U.S. Department of the Interior
U.S. Geological Survey

Scientific Investigations Map 2881
Sheet 1 of 2

Nitrate in Ground Water:

Using a model to simulate the probability of nitrate contamination of shallow ground water in the conterminous United States

By Kerie J. Hitt and Bernard T. Nolan
National Water-Quality Assessment Program
2005

Introduction

Nitrate comes from nitrogen, an essential plant nutrient supplied to cultivated plants mostly in inorganic fertilizers and animal manure. An elevated level of nitrate in drinking water, however, poses possible health risks. Elevated nitrate levels in ground water used as drinking supplies also may indicate the presence of additional contaminants. Among mixtures detected at least 1 percent of the time in ground-water samples collected from domestic and public supply wells in the United States, nitrate frequently occurred with atrazine (herbicide), simazine (herbicide), and deethylatrazine (breakdown product of atrazine) (Squillace and others, 2002). In addition to the human health concerns, nitrate in ground water can affect nitrate levels in surface water, especially during periods of low streamflow when ground water may discharge to a stream. Nutrient enrichment (eutrophication) of surface water—commonly caused by nitrate and phosphate—can degrade water quality by triggering excessive growth of nuisance plants. Excess nitrate is a key factor in eutrophication of coastal waters. Reducing the concentration of nitrate in ground water requires expensive, specialized treatment measures. Aquifers with nitrate concentrations greater than the U.S. Environmental Protection Agency's drinking-water standard of 10 milligrams per liter (mg/L) are very difficult to remediate. As a result, identifying areas at risk of contamination and preventing contamination are more effective than trying to clean up aquifers after contamination occurs.

The National Water-Quality Assessment (NAWQA) Program has collected information on nitrate and other water-quality constituents in 51 major river basins and ground-water systems across the Nation. Ancillary information on hydrology, land use, chemical use, and natural features, such as geology, soils, and climate, also has been collected at monitored sites to help explain where and why certain water conditions exist and why some areas are at higher risk for contamination. These assessments of monitoring data provide the needed understanding for estimating nitrate occurrence at unmonitored sites under a wide range of conditions. Statistical predictive models provide one means for estimating nitrate occurrence at unmonitored sites. These models are cost-effective management tools that resource managers and scientists can use to assess the vulnerability and sustainability of water resources for future supply across broad regions.

What follows is a national map of the probability of nitrate contamination of shallow ground water (typically 5 meters deep or less), which was produced using a logistic regression model based on six explanatory variables. This approach relates the probability that nitrate concentration exceeds a certain threshold (4 mg/L) to those natural or human factors that influence nitrate concentrations. The outcome predicted by the model—the probability of nitrate contamination—is made by associating nitrate measurements with known nitrogen input sources (such as fertilizer applications), land use (such as agricultural land use and population density), and characteristics of soils and aquifers, which control the transport of nitrate to ground water.

Sheet 2 (see reverse side) contains supplementary information about the importance of ground water, general factors influencing nitrate contamination of ground water, and fundamentals of designing a model. Technical details of the model and explanatory variables are provided in "Probability of nitrate contamination of recently recharged groundwaters in the conterminous United States" by Nolan and others (2002).

Findings

- *The map highlights general areas where nitrate more likely occurs in shallow, recently recharged ground water. The map provides a focus for resource management strategies to prioritize areas for monitoring, clean up, or restoration.*
- *The likelihood of nitrate contamination of shallow ground water is greatest in areas with high nitrogen input and well-drained surficial soils that overlie unconsolidated sand and gravel aquifers. (For example, the High Plains of northeastern Nebraska and northwestern Texas.)*
- *Of the many variables examined, six were determined as most likely to affect the probability of nitrate contamination of shallow ground water. As the following factors increase, the likelihood of contamination increases:*

