USGS
science for a changing world

Prepared in cooperation with the
New Jersey Department of Environmental Protection

Median Nitrate Concentrations in Groundwater in the New Jersey Highlands Region Estimated Using Regression Models and Land-Surface Characteristics

Scientific Investigations Report 2015–5075
Version 1.1, August 2015

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

Cover: All photos provided by Ronald J. Baker, U.S. Geological Survey. Upper left, upper right, lower left, show farms in Long Valley, N.J., and lower right a mill in historic Waterloo Village Park, Stanhope, N.J.

Median Nitrate Concentrations in Groundwater in the New Jersey Highlands Region Estimated Using Regression Models and Land-Surface Characteristics

By Ronald J. Baker, Mary M. Chepiga, and Stephen J. Cauller

Prepared in cooperation with the
New Jersey Department of Environmental Protection

Scientific Investigations Report 2015–5075
Version 1.1, August 2015

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

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2015
Revised: August 2015 (ver. 1.1)

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit http://www.usgs.gov/pubprod/.

Suggested citation:
Baker, R.J., Chepiga, M.M., and Cauller, S.J., 2015, Median nitrate concentrations in groundwater in the New Jersey Highlands Region estimated using regression models and land-surface characteristics (ver. 1.1, August 2015): U.S. Geological Survey Scientific Investigations Report 2015–5075, 27 p., http://dx.doi.org/10.3133/sir20155075.

ISSN 2328-0328 (online)

Acknowledgments

The authors gratefully acknowledge the New Jersey Highlands Council for their interest and participation in this work; and Jeff Hoffman, Nick Procopio, and Sandra Goodrow of the New Jersey Department of Environmental Protection for their many contributions and constructive suggestions during report development and review.

Contents

Figures

Tables

Conversion Factors, Datums

Inch/Pound

Multiply	By	To obtain
	Area	
square kilometer (km^2)	247.1	acre
square kilometer (km^2)	0.3861	square mile (mi^2)
	Volume	
liter (L)	33.82	ounce, fluid (fl. oz)
liter (L)	2.113	pint (pt)
liter (L)	1.057	quart (qt)
liter (L)	0.2642	gallon (gal)
liter (L)	61.02	cubic inch (in^3)
	Mass	
gram (g)	0.03527	ounce, avoirdupois (oz)

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88).

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Abbreviations

ANOVA	Analysis of variance
EPA	U.S. Environmental Protection Agency
GPS	Global positioning system
MCL	Maximum contaminant level
MDL	Minimum detection limit
MLE	Maximum likelihood estimate
NJDEP	New Jersey Department of Environmental Protection
NWIS	National Water Information System
PWTA	Private Well Testing Act
USGS	U.S. Geological Survey

Median Nitrate Concentrations in Groundwater in the New Jersey Highlands Region Estimated Using Regression Models and Land-Surface Characteristics

By Ronald J. Baker, Mary M. Chepiga, and Stephen J. Cauller

Abstract

Nitrate-concentration data are used in conjunction with land-use and land-cover data to estimate median nitrate concentrations in groundwater underlying the New Jersey (NJ) Highlands Region. Sources of data on nitrate in 19,670 groundwater samples are from the U.S. Geological Survey (USGS) National Water Information System (NWIS) and the NJ Private Well Testing Act (PWTA).

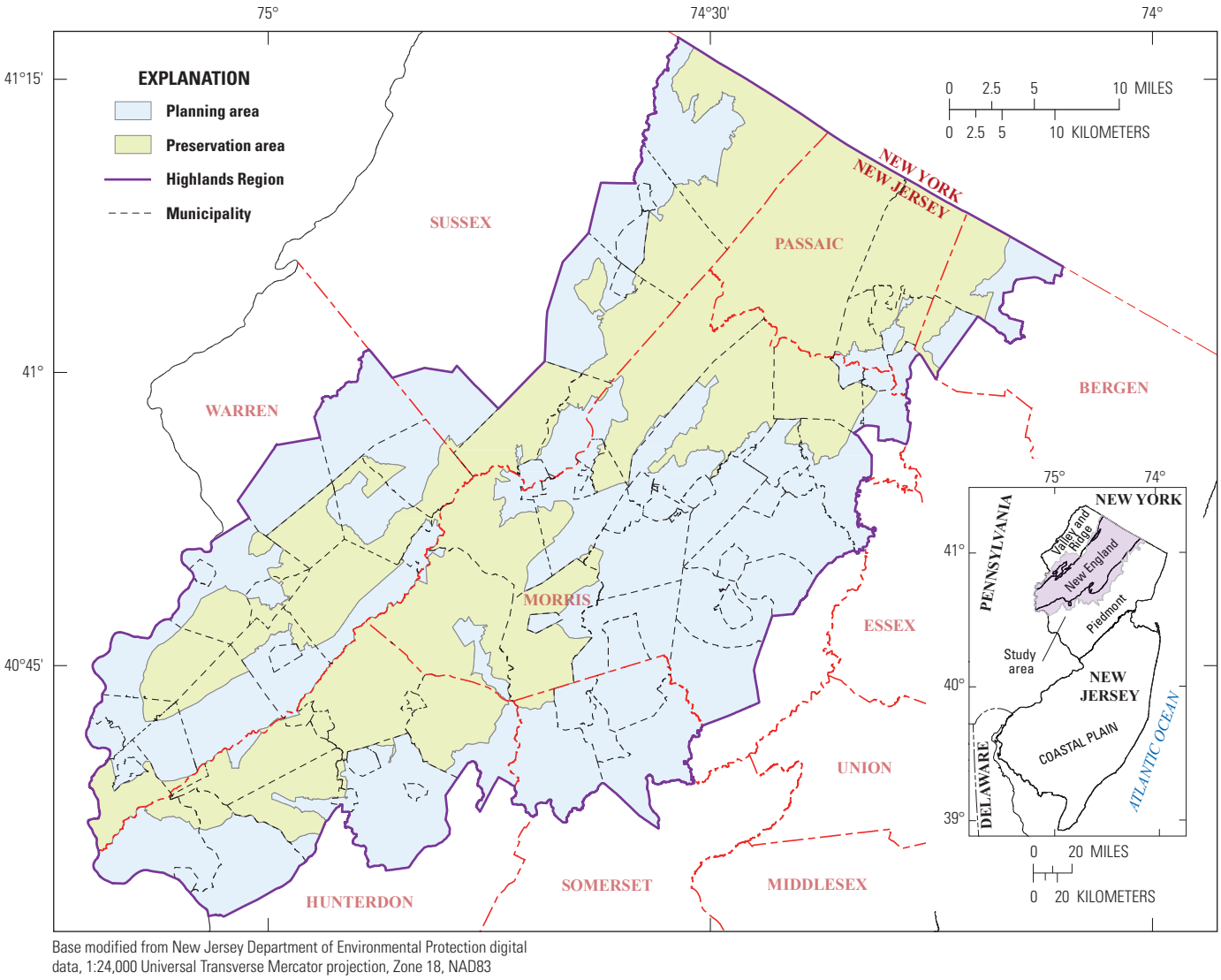
In a study conducted by the USGS, in cooperation with the New Jersey Department of Environmental Protection, logistic regression was used to relate measured nitrate concentrations to five explanatory variables (percent urban and agricultural land use, septic-system density, total length of streams, and number of known contaminated sites) quantified in 610-meter-square grid cells). A method for calculating the median concentrations of nitrate from a series of logistic regression models was developed. Two calibration and two validation procedures showed that the logistic-regression-based method can estimate groundwater-nitrate concentrations in the Highlands Region accurately to within 0.1 milligram per liter as nitrogen (mg/L as N). Limitations of the logistic-regression-based method include the inability to select a logistic model with exactly 0.5 probability of exceeding the threshold value and lack of an algorithm to directly calculate the median value. Quantile regression was evaluated as a suitable alternative and was slightly less accurate than the logistic-regression method in estimating median groundwater nitrate concentrations in the Highlands Region.

Multiple-linear regression with log-transformed nitrate-concentration data and the same five explanatory values was less accurate than either logistic or quantile regression in estimating median nitrate concentrations. On the basis of 4,516 2000 x 2000 foot grid cells that contain wells with data stored in NWIS and the PWTA database, the estimated median nitrate concentration for the entire Highlands Region is about 1.25 mg/L as N, and estimated median concentrations range from about 1.05 to 1.78 mg/L as N among 11 smaller administratively defined areas within the Highlands Region that vary in percentages of urban land use, agricultural land use, and septic-system density.

The Kaplan-Meier method of estimating summary statistics from left-censored data was applied in order to include nondetects (left-censored data) in median nitrate-concentration calculations. Median concentrations also were determined using three alternative methods of handling nondetects. Treatment of the 23 percent of samples that were nondetects had little effect on estimated median nitrate concentrations because method detection limits were mostly less than median values.

Introduction

Monitoring and assessment of groundwater quality is important for the management of groundwater resources from a public health and an ecological perspective. Groundwater quality and anticipated water-quality changes in the New Jersey (NJ) Highlands Region, which is distinct from but overlaps much of the Highlands Physiographic Province (fig. 1), are used by government agencies as regulatory criteria for land-use decisions. Nitrate (NO_3) concentrations are used as an indicator of overall water quality (New Jersey Highlands Water Protection and Planning Council, 2008) One objective of the Highlands Regional Master Plan is "to determine the amount and type of human development and activity which the ecosystem of the Highlands Region can sustain." This objective has zoning and building-restriction implications. The first step in observing changes in groundwater nitrate concentrations is to characterize pre-regulatory (pre-2008) nitrate concentrations, which are used as a "baseline" for comparison with present (2014) and future nitrate concentrations. Although the groundwater nitrate concentration at a location (for example, a single house or public supply well) can be reliably determined by sampling and analyzing the well water, determining the central tendency and range of nitrate concentrations for an entire region such as the Highlands Region is problematic. Therefore, the U.S. Geological Survey (USGS), in cooperation with the New Jersey Department of



Base modified from New Jersey Department of Environmental Protection digital data, 1:24,000 Universal Transverse Mercator projection, Zone 18, NAD83

Figure 1. New Jersey Highlands Region with Planning and Preservation Areas.

Environmental Protection, conducted a study to determine the best method for use in estimating nitrate concentrations in the Highlands Region.

The range of nitrate concentrations in groundwater in an area can be quantified by sampling water from a representative number of randomly distributed wells. The range, however, will be biased if the wells are not uniformly distributed throughout the study area. Public supply, industrial, agricultural, domestic, and observation wells used for groundwater sampling tend to be installed in land-use areas that are urban, suburban, and agricultural. Wells are less frequently installed in forested, barren, and wetlands areas. Previous investigations (Wakida and Lerner, 2005; Nolan and others, 1998; Dubrovsky and Hamilton, 2010) report that nitrate concentrations in groundwater in urbanized, industrialized, and agricultural areas are consistently greater than those in forested and wetland areas. Therefore, median nitrate concentrations determined in samples from wells in urban and other developed areas likely will be higher than the median nitrate concentrations in samples from wells in forested and wetlands areas. One remedy for such bias is to use data from a subset of wells that are uniformly distributed geographically. This would eliminate geographic bias at a cost of decreasing the data density, reducing confidence in the statistical analyses, and possibly introducing additional bias from the well-selection criteria. An alternative method to eliminate bias is to relate nitrate concentrations to explanatory variables such as land use, surface activities, soil characteristics, hydrology, and population, then use these relations to estimate nitrate concentrations for each area of interest. The alternative method was used in this study to estimate baseline groundwater median nitrate concentrations in the NJ Highlands Region and areas within the Highlands Region.

Regression models can be used to relate water-quality characteristics, such as nitrate concentrations, to independent (explanatory) variables, such as percentages of different

land-use categories and septic-system density. Logistic-regression models typically are used to relate a set of explanatory variables to the probability of exceeding a threshold value of a water-quality characteristic (Greene and others, 2005; Huang and others, 2013; Tu, 2008; Eckhardt and Stackelberg., 1995; Nolan, 2001; Gardner and Vogel, 2005; Tesoriero and Voss, 2005; Gurdak and Qi, 2012. Typically, the threshold value is a concentration of interest, for example 2 milligrams per liter (mg/L) nitrate as nitrogen (N) in groundwater (Gardner and Vogel, 2005), which was suggested by Mueller and Helsel (1996) as a conservative value to indicate anthropogenic effects. For this study, rather than calculating the probability of exceeding a threshold value, the probability was set in advance (50%, which corresponds to the probability of exceeding the median value), and the logistic-regression model and corresponding threshold concentration (which is the median value) was then determined.

Quantile regression and multiple-linear regression (MLR) were tested as alternatives to the logistic-regression method for estimating median nitrate concentrations in the Highlands Region. Quantile regression (Kroenker and Hallock, 2001) is used to relate one or more independent variables to the value of one dependent variable that corresponds to specified quantiles of the range of dependent-variable values. Multiple linear regression (MLR) is commonly used to relate sets of explanatory variables to water-quality constituents (Helsel and Hirsch, 2002). For example, Sando and others (2014) relate log-transformed concentrations of water-quality constituents to time, streamflow, and season. MLR was similarly applied in this study. Median calculations based on logistic regression, quantile regression, and multiple-linear regression were all subjected to calibration checks and validation by comparing medians of lab-measured nitrate concentrations to calculated concentrations. The best estimates of median nitrate concentrations for the entire New Jersey Highlands Region and smaller administratively defined areas were then calculated.

