# USGS
science for a changing world

National Water-Quality Assessment Program

The Use of Process Models to Inform and Improve Statistical Models of Nitrate Occurrence, Great Miami River Basin, Southwestern Ohio

EXPLANATION

Difference in probability of exceeding a concentration of 4 mg/L, in percent, from 500-meter circular buffers and modeled recharge areas

27

-9

No difference

Area excluded

No model solution

Scientific Investigations Report 2012–5001

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

The Use of Process Models to Inform and Improve Statistical Models of Nitrate Occurrence, Great Miami River Basin, Southwestern Ohio

By Donald A. Walter and J. Jeffrey Starn

National Water-Quality Assessment Program

Scientific Investigations Report 2012–5001

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: 2013

Foreword

The U.S. Geological Survey (USGS) is committed to providing the Nation with reliable scientific information that helps to enhance and protect the overall quality of life and that facilitates effective management of water, biological, energy, and mineral resources (http://www.usgs.gov/). Information on the Nation's water resources is critical to ensuring long-term availability of water that is safe for drinking and recreation and is suitable for industry, irrigation, and fish and wildlife. Population growth and increasing demands for water make the availability of that water, measured in terms of quantity and quality, even more essential to the long-term sustainability of our communities and ecosystems.

The USGS implemented the National Water-Quality Assessment (NAWQA) Program in 1991 to support national, regional, State, and local information needs and decisions related to water-quality management and policy (http://water.usgs.gov/nawqa). The NAWQA Program is designed to answer: What is the quality of our Nation's streams and groundwater? How are conditions changing over time? How do natural features and human activities affect the quality of streams and groundwater, and where are those effects most pronounced? By combining information on water chemistry, physical characteristics, stream habitat, and aquatic life, the NAWQA Program aims to provide science-based insights for current and emerging water issues and priorities. From 1991 to 2001, the NAWQA Program completed interdisciplinary assessments and established a baseline understanding of water-quality conditions in 51 of the Nation's river basins and aquifers, referred to as Study Units (http://water.usgs.gov/nawqa/studies/study_units.html).

In the second decade of the Program (2001–2012), a major focus is on regional assessments of water-quality conditions and trends. These regional assessments are based on major river basins and principal aquifers, which encompass larger regions of the country than the Study Units. Regional assessments extend the findings in the Study Units by filling critical gaps in characterizing the quality of surface water and groundwater, and by determining water-quality status and trends at sites that have been consistently monitored for more than a decade. In addition, the regional assessments continue to build an understanding of how natural features and human activities affect water quality. Many of the regional assessments employ modeling and other scientific tools, developed on the basis of data collected at individual sites, to help extend knowledge of water quality to unmonitored, yet comparable areas within the regions. The models thereby enhance the value of our existing data and our understanding of the hydrologic system. In addition, the models are useful in evaluating various resource-management scenarios and in predicting how our actions, such as reducing or managing nonpoint and point sources of contamination, land conversion, and altering flow and (or) pumping regimes, are likely to affect water conditions within a region.

Other activities planned during the second decade include continuing national syntheses of information on pesticides, volatile organic compounds (VOCs), nutrients, trace elements, and aquatic ecology; and continuing national topical studies on the fate of agricultural chemicals, effects of urbanization on stream ecosystems, bioaccumulation of mercury in stream ecosystems, effects of nutrient enrichment on stream ecosystems, and transport of contaminants to public-supply wells.

The USGS aims to disseminate credible, timely, and relevant science information to address practical and effective water-resource management and strategies that protect and restore water quality. We hope this NAWQA publication will provide you with insights and information to meet your needs, and will foster increased citizen awareness and involvement in the protection and restoration of our Nation's waters.

The USGS recognizes that a national assessment by a single program cannot address all water-resource issues of interest. External coordination at all levels is critical for cost-effective management, regulation, and conservation of our Nation's water resources. The NAWQA Program, therefore, depends on advice and information from other agencies—Federal, State, regional, interstate, Tribal, and local—as well as nongovernmental organizations, industry, academia, and other stakeholder groups. Your assistance and suggestions are greatly appreciated.