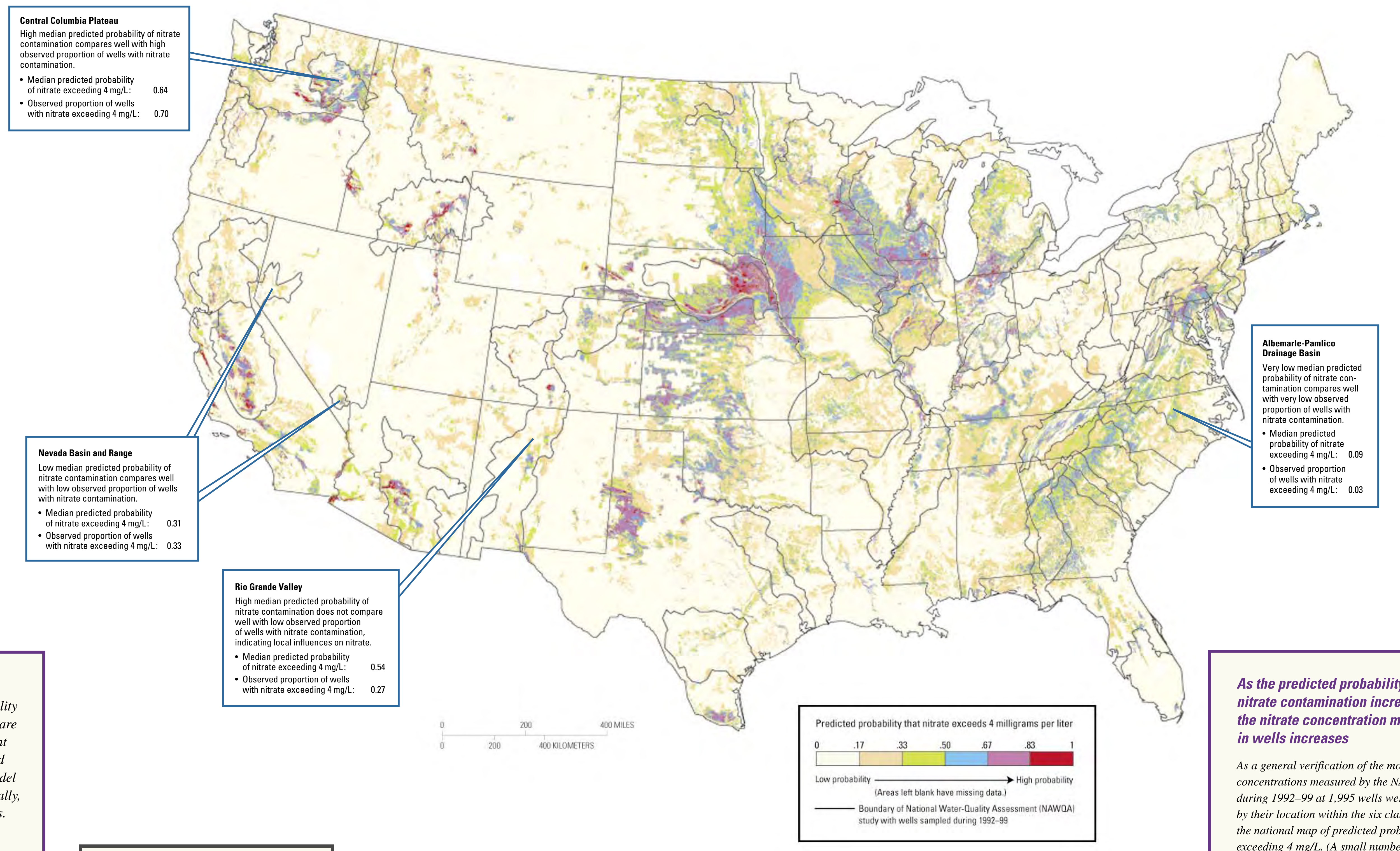
Nitrogen input
- *nitrogen fertilizer applications*
- *agricultural land use*
- *population density*

Aquifer susceptibility
- *well-drained soils*
- *depth to seasonally high water table*[1]
- *unconsolidated sand and gravel aquifers*

---

[1] Depth to seasonally high water table represents the minimum thickness of the unsaturated zone.

Probability of Nitrate Contamination of Shallow Ground Water



Central Columbia Plateau
High median predicted probability of nitrate contamination compares well with high observed proportion of wells with nitrate contamination.
- Median predicted probability of nitrate exceeding 4 mg/L: 0.64
- Observed proportion of wells with nitrate exceeding 4 mg/L: 0.70

Nevada Basin and Range
Low median predicted probability of nitrate contamination compares well with low observed proportion of wells with nitrate contamination.
- Median predicted probability of nitrate exceeding 4 mg/L: 0.31
- Observed proportion of wells with nitrate exceeding 4 mg/L: 0.33

Rio Grande Valley
High median predicted probability of nitrate contamination does not compare well with low observed proportion of wells with nitrate contamination, indicating local influences on nitrate.
- Median predicted probability of nitrate exceeding 4 mg/L: 0.54
- Observed proportion of wells with nitrate exceeding 4 mg/L: 0.27

Albemarle-Pamlico Drainage Basin
Very low median predicted probability of nitrate contamination compares well with very low observed proportion of wells with nitrate contamination.
- Median predicted probability of nitrate exceeding 4 mg/L: 0.09
- Observed proportion of wells with nitrate exceeding 4 mg/L: 0.03

The nitrate level observed for wells generally follows the probability of nitrate contamination predicted for the wells, but local conditions must be considered when analyzing water quality

Comparisons of median predicted probability with the observed proportion of wells with nitrate exceeding 4 mg/L for NAWQA study areas in the Albemarle-Pamlico Drainage Basin, Nevada Basin and Range, Central Columbia Plateau, and Rio Grande Valley are shown around the national probability map.