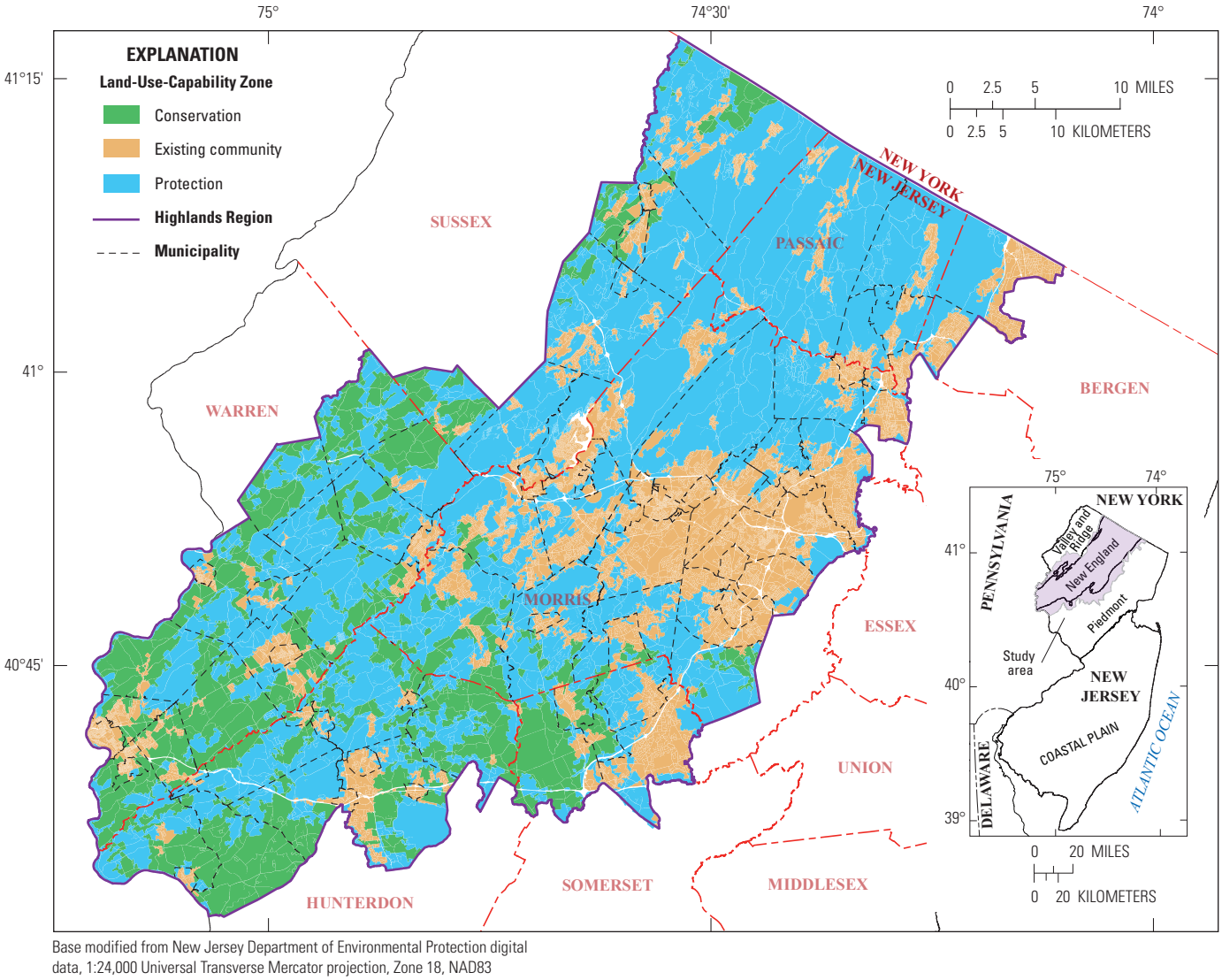
Purpose and Scope

The purpose of this report is to present the methods used to quantify median groundwater nitrate concentrations in the NJ Highlands Region and to estimate median concentrations for the entire Highlands Region and for selected areas within the Highlands Region. Criteria for selecting explanatory variables and developing logistic-regression models are presented, and statistical tests used to evaluate the logistic regression models also are presented. Benefits and limitations of the methods used are described. Comparisons are made between measured and estimated median values. Nitrate concentrations were less than the method detection limit (MDL) in 23 percent of the groundwater samples tested. Nondetects are water samples in which the constituent of interest, in this case nitrate, was not detected. The Kaplan-Meier method of including nondetects (also referred to as left-censored data) was selected over three other methods (assigning nondetects a value of zero, one-half the MDL, and the MDL), though the estimated median nitrate concentrations determined using the other three nondect methods also are presented. The effect of applying each of these four methods on the estimated median nitrate concentrations is described. The estimated median nitrate concentrations were determined for the entire Highlands Region and selected areas within the Highlands Region, and these values can be used as baseline conditions for comparison with future concentrations as land use and other surface characteristics change over time.

Description of Study Area

New Jersey is divided into four physiographic provinces: the Coastal Plain, Piedmont, Highlands, and Valley and Ridge (Dalton, 2003). The Piedmont, Highlands, and Valley and Ridge consist mostly of a series of discontinuous, rounded ridges separated by deep, narrow valleys and occupy the northern one-third of the area of New Jersey. The NJ Highlands Physiographic Province is part of the Highlands that extends to Connecticut, New York, and Pennsylvania (U.S. Forest Service, 2014).

The NJ Highlands Region is an administratively (not geologically) defined area that overlaps, but is distinct from, the Highlands Physiographic Province. The New York-NJ Highlands Region was first delineated in a study by the U.S. Forest Service (USFS) (Michaels and others, 1992). The USFS described it as an area of national significance that largely consists of forests and farms but needing protection from encroaching urban sprawl. The Highlands Region was later expanded to include areas of Pennsylvania and Connecticut. The NJ Highlands Region encompasses 2,505 square kilometers (km^2) of the Highlands Physiographic Province and also includes 654 km^2 of the Piedmont and 316 km^2 of the Valley and Ridge Physiographic Provinces. The Highlands Region covers 3,474 km^2 and includes parts of Hunterdon, Somerset, Sussex, Warren, Morris, Passaic and Bergen Counties (New Jersey Highlands Council, 2008). The protection of forests and wetlands, preservation of farmland, and permitting of additional urbanization in existing community areas are objectives of the NJ Highlands Water Protection and Planning Council. The Council was formed in 2004 as a provision of the Highlands Water Protection and Planning Act (N.J.S.A. 58:12A-26 et seq.; New Jersey Highlands Water Protection and Planning Council, 2008), which was primarily enacted to protect the drinking-water source for 5.4 million residences of New Jersey and New York. The NJ Highlands Region is divided into the Planning Area, administered by the New Jersey Highlands Council, in which conformance with the Regional Master Plan is voluntary; and the Preservation Area, administered by the New Jersey Department of Environmental Protection, in which conformance with the Regional Master Plan administered by the Highlands Council is mandatory. The Planning and Preservation Areas are each further divided into three Land-Use-Capability Zones: the Conservation Zone, Existing Community Zone, and Preservation Zone (fig. 2, table 1).



Base modified from New Jersey Department of Environmental Protection digital data, 1:24,000 Universal Transverse Mercator projection, Zone 18, NAD83

Figure 2. NJ Highlands Region with Land-Use Capability Zones.

Land use in the NJ Highlands Region, as determined from 2007 land use-land cover data (New Jersey Department of Environmental Protection (NJDEP) 2010), consists of about 44 percent of forest land; 12 percent of agricultural land; 26 percent of urban land; and 15 percent of barren, wetlands, and water (fig. 3). Percentages of the six major land-use categories in the Highlands Region vary by Planning Area and Preservation Area and among the three Land-Use-Capability Zones within each Area are shown in table 1.

The Preservation Area consists of 18,000 km^2, and the Planning area is slightly larger, about 19,000 km^2. There is more urban and agriculture land use in the Planning Area than in the Preservation Area (table 1). Urban expansion is limited in the Protection Zones and Conservation Zones. Urban expansion is allowed to a greater extent in the Existing Community Zones but only as is "compatible with the protection and character of the Highlands environment, at levels that are appropriate to maintain the character of established communities," as stated in the Master Plan (New Jersey Highlands Council, 2008). The Conservation Zone consists of highly agricultural land, and urban expansion is limited to protect the resources and character of this Zone (New Jersey Highlands Council, 2008). In the Preservation Area and Planning Area, forest and wetland land uses dominate the Protection Zone, agriculture dominates the Conservation Zone, and urban land use dominates the Existing Community Zone.

Previous Investigations

Previous investigations addressed nitrate in the study area and relations between land-use patterns and nitrate in groundwater. Agricultural and domestic fertilizers and septic systems are acknowledged as substantial sources of nitrate to groundwater (Nolan and others, 2002). Nicholson and others (1996) studied the hydrogeology, groundwater flow, and nitrates in the NJ Highlands. Glacial valley-fill, carbonate

Table 1. Land use in the NJ Highlands Region.

Area and zone	Land use, in percent[1,2]					
	Urban	Agricultural	Forest	Wetlands	Barren	Water
Entire NJ Highlands	27.0	12.3	45.6	10.3	2.1	2.7
Planning Area	37.8	16.9	33.0	11.3	1.0	3.6
Conservation Zone	15.89	44.48	27.46	10.32	0.74	1.11
Existing Community Zone	64.16	2.83	20.97	6.64	1.11	4.29
Protection Zone	22.98	6.54	49.54	15.77	0.92	4.25
Preservation Area	16.9	8.1	60.3	9.6	0.6	4.6
Conservation Zone	16.81	38.12	33.80	10.04	0.35	0.89
Existing Community Zone	54.64	3.56	28.00	7.04	0.77	5.99
Protection Zone	13.56	3.71	67.32	9.82	0.61	4.98
All grid cells with sampled wells	32.92	13.26	41.38	8.82	0.57	2.80
All grid cells with no sampled wells	21.75	11.47	49.61	11.08	0.93	4.89

[1]Calculated from New Jersey Department of Environmental Protection digital data (New Jersey Department of Environmental Protection, 2010)

[2]Percentages do not sum to 100 because of rounding.

rock, and gneissic rock aquifers were identified as the major sources of water. Human activities affected water resources by increasing the concentrations of volatile organic compounds, iron, and nitrate and in groundwater and surface water, and by consumptive use, resulting in decreasing discharge to streams and lower water tables. The maximum nitrate concentration in water samples collected from 73 wells completed in three aquifers was 9.5 milligrams per liter (mg/L) as nitrogen (N), with the highest nitrate concentrations in samples from present or previous agricultural areas and urbanized areas. The distribution of nitrate concentrations in forested and wetland areas indicate that background concentrations are less than 1 mg/L as N. Concentrations varied among three aquifers (table 2); glacial valley-fill aquifers had higher median concentrations of nitrate than carbonate and gneissic aquifers.

Serfes (1994) studied the natural groundwater quality in bedrock aquifers of the Newark Basin. The Newark Basin is synonymous with the Piedmont Physiographic Province and is adjacent to and southeast of the Highlands Physiographic Province. Nitrate concentrations in 55 well-water samples ranged from 0.1 to 7.4 mg/L as N with a median concentration of 1.6 mg/L as N.

Clawges and Vowinkel (1996) evaluated the roles of well construction, hydrogeology, and land use in the susceptibility of groundwater bedrock aquifers in the Newark Basin to nitrate contamination. Nitrate was the dominant form of nitrogen measured in samples from 132 wells. Nitrate concentrations ranged from less than 0.1 to 9.5 mg/L as N with a median concentration of 1.6 mg/L as N, similar to that in the Piedmont Physiographic Province (Serfes, 1994). Shallow wells and wells with shallow open intervals were associated with higher nitrate concentrations, indicating a greater effect from agriculture and urbanization. Groundwater nitrate concentrations were statistically lower in forested areas and wetlands than in agricultural and urban areas. The highest concentration measured was 9.5 mg/L as N in a groundwater sample from an agricultural area.

Serfes (2004) summarized the groundwater quality in the bedrock aquifers of the Highlands and Valley and Ridge Physiographic Provinces in New Jersey (fig. 1 inset) from 97 well-water samples collected from the Middle Proterozoic, Kittatinny Supergroup, and Martinsburg Formations. Nitrate concentrations ranged from less than (<) 0.05 to 5.7 mg/L as

Table 2. Statistical summary of nitrate in groundwater samples from the glacial valley-fill, carbonate-rock, and gneissic-rock aquifer systems in the NJ Highlands Region.

