicting true negative measurements) based on a cut point of 0.5 for the predicted probability. Thus, the

value 0 was assigned if the prediction  $\hat{P}(Y_j = 1)$  was  $< 0.5$  and the value 1 if  $\hat{P}(Y_j = 1)$  was  $\geq 0.5$ . A calibration data set of 2090 (85%) arsenic measurements was used to develop the initial model, and we evaluated it with a random validation data set of 380 (15%) measurements. The calibration and validation data were combined (2470 observations) for the final model.

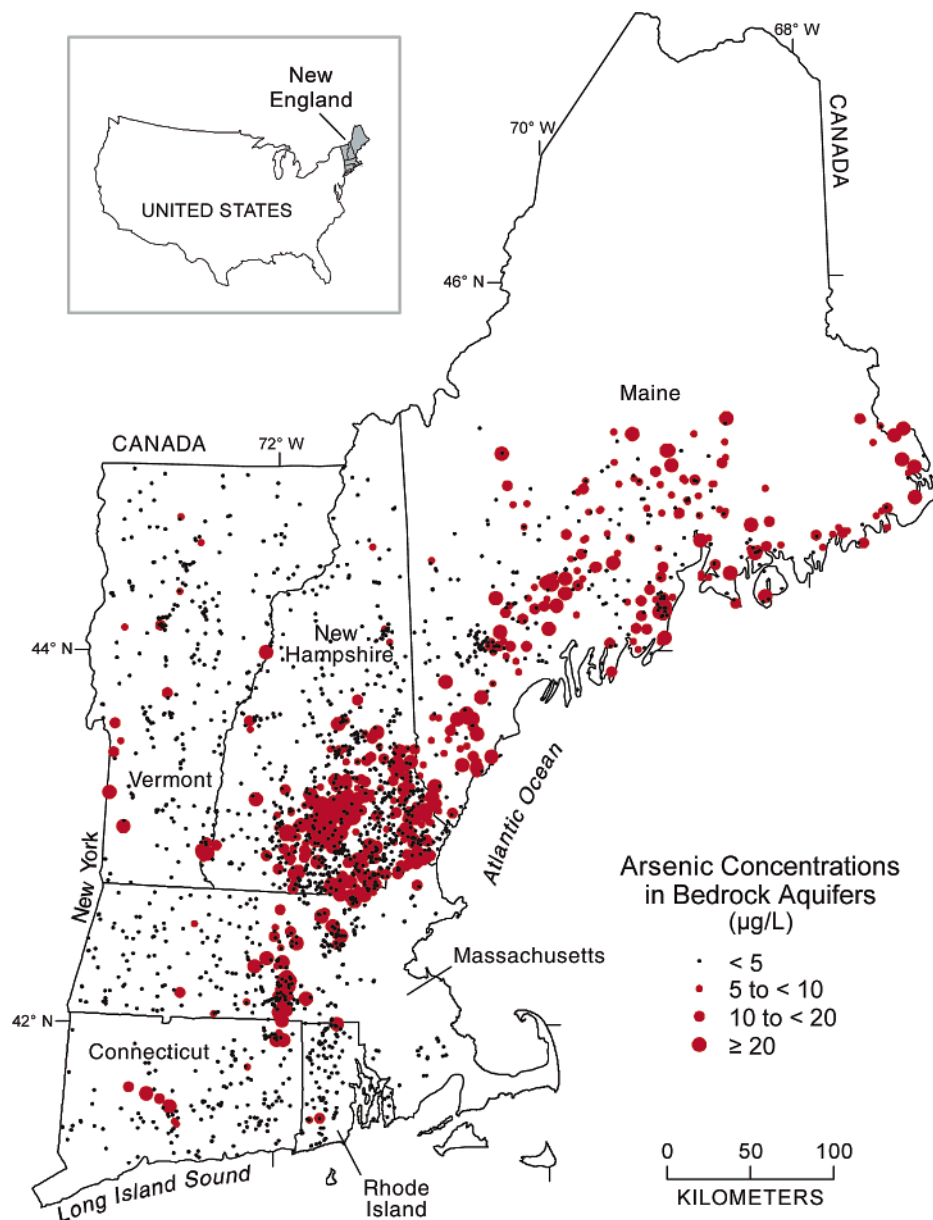
**Dependent Variable.** The dependent variable data consisted of arsenic concentrations measured in water samples collected between 1995 and 2003 from 2470 bedrock-aquifer wells in New England (Figure 1). Sixty-four percent of the measurements were from single-well bedrock public supply systems collected for compliance with USEPA Safe Drinking Water Act requirements, and the remainder were from surveys of private domestic bedrock wells (7, 18, 37). Compliance monitoring data were reported at LRLs (similar to the limit of quantitation (LOQ)) of 0.1–5  $\mu\text{g/L}$  and were analyzed according to standardized USEPA methods using graphite furnace atomic absorption spectrometry (method 200.9) and inductively coupled plasma–mass spectrometry (method 200.8). Information regarding other survey data has been reported previously (7, 37). The distribution of the data is shown in Appendix 2, Supporting Information. A total of 1902 (77%) measurements of arsenic were less than 5  $\mu\text{g/L}$ ; 267 (11%) were 5–10  $\mu\text{g/L}$ ; and 301 (12%) were  $> 10 \mu\text{g/L}$ .