William H. Werkheiser
USGS Associate Director for Water

Contents

Figures

Table

Conversion Factors, Datum, and Abbreviations

Inch/Pound to SI

Multiply	By	To obtain
Length		
foot (ft)	0.3048	meter (m)
Area		
acre	0.4047	hectare (ha)
square mile (mi^2)	259.0	hectare (ha)
Flow rate		
foot per day (ft/d)	0.3048	meter per day (m/d)
million gallons per day (Mgal/d)	0.04381	cubic meter per second (m^3/s)
Hydraulic conductivity		
foot per day (ft/d)	0.3048	meter per day (m/d)

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

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

Altitude, as used in this report, refers to distance above the vertical datum.

Concentrations of chemical constituents in water are given either in milligrams per liter (mg/L) or moles per liter (M/L).

Application rates are given in kilograms per year (kg/yr).

Abbreviations

NEU	nitrogen uptake efficiency
NAWQA	National Water-Quality Assessment Program
NLCDE92	Enhanced National Land Cover Data 1992
STATSGO	U.S. Department of Agriculture State Soil Geographic Database
USGS	U.S. Geological Survey

The Use of Process Models to Inform and Improve Statistical Models of Nitrate Occurrence, Great Miami River Basin, Southwestern Ohio

By Donald A. Walter and J. Jeffrey Starn

Abstract

Statistical models of nitrate occurrence in the glacial aquifer system of the northern United States, developed by the U.S. Geological Survey, use observed relations between nitrate concentrations and sets of explanatory variables—representing well-construction, environmental, and source characteristics—to predict the probability that nitrate, as nitrogen, will exceed a threshold concentration. However, the models do not explicitly account for the processes that control the transport of nitrogen from surface sources to a pumped well and use area-weighted mean spatial variables computed for a circular buffer around the well as a simplified source-area conceptualization. The use of models that explicitly represent physical-transport processes can inform and potentially improve these statistical models. Specifically, groundwater-flow models simulate advective transport—predominant in many surficial aquifers—and can contribute to the refinement of the statistical models by (1) providing for improved, physically based representations of a source area to a well, and (2) allowing for more detailed estimates of environmental variables.

A source area to a well, known as a contributing recharge area, represents the area at the water table that contributes recharge to a pumped well; a well pumped at a volumetric rate equal to the amount of recharge through a circular buffer will result in a contributing recharge area that is the same size as the buffer but has a shape that is a function of the hydrologic setting. These volume-equivalent contributing recharge areas will approximate circular buffers in areas of relatively flat hydraulic gradients, such as near groundwater divides, but in areas with steep hydraulic gradients will be elongated in the upgradient direction with respect to the corresponding circular buffers.

The degree to which contributing recharge areas estimated by process models, which simulate advective transport and therefore account for local hydrologic settings, would inform and improve the development of statistical models can be implicitly estimated by evaluating the differences between explanatory variables estimated from the contributing recharge areas and from the circular buffers used to develop existing statistical models. The larger the difference in estimated variables, the more likely that statistical models would be changed, and presumably improved, if explanatory variables estimated from contributing recharge areas were used in model development. Comparing model predictions from the two sets of estimated variables would further quantify—albeit implicitly—how an improved, physically based estimate of explanatory variables would be reflected in model predictions. Differences between the two sets of estimated explanatory variables and resultant model predictions vary spatially; greater differences are associated with areas of steep hydraulic gradients. A direct comparison, however, would require the development of a separate set of statistical models based on explanatory variables from contributing recharge areas.

In the Great Miami River Basin, a buried-valley glacial aquifer in southwestern Ohio, the agreement (measured as the coincident or overlapping area) between circular buffers and contributing recharge areas in a network of 21,774 uniformly spaced hypothetical wells ranged from less than 5 to more than 95 percent and averaged 32 percent. In this study, source areas represented as circular buffers generally differed substantially from those represented as contributing recharge areas of equivalent size. This difference indicates that incorporating an improved understanding of groundwater processes through process models can inform statistical models of nitrate occurrence in glacial environments.