[mg/L, milligrams per liter; N, nitrogen; <, less than]

Aquifer type	Nitrate concentration (mg/L as N)[1]			
	Samples	Minimum	Median	Maximum
Glacial Valley-Fill	27	<0.10	1.40	6.10
Carbonate rock	30	<0.10	1.00	9.50
Gneissic rock	16	<0.10	0.38	2.00

[1] From Nicholson and others, 1996

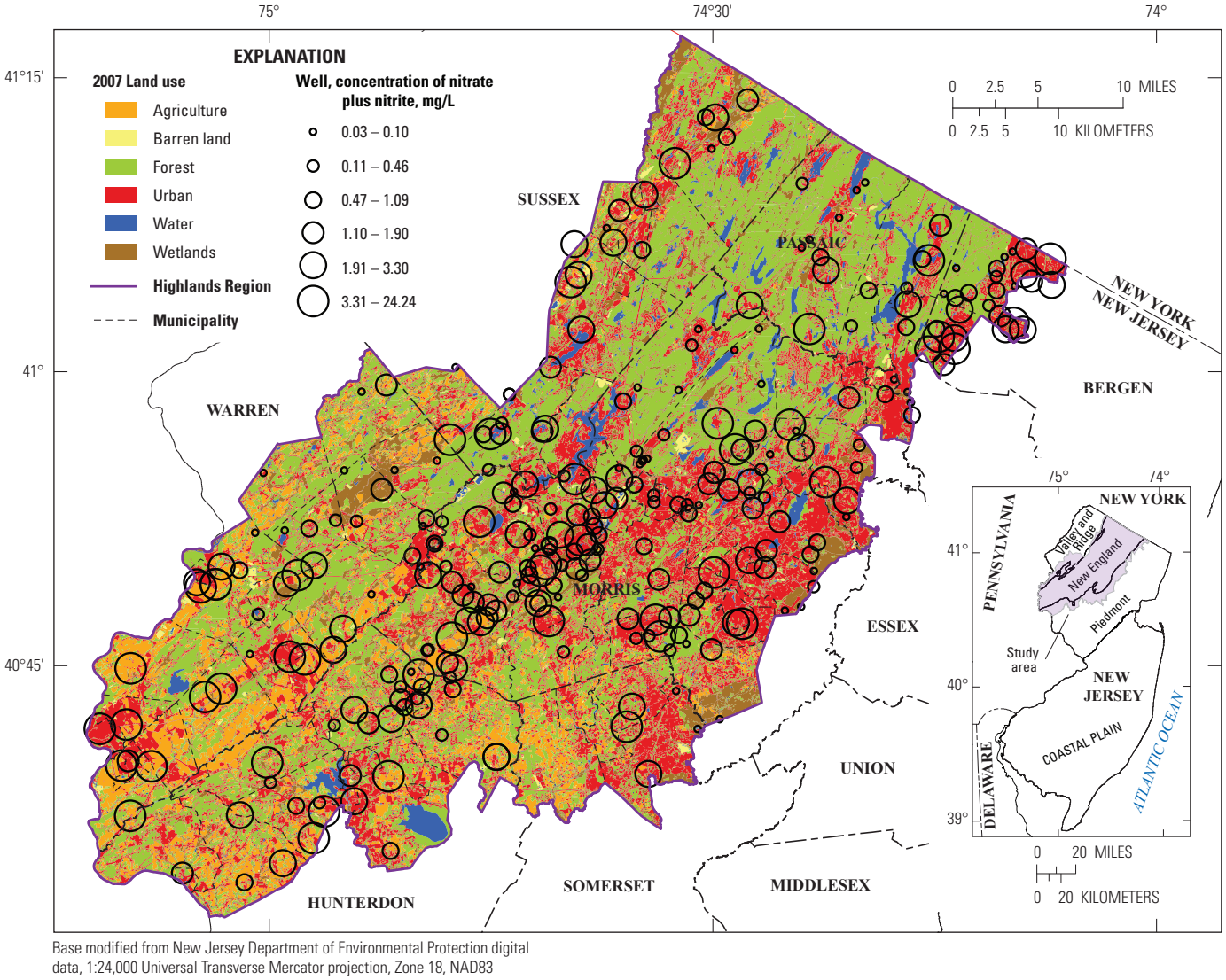


Figure 3. Land-use patterns, locations of wells with data in the U.S. Geological Survey National Water Information System, and nitrate concentrations in groundwater in the New Jersey Highlands Region. (mg/L, milligrams per liter)

N in these samples (table 3). The median nitrate concentration among the 97 samples was 0.41 mg/L as N.

Hoffman and Canace (2004) published a method to estimate nitrate concentrations in groundwater near septic systems and the average land area required to sufficiently dilute nitrate in septic-system effluent to avoid exceeding maximum nitrate concentration goals for potable water supply. The method is based on a mass-dilution model (Trela and Douglas, 1978) and a groundwater-recharge model (Charles and others, 1993). This method is currently (2015) used in the New Jersey Pinelands to protect groundwater resources by limiting the density of new urban construction.

Dubrovsky and others (2010) reviewed water-quality data from 1992 to 2004 for streams and groundwater throughout the United States. Nationally, nitrate concentrations in groundwater appear to be increasing slowly. Data from more than 5,000 wells showed that nitrate concentrations in groundwater in urban and agricultural areas are substantially greater than concentrations in forested and wetland areas and are greater than the estimated 1 mg/L as N that occurs as a natural background level. Among 495 wells that were sampled during 1993–2003, median nitrate concentrations increased from 3.2 to 3.4 mg/L as N, and the exceedance of the Federal and State health-based maximum contaminant level (MCL; 10 mg/L as N) increased from 16 to 21 percent.

Table 3. Statistical summary of nitrate concentrations in groundwater samples from the Middle Proterozoic bedrock, Kittatinny Supergroup, and Martinsburg Formation in the Highlands and Valley and Ridge Physiographic Provinces in NJ.

[mg/L, milligrams per liter; N, nitrogen; <, less than]

Geologic Formation	Nitrate concentration (mg/L as N)[1]		
	Minimum	Median	Maximum
Middle Proterozoic (45 samples)	<0.10	0.76	5.7
Kittatinny Supergroup (26 samples)	<0.05	0.39	5.6
Martinsburg Formation (26 samples)	<0.05	0.16	5.3

[1]From Serfes, 2004

Method of Study

The method for determining the median nitrate concentration in groundwater of the NJ Highlands Region and Areas and Zones within the Highlands Region required the following steps:

1. Obtain available nitrate-concentration data from groundwater samples from the NJ Highlands Region.
2. Develop a method of estimating the median value of a dependent variable from a series of logistic-regression relations and a set of explanatory (independent) variables. In this case, nitrate concentration is the dependent variable, and the independent variables are the land characteristics responsible for, or otherwise related to, nitrate concentrations.
3. Compile a comprehensive set of potential explanatory variables, which may be related to nitrate concentration, to be used in regression models.
4. Quantify the explanatory variables for areas surrounding wells for which nitrate data are available.
5. Identify an optimum set of explanatory variables for a series of logistic-regression equations based on threshold values that represent the range of measured nitrate concentrations in the NJ Highlands Region. Develop logistic-regression models.
6. Estimate median nitrate concentrations for the NJ Highlands Region and the Planning and Preservation Areas; Conservation, Existing Community, and Protection Zones; and each Area:Zone combination.
7. Evaluate logistic-regression model performance.
8. Compare median nitrate concentrations to concentrations obtained using quantile regression and multiple-linear regression.
9. Analyze the effects of using alternative methods that include nondetects in the model development and median nitrate calculations.

Nitrate-Concentration Data

Two independent sources of groundwater nitrate data were used for this study (table 4). The first dataset is a subset consisting of 782 wells in the Highlands Physiographic Province with data available from the USGS National Water Information System (NWIS) (http://waterdata.usgs.gov/nwis). The second dataset consists of 19,369 wells in the Highlands Physiographic Province with data available from the NJ Private Well Testing Act (PWTA; New Jersey Department of Environmental Protection, 2003).

National Water Information System Data

Wells in the NWIS database were installed for diverse purposes, including water supply, observation, and agriculture. The data are of exceptional quality because samples were analyzed by USGS laboratories and extensively reviewed. This dataset was used in a previous study to quantify median nitrate concentrations in the Highlands Region (New Jersey Highlands Council, 2008).

The 782 wells in the NWIS database sampled between 1983 and 2004 were evaluated to identify well clustering and to remove the wells with substantially overlapping 500-meter-radius buffers in order to avoid duplicate representation of areas (Barringer and others, 1990). A subset of 352 wells within the Highlands Region (fig. 3) was identified with minimum buffer overlapping, and nitrate data from this subset were used to identify the five explanatory variables used in all logistic models and for the initial estimates of median nitrate concentrations in Highlands groundwater (New Jersey Highlands Water Protection and Planning Council, 2008). Generally, NWIS wells are clustered in urban land-use areas. Water samples with the highest nitrate concentrations occurred mostly in areas dominated by urban and agricultural land use. Nitrate concentrations ranged from 0.03 to 24.2 mg/L as N with a median value of 0.98 mg/L as N, and 48 values (16 percent) were nondetects.

Circular well buffers are widely used in spatial groundwater-quality investigations. The 500-meter radius was selected on the basis of an evaluation by Koterba (1998), which stated that the best compromise for defining land-use characteristics around wells in a wide variety of hydrogeologic settings across the Nation would be a circular buffer with a 500-meter radius from the well. This was further supported by Johnson and Belitz (2009). In this investigation, it was assumed that the area within the 500-meter circular well buffer represents land use and surface characteristics, such as

Table 4. Sources of data on nitrate in groundwater in the New Jersey Highlands Region.

[NWIS, U.S. Geological Survey (USGS) National Water Information System; PWTA, New Jersey Private Well Testing Act; mg/L, milligrams per liter; --, no information; <, less than; NO_3, nitrate; QA, quality assurance]

Data source:	NWIS	PWTA	NWIS and PWTA
Number of groundwater samples	300	19,369	19,669
Sampling dates	12/02/1983–06/23/2004	10/14/2001–01/20/2011	12/02/1983–01/20/2011
Nitrate concentration (mg/L)			
Minimum (mg/L)	<0.03	<0.02	<0.02
Median (mg/L)	0.94	1.79	1.80
Maximum (mg/L)	24.20	153	153
Analytical methods, QA	Multiple USGS methods, USGS sampling and QA	As specified by each analytical method	--
Total number of grid cells	9,745	9,745	9,745
Number of grid cells with NO_3 data	284	4,379	4,516

septic-system inputs and fertilizer application, that may affect the local groundwater quality.

Private Well Testing Act Data

The New Jersey PWTA, which became effective in September 2002, requires water-quality sampling of domestic wells at the time of the sale of a home (Atherholt, 2009; NJDEP, 2003). PWTA water-quality data and the global positioning system location data are compiled by the NJDEP. The PWTA data are extensive, but water samples are collected only from domestic supply wells. PWTA data do not contain the information available for NWIS wells, such as well depth and aquifer identification. The PWTA specified a list of 12 approved analytical methods for analysis of nitrate (table 5). All samples were analyzed by NJ State certified laboratories, which are required to follow quality assurance/quality control protocols specified by the published analytical methods. Analysis of variance (ANOVA) and graphical analysis showed no spatial bias in either analytical methods used to analyze PWTA samples or in the nitrate method detection limit (MDL).

PWTA data for 19,369 wells sampled during October 2002–January 2011 within the NJ Highlands Region boundary were used in this study. PWTA rules mandate a level of anonymity associated with well locations. NJDEP's method to obscure the well location is to create a grid of square cells that are 610 m (2,000 ft) per side. Wells in each grid cell are plotted at the center of each grid. Therefore, the location of each well was generalized to within plus or minus 431 m (1,414 ft.) of the actual location. The number of groundwater samples collected in each grid cell is shown on a map of the Highlands Region in fig. 4. As with NWIS wells, PWTA wells are clustered in urban land-use areas.

The grid of 610-m-square cells (total of 9,745 cells) was generated using GIS software for the NJ Highlands Region boundary, and a unique identifier was assigned to each grid cell (Grid ID). Nitrate concentrations from all well samples within a grid cell were compiled, and the median measured nitrate concentration was calculated where grid cells contained more than one well. Of the 9,745 cells that make up the area of the NJ Highlands Region and the 19,369 wells that were sampled, the PWTA dataset provided a median nitrate concentration for each of 4,379 grid cells and no data for 5,366 grid cells. Nitrate concentrations for all PWTA samples ranged from 0.02 to 153 mg/L as N, and the median concentration was 1.79 mg/L.

Combined NWIS and PWTS Data

The NWIS and PWTA water-quality datasets were combined to optimize the use of all available data, which provided measured nitrate concentrations in 4,516 grid cells, and 5,228 grid cells with no nitrate data. The number of samples per cell with nitrate data ranged from 1 to 114 with a median of 3 samples and an average of 4.3 samples. As with the separate NWIS and PWTA data, most sampled wells tended to be situated in areas with urban or agricultural land. The median nitrate concentration in each cell ranges from 0.027

Table 5. Approved methods for nitrate analysis of New Jersey Private Well Testing Act samples.

[EPA, U.S. Environmental Protection Agency; ASTM, American Society for Testing and Materials; Cd, cadmium; --, no information; mg/L, milligrams per liter; N, nitrogen]

Methodology	Typical MDL[1] (mg/L as N)	EPA method	ASTM method	Standard methods	Other methods	Number of samples
Automated Cd reduction	0.05	353.2	D3867-90A	4500-NO3-F	--	1,009
Ion chromatography	0.01	300.0	D4327-91	4110B	B-1011 (Millipore)	9,032
Ion selective electrode	0.14	--	--	4500-NO3-D	601 (ATI Orion)	3,000
Manual Cd reduction	0.01	--	D3867-90B	4500-NO3-E	--	0
Unspecified method	--	--	--	--	--	4,864
Flow Injection/Cd reduction	0.01	--	--	--	10-107-04-1A (Lachat)	977

[1]Method detection limits (MDL) vary among labs and over time in individual labs

to 26.2 mg/L as N with an overall median concentration of 1.50 mg/L as N.