**Explanatory Variables.** We assembled data representing more than 50 variables from 13 sources (7–10, 38) (Appendix 3, Supporting Information). The variables fall into three broad categories: (1) geologic and anthropogenic sources of arsenic; (2) geochemical processes; and (3) hydrogeologic and land-use factors, mainly related to groundwater flow. Arsenic concentrations in groundwater are high in wells in many of the metamorphic rocks (metasedimentary rocks derived from calcareous protoliths) in eastern New England, as well as near specific bedrock formations such as the Concord Granite (7, 8, 18, 37). Stream sediment arsenic concentrations were also evaluated, and these data potentially integrate both natural and anthropogenic sources (39).

Arsenic mobility in bedrock groundwater depends on aquifer geochemistry and redox conditions, although few of these factors have been mapped at the regional scale. Dissolved oxygen concentrations and the pH of water are redox indicators and correlate with the presence of arsenic (7, 9, 12, 38) but require direct measurement in the well. In modeling, surrogate variables were evaluated for geochemical factors including areas of Pleistocene marine inundation where ion exchange may increase groundwater pH (7) and areas near intrusive granitic plutonic rocks, since they may have caused arsenic mineral migration during late-stage pegmatite formation (8). We also evaluated generalized streamwater pH (39), alkalinity (39, 40), and information on soil characteristics (41) including permeability, percent organic matter, and texture, since soil features are factors in other parts of the world (25).

Data on precipitation, elevation, slope characteristics, recharge, well-yield, land use, and other factors can correlate with hydrologic processes such as groundwater residence time in the aquifer and transmissive properties of the aquifer (12, 42). The distance of wells to lineaments (potential bedrock fracture zones mapped from variable-scale imagery) and predicted well-yields may also relate to groundwater residence time.

Land development, agriculture, and physical landscape features may influence hydrologic processes that affect arsenic concentrations in groundwater and may indicate anthropogenic sources of arsenic. The National Land Cover Data set characterizes land-use and land-cover data from the early 1990s (43). We tested variables for categories including developed, agricultural, forest, vegetative covers, and water/wetlands. Land-use data were evaluated as



**FIGURE 1.** Locations of wells and concentrations of arsenic in water from bedrock aquifer wells in New England. The concentration data are shown with circles sized by concentration ranges. Number of samples = 2470.

percentages in a 1000 m radius circle around the well. Data on population density were from the 2000 U.S. census (44). Data on past agricultural activities were from historic county agricultural reports (45).

## Results

Arsenic concentrations in groundwater varied widely among the states. They were highest in Maine and New Hampshire and lowest in Rhode Island and Vermont. Table 1 shows estimates of the arithmetic mean (AM), geometric mean (GM), and geometric standard deviation (GSD) of the arsenic concentrations (46).

The final logistic regression model consists of 28 explanatory variables (Table 2) representing geologic arsenic sources, geochemical processes, hydrogeologic processes, and land use. The geologic variables (20 of the 28) are multi-level indicator variables. Several geologic provinces (e.g., the Waits River Province,  $\exp(\beta) = 0.054$ ; Bronson Hill,  $\exp(\beta) = 0.047$ ; Avalon Belt,  $\exp(\beta) = 0.210$ ) identified areas of reduced probability (i.e.,  $\exp(\beta) < 1$ ) of arsenic occurrence (Table 2).

Individual bedrock geologic units, such as the Massabesic Gneiss Complex ( $\exp(\beta) = 0.116$ ;  $p < 0.0001$ ) also lacked high arsenic concentrations in groundwater (7, 37), although the water chemistry is otherwise generally similar to that from wells in other bedrock units. The presence of an arsenic source was associated primarily with local-scale geologic variables, including a 13-fold greater odds of arsenic  $\geq 5\ \mu\text{g/L}$  in wells located in the Eliot Formation (SOec, Calef Member) in New Hampshire, as well as 6-fold greater odds for the Kittery (SZk) and 5-fold greater odds for the Madrid (DSm) Formations in Maine. Arsenic in stream sediments increased the odds of elevated arsenic concentrations in groundwater by 67% for each  $\text{mg kg}^{-1}$  ( $\exp(\beta) = 1.67$ ;  $p < 0.0001$ ).

For geochemical processes variables, we found 2-fold greater odds in areas of Pleistocene marine inundation ( $p < 0.0001$ ) and a 44% increase in odds associated with wells within 3 km of intrusive igneous plutons ( $p = 0.0069$ ). Statistically significant hydrologic variables included precipitation, elevation, and percent area of water bodies in a 1000m radius of a well. Significant anthropogenic variables

**TABLE 1. Well Water Arsenic Concentrations by State<sup>a</sup>**

state	arsenic concentrations ( $\mu\text{g/L}$ ) in well measurements					maximum
	N	AM	GM	GSD	% $\geq 5 \mu\text{g/L}$	
Connecticut	176	5.4	0.1	18.7	8.5	33
Maine	471	4.7	0.5	8.5	44.7	217
Massachusetts	473	9.0	3.0	4.4	12.1	1100
New Hampshire	937	9.2	1.5	6.7	28.0	300
Rhode Island	144	2.0	0.2	8.2	6.9	46
Vermont	269	2.1	0.2	8.5	6.7	156
New England	2470	n/a	n/a	n/a	23.0	1100

<sup>a</sup> N, number of samples; AM, arithmetic mean; GM, geometric mean; GSD, geometric standard deviation; n/a, not applicable. The AM, GM, and GSD are estimates based on a fitted log-normal distribution within state (46).