Area-weighted means of three environmental variables—silt content, alfisol content, and depth to water from the U.S. Department of Agriculture State Soil Geographic (STATSGO) database—and one nitrogen-source variable (fertilizer-application rate from county data mapped to agricultural land use within the Enhanced National Land Cover Data 1992 (NLCDe92) database can vary substantially between circular buffers and volume-equivalent contributing recharge areas and among contributing recharge areas for different sets of well variables. The differences in estimated explanatory variables are a function of the same factors affecting the contributing recharge areas as well as the spatial resolution and local distribution of the underlying spatial data. As a result, differences

in estimated variables between circular buffers and contributing recharge areas are complex and site specific as evidenced by differences in estimated variables for circular buffers and contributing recharge areas of existing public-supply and network wells in the Great Miami River Basin. Large differences in area-weighted mean environmental variables were observed at the basin scale, determined by using the network of uniformly spaced hypothetical wells; the differences have a spatial pattern that generally is similar to spatial patterns in the underlying STATSGO data. Generally, the largest differences were observed for area-weighted nitrogen-application rates from county and national land-use data; the basin-scale differences ranged from -1,600 (indicating a larger value for the area within the volume-equivalent contributing recharge area) to 1,900 kilograms per year; the range in the underlying spatial data was 0 to 2,200 kilograms per year. Silt content, alfisol content, and nitrogen-application rate are defined by the underlying spatial data and are external to the groundwater system; however, depth to water is an environmental variable that can be estimated in more detail and, presumably, in a more physically based manner by using a groundwater-flow model instead of the spatial data. Model-calculated depths to water within circular buffers in the Great Miami River Basin differed substantially from values derived from the spatial data and had a much larger range.

Differences in estimates of area-weighted spatial variables result in corresponding differences in predictions of nitrate occurrence in the aquifer. In addition to the factors affecting contributing recharge areas and estimated explanatory variables, differences in predictions also are a function of the specific set of explanatory variables used and the fitted slope coefficients in a given model. For models that predicted the probability of exceeding 1 and 4 milligrams per liter as nitrogen (mg/L as N), probabilities based on variables estimated from circular buffers and contributing recharge areas generally were correlated but differed significantly at the local and basin scale. The scale and distribution of prediction differences can be explained by the underlying differences in the estimated variables and the relative weight of the variables in the statistical models. Differences in predictions of exceeding 1 mg/L as N on the basis of only environmental variables generally were correlated with the underlying differences in STATSGO data, whereas differences in predictions of exceeding 4 mg/L as N were more spatially extensive because that model included environmental and nitrogen-source variables. Using depths to water from within circular buffers derived from the spatial data and depths to water within the circular buffers calculated from the groundwater-flow model, restricted to the same range, resulted in large differences in predicted probabilities. The differences in estimated explanatory variables between contributing recharge areas and circular buffers indicate that incorporation of physically based contributing recharge areas likely would result in a different set of explanatory variables and an improved set of statistical models.