In three of the four examples (all but Rio Grande Valley), model predictions are comparable to nitrate measurements. Central Columbia Plateau illustrates the situation in which high median predicted probability of nitrate contamination corresponds to a high observed proportion of wells with nitrate concentrations greater than 4 mg/L.

Model predictions, however, do not compare as well for the Rio Grande Valley, where nitrogen input and the predicted probability of contamination are high, but the measured nitrate concentration generally is low. Leaching of nitrate is locally controlled and highly variable in the Rio Grande Valley, and depends on factors such as the timing of fertilizer and irrigation and recharge rate to the aquifer. In addition, evapotranspiration is high in the southern part of the Rio Grande Valley, which can limit nitrate leaching. These local conditions might explain the tendency of the model to overpredict the probability of nitrate contamination in this area.

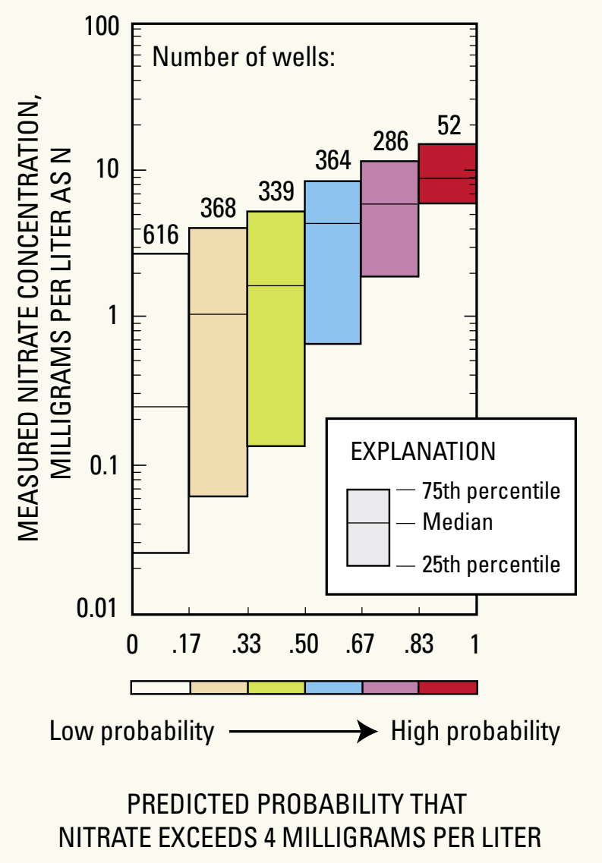
The national probability map is intended for regional (multi-county) use and has several limitations. Areas of high probability on the map have high potential for nitrate contamination, but are not necessarily contaminated. Variables that are not significant at the national scale (such as percentage of artificially drained soils) or were not considered or were not available during model calibration (such as irrigation) can affect nitrate leaching locally, so the map should not be used for local management decisions. Also, variations in local hydrogeologic conditions can cause variations in water quality that are inconsistent with mapped probabilities. For example, sinkholes in karst areas can facilitate nitrate leaching to ground water, but karst features could not be mapped at the national scale.

Logistic regression identified six explanatory variables that significantly affect nitrate contamination of shallow ground water.

The six variables were used as input to the logistic regression equation to produce the national probability map.

As the predicted probability of nitrate contamination increases, the nitrate concentration measured in wells increases

As a general verification of the model, nitrate concentrations measured by the NAWQA Program during 1992–99 at 1,995 wells were categorized by their location within the six classes on the national map of predicted probability of exceeding 4 mg/L. (A small number of wells fell in areas of missing data on the map and were not assigned a predicted probability.) The NAWQA data, which strongly supported the modeled probabilities, showed that the median concentration of nitrate increased from 0.24 mg/L in the geographic areas of lowest probability of contamination to more than 8 mg/L in those areas of highest probability of contamination.



VARIABLES REPRESENTING NITROGEN INPUT

(1) 1991–93 nitrogen fertilizer applications

Farm fertilizer applications were allocated to agricultural lands, and non-farm fertilizer applications were allocated to residential lands to account for spatial variations in land use within counties.