**TABLE 2. Explanatory Variables and Their Characteristics Used in the Final Logistic-regression Model for New England Bedrock Aquifers<sup>a</sup>**

variable	type	GIS Format	data summary <sup>b</sup>	exp( $\beta$ )	p-value
intercept				12.25	0.0037
<b>arsenic sources – geologic provinces<sup>c</sup></b>					
Avalon Belt	binary	polygon	14.0%	0.210	<0.0001
Bronson Hill Belt	binary	polygon	4.0%	0.047	0.0026
Eugeosyncline Sequence	binary	polygon	5.1%	0.339	0.0096
Mesozoic Basin	binary	polygon	1.4%	2.627	0.0443
Waits River Basin	binary	polygon	3.8%	0.054	0.0042
<b>arsenic sources – lithochemistry<sup>c,d</sup></b>					
pelitic rocks (Bronson Hill)	binary	polygon	1.0%	0.157	0.0748
peraluminous granite (New Hampshire Maine Sequence)	binary	polygon	4.5%	0.334	0.0030
mafic rocks (Narragansett Basin)	binary	polygon	0.2%	12.84	0.0040
<b>arsenic sources – bedrock geologic units<sup>c,e</sup></b>					
Concord Granite (Dc1m, granite)	binary	polygon	3.2%	1.878	0.0134
Madrid Fm. (DSm, metamorphic)	binary	polygon	0.5%	4.515	0.0244
Rindgemere Fm., lower member (DSrb, metamorphic)	binary	polygon	0.5%	3.821	0.0343
Berwick Fm., calcareous member (SObc, Calcipelite)	binary	polygon	1.9%	3.770	<0.0001
Eliot Fm., Calef member (SOec, metamorphic)	binary	polygon	0.2%	13.286	0.0281
Kittery Fm., (SOk, metamorphic)	binary	polygon	0.8%	6.176	0.0695
Perry Mountain Fm. (Sp, metamorphic)	binary	polygon	3.4%	2.093	0.0095
Rangeley Fm., lower part, (Srl, metamorphic)	binary	polygon	3.4%	2.320	0.0010
Sangerville Fm., (Sspm, metamorphic)	binary	polygon	0.8%	3.499	0.0139
Waterville Fm., (Sw, metamorphic)	binary	polygon	1.1%	2.838	0.0080
Kittery Fm., (SZk, metamorphic)	binary	polygon	0.5%	6.176	0.0241
Massabesic Gneiss Complex (Zmz, granite)	binary	polygon	2.1%	0.116	0.0033
<b>arsenic sources – integrated natural and anthropogenic</b>					
stream sediment arsenic, (ln) $\text{mg kg}^{-1}$	continuous	grid	GM = 3.4; GSD = 2.3; AM = 4.9	1.668	<0.0001
<b>geochemical processes<sup>c</sup></b>					
Pleistocene marine inundation	binary	polygon	19.0%	2.077	<0.0001
intrusive granitic pluton category (within 3 km of pluton)	binary	polygon	61.3%	1.439	0.0069
<b>hydrologic processes and land use</b>					
developed land flag (cut point of 33%) <sup>c</sup>	binary	grid	22.3%	0.597	0.0007
elevation (1:24,000 scale DEM, m)	continuous	grid	GM = 121; GSD = 2.3; AM = 162	0.997	0.0005
population density (persons $\text{km}^{-2}$ )	continuous	grid	GM = 40.3; GSD = 3.2; AM = 77.4	0.999	0.0354
precipitation, $\text{mm yr}^{-1}$	continuous	grid	GM = 1,131; GSD = 1.1; AM = 1,136	0.997	<0.0001
water bodies (% area in 1000 m radius buffer)	continuous	grid	GM = 6.6; GSD = 2.6; AM = 9.8	0.980	0.0023