The use of a groundwater-flow model to improve representations of source areas or to provide more detailed estimates of specific explanatory variables includes a number of limitations and technical considerations. Assumptions in these analyses are that (1) there is a state of mass balance between recharge and pumping, and (2) transport to a pumped well is under a steady-state flow field. Comparison of volume-equivalent contributing recharge areas under steady-state and transient transport conditions at a location in the southeastern part of the basin shows that the steady-state contributing recharge area is a reasonable approximation of the transient contributing recharge area after 10 to 20 years of pumping. The first assumption is a more important consideration for this analysis. A gradient effect refers to a condition where simulated pumping from a well is less than recharge through the corresponding contributing recharge area. This effect generally takes place in areas with steep hydraulic gradients, such as near discharge locations, and can be mitigated by using a finer model discretization. A boundary effect refers to a condition where the recharge rate through the contributing recharge area is less than the volumetric pumping rate. This effect indicates that other sources could be contributing water to the simulated well. In the Great Miami River Basin, large gradient effects, where the volumetric recharge rate is more than double the corresponding pumping rate, and large boundary effects, where the volumetric recharge rate is less than half the corresponding pumping rate, occurs in 5 and 14 percent of the modeled basin, respectively. The agreements between circular buffers and volume-equivalent contributing recharge areas, differences in estimated variables, and the effects on statistical-model predictions between the population of wells with balances between pumping and recharge rates within 10 percent and the population of all wells were similar. These agreements indicated that process-model limitations did not affect the overall findings for the Great Miami River Basin; however, such a conclusion would be model specific, and prudent use of a process model needs to entail a limitations analysis and, if necessary, alterations to the model.

Introduction

Unconsolidated glacial sediments cover about 722,000 square miles (mi^2) and portions of 25 states in the northern contiguous United States (fig. 1A). The glacial sediments, which consist of gravel, sand, silt, and clay and exceed 1,000 feet (ft) in thickness in some areas, compose an important regional surficial aquifer system. The glacial aquifer system, which is defined as all unconsolidated sediments overlying bedrock in glaciated terrain, was a source of water for about 41 million people in 2000; about 40 percent of the water was pumped from private wells, primarily in rural areas (Warner and Arnold, 2006). About half (51 percent) of the land area overlying the glacial aquifer system is agricultural land that receives large amounts of applied fertilizer, and wells
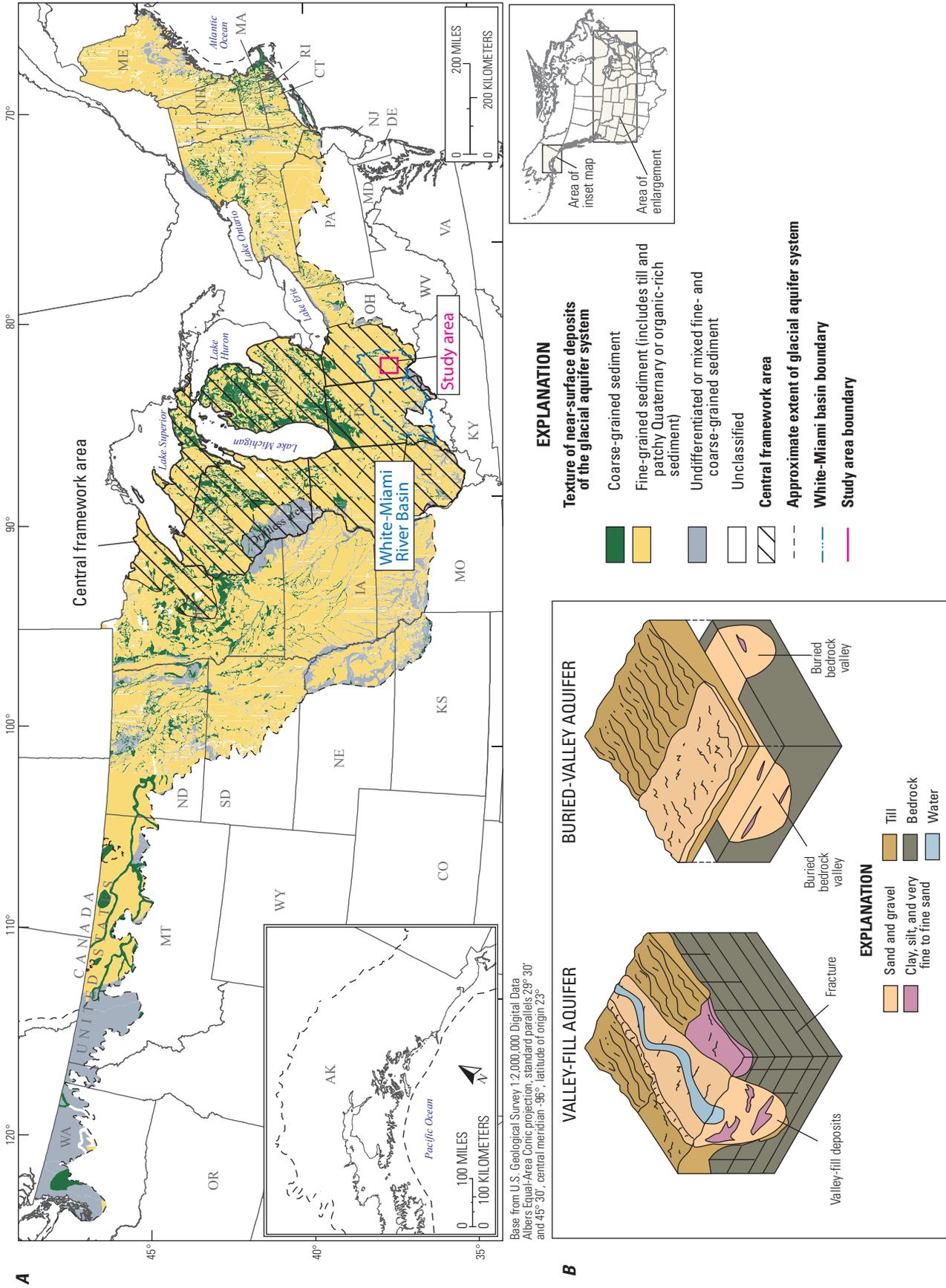
Figure 1. *A,* Extent of the glacial aquifer system, texture of near-surface glacial deposits, the central framework area, and location of the study area within the upper part of the Great Miami River Basin and *B,* block diagrams of valley-fill and buried-valley aquifers commonly found in the central framework area. (Modified from Warner and Arnold, 2010).