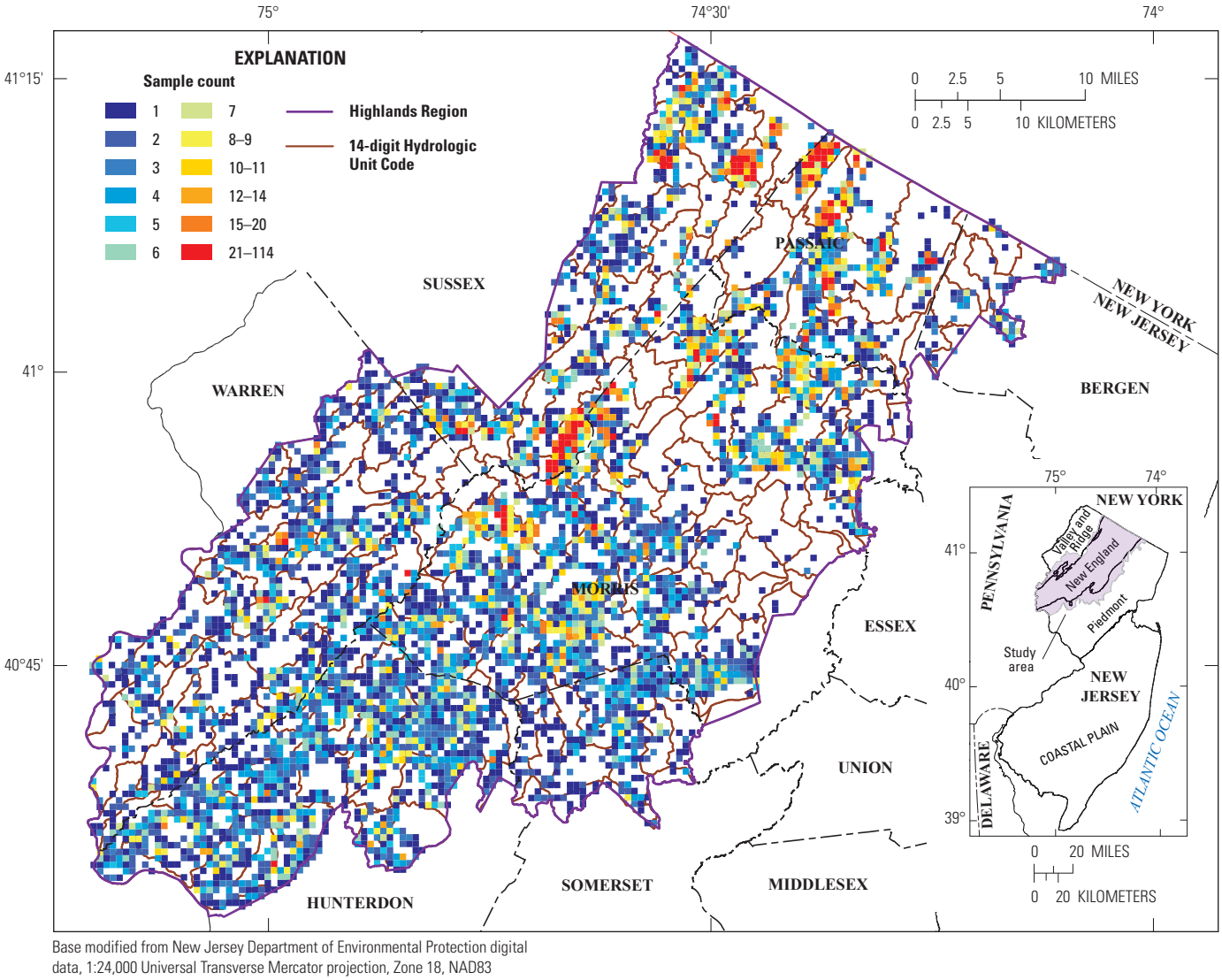
Of the 19,670 PWTA and NWIS samples, 511 (3 percent) had concentrations greater than the State and Federal Maximum Contaminant Level (MCL) for nitrate of 10 mg/L as N. A total of 4,519 (23 percent) samples had concentrations less than the MDL, which ranged from 0.020 to 10.0 mg/L as N, and are categorized as nondetects. The MDL varied among samples because of differences among laboratories and analytical methods used.

Nondetect Data

Among the 23 percent of samples that were nondetects, a disproportionally large percentage occurred in forested land and wetlands. The Preservation Area had a higher percentage of nondetects (25.8 percent) than the Planning Area (19.9 percent). Similarly, the Protection Zone had a higher percentage (27.0 percent) than either the Conservation or Existing Community Zones (18.3 and 19.6 percent, respectively). This result is expected, as groundwater that is not affected by anthropogenic surface activities tends to have lower concentrations of nitrate and therefore more nondetects.

Methods of analyzing data that contain nondetects (left-censored data) include assigning a value such as zero, the MDL, or some fraction (such as one-half) of the MDL. These substitution methods, though prevalent in published literature, have no sound basis because no substituted value between zero and the MDL can be argued to be more valid than any other (Helsel, 2005). An alternative method is to conduct the data analysis without assigning specific values to nondetects, such as the Maximum Likelihood Estimate (MLE), where a distribution (known or assumed) is assigned to the data, and statistics based on that distribution, not on individual data, are calculated. The MLE was not used here to estimated median nitrate concentrations in grid cells because the method does not perform well for sample sizes less than 30, and few grid cells contain 30 or more sampled wells. The Kaplan-Meier method is recommended for estimating summary statistics for censored data (Helsel, 2005). It is nonparametric and therefore does not require information or assumptions about data distribution; also it has no minimum sample-size limitations. Therefore, this method was used to calculate the median nitrate concentration for each grid cell. Although it is the best choice for this dataset and analysis, the Kaplan-Meier method is not without limitation. In grid cells that provided a single value, and that value is a nondetect, the default median value was the MDL. This circumstance applied to 385 nitrate concentrations and affected 8.5 percent of grid cells with a single nitrate concentration value. Only 152 (3.3 percent) of those MDL values were greater than or equal to (≥)0.5 mg/L as N, and only those could affect the median concentration in the entire Highlands Region, Planning and Preservation Land-Use-Capability Zone, or Area:Zone combination. Thus, for at least 96.7 percent of grid cells with nitrate data, either the Kaplan-Meier method was applied as designed or the method detection limit was substantially less than the estimated median nitrate concentration in the Highlands Region.

For comparison, median concentrations also were calculated with nondetects set to zero, one-half the MDL, or the MDL; these are discussed in the section “Four Methods of Including Nondetects.” This gives the full range of variability in nitrate concentrations resulting from the choice of methods for handling nondetects.



Base modified from New Jersey Department of Environmental Protection digital data, 1:24,000 Universal Transverse Mercator projection, Zone 18, NAD83

Figure 4. Numbers of groundwater samples collected in each model grid cell for the New Jersey Highlands Region. (Data from the U.S. Geological Survey National Water Information System database and the New Jersey Private Well Testing Act database)

Logistic Regression Model Development

The logistic model, as presented by Greene and others (2005), is of the form

$$p_i = P(Y = 1 \mid X_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ..., + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ..., + \beta_k x_{ik})} \quad (1)$$

or (equivalently, the logit form)

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ..., + \beta_k x_{ik} \quad , \quad (2)$$

where

p_i is the probability of the binary response variable Y_i (which can only have values of 0 or 1) being equal to 1;

β_0 is the intercept;

$\beta_1 \ldots \beta_k$ are regression coefficients for each explanatory variable of the regression equation;

$x_{i1}, x_{i2}, ..., x_{ik}$ are values of explanatory variables; and X_i refers to the set of values of all explanatory variables $(1, x_{i1}, x_{i2}, \ldots, x_{ik})$.

The quantity $\frac{p_i}{1 - p_i}$ is referred to as the "odds ratio" and is equivalent to $\frac{p(Y = 1)}{p(Y = 0)}$; it is used to express the probability that an event will occur. For example, if the

probability of nitrate concentration in a water sample exceeding 2.0 mg/L as N is 0.25, then the odds ratio is $\frac{p(Y = 1) = 0.25}{p(Y = 0) = 0.75}$, or 1 to 3. Model coefficients and other parameters were calculated with an iterative maximum-likelihood algorithm by S-Plus (Insightful Corp., 2003). Input scripts for developing multiple logistic-regression models are shown in Appendix 1.

In this investigation, the binary variable Y represents the two cases: Y = 0 where the nitrate concentration of a water sample is less than a threshold concentration C_{Ti}, or Y = 1 where the concentration is greater than or equal to C_{Ti}. Thus, increasing the value of C_{Ti} would increase the probability that Y = 0.

Estimation of Median Nitrate Concentrations

The value of p_i obtained from a logistic-regression model represents a quantile of the range of possible values of the dependent variable. Thus, $p_i = 0.5$ at the 0.5 quantile, which is the median value, of the dependent variable (nitrate concentration). The condition where $p_i = 0.5$ can be expressed as

$$\ln\left(\frac{p_i}{1-p_i}\right) = \ln\left(\frac{0.5}{1-0.5}\right) = \ln(1) = 0 = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ..., + \beta_k x_{ik}, \quad (3)$$

and the median value of the dependent variable is equal to the threshold value (C_{Ti}) of the logistic regression model shown in equation (3). This is the property of logistic regression that enables it to be used for estimating median values. The method applied here to estimate a median value is basically to identify two logistic models for which p_i is slightly greater than and slightly less than 0.5 and to assign the median value of the dependent variable by using an interpolation process described below.

The threshold value C_{Ti} for $p_i = 0.5$ is not directly calculable but can be selected from a series of logistic equations with a range of C_{Ti} values. In this investigation, logistic models with C_{Ti} = 0.05, 0.1, 0.15…1.0, 1.1, 1.2,…10.0 were developed (a total of 110 models). Ideally, a logistic model for which $p_i = 0.5$ would be identified, and the corresponding C_{Ti} value would be assigned as the median value. In practice, the set of C_{Ti} values is incremental and not continuous, and therefore, identifying the logistic model for which p_i = exactly 0.5 is unlikely. Linear interpolation was used to calculate between C_{T1} and C_{T2} of two logistic models M_1 and M_2 that have values of p_1 and p_2 which are nearest to 0.5, where p_1 is less than 0.5 and p_2 is greater than 0.5:

$$\text{Median } [N0_3 - N] = \frac{C_{T1}(p_2 - 0.5)}{(p_2 - p_1)} + \frac{C_{T2}(0.5 - p_1)}{(p_2 - p_1)}. \quad (4)$$

Thus, the median nitrate concentration in groundwater underlying an area, such as that within a circular well buffer or grid cells, is determined from a set of explanatory variable values, and the two logistic regression equations from which the probability of exceeding the nitrate concentration is 0.5 are determined by interpolation. The median concentration over a larger area, such as the entire Highlands Region or an area within the Highlands Region is then calculated as the median of the median concentrations of all the smaller areas (well buffers or grid cells).

Explanatory Variables

A total of 320 geographic and environmental characteristics are potential explanatory variables (Appendix 2) related to median nitrate concentrations in groundwater and were compiled in a previous investigation (New Jersey Highlands Council, 2008). The variables include land-use/land-cover characteristics as defined by the Anderson system (Anderson and others 1976). Land-use data for 1986 (NJDEP, 1986), 1995 (NJDEP, 2001), 2002 (NJDEP, 2008), and 2007 (NJDEP, 2010) were used in this investigation so that median nitrate concentrations were evaluated with land-use patterns during similar time periods. Land-use characteristics included percentages of each land use within well buffers and distances between the well and the nearest land-use type. Anderson Level 1 land-use categories in New Jersey are urban, agricultural, rangeland, forest, water, wetlands, and barren land. Subcategories (Anderson Level 2) include mixtures of land use such as, but not limited to, urban/residential, urban/industrial, agricultural/cropland, and agricultural/confined feeding operations. Level 2 categories that are in the Highlands Region are included in the list of 320 potential explanatory variables. Other characteristics listed in Appendix 2 include soil properties, transportation (length and number of roads and railroads), population, hydrology, water-quality characteristics (concentrations of chemical species), and well depth.

To relate independent variables to median nitrate concentrations at individual wells, the value of each variable within an area surrounding the well was determined. This was done for wells in the NWIS database by calculating the value of each of the 320 variables within the 500-meter-radius circular buffer of each well. This was not done for the PTWA wells because the exact location of each well is unknown. The explanatory variables that are identified as the best predictors of median nitrate concentrations in water samples from wells in the NWIS database were then used to determine median nitrate concentrations in the grid cells.

The best predictor variables from the list of 320 geographic and environmental characteristics (Appendix 2) were identified by applying a step-wise regression procedure to obtain a series of five-variable models that best fit the measured nitrate concentrations in groundwater samples from the set of wells in the NWIS database, as described in the Highlands Regional Master Plan (New Jersey Highlands Council, 2008). First, Spearman's Rho nonparametric correlation coefficients (Spearman, 1904) were calculated using the

value of each variable and nitrate concentration in each well. This procedure, previously used by Kolpin (1997), provided a nonparametric, univariate assessment of the regression relation between nitrate concentration and each potential explanatory variable. Next, univariate logistic regression relations were developed for each potential explanatory variable for C_{Ti} values of 0.1, 0.3, 1.0, 3.0, 5.0, and 10 mg/L as N. Potential variables that had significant Spearman's rho (>0.105 or <-0.105 for p = 0.05) or significant t values in one or more univariate logistic models (>1.96 for p = 0.05) were used in two-variable logistic models. The process was continued for 3-, 4-, and 5-variable models until all possible combinations of variables that were significant in the 4-variable models were used to generate a set of 5-variable models. Selection of the final set of five variables included numerical and subject criteria. Selection was based on the sum of all six t statistic values for the model (including the intercept), minimum of expected colinearity among variables, and maximum spatial representation of the variables. The sum of t statistic values provides an objective, numerical assessment of the overall significance of the logistic model. This metric varies, depending upon the value of the threshold nitrate concentration C_{Ti} associated with the regression equation. C_{Ti} is shown in relation to t value in fig. 5, where t for each of five explanatory variables varies as a function of C_{Ti}. Collinearity occurs between related variables, such as population and percent urban land use, or distance to nearest agriculture and percent agricultural land use. Models with two or more variables that are expected to be collinear were not considered in the selection of the final five variables. Where a choice between two similar or related variables had to be made, the variable with the greatest spatial representation was selected. For example, in selecting between percent urban residential and percent total urban land use, the total urban variable was selected. The final set of explanatory variables consists of urban land use, agricultural land use, septic-system density, total length of streams, and the number of known contaminated sites in the well buffer. The sum of t statistic values for these five variables was large for all six C_{Ti} values. Collinearity was expected to be minimal, and all five variables have greater than or equal spatial representation compared to related variables. This set of variables was used for all future logistic, quantile, and multiple-linear regression model development.

Spreadsheet Design for Estimating Median Nitrate Concentrations

A Microsoft Excel spreadsheet was used to calculate the median nitrate concentration for each area of interest (well buffer, grid cell, or Area:Zone combination). An example of the spreadsheet is shown as Appendix 1. The spreadsheet is arranged in five sections.

1. A list of 110 logistic regression models with model parameters (regression coefficients and intercept), t-statistic values, and standard errors, with documentation of data files used to develop the model.
2. Fields that contain information about the wells and buffer or grid cell. This includes location within Highlands areas and zones, grid or well identification number, lab-measured nitrate concentration (if applicable), and values of the five explanatory variables.
3. Fields in which the probability of exceeding the threshold nitrate concentration is calculated for each well or grid cell for each of the 110 logistic-regression relations.
4. Fields in which the results of (3) are used to estimate the median nitrate concentration for each well or grid cell on the basis of equations 3 and 4.
5. A field in which the overall estimated median nitrate concentration for the area of interest (entire Highlands Region or smaller area within the Highlands) is shown.