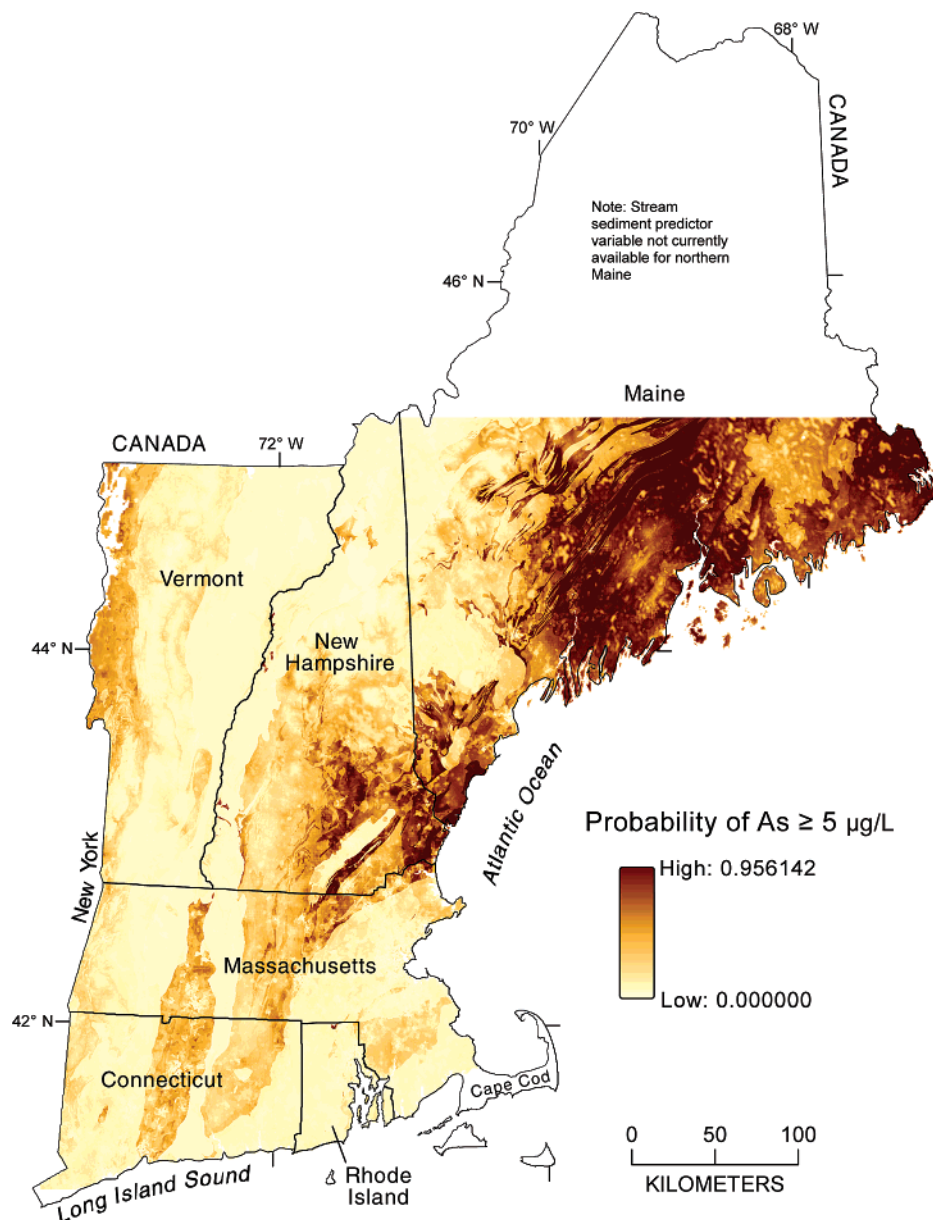
<sup>a</sup> Data set includes 2470 observations. <sup>b</sup> Percentage with attribute or the geometric mean (GM), geometric standard deviation (GSD), and arithmetic mean (AM) estimated based on a log-normal distribution within each state (46). <sup>c</sup> Data included in the model using zero-one indicator variables. <sup>d</sup> Bedrock geologic unit names are identified in Appendix 3, Supporting Information. <sup>e</sup> Capital letters in formation code (in parentheses) refer to the age of the unit. Generalized lithology follows formation code, see Appendix 1, Supporting Information(47).

included population density and percent of developed land within a 1000 m radius of a well.

Several variables were not retained in the final model. Streamwater pH (48) and alkalinity data (40, 48) were not related to bedrock groundwater arsenic concentration. Surprisingly, variables reflecting current (43) and past

agricultural land use (45) were not statistically significant and did not improve the model fit. Glacial stratified drift, a binary variable that reflects major differences in surficial geologic material (49), was also not significant.

We extensively evaluated interactions among variables. Only the interaction of the stream-sediment arsenic variable



**FIGURE 2. Predicted probability of arsenic concentrations  $\geq 5 \mu\text{g/L}$  in groundwater wells in bedrock aquifers. Truncation at about 45 degrees north latitude reflects absence of stream-sediment data. Bedrock wells are not generally used on Cape Cod.**

and the intrusive igneous pluton variable was marginally statistically significant. None of the interactions improved the model fit significantly and thus were not included in the final model.

The rate of total correct classification for the model is 79.8%, with 92.5% of the arsenic concentrations  $< 5 \mu\text{g/L}$  (specificity) and 37.1% of the arsenic concentrations  $\geq 5 \mu\text{g/L}$  (sensitivity) being correctly classified. Classification results were similar to the results for the validation data set (Appendix 4, Supporting Information). The c statistic was 0.82 for our final model, indicating excellent model discrimination (31). The  $p$ -values associated with the HL statistic (0.6735) indicated that overall model fit was very good. (Higher HL  $p$ -values indicate better fit because the null hypothesis is that the model fits the data.)