in agricultural areas are susceptible to contamination from fertilizer-derived nitrate (DeSimone and others, 2009). Nitrate, which can adversely affect human health, is conservative in oxic environments and, under such conditions, is readily transported in groundwater to pumped wells (Hem, 1985).

Since 1991, the U.S. Geological Survey (USGS) has collected water-quality data for the Nation's principal aquifers, including the glacial aquifer system, as part of the National Water-Quality Assessment (NAWQA) Program. DeSimone and others (2009) found that nitrate concentrations in private wells screened in the glacial aquifer system exceeded the U.S. Environmental Protection Agency Maximum Concentration Limit (MCL) of 10 milligrams per liter as nitrogen (mg/L as N) in about 3.8 percent of sampled wells. Concentrations exceeded the MCL in about 37.5 percent of private wells in agricultural areas; maximum concentrations were as high as 77 mg/L as N. Background concentrations of nitrate in the glacial aquifer system, where the aquifer is unaffected by anthropogenic sources, generally are less than 1 mg/L as N (Nolan and Hitt, 2003). Warner and Arnold (2010) identified relations between nitrate occurrence in the glacial aquifer system and a set of explanatory well-construction, environmental, and source variables and used the relations to develop statistical models of nitrate occurrence. The models predict the probability of nitrate occurrence above a set of threshold concentrations by using different combinations of well-construction and spatial variables; the latter set of variables was estimated by using a circular buffer around a well as a simplified representation of the source area to the well.

Although the statistical models may implicitly represent transport and attenuation processes through inclusion of environmental characteristics, they do not explicitly represent the physiochemical processes controlling nitrate concentrations in groundwater and do not account for many of the complexities of natural systems. Numerical process models simulating groundwater flow more explicitly represent areas contributing recharge to pumped wells and have the capacity to represent, to an extent, complexities of advective transport in natural systems.