Methods Used to Evaluate Logistic Regression Models

Four methods were used to assess the statistical significance of explanatory variables and evaluate the logistic-regression models: the t statistic of each logistic-regression coefficient, which is equivalent to the Wald statistic as calculated by Hosmer and Lemeshow (2000); Press's Q; a test of the correlation between estimated and measured median nitrate; and a test of the overall accuracy of the method to estimate the median values of measured nitrate concentrations.

The t statistic is calculated as the ratio of the maximum likelihood estimate of the slope parameter to an estimate of its standard error:

$$t = W = \beta_i / (\text{standard error of } \beta_i) \quad , \qquad (5)$$

where

β_i represents the coefficient of explanatory variable i in the logistic-regression model.

The value of t for each variable indicates the significance of that variable. Variables with values of t for large samples (where the t distribution is indistinguishable from the standard normal distribution) of greater than 1.96 are significant at the 0.05 level and contribute significantly to the model.

Press's Q statistic is a function of the model's ability to correctly categorize data that were used to develop the model. The two categories are p>0.5 and p<0.5. A value, for example a lab-measured nitrate concentration, is correctly categorized where p>0.5 and the value is greater than the threshold value

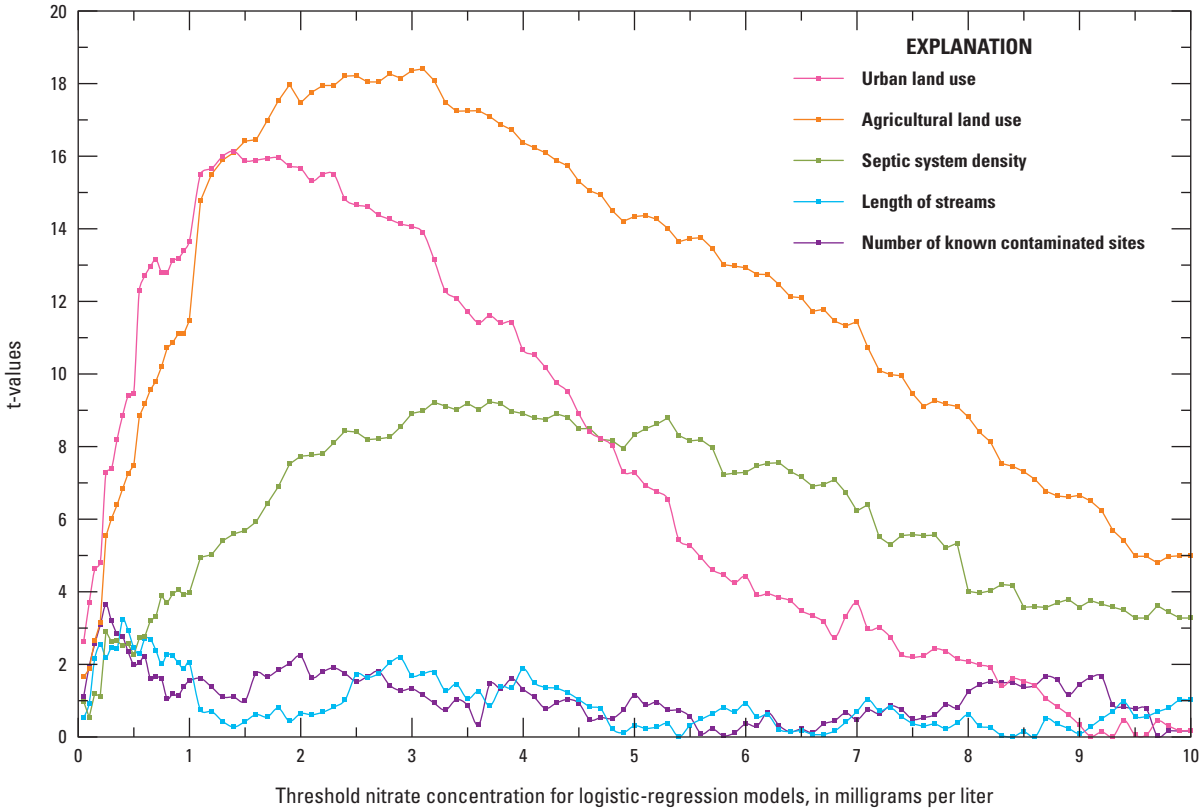


Figure 5. Values of the t statistic for five explanatory variables in logistic-regression equations for nitrate-threshold concentrations 0.05–10.0 milligrams per liter as nitrogen.

or p<0.5 and the value is less than the threshold value. The percent of correct classifications is used to calculate Press's Q:

$$Q = [N - (n*K)] \text{ ^2/N*}(K - 1) \quad , \qquad (6)$$

where

N is sample size,
n is number of correct classifications, and
K is number of groups (2).

The correlation between deciles of estimated and measured nitrate concentrations also was applied on the basis of methods of Greene and others (2005) and Nolan (2001), where the relation between the calculated probability of exceeding a threshold value and the fraction of measured data that exceeded the threshold value for each quantile of calculated values was determined. Instead of probabilities, the estimated median nitrate concentrations for each quantile were compared to the median measured concentration for the same quantile. This provided an overall assessment of the method's ability to accurately estimate median values and information about the relative magnitude of error over a range of nitrate concentrations.

Two additional model diagnostics were used to evaluate and validate the logistic, quantile, and multiple-linear regression methods of calculating median nitrate concentrations. In the first, the median of measured concentrations was compared to the median concentration determined with the regression methods for the same set of grid cells (all of those with measured nitrate data) as a calibration check. In the second, model validation was conducted by developing regression models using data from half of the grid cells, calculating the median nitrate concentration of the other half with the regression methods, and comparing those estimates to the median of measured nitrate concentration values.

Median Nitrate Concentrations in Groundwater

Median values of lab-measured and estimated nitrate concentrations are discussed in this section (table 6). Medians of lab-measured nitrate concentrations were determined in two ways: as the median of all measured nitrate-concentration values in the combined NWIS-PWTA dataset, and as the median concentration at the grid-cell level. Median nitrate concentrations were determined for the entire Highlands Region, for the

Table 6. Measured and estimated nitrate concentrations in groundwater from the New Jersey Highlands Region.

[mg/L, milligrams per liter; N, nitrogen]

Area within the NJ Highlands	Median nitrate concentration (mg/L as N)		
	Median of measured concentrations		Estimated median concentrations for all grid cells
	Individual water samples	Grid-cell level[1]	
Entire Highlands Region	1.79	1.50	1.25
Planning Area	2.16	1.78	1.55
Preservation Area	1.42	1.25	1.08
Conservation Zone	2.17	2.02	1.76
Existing Community Zone	2.51	2.14	1.78
Protection Zone	1.28	1.10	1.07
Planning Area:Conservation Zone	2.40	2.15	1.78
Planning Area:Existing Community Zone	2.71	2.17	1.78
Planning Area:Protection Zone	1.50	1.27	1.19
Preservation Area:Conservation Zone	1.85	1.93	1.64
Preservation Area:Existing Community Zone	2.26	2.07	1.79
Preservation Area:Protection Zone	1.18	1.02	1.05

[1]For the models, the New Jersey Highlands Region is divided into a grid of 9,745 610-meter-square cells.

Planning and Preservation Areas, for each of the three Land-Use-Capability Zones, and for each Area:Zone combination.

Median of Measured Nitrate Concentrations in the NJ Highlands Region

The median measured nitrate concentration among all water samples in the combined NWIS-PWTA dataset was 1.79 mg/L as N, and concentrations in the 2 Areas, 3 Land-Use Capability Zones, and 6 Area:Zone combinations range from 1.18 to 2.71 mg/L as N (table 6). Concentrations were higher where agricultural or urban land use is more prevalent, such as the Conservation and Existing Community Zones, and lower where land use is predominantly forested land, such as the Protection Zone. There is spatial bias in well locations because many sampled wells are located in urban areas; thus, a bias in median nitrate concentrations was expected. Over-representation of urban and possibly agricultural areas and under-representation of forested areas in the combined NWIS-PWTA database must, therefore, result in higher median nitrate concentrations for all water samples than the actual median concentration for groundwater underlying the entire Highlands Region or any Area, Zone, or Area:Zone combination.

The median nitrate concentrations for the Highlands at the grid-cell level was 1.50, and concentrations in the 2 Areas, 3 Land-Use Capability Zones, and 6 Area:Zone combinations range from 1.02 to 2.17 mg/L as N (table 6). Spatial bias in well locations was reduced by calculating a single nitrate concentration for each grid cell, then calculating the median concentration at the grid-cell level. Each grid cell that contained wells in the combined NWIS-PWTA database received equal

weight in all calculations. The remaining spatial bias is caused by the lack of nitrate data for about one-half the grid cells; those grid cells tended to have a larger percentage of forested land use (table 1). Therefore, although median concentration at the grid-cell level are subject to less spatial bias than those calculated from individual nitrate concentrations, some spatial bias remains and leads to over-estimation of median nitrate concentrations.

Median of Estimated Nitrate Concentrations

The estimated median nitrate concentration for all grid cells in the entire Highlands Region estimated using the logistic-regression method is 1.25 mg/L as N, and estimated median concentrations range from 1.05 to 1.79 mg/L as N among the Area, Zone, and Area:Zone combinations (table 6). Spatial distribution of estimated nitrate concentrations is shown on a map of the Highlands Region (fig. 6). A comparison of figs. 3 and 6 shows that forested areas correspond to lower median nitrate concentrations, and urban areas correspond to higher median concentrations. Estimated median nitrate concentrations are lower for the entire Highlands Region than the median of measured nitrate concentrations at the individual-sample and grid-cell scales (table 6). This is consistent with lower nitrate concentrations occurring in groundwater underlying grid cells dominated by forested and wetland areas, which are under-represented in the database of nitrate concentrations. Median values of the five explanatory variables used in the logistic models are shown in table 7. Urban and agricultural land uses and septic-system density are greater in the grid cells with sampled wells than in those without. The non-random distribution of wells is apparent when considering that the average percentage of non-urban, non-agricultural land in areas with sampled wells is nearly 15 percent greater than in areas without sampled wells, and there are on average 41 percent more septic systems per unit area in areas with sampled wells than in areas without sampled wells. The average number of known contaminated sites is 21 percent greater in grid cells with sampled wells. It is clear, therefore, that a median nitrate concentration calculated directly from water-sample data would have over-estimated the nitrate concentration of the underlying groundwater for the entire Highlands Region or any Area or Zone.

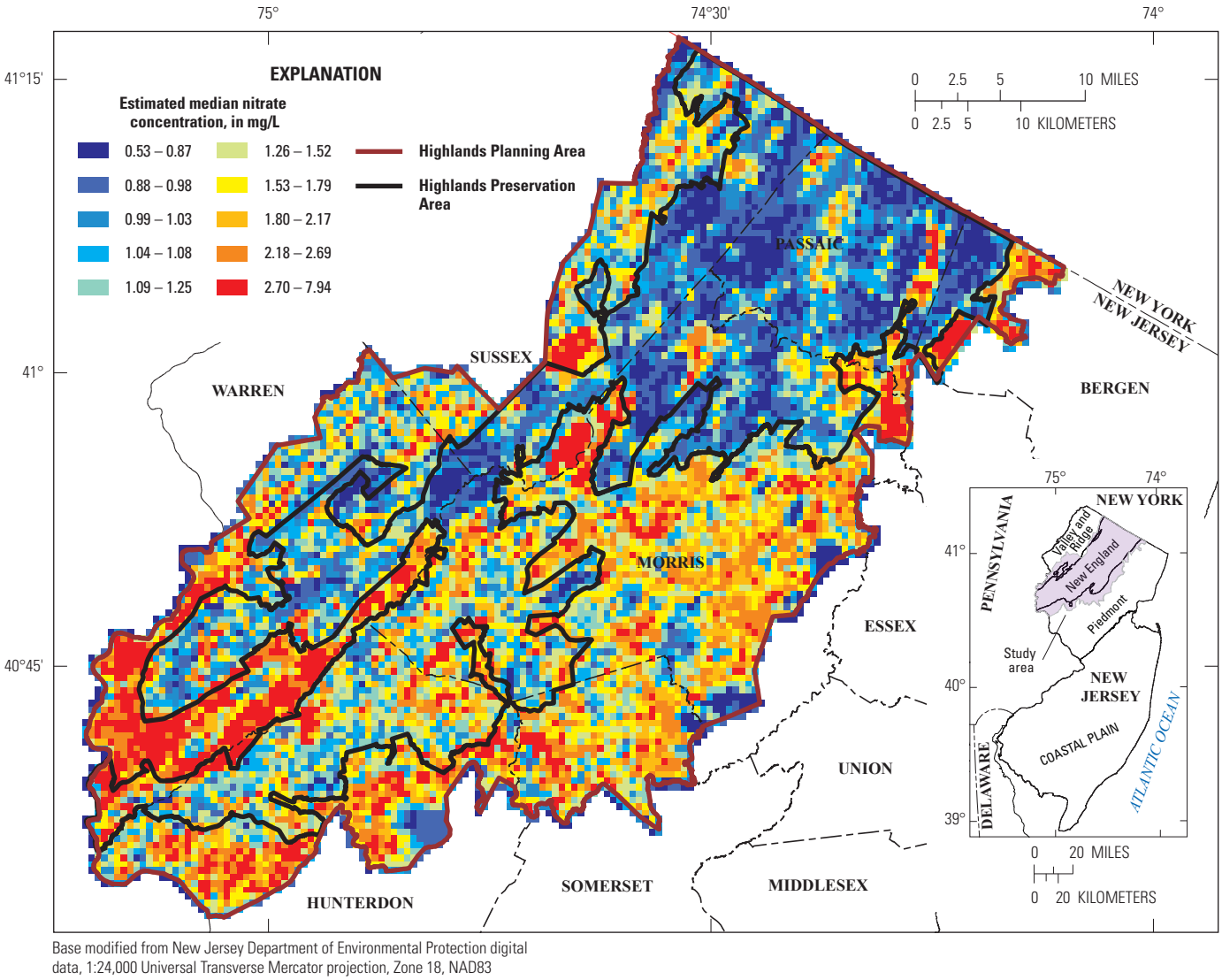
Fit and Validation of Logistic-Regression Models

Four tests demonstrated that the logistic-regression models generally contained significant explanatory variables, had significant predictive power, and can reliably estimate nitrate concentrations for grid cells in which no wells were sampled. Values of the t statistic for the five explanatory variables are shown in fig. 5. Urban and agricultural land use and septic-system density are the most significant variables over most of the range of threshold nitrate concentrations. The 5 variables decline in significance at the low and high extremes of the range of threshold values, though at least 3 variables are significant (p=0.05) for the nitrate threshold range of 0.1–8.3 mg/L as N. All five variables are significant for a nitrate-concentration threshold range of 0.25–0.60 mg/L as N, which includes a large portion of the measured nitrate concentrations. A case could be made for discarding the two weakest variables, known contaminated sites and total length of streams. However, these variables were significant at low concentration thresholds where land-use and septic-system-density variables were depressed. Using or not using the less-than-significant variables in the models has little effect on the calculated probability values, and therefore they were retained for the value they add to the models at the low range of the threshold values. Also, the same five explanatory variables were used in all logistic models so that probabilities among threshold values would be comparable.