Our model-based predicted probabilities of arsenic concentrations  $\geq 5 \mu\text{g/L}$  for groundwater in New England (at a resolution of 30 m) are shown in Figure 2. The probability map emphasizes the spatial variability of elevated arsenic in New England and shows that Maine and New Hampshire have the highest probabilities. Pearson residuals of the relative

errors of prediction are small in most regions, ranging from  $-2.2$  to  $+7.2$ , with a 5th and 95th percentile range of  $-1.0$  to  $+1.8$  and a median value of  $-0.25$  (Appendix 5, Supporting Information). The largest residuals generally occur in areas where arsenic concentrations are most variable, indicating that arsenic concentrations can vary substantially over relatively short distances.

## Discussion

Direct measurement of inorganic arsenic in drinking water from private wells provides the best basis for estimating individual exposure to arsenic from this source. However, such measurements often are not available or feasible (30). Several studies have identified localized contamination, but none have related arsenic occurrence to predictor variables in New England. This paper presents a process-based model of inorganic arsenic in bedrock aquifer wells in New England. It indicates that geologic source, geochemistry, and hydrologic and land-use variables are significant predictors (Table 2) and suggests that historic use of arsenical pesticides in

agriculture is not related to current levels of arsenic in bedrock groundwater.

Stream sediment arsenic concentration is a strong predictor of arsenic probability in groundwater, but iron and manganese concentrations in stream sediments, which primarily occur as metal oxyhydroxide coatings, are not. One explanation for the lack of predictive significance of iron and manganese is that stream sediment arsenic is not well correlated with stream sediment iron or manganese (50). Arsenic on stream sediment may represent long-term accumulation of natural and, possibly, anthropogenic sources of arsenic.

Areas of Pleistocene marine inundation are associated with a higher probability of arsenic concentrations equal to or exceeding 5  $\mu\text{g/L}$ . The effect of this factor ( $\exp(\beta) = 2.077$ ;  $p < 0.0001$ ) may relate to the potential for arsenic mobilization due to ion-exchange-related increases in pH (7). Increases in pH due to ion exchange with sodium have been documented in parts of the central United States (51, 52). Increased pH can mobilize arsenic (10, 24) and is closely related to high concentrations of groundwater arsenic in areas of New England (7).

The geochemical effects of granitic plutons intruding the surrounding rocks may enhance arsenic occurrence in groundwater wells (8). In our model, well proximity (within 3000 m) to mapped intrusive granitic plutons increases the odds of arsenic exceeding 5  $\mu\text{g/L}$  by 44% ( $p = 0.0069$ ), supporting the hypothesis that late-stage magmatic and hydrothermal fluids associated with intrusive plutonic activity may concentrate arsenic in pegmatites and along fractures in adjacent bedrock.

Based on our model, the effects of precipitation, elevation, and percentage of water bodies (within a 1000 m radius of a well) suggest that arsenic is positively related to groundwater residence time (i.e., groundwater with relatively short, shallow flow paths or relatively rapid velocity has reduced arsenic concentrations). Our finding differs from a bedrock aquifer study in eastern New England (7), where arsenic concentration was not related to average groundwater age. Deeper, older groundwater may be in contact with naturally occurring sources of contaminants for longer periods of time than shallow groundwater (53), increasing the chance of mobilizing arsenic.

The effects of developed land and population density ( $\exp(\beta) < 1$ ) are less clear. Areas with high population density may have increased groundwater nitrate (54), from such sources as septic systems, lawn fertilization, sewer lines, and atmospheric deposition. Data from recent regional studies in eastern New England show that arsenic is almost never present above 1  $\mu\text{g/L}$  in groundwater where nitrate concentrations exceed 1  $\text{mg/L}$  (55). One hypothesis is that arsenic sorbs to aquifer materials in the presence of nitrate (a potential terminal electron acceptor for microbial oxidation of arsenite to arsenate) (27, 28); another is that nitrate is simply an indicator of oxidizing conditions, which limits arsenic mobility.

In addition, groundwater arsenic concentrations in the bedrock aquifer do not correlate with some factors that affect other groundwater contaminants. Well depth, for example, is not correlated with arsenic concentrations in groundwater from wells as it is with nitrate in other hydrologic systems (54, 56). Some have hypothesized that historical arsenical pesticide use may affect groundwater arsenic concentrations (19–21, 23). However, our study suggests that this is not the case for bedrock aquifers. We found no significant association between bedrock groundwater arsenic and surrogates for historical arsenical pesticide use, including an index of agricultural intensity, class of agriculture, and percent of cropped lands and orchards. The relevance of this finding for past bedrock aquifer arsenic concentrations, however, is unclear.

Our model is restricted to groundwater from fractured crystalline bedrock in New England because concentrations of arsenic are significantly higher in these aquifers than in the unconsolidated aquifers (7). The model indicates that bedrock geology is a major factor in predicting arsenic occurrence, includes multiple factors that are conceptually related to arsenic occurrence, and provides evidence for mobilization of arsenic by desorption. Last, our model is being used for exposure assessment in an epidemiologic study of bladder cancer. However, despite an improved understanding of controls on arsenic in this region, arsenic prediction in individual wells remains problematic. Additional data, such as the proximity of sampled wells to rock fractures, may account for additional variation in local hydrogeologic and geologic conditions and improve the model prediction. Also, the inclusion of improved geologic variables at higher spatial resolution may ultimately lead to more precise modeling.