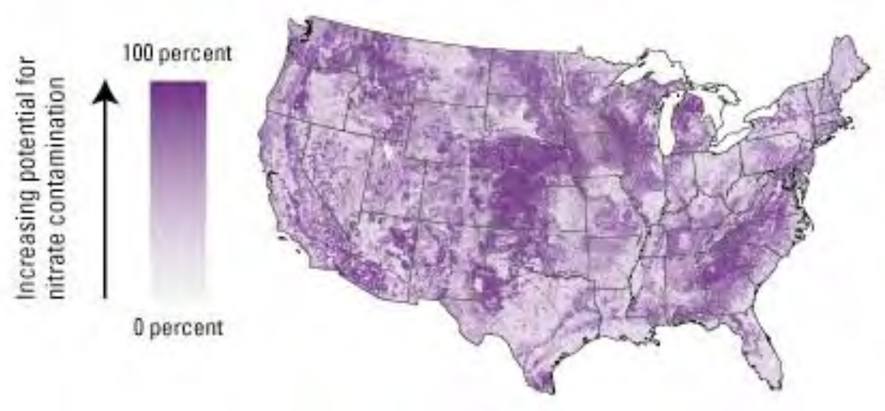
(2) Percentage of agricultural land use

A high percentage of cropland and pasture agricultural land use (row crops, small grains, fallow land, and pasture/hay) indicates a high intensity of agriculture and also represents high contributions of nitrogen from manure and rural septic systems.

(3) 1990 population density

In more urbanized areas, population density is a surrogate for non-agricultural sources of nitrogen, such as septic systems, sewers, industrial emissions, and pet waste.

VARIABLES REPRESENTING AQUIFER SUSCEPTIBILITY



(4) Percentage of well-drained soils

Infiltration characteristics of the upper 1.5 meters of soil indicate the potential for nitrogen to leach into ground water. Soils with high infiltration rates such as coarse-grained sands and gravels allow a ready pathway for water and nitrate from the land surface to enter ground water. Poorly drained soils hinder the movement of nitrate into the water table.

(5) Depth to seasonally high water table

Depth to seasonally high water table represents the minimum thickness of the unsaturated zone. Soils with a shallow depth to water table (waterlogged soils) have a high potential for denitrification and a low likelihood of nitrate contamination.

(6) Unconsolidated sand and gravel aquifers

The presence or absence of unconsolidated sand and gravel aquifers combined with variables 4 and 5 indicates the overall susceptibility of shallow aquifers to contaminants leaching from the land surface. Sand and gravel aquifers are more susceptible to nitrate leaching from the land surface.

Model development

The risk of ground-water contamination by nitrate depends on many human and natural factors, which can be grouped into two broad categories: 1) how much nitrogen is put on the land surface *(nitrogen input)* and 2) how susceptible the ground water is to nitrate leaching and accumulation *(aquifer susceptibility)*. Aquifer vulnerability depends on the combined effects of nitrogen input and the intrinsic susceptibility of the aquifer.

Logistic regression was used to evaluate 13 possible explanatory variables representing nitrogen input and aquifer susceptibility that potentially influence nitrate concentration in shallow ground water. Logistic regression relates the probability of an outcome (for example, nitrate greater than 4 mg/L) to explanatory variables using the method of "maximum likelihood estimation." Maximum likelihood estimation is used to optimize the likelihood of observing the data on which the model is based.

The model was calibrated using measured nitrate concentrations in 1,280 wells sampled in 20 NAWQA study areas during 1992–95. The model results were validated by comparing modeled predictions to additional nitrate data measured in 736 different wells sampled in 16 other NAWQA study areas during 1996–99.

Results of the model identified three nitrogen input variables *(fertilizer applications, agricultural land use, and population density)* and three aquifer susceptibility variables *(soil drainage, depth to seasonally high water table, and sand and gravel aquifers)* as the most significant factors that explain the probability of nitrate contamination of shallow ground water. The six variables are highly significant, and the model correctly predicted nitrate status in 68 percent of the wells used for calibration and in 75 percent of the wells used for validation, indicating that the model fits the data well.

A threshold of 4 mg/L was used in the model because this concentration indicates contributions of nitrate from human activities rather than what may occur naturally (Nolan and Hitt, 2003). In addition, 4 mg/L has been associated with increased risk of non-Hodgkins lymphoma in Nebraska (Ward and others, 1996). The drinking-water standard of 10 mg/L was not used as the logistic regression threshold because it was considered to be too high to protect drinking-water source areas.

Shallow wells (for which depth from land surface to water typically is less than 5 meters) were used in this study to represent the effects of land use on recently recharged ground water. Deep wells containing old ground water, which might have yielded samples with low nitrate concentration, were not considered.

For building the model and predicting the probability of contamination for the wells, national data sets describing the explanatory variables were assembled from various sources and at different spatial resolutions. For making the probability map, significant variables were recompiled at a uniform 1-square kilometer level of detail. Nolan and others (2002) describe more details on the input data sets.

Values of the six explanatory variables identified as most likely to influence nitrate contamination of shallow ground water were used as input to the logistic regression model to predict the probability of nitrate exceeding 4 mg/L in unsampled areas of the Nation—a total of more than 7 million 1-square-kilometer grid cells spanning the conterminous United States. The results are depicted on the national map of probability of nitrate contamination of shallow ground water.

Contents