Press's Q and the percent of correctly classified samples are shown in fig. 7. More than 60 percent of nitrate concentrations at the grid-cell level were correctly classified as greater than or less than the threshold value for the 110 logistic-regression models. Results of this statistical analysis are most meaningful where the median nitrate concentration is near the threshold value and random selection would result in 50 percent of values being incorrect. At the high and low extremes of values, a large fraction of values would be categorized incorrectly only if the model had no predictive power. This is not the case here (fig. 7) because greater than 90 percent of values were categorized correctly. Similarly, values of Press's Q (equation 6, fig. 7) are significantly greater than the critical value for logistic-regression models where the median nitrate concentration is near the threshold value. The "dip" in the curve (fig. 7) occurs because incorrectly expecting a calculated value to be above or below the threshold value is more likely to occur near the threshold value. The large values of Q reflect the large sample size for each model, which enables all models to accurately categorize samples as greater than or less than the critical value on the basis of the values of the explanatory variables.

Comparisons of Median Measured Nitrate Concentrations and Estimated Median Nitrate Concentrations

Two simulation scenarios were developed to assess the accuracy of the logistic regression method for estimating median nitrate concentrations of the entire Highlands Region and Areas and Zones within the Highlands Region (table 8). In scenario 1, the medians of lab-measured concentrations were compared to the estimated nitrate concentrations for 4,516 grid cells. The set of 110 logistic-regression equations were prepared from values of the five explanatory variables and the median of measured nitrate values for those 4,516 cells. The estimated (1.49 mg/L as N) and measured (1.50 mg/L as N) median concentrations were nearly identical.



Base modified from New Jersey Department of Environmental Protection digital data, 1:24,000 Universal Transverse Mercator projection, Zone 18, NAD83

Figure 6. Estimated median nitrate concentration in model grids for the New Jersey Highlands Region. (mg/L, milligrams per liter)

This complements the t (equivalent to the Wald) and Press's Q statistics, which indicate that the selection of explanatory variables was appropriate and that the results of the logistic-regression method are generally significant.

The median estimation method based on logistic regression was further tested by sorting the grid cells in Scenario 1 by estimated concentration (high to low) and dividing the population by 10 quantiles (deciles). The median of estimated nitrate values in each quantile (decile) was then compared to that of the lab-measured values (fig. 8). Median estimated concentrations in the quartiles ranged from 1.01 to 3.24 mg/L as N. The median of estimated nitrate concentrations for the tenth quantile (decile) (3.24 mg/L as N) was the most dissimilar from the median of measured values (3.58 mg/L as N) and was lower by 9.5 percent. The average deviation between quantile (decile) median nitrate concentrations on the basis of estimated versus lab-measured was 4.4 percent. This shows that medians of lab-measured nitrate concentrations were predicted with reasonable accuracy over the range of threshold concentrations.

Model validation was accomplished by developing the set of 110 logistic-regression models using only data associated with half of the grid cells and using those models to estimate median nitrate concentrations for the remaining half of the grid cells, then comparing those median values to the corresponding median concentrations (Scenario 2). As shown in figure 9, the relation between the medians of estimated and lab-measured nitrate concentrations is similar to that for Scenario 1, although the deviation in the 10th decile was slightly higher (10.3 percent) as was the average deviation (5.4 percent).

Figures 8 and 9 show the error trend in the estimated values for Scenarios 1 and 2. Scenario 1 tested the capability of a set of logistic-regression equations to predict the nitrate

Table 7. Summary statistics for explanatory variables used in logistic-regression models to calculate median nitrate concentration in groundwater from in the NJ Highlands Region.

	Grid cells[1] with sampled wells[2]	Grid cells with no sampled wells
Percent urban land use[3]		
Minimum	0	0
Mean	32.9	20.2
Median	27.9	6.6
Maximum	100	100
Percent agricultural land use		
Minimum	0	0
Mean	13.2	11.5
Median	2.3	0
Maximum	97.1	100
Total length of streams		
Minimum	0	0
Mean	1,613	1,628
Median	1,263	1,172
Maximum	13,707	12,942
Septic-system density		
Minimum	0	0
Mean	41.6	29.5
Median	24.2	18.3
Maximum	843	669
Number of known contaminated sites		
Minimum	0	0
Mean	0.17	0.14
Median	0	0
Maximum	7	9

[1]The Highlands Region is divided into a grid of 9,745 610-meter-square cells.

[2]Wells with results inventoried in U.S. Geological Survey National Water Information System or sampled as a requirement of the New Jersey Private Well Testing Act.

[3]Calculated from NJ Department of Environmental Protection digital data (NJ Department of Environmental Protection, 2010).

concentrations at various quantiles, including quartiles for the data used in developing the logistic equations. Scenario 2 was a more realistic test in which median-nitrate-concentration and land-use data from half the grid cells were used to predict the median concentrations in the other half. The analysis of error in the estimated nitrate concentrations (figs. 10 and 11) shows that most estimates were accurate within 10 percent and that the interquartile range of concentrations was estimated accurately. The first and third quantile errors of the estimate for Scenario 2 were 8.3 and 5.5 percent, respectively. The largest errors occurred for the 5th and 10th percentiles (nearly 30% in Scenario 2), indicating that there is greater error, in terms of percent, in the lower concentrations than in the higher concentrations.

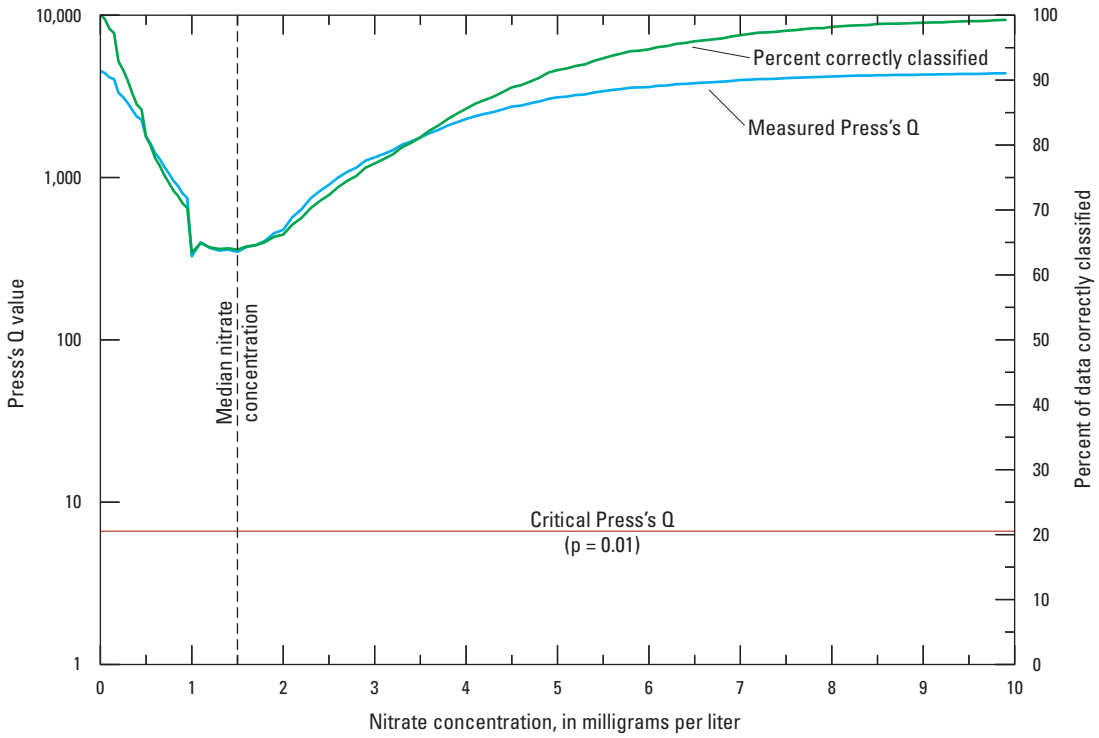


Figure 7. Values of the Press's Q statistic for logistic-regression models with nitrate-threshold concentrations of 0.05–10.0 milligrams per liter of nitrate as nitrogen.

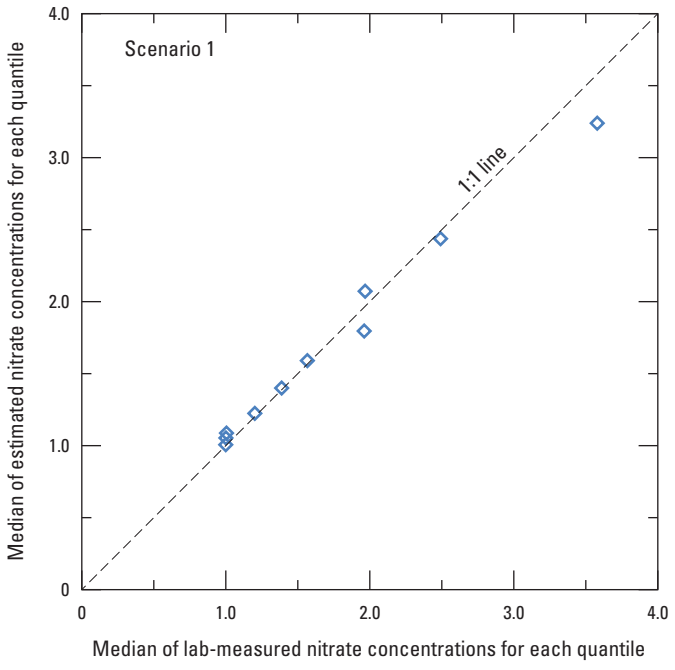


Figure 8. Median measured nitrate concentration in relation to estimated median nitrate concentration for each of 10 quantiles in Scenario 1, in milligrams per liter as nitrogen.

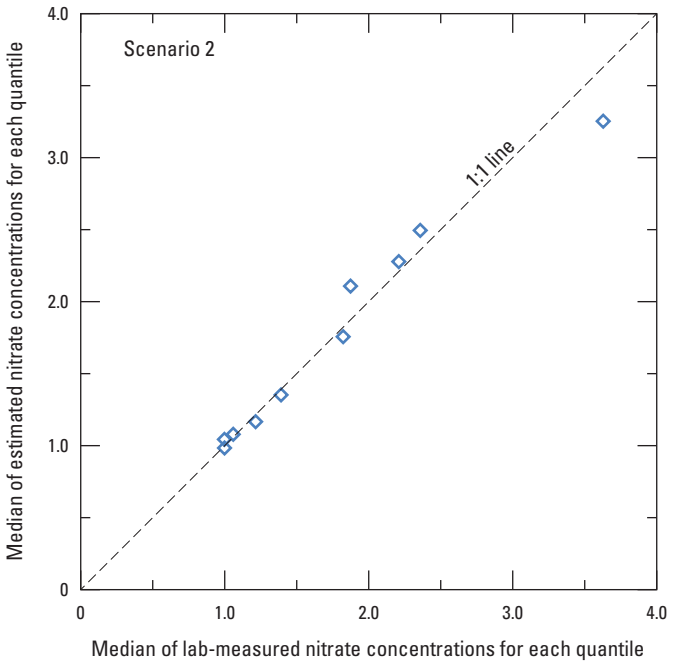


Figure 9. Median measured nitrate concentration in relation to estimated median nitrate concentration for each of 10 quantiles in Scenario 2, in milligrams per liter as nitrogen.

Table 8. Simulation scenarios for logistic-regression model validation: comparisons between lab-measured and estimated median nitrate concentrations.

[mg/L, milligrams per liter; N, nitrogen]

Validation scenario number and description	Number of grid cells[1]	Median of lab-measured nitrate concentrations (mg/L as N)	Median of estimated nitrate concentrations (mg/L as N)	Percent difference
1. Comparison between medians of lab-measured and estimated nitrate concentrations for the same set of grid cells	4,516	1.50	1.49	0.7
2. Comparison between medians of lab-measured and estimated nitrate concentrations for the same set of grid cells sorted by Highlands Administrative Area and Land-Use Capability Zone[2]				
a. Planning Area	2,300	1.78	1.75	1.7
Conservation Zone	732	2.14	2.04	4.7
Existing Community Zone	759	2.17	2.10	3.0
Protection Zone	809	1.28	1.28	0.0
b. Preservation Area	2,152	1.25	1.25	0.0
Conservation Zone	336	1.90	1.88	1.1
Existing Community Zone	275	2.07	1.95	5.8
Protection Zone	1,541	1.05	1.09	3.8
3. Comparison between median of lab-measured values in 2,258 randomly selected grid cells and estimated values for the remaining cells that have sampled wells	2,258	1.50	1.45	3.3

[1]The Highlands Region is divided into a grid of 9,745 610-meter-square grid cells. Median groundwater-nitrate concentration, land use, and other variables used in logistic regression models are calculated for each grid cell.