## Acknowledgments

We thank the private citizens; the New Hampshire Department of Environmental Services; the Vermont Department of Environmental Conservation; the Maine Bureau of Health, Geological Survey, and Department of Conservation; the Massachusetts Department of Environmental Quality; the Rhode Island Department of Health; and the Connecticut Department of Health for arsenic data used in this study. Alan H. Welch and Dennis R. Helsel of the U.S. Geological Survey and Sarah J. Ryker of Carnegie-Mellon University provided conceptual input. This research was supported, in part, by the Intramural Program of NIH (NCI) and by the National Water-Quality Assessment Program of the U.S. Geological Survey.

## Supporting Information Available

Maps of New England showing geologic provinces and generalized lithology, a probability map emphasizing the spatial variability of elevated arsenic in New England, classification results, explanatory variables, and survey data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## Literature Cited

- 1) Devesa, S. S.; Grauman, D. G.; Blot, W. J.; Pennello, G.; Hoover, R. N.; Fraumeni, J. F., Jr. *Atlas of cancer mortality in the United States, 1950–94*; NIH Publication no. (NIH) 99-4564; U.S. Govt. Printing Office: Washington, D.C., 1999.
- 2) Silverman, D. T.; Morrison, A. S.; Devesa, S. S. Bladder Cancer, In *Cancer Epidemiology and Prevention*, 2nd ed.; Schottenfeld, D., Fraumeni, J. F., Jr., Eds.; Oxford University Press: New York, 1996; pp 1156–1179.
- 3) Hopenhayn-Rich, C.; Biggs, M. L.; Fuchs, A.; Bergoglio, R.; Tello, E. E.; Nicolli, H.; Smith, A. H. Bladder cancer mortality associated with arsenic in drinking water in Argentina. *Epidemiology* **1996**, *7* (2), 117–124.
- 4) Karagas, M. R.; Tosteson, T. D.; Blum, J.; Morris, J. S.; Baron, J. A.; Klaue, B. Design of an epidemiologic study of drinking water arsenic exposure and skin and bladder cancer risk in a U.S. population. *Environ. Health Perspect.* **1998**, *106* (Supplement 4), 1047–1050.
- 5) Smith, A. H.; Goycolea, M.; Haque, R.; Biggs, M. L. Marked increase in bladder and lung cancer mortality in a region of Northern Chile due to arsenic in drinking water. *Am. J. Epidemiol.* **1998**, *147* (7), 660–669.
- 6) Wu, M. M.; Kuo, T. L.; Hwang, Y. H.; Chen, C. J. Dose-response relation between arsenic concentration in well water and mortality from cancers and vascular diseases. *Am. J. Epidemiol.* **1989**, *130* (6), 1123–1132.
- 7) Ayotte, J. D.; Montgomery, D. L.; Flanagan, S. F.; Robinson, K. W. Arsenic in ground water in eastern New England: Occurrence, controls, and human health implications. *Environ. Sci. Technol.* **2003**, *37* (10), 2075–2083.