In 2008, the USGS began a study to evaluate how incorporating an understanding of transport processes can inform or improve statistical models of nitrate occurrence. The area of investigation is within the upper part of the Great Miami River Basin in southwestern Ohio (fig. 2). The area is within the central framework region of the glacial aquifer system (fig. 1A), which is characterized by a predominance of buried-valley and valley-fill glacial aquifers that can have thick sequences of sand and gravel and high yields of water (fig. 1B) (Warner and Arnold, 2010). This basin, which is in the eastern part of the broader White-Miami River Basin encompassing parts of southern Ohio and Indiana (fig 1A), was chosen for study for a number of reasons: (1) the part of the basin used in this effort includes a mix of land uses typical of the broader glacial aquifer system; (2) environmental variables used in statistical models of nitrate occurrence had

already been defined for the basin, and data from sampled wells in the basin were used in developing the models; and (3) a numerical groundwater-flow model of the area had been developed and maintained that can simulate the physical process of advective transport in the underlying glacial aquifer system (Sheets, 2007) (fig. 2).

Purpose and Scope

This report documents the capabilities of a process model, specifically a numerical groundwater-flow model, to inform and improve statistical models of nitrate occurrence in the glacial aquifer system. The capabilities of groundwater-flow models to represent the advective transport of nitrate as nitrogen are evaluated, including the ability to (1) provide an improved representation of the source area to a pumped well, and (2) predict time-varying concentrations of nitrate in a well arising from the process of advective transport. Results from process and statistical models are compared to evaluate how incorporation of a process-level understanding of groundwater flow and nitrate transport can inform the development and use of statistical models.

Factors affecting the size, shape, and location of simulated contributing recharge areas and differences between contributing recharge areas and circular buffers at specific locations are evaluated, including well depth and hydrologic setting. A range of pumping rates is evaluated— from rates consistent with public-supply wells to those consistent with domestic wells. The role of hydrologic setting on the distribution of steady-state traveltimes and transient contributing recharge areas is evaluated and related to steady-state contributing recharge areas. The use of a process model to estimate nitrate concentrations in pumped wells and the factors that can affect those predictions are presented.