[2]The NJ Highlands Region is divided into the Planning Area, administered by the New Jersey Highlands Council, in which conformance with the Regional Master Plan is voluntary; and the Preservation Area, administered by the New Jersey Department of Environmental Protection, in which conformance with the Regional Master Plan administered by the Highlands Council is mandatory.

Comparison among Estimated Median Nitrate Concentrations Obtained with Logistic, Quantile, and Multiple-Linear Regression Methods

Estimated median nitrate concentrations for the Highlands Region, Planning and Preservation Areas, Land-Use-Capability Zones, and Area:Zone combination determined by logistic, quantile, and multiple-linear regressions are shown in table 9. Although validation results showed that the logistic-regression method was able to estimate median nitrate concentrations slightly more accurately than quantile regression and substantially more accurately than MLR, median concentrations determined using the three methods were similar. The average difference between medians determined with logistic and quantile regressions was less than 0.1 mg/L as N, and the average difference between medians from logistic regression and MLR was 0.15 mg/L as N. As these are all regression methods based on minimizing residual error and all were developed using the same five explanatory variables and nitrate data, similarity among the resulting estimated median nitrate concentrations is not surprising. All three methods effectively remove the spatial bias caused by systematically larger percentages of urban land use and higher septic-system density in grid cells that contain NWIS and PWTA wells. The decision about which regression method to select rests on whether a higher priority is placed on the use of a well-established, proven, accepted method (quantile regression or MLR) that is slightly less accurate according to validation results or the unconventional use of logistic regression with slightly more accurate estimates.

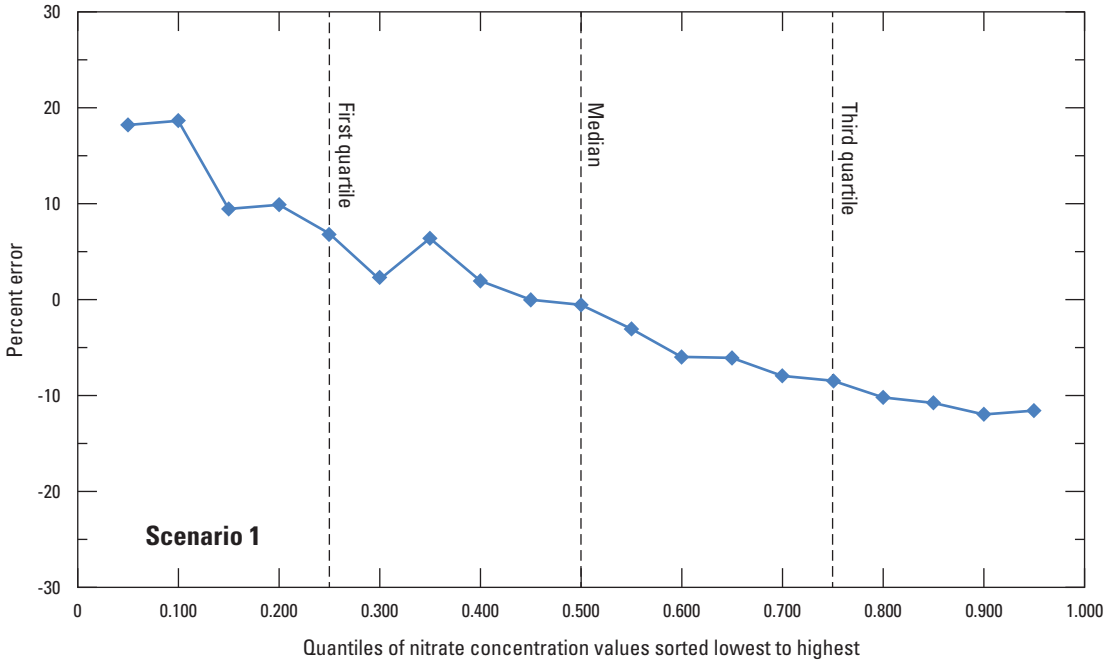


Figure 10. Percent error in estimates of quantiles of nitrate concentration in Scenario 1.

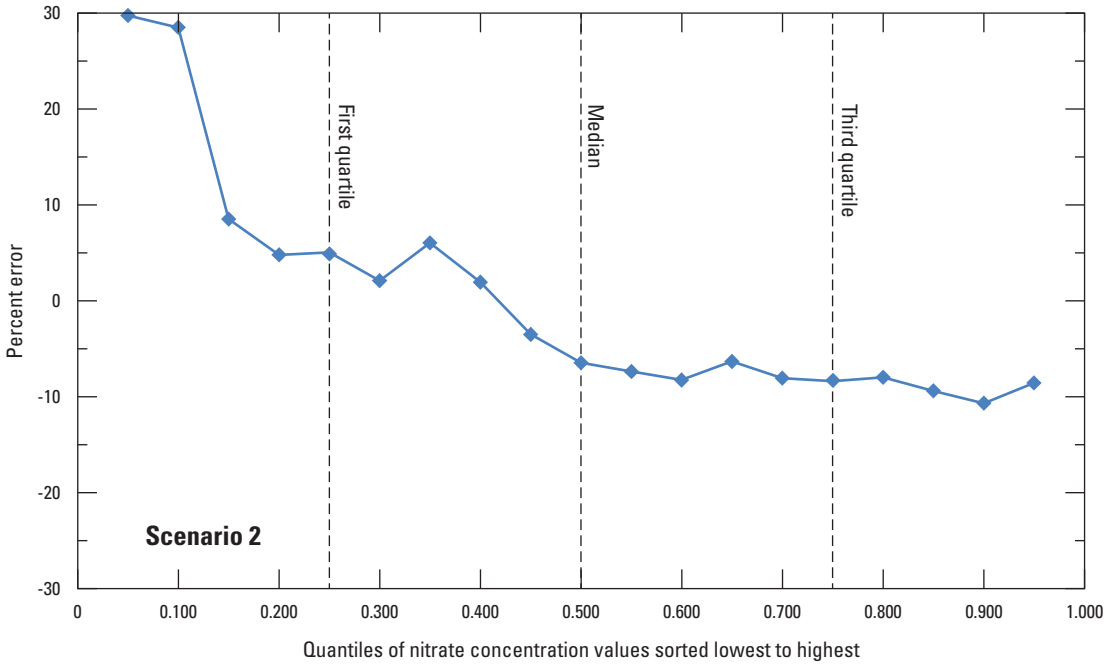


Figure 11. Percent error in estimates of quantiles of nitrate concentration in Scenario 2.

Table 9. Estimated median nitrate concentrations based on logistic regression, quantile regression, and multiple-linear regression models of the NJ Highlands Region.

[mg/L, milligrams per liter; N, Nitrogen; NJ, New Jersey]

Area within the NJ Highlands	Estimated median nitrate concentrations (mg/L as N) assigned to nondetect samples		
	Method of calculating the median value		
	Logistic regression	Quantile regression	Multiple-linear regression
Entire Highlands Region	1.25	1.37	1.24
Planning Area	1.55	1.67	1.38
Preservation Area	1.08	1.08	1.12
Conservation Zone	1.76	1.94	1.52
Existing Community Zone	1.78	1.83	1.48
Protection	1.07	1.05	1.09
Planning Area:Conservation Zone	1.78	1.97	1.52
Planning Area:Existing Community Zone	1.78	1.83	1.47
Planning Area:Protection Zone	1.19	1.28	1.17
Preservation Area:Conservation Zone	1.64	1.87	1.49
Preservation Area:Existing Community Zone	1.79	1.84	1.50
Preservation Area:Protection Zone	1.05	0.96	1.05

Four Methods of Including Nondetects

The four methods that include nondetects in the median nitrate concentration calculations are (1) simple substitution of zero, (2) substitution of one-half the detection limit, (3) substitution of the detection limit, and (4) estimation based on the Kaplan-Meier method. Although substitution is discouraged in the statistical research literature (Helsel, 2005), there is historical precedent and some conditions under which this approach is recommended (U.S. Environmental Protection Agency, 2009). Substituting zero may be appropriate in cases where nondetects represent an absence of the contaminant being measured. Substituting half the detection limit may be reasonable if it is assumed that the population of nondetects is uniformly distributed along the interval of zero and the detection limit. Substituting the detection limit is the most conservative approach, as it is known that (within the precision of the data-collection methods) the concentration does not exceed that value.

The variability in estimated median nitrate concentrations resulting from the four methods is shown in fig. 12 and table 10. The variability in the Planning Area for each Land-Use-Capability Zone is small. Variability in the Preservation Area, where nitrate concentrations are generally lower, is greater. This is because a greater portion of nitrate concentrations in grid cells are shifted from above the median to below the median when a smaller value is assigned to each nondetect, thus shifting the recalculated median lower to a greater extent. Thus, the median nitrate concentration in the Preservation Area is decreased by 0.11 mg/L as N when the assignment of nondetects is changed from the MDL to zero. This effect is increased when the areas are parsed into the three zones (fig. 12). The choice of nondetect assignment is unimportant for the Conservation Zone within the Planning Area but is significant for the Protection Zone in the Preservation Area where the estimated median nitrate concentration decreases by 0.26 mg/L as N when the nondetect assignment is similarly changed. Other sources of error, which include uncertainty about well location within grid cells and direction and flow rate of groundwater, changing land-use percentages over time, analytical precision, sampling procedures, and the selected set of explanatory variables can add substantially to the error in estimates of median nitrogen concentration.

In summary, the handling of nondetect values is important in estimating median nitrate concentrations where a large fraction of values are nondetects or where the MDL is greater than the median value but has little effect where only a small fraction of values are nondetects or where the MDL is substantially less than the median value. For estimating the median nitrate concentration of groundwater in Highlands Region,

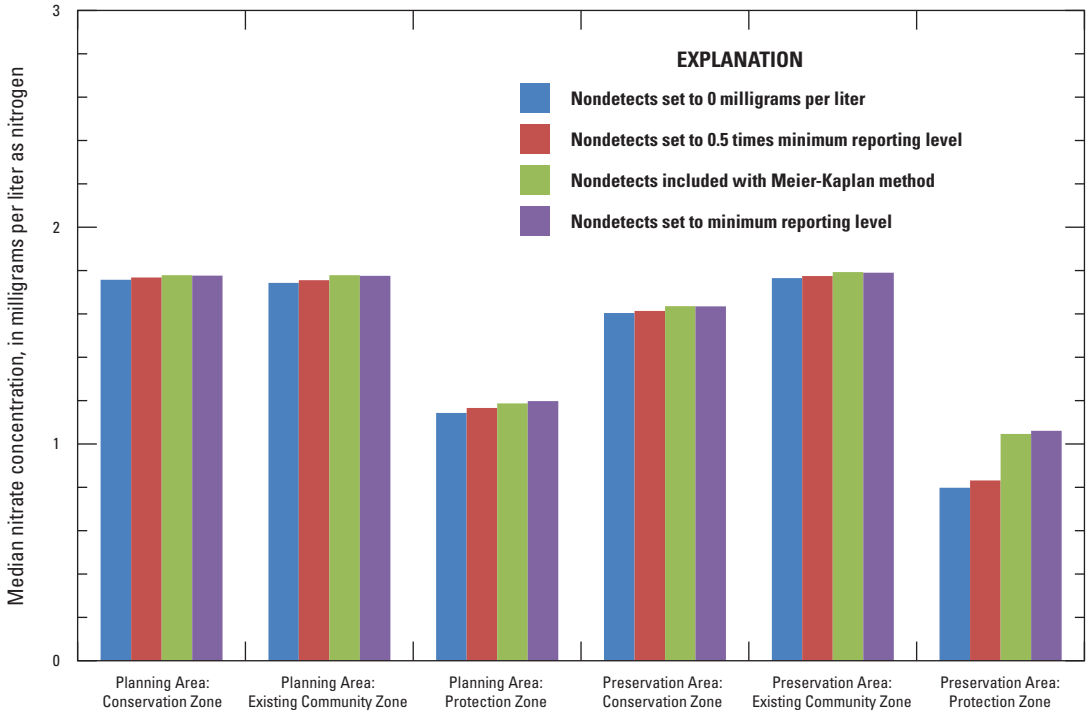


Figure 12. Median estimated groundwater-nitrate concentrations aggregated by grid cell for areas and Land-Use Capability Zones calculated with four methods of including nondetects.

Table 10. Estimated median nitrate concentrations based on logistic-regression models of the NJ Highlands Region calculated with four methods of assigning values to nondetects.

[mg/L, milligrams per liter; MDL, minimum detection limit; N, nitrogen; NJ, New Jersey]

Area within the NJ Highlands	Estimated median nitrate concentrations (mg/L as N)			
	Value assigned to nondetect			
	Zero[1]	0.5 x MDL[2]	Kaplan-Meier[3]	MDL[4]
Entire Highlands	1.21	1.23	1.25	1.25
Planning Area	1.52	1.53	1.55	1.55
Preservation Area	0.95	0.98	1.08	1.09
Conservation Zone	1.74	1.75	1.76	1.76
Existing Community Zone	1.75	1.76	1.78	1.78
Protection	0.89	0.93	1.07	1.08
Planning Area:Conservation Zone	1.76	1.77	1.78	1.78
Planning Area:Existing Community Zone	1.74	1.76	1.78	1.78
Planning Area:Protection Zone	1.14	1.17	1.19	1.20
Preservation Area:Conservation Zone	1.60	1.61	1.64	1.63
Preservation Area:Existing Community Zone	1.77	1.77	1.79	1.79
Preservation Area:Protection Zone	0.80	0.83	1.05	1.06

[1] Values of all nondetect samples set to zero.

[2] Values of all nondetect samples set to ½ the MDL.

[3] Kaplan-Meier method (Helsel, 2005) used to assign values to nondetect samples.