- (8) Peters, S. C.; Blum, J. D.; Klaue, B.; Karagas, M. R. Arsenic occurrence in New Hampshire drinking water. *Environ. Sci. Technol.* **1999**, *33* (10), 1328–1333.
- (9) Welch, A. H.; Westjohn, D. B.; Helsel, D. R.; Wanty, R. B. Arsenic in ground water of the United States: Occurrence and geochemistry. *Ground Water* **2000**, *38* (4), 589–604.
- (10) Welch, A. H., Stollenwerk, K. G., Eds.; *Arsenic in ground water: geochemistry and occurrence*; Kluwer Academic Publishers: Boston, 2003.
- (11) U.S. Bureau of the Census. Historical Census of Housing Tables—Sources of Water. 1999; <http://www.census.gov/hhes/www/housing/census/historic/water.html> (accessed Oct 2002).
- (12) Rogers, R. J. Geochemical comparison of ground water in areas of New England, New York, and Pennsylvania. *Ground Water* **1989**, *27*, 690–712.
- (13) Hansen, B. P.; Simcox, A. C. *Yields of bedrock wells in Massachusetts*; Water-Resources Investigations Report 93-4115; U.S. Geological Survey: 1994.
- (14) Drew, L. J.; Karlinger, M. R.; Schuenenmeyer, J. H.; Armstrong, T. R. Hydrogeologic significance of the association between well-yield variability and bedrock geologic structures, Pinardville Quadrangle, New Hampshire. *Nat. Resour. Res.* **2002**, *12* (1), 79–91.
- (15) Drew, L. J.; Schuenenmeyer, J. H.; Armstrong, T. R.; Sutphin, D. M. Initial yield to depth relation for water wells drilled into crystalline bedrock—Pinardville Quadrangle, New Hampshire. *Ground Water* **2001**, *39* (5), 676–684.
- (16) Shapiro, A. M. Cautions and suggestions for geochemical sampling in fractured rock. *Ground Water Monit. Rem.* **2002**, *22* (3), 151–164.
- (17) Shapiro, A. M. *Characterizing fractured rock for water supply: From the field to the watershed*; Fractured Rock Aquifers 2002, Proceedings, Denver, March 13–15, 2002; National Ground Water Association: Westerville, OH, 2002.
- (18) Ayotte, J. D.; Nielsen, M. G.; Robinson, G. R., Jr.; Moore, R. B. *Relation of arsenic, iron, and manganese in ground water to aquifer type, bedrock lithochemistry, and land use in the New England Coastal Basins*; Water-Resources Investigations Report 99-4162; U.S. Geological Survey: Pembroke, 1999; <http://pubs.water.usgs.gov/wri994162>.
- (19) Marvinney, R. G.; Loiselle, M. C.; Hopeck, J. T.; Braley, D.; Krueger, J. A. Arsenic in Maine ground water: an example from Buxton, Maine; 1994 Focus Conference on Eastern Regional Groundwater Issues; National Groundwater Association: Westerville, OH, 1994.
- (20) Boudette, E. L.; Canney, F. C.; Cotton, J. E.; Davis, R. I.; Ficklin, W. H.; Motooka, J. M. *High levels of arsenic in the ground waters of southeastern New Hampshire—a geochemical reconnaissance*; Open-File Report 85-202; U.S. Geological Survey: Bow, NH, 1985.
- (21) Chormann, F. H., Jr., Masters thesis, Department of Earth Sciences, University of New Hampshire, Durham, 1985.
- (22) D'Angelo, D.; Norton, S. A.; Loiselle, M. C. *Historical uses and fate of arsenic in Maine*; University of Maine Water Research Institute Completion Report: Orono, ME, 1996.
- (23) Zeuna, A. J.; Keane, N. W. Arsenic contamination of private potable wells; U.S. Environmental Protection Agency National Conference on Environmental Engineering, Northeastern University, Boston, MA, July 1–5, 1985; American Society of Civil Engineers: Reston, VA, 1985.
- (24) Smedley, P. L.; Kinniburgh, D. G. Arsenic in groundwater and the environment, In *The essentials of medical geology: Impacts of the natural environment on public health*; Selinus, O., Ed.; Elsevier: Boston, 2005; pp 263–299.
- (25) Smedley, P. L. Arsenic in groundwater—south and east Asia, In *Arsenic in groundwater: geochemistry and occurrence*; Welch, A. H., Stollenwerk, K. G., Eds.; Kluwer Academic Publishers: Boston, 2003; pp 179–209.
- (26) Stollenwerk, K. G. Geochemical processes controlling transport of arsenic in groundwater: a review of adsorption, In *Arsenic in Ground Water: Geochemistry and Occurrence*; Welch, A. H., Stollenwerk, K. G., Eds.; Kluwer Academic Publishers: Boston, 2003; pp 67–100.
- (27) Harvey, C. F.; Swartz, C. H.; Badruzzaman, A. B. M.; Keon-Blute, N.; Yu, W.; Ali, M. A. Arsenic mobility and groundwater extraction in Bangladesh. *Science* **2002**, *298* (5598), 1602–1606.
- (28) Oremland, R. S.; Stoltz, J. F. Arsenic, microbes, and contaminated aquifers. *Trends Microbiol.* **2005**, *13* (2), 45–49.
- (29) Polizzotto, M. L.; Harvey, C. F.; Sutton, S. R.; Fendorf, S. Processes conducive to the release and transport of arsenic into aquifers of Bangladesh. *Proc. Natl. Acad. Sci. USA* **2005**, *102* (52), 18819–18823.
- (30) Colt, J. S.; Baris, D.; Clark, S. F.; Ayotte, J. D.; Ward, M.; Nuckols, J. R.; Cantor, K. P.; Silverman, D. T.; Karagas, M. R. Sampling private wells at past homes to estimate arsenic exposure: a methodologic study in New England. *J. Exposure Anal. Environ. Epidemiol.* **2002**, *12*, 329–334.
- (31) Hosmer, D. W.; Lemeshow, S. *Applied Logistic Regression*; 2nd ed.; John Wiley and Sons: New York, 2000.
- (32) Helsel, D. R.; Hirsch, R. M. *Statistical Methods in Water Resources*; Elsevier Science Company, Inc.: New York, 1992.
- (33) Helsel, D. R. *Nondetects and data analysis: statistics for censored environmental data*; First ed.; John Wiley & Sons: New York, 2005.
- (34) Harris, D.; Zurcher, L.; Stanley, M.; Marlow, J.; Pan, G. A comparative analysis of favorability mappings by weights of evidence, probabilistic neural networks, discriminant analysis, and logistic regression. *Nat. Resour. Res.* **2003**, *12* (4), 241–255.
- (35) SAS Institute. *SAS/STAT User's Guide, Version 6*; 4 ed.; SAS Institute Inc.: Cary, 1990; Vol. 1 and 2.
- (36) Menard, S. *Applied Logistic Regression Analysis*; Sage Publications: Thousand Oaks, CA, 2002.
- (37) Montgomery, D. L.; Ayotte, J. D.; Hamlin, P.; Carroll, P. R. *Arsenic concentrations in private bedrock wells in Southeastern New Hampshire*; Fact Sheet FS-051–03; U.S. Geological Survey: Pembroke, NH, 2003; <http://pubs.water.usgs.gov/fs-051-03/>.
- (38) Alley, W. M. *Regional Ground Water Quality*; Van Nostrand Reinhold: New York, 1993.
- (39) Robinson, G. R., Jr.; Kapo, K. E.; Grossman, J. N. *Chemistry of stream sediments and surface waters in New England*; Open-File Report 04-1026; U.S. Geological Survey: Reston, VA, 2004; <http://pubs.usgs.gov/of/2004/1026/>.
- (40) Omernik, J. M.; Kinney, A. J. *Total alkalinity of surface waters—a map of the New England and New York region*; EPA-600/D-84–216; U.S. Environmental Protection Agency: Corvallis, OR, 1985.
- (41) U.S. Department of Agriculture Natural Resources Conservation Service. State Soil Geographic (STATSGO) database for New England. 1994; <http://www.ncgc.nrcs.usda.gov/branch/ssb/products/statsgo/index.html> (accessed Apr. 13, 2004).
- (42) Moore, R. B.; Schwarz, G. E.; Clark, S. F., Jr.; Walsh, G. J.; Degnan, J. R. *Factors related to well yield in the fractured-bedrock aquifer of New Hampshire*; Professional Paper 1660; U.S. Geological Survey: Pembroke, 2002; <http://pubs.water.usgs.gov/pp1660>.
- (43) U.S. Geological Survey. National land cover data (NLCD) 1992. 2000; <http://landcover.usgs.gov/nationallandcover.asp> (accessed Oct. 27, 2003).
- (44) U.S. Bureau of the Census. 2000 Census of Population and Housing (SF1). 2000; [http://arcdata.esri.com/data/tiger2000/tiger\\_download.cfm](http://arcdata.esri.com/data/tiger2000/tiger_download.cfm) (accessed Nov. 17, 2003).
- (45) Robinson, G. R., Jr. Personal communication, 2003.
- (46) Lubin, J. H.; Colt, J. S.; Camann, D.; Davis, S.; Cerhan, J. R.; Severson, R. K.; Bernstein, L.; Hartge, P. Epidemiologic evaluation of measurement data in the presence of detection limits. *Environ. Health Perspect.* **2004**, *112* (17), 1691–1696.
- (47) Robinson, G. R., Jr.; Kapo, K. E. *Generalized lithology and lithochemical character of near-surface bedrock in the New England region*; Open-File Report 03-225; U.S. Geological Survey: Reston, VA, 2003; <http://pubs.usgs.gov/of/2003/of03-225/of03-225.pdf>.
- (48) Smith, S. M. National geochemical database (reformatted data from the National Uranium Resource Evaluation Hydrogeochemical and Stream Sediment Reconnaissance program), 2001; <http://pubs.usgs.gov/of/1997/ofr-97-0492> (accessed July 20, 2004).
- (49) Flanagan, S. M.; Nielsen, M. G.; Robinson, K. W.; Coles, J. F. *Water quality assessment of the New England Coastal Basins in Maine, Massachusetts, New Hampshire, and Rhode Island—environmental settings and implications for water quality and aquatic biota*; Water-Resources Investigations Report 98-4249; U.S. Geological Survey: Pembroke, NH, 1999.
- (50) Robinson, G. R., Jr.; Ayuso, R. A. Use of spatial statistics and isotopic tracers to measure the influence of arsenical pesticide use on stream sediment chemistry in New England. *Appl. Geochem.* **2004**, *19*, 1097–1110.
- (51) Schlottmann, J. L.; Breit, G. N. Mobilization of arsenic and uranium in the central Oklahoma aquifer, In *Water-Rock Interaction*; Kharaka, Y. K., Maest, A. S., Eds.; A. A. Balkema: Rotterdam, 1992; pp 835–838.
- (52) Appelo, C. A. J.; Postma, D. *Geochemistry, Groundwater and Pollution*; A. A. Balkema: Rotterdam, 1993.

- (53) Focazio, M. J.; Reilly, T. E.; Rupert, M. G.; Helsel, D. R. *Assessing groundwater vulnerability to contamination: providing scientifically defensible information for decision makers*; Circular 1224; U.S. Geological Survey: Reston, VA, 2002; <http://pubs.water.usgs.gov/circ1224>.
- (54) Nolan, B. T.; Hitt, K. J.; Ruddy, B. C. Probability of nitrate contamination of recently recharged groundwaters in the United States. *Environ. Sci. Technol.* **2002**, *36* (10), 2138–2145.
- (55) U.S. Geological Survey. National Water-Quality Assessment program, data warehouse. 2005; <http://water.usgs.gov/nawqa/data> (accessed Jan. 4, 2005).
- (56) Tesoriero, A. J.; Voss, F. D. Predicting the probability of elevated nitrate concentrations in the Puget Sound basin: Implications for aquifer susceptibility and vulnerability. *Ground Water* **1997**, *35* (6), 1029–1039.

*Received for review October 4, 2005. Revised manuscript received March 17, 2006. Accepted March 24, 2006.*

ES051972F