[4] Values of all nondetect samples set to the MDL.

the choice of method to include nondetects in the calculation makes a significant difference only for the Protection Zone within the Preservation Area. Assigning the MDL as the value for all nondetects would produce a "worst-case scenario" median concentration and likely would overstate the median nitrate concentration. Selecting either 0.5 x MDL or applying the Kaplan-Meier method would most likely increase the accuracy of the median estimate, but there is no justification for using 0.5 x MDL and no justification for assigning zero as the concentrations of nondetects. Therefore, Kaplan-Meier method remains the most reasonable choice for handling nondetects.

Summary and Conclusions

Nitrate-concentration data were used in conjunction with variables related to land use and land-surface characteristics to estimate median nitrate concentrations in groundwater underlying the New Jersey (NJ) Highlands Region in a study conducted by the U.S. Geological Survey (USGS) in cooperation with the New Jersey Department of Environmental Protection. Sources of nitrate data were the USGS National Water Information System (NWIS) and the New Jersey Private Well Testing Act (PWTA). Spearman's nonparametric correlation coefficient and a step-wise logistic-regression procedure were used to identify five independent (explanatory) variables that produce highly correlated logistic models that are based on a range of nitrate-concentration thresholds—0.1, .03, 1.0, 3.0, 5.0, and 10 milligrams per liter as nitrogen (mg/L as N). The explanatory variables that were found to be significant in logistic-regression models that used NWIS data are percent urban land use, agricultural land use, length of streams, septic-system density, and number of known contaminated sites. Each explanatory variable was quantified within 610-meter x 610-meter grid cells. A series of 110 logistic models with thresholds ranging from 0.05 to 10 mg/L as N were developed, and the probability of a nitrate concentration exceeding a designated threshold concentration in groundwater underlying a grid cell was calculated. For each grid cell, the median concentration was determined by identifying the two logistic models for which the probability of exceedance was nearest to 50 percent, and the corresponding thresholds were the two nitrate concentrations nearest to the median by definition. Linear interpolation was used to calculate the actual median nitrate concentration for each grid cell. A series of evaluation methods was applied to the logistic models. Twenty-three percent of the nitrate data were left-censored (included nondetect values), and the Kaplan-Meier method of including nondetects in the median calculation was applied to estimate the median nitrate concentration in each grid cell. Three additional methods of assigning values to nondetects were explored. Little difference in median nitrate concentration was noted for the Highlands Region and most areas within the Highlands Region regardless of which method was used to handle nondetects. Median nitrate concentrations within the Highlands Region correlated positively with percentages of urban land use, agricultural land use, and septic-system density. Model validation showed that the logistic-regression approach was able to accurately calculate median concentrations with a maximum error of less than 0.1 mg/L as N. Median estimated nitrate concentrations based on quantile regression were slightly less accurate, and those based on multiple-linear regression were substantially less accurate. Although logistic regression produced accurate estimates of median nitrate concentrations for the NJ Highlands Region, the more conventional quantile regression method would be the favored alternative for future similar studies over the somewhat cumbersome logistic-regression method used here. An additional benefit of quantile regression is that it generates an estimate of the value of the dependent variable (for example, nitrate concentration) for any quantile.

The estimated median nitrate concentration in groundwater in the NJ Highlands Region was 1.25 mg/L as N. The estimated median concentrations were highest in the Preservation Area/Existing Community Zone (1.79 mg/L as N) and lowest in the Preservation Area/Protection Zone (1.05 mg/L as N) using the logistic-regression method.

This application of logistic regression to determine the median value of a dependent variable requires a dataset sufficiently large to represent conditions in the area of study, logistic models that are appropriate for evaluating the phenomenon in question with closely spaced threshold values that bracket the range of expected values for the dependent variable, and explanatory variables that are significant across the range of threshold values. The large database provided by the New Jersey Private Well Testing Act coupled with extensive land-use data and other geo-referenced data was ideal for this application of logistic regression. With the spatial bias of well distribution removed, estimates of median nitrate concentration are more representative of the Highlands Region than are median nitrate concentrations for analyzed water samples.

References Cited

Anderson, J.R., Hardy, E.E., Roach, J.T., and Witmer, R.E., 1976, A land use and land cover classification system for use with remote sensor data: U.S. Geological Survey Professional Paper 964, 41 p.

Atherholt, T.B., Louis, J.B., Shevlin, J., Fell, K., Krietzman, S., 2009, The New Jersey Private Well Testing Act: An overview: New Jersey Department of Environmental Protection, Division of Science, Research and Technology, accessed October 31, 2014, at http://www.state.nj.us/dep/dsr/research/pwta-overview.pdf.

Barringer, T., Dunn, D., Battaglin, W., and Vowinkel, V., 1990, Problems and methods involved in relating land use to groundwater quality: Water Resources Bulletin, American Water Resources Association, v. 26, no. 1, 9 p.

Charles, E.G., Behroozi, Cyrus, Schooley, Jack, and Hoffman, J.L., 1993, A method for evaluating ground-water-recharge areas in New Jersey: Trenton, N.J., New Jersey Geological Survey Report GSR-32, 95 p.

Clawges, R.M., and Vowinkel, E.F., 1996, Variables indicating nitrate contamination in bedrock aquifers, Newark Basin, NJ: Journal of the American Water Resources Association, v. 32 no. 5, p. 1055–1066.

Dalton, 2003, Physiographic provinces of New Jersey: Trenton, N.J., New Jersey Geological Survey Circular, 2 p.

Dubrovsky, N.M., Burow, K.R., Clark, G.M., Gronberg, J.M., Hamilton, P.A., Hitt, K.J., Mueller, D.K., Munn, M.D., Nolan, B.T., Puckett, L.J., Rupert, M.G., Short, T.M., Spahr, N.E., Sprague, L.A., and Wilber, W.G., 2010, The quality of our nation's water—Nutrients in the nation's streams and groundwater, 1992–2004: U.S. Geological Survey Circular 1350, 176 p.

Dubrovsky, N.M., and Hamilton, P.A., 2010, Nutrients in the Nation's streams and groundwater: National findings and implications: U.S. Geological Survey Fact Sheet 2010–3078, 6 p.

Eckhardt, D.A.V., and Stackelberg, P.E., 1995, Relation of ground-water quality to land use on Long Island, New York: Ground Water, v. 33, p. 1019–1033.

Gardner, K.K., and Vogel, R.M., 2005, Predicting groundwater nitrate concentration from land use: Ground Water, v. 43, no. 3, p. 343–352.

Greene, E.A., LaMotte, A.E., and Cullinan, K., 2005, Groundwater vulnerability to nitrate contamination at multiple thresholds in the Mid-Atlantic Region using spatial probability methods: U.S. Geological Survey Scientific Investigation Report 2004–5118. 24 p.

Gurdak, J.J., and Qi., S.L., 2012, Vulnerability of recently recharged groundwater in principle aquifers of the United States to nitrate contamination: Environmental Science and Technology, v. 46, p. 6004–6012.

Helsel, D.R., 2005, Nondetects and data analysis: statistics for censored environmental data: Hoboken, N.J., Wiley-Interscience, 250 p.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources, *in* U.S. Geological Survey, Techniques of Water-Resources Investigations of the United States Geological Survey, Hydrologic analysis and interpretation, book 4, chap. A3.

Hoffman, J.L., and Canace, R.J., 2004, A recharge-based nitrate-dilution model for New Jersey: New Jersey Geological Survey Open File Report OFR 04–1, 27 p.

Hosmer, D.W., and Lemeshow, S., 2000, Applied logistic regression: Hoboken, N.J., Wiley-Interscience, 373 p.

Huang, J., Zhan, J., Yan, H., Wu, F., and Deng, X, 2013, Evaluation of the impacts of land use on water quality: A case study in the Chaohu Lake Basin: The Scientific World Journal, v. 2013, 7 p.

Insightful Corp., 2003, S-PLUS version 6.2: Seattle, Wash., USA.

Johnson, T.D., and Belitz, K., 2009, Assigning land use to supply wells for the statistical characterization of regional groundwater quality: Correlating urban land use and VOC occurrence: Journal of Hydrology, v. 370, p. 100–108.

Kolpin, D.W., 1997, Agricultural chemicals in groundwater of the Midwestern United States: Relations to land use: Environmental Science and Technology, v. 26, p. 1025–1037.

Koterba, M.T., 1998, Ground-water data-collection protocols and procedures for the National Water-Quality Assessment Program-Collection, documentation, and compilation of required site, well, subsurface, and landscape data for wells: U.S. Geological Survey Water-Resources Investigations Report 98–4107, 91 p.

Koenker, R., and Hallock, K., 2001, Quantile regression: Journal of Economic Perspectives, v. 15, no. 4, p. 143–156.

Michaels, J.A., Neville, L.R., Edelman, D., Sullivan, T., and DiCola, L.A., 1992. New York–New Jersey Highlands Regional Study: Radnor, Pa., U.S. Department of Agriculture, Forest Service, Northeastern Area State and Private Forestry, 130 p.

Mueller, D.K., and D.R. Helsel, 1996, Nutrients in the nations waters—too much of a good thing?: U.S. Geological Survey Circular 1136, 24 p.

New Jersey Department of Environmental Protection (NJDEP), 1986, 1986 Land use/land cover, accessed April 14, 2015, at http://www.nj.gov/dep/gis/lulcshp.html.

New Jersey Department of Environmental Protection (NJDEP), 2001, 1995/97 Land use/land cover by Watershed Management Area (WMA), accessed April 14, 2015, at http://www.state.nj.us/dep/gis/lulc95shp.html.

New Jersey Department of Environmental Protection (NJDEP), 2003, Private well testing act program electronic data deliverable manual: Trenton, N.J., Bureau of Safe Drinking Water, 70 p.

New Jersey Department of Environmental Protection (NJDEP), 2008, 2002 Land use/land cover by Watershed Management Area (WMA), accessed April 14, 2015, at http://www.state.nj.us/dep/gis/lulc02shp.html.

New Jersey Department of Environmental Protection (NJDEP), 2010, NJDEP 2007 Land use/land cover update, accessed April 16, 2015, at http://www.state.nj.us/dep/gis/lulc07shp.html.

New Jersey Highlands Water Protection and Planning Council, 2008, Highlands Master Plan, Technical report: Water resources volume I, watersheds and water quality, p. 114–173, accessed April 16, 2015, at http://www.highlands.state.nj.us/njhighlands/master/tr_water_res_vol_1.pdf.

Nicholson, R.S., McAuley, S.D., Barringer, J.L., and Gordon, A.D., 1996, Hydrogeology of, and ground-water flow in, a valley-fill and carbonate rock aquifer system near Long Valley in the New Jersey Highlands: U.S. Geological Survey Water-Resources Investigation Report 93–4157, 159 p., 3 pl.

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

Nolan, B.T., Ruddy, B.C., Hitt, K.J., and Helsel, D.R., 1998, A national look at nitrate contamination in ground water: Water Conditioning and Purification, v. 39, no. 12, p. 76–79.

Nolan, B.T., Hitt, K.J., and Ruddy, B.C., 2002. Probability of nitrate contamination of recently recharged groundwaters in the conterminous United States: Environmental Science and Technology, v. 36, no. 10, p. 2138–2145.

Sando, S.K., Vecchia, A.V., Lorenz, D.L., and Barnhart, E.P., 2014, Water-quality trends for selected sampling sites in the upper Clark Fork Basin, Montana, water years 1996–2010: U.S. Geological Survey Scientific Investigations Report 2013–5217, 162 p., with appendixes, http://dx.doiorg/10.3133/sir20135217.

Serfes, M.E., 1994, Natural ground-water quality in bedrock of the Newark Basin, New Jersey: New Jersey Geological Survey Report no. 35, 29 p.

Serfes, M.E., 2004, Ground water quality in the bedrock aquifers of the Highlands and Valley and Ridge Physiographic Provinces of New Jersey: New Jersey Geological Survey Report 39, 29 p.

Spearman, C., 1904, The proof and measurement of association between two things: The American Journal of Psychology, v. 15, no. 1, p. 72–101.

Tesoriero, A.J., and Voss, F.D., 2005, Predicting the probability of elevated nitrate concentrations in the Puget Sound Basin: Implications for aquatic susceptibility and vulnerability: Groundwater, v. 35, no. 6, p. 1029–1039.

Trela, J.J., and Douglas, L.A., 1978, Soils, septic systems and carrying capacity in the NJ Pine Barrens: Paper presented at the First Annual Pine Barrens Research Conference, Atlantic City, N.J., May 22, 1978, 34 p.

Tu, J., 2008, Assessing the impact of long-term land use changes on water quality in Eastern Massachusetts: Ann Arbor, Mich., ProQuest, 271 p.

U.S. Forest Service, 2014, The Highlands of Connecticut, New Jersey, New York and Pennsylvania, accessed April 28, 2015, at http://na.fs.fed.us/highlands/about/index.shtm.

U.S. Environmental Protection Agency, 2009, Statistical analysis of groundwater monitoring data at RCRA facilities. Unified guidance, EPA 530/R-09-007, 888 p. accessed April 28, 2015, at http://www.epa.gov/osw/hazard/correctiveaction/resources/guidance/sitechar/gwstats/unified-guid.pdf.

Wakida, F.T., and Lerner, D.N., 2005, Non-agricultural sources of groundwater nitrate: A review and case study: Water Research, v. 39, no.1, p. 3–16.

Appendixes 1 and 2

Appendix 1

Example spreadsheet for calculating median nitrate concentrations with logistic-regression models. (Appendix 1 available at http://dx.doi.org/10.3133/sir20155075)

Appendix 2

Geographic and environmental characteristics evaluated as possible explanatory variables in models of median nitrate concentrations in groundwater in the NJ Highlands Region. (Appendix 2 available at http://dx.doi.org/10.3133/sir20155075)

Prepared by the West Trenton Publishing Service Center

For more information, contact:
New Jersey Water Science Center
U.S. Geological Survey
3450 Princeton Pike, Suite 110
Lawrenceville, NJ 08648

http://nj.usgs.gov/

ISSN 2328-0328 (online)
http://dx.doi.org/10.3133/sir20155075