Sets of environmental variables estimated from circular buffers, as used in the development of statistical models, are compared to those estimated from model-calculated contributing recharge areas. Comparisons are made for existing public-supply and NAWQA network wells as well as for hypothetical wells in different hydrologic settings and at different depths and pumping rates. The sets of estimated environmental variables are used to calculate and compare probabilities of nitrate occurrence above selected threshold concentrations predicted by using the statistical models developed by Warner and Arnold (2010). These comparisons are used to quantify the degree to which incorporation of groundwater-flow processes can affect predictions of nitrate occurrence in pumped wells. The same set of analyses is presented for a network of uniformly spaced hypothetical wells to evaluate differences between statistical and process models at the basin scale. The effect of hydrologic setting on physical representations of source areas and the potential effect of an improved representation of the source on statistical models across the basin are evaluated.
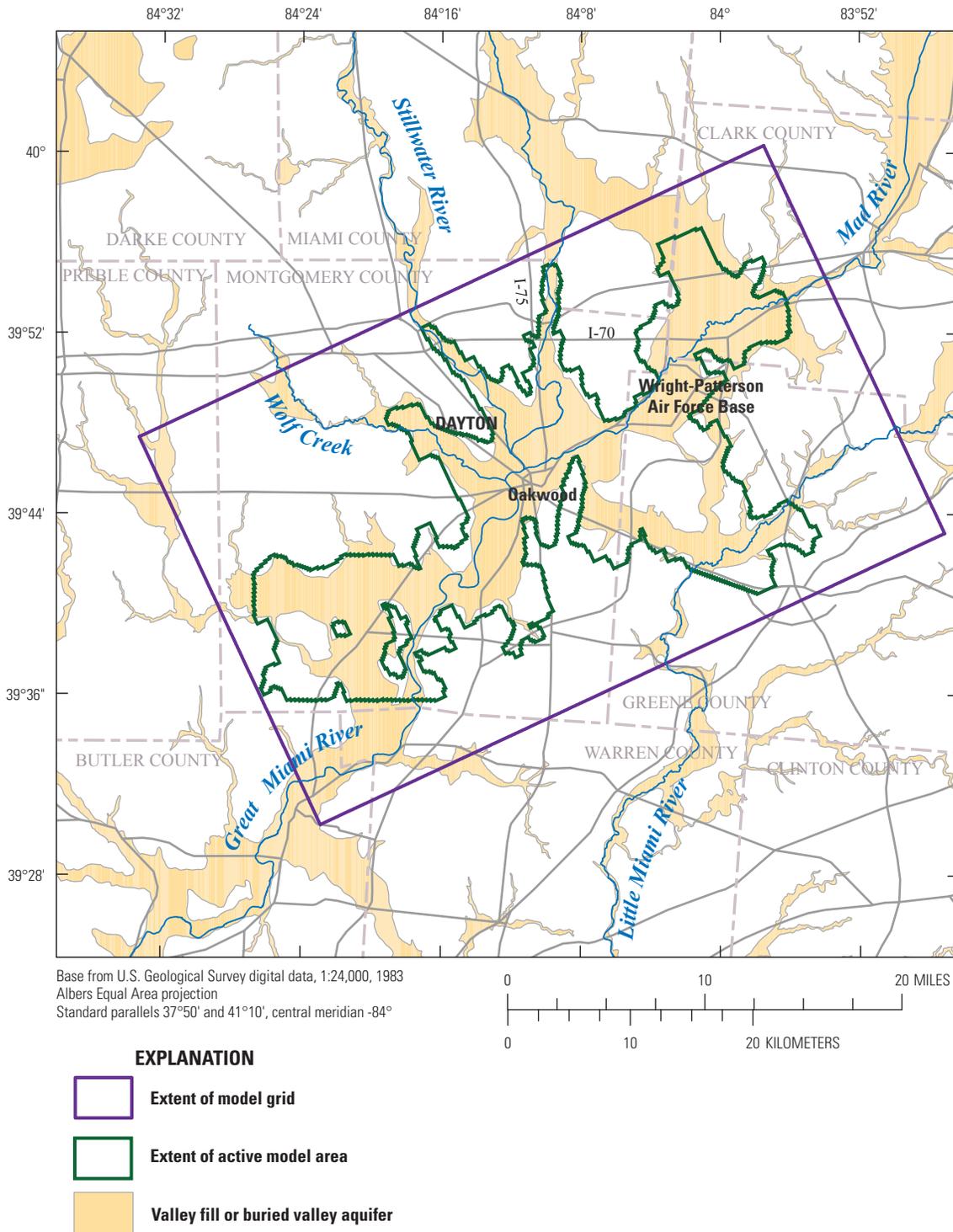
Figure 2. Extent of buried-valley glacial aquifers within the study area near Dayton, Ohio, and extent of regional groundwater-flow-model grid, southwestern Ohio. (Modified from Sheets, 2007).

Limitations and technical considerations for use of process models to inform statistical models that are applicable to other basins are presented. Potential effects of process-model design on physically based source areas and how those effects can, in turn, affect differences in statistical and process-model predictions are evaluated. The evaluation of model limitations presents considerations for use in developing future process models or modifying existing process models to better facilitate their use in improving statistical models. It should be noted that the analyses presented in the report require existing statistical and process models. Although statistical models have been developed that are applicable to the entire glacial aquifer system, currently (2011), areas represented by numerical process models encompass a relatively small part of that system.

Methods of Analysis

This effort utilized statistical and process models developed for the glacial aquifer system as part of previous efforts by the USGS. Warner and Arnold (2010) developed statistical models—using linear and logistic regression—to relate nitrate concentrations and occurrence to sets of well-construction and environmental explanatory variables. Sheets (2007) modified an existing numerical groundwater-flow model of a glacial aquifer system near Dayton, Ohio, within the Great Miami River Basin; the model allows for the simulation of groundwater-flow processes and the advective transport of conservative solutes.

Statistical Models of Nitrate Occurrence

Warner and Arnold (2010) identified a number of correlations between the occurrence of nitrate and a number of well-construction and environmental variables; the study used water-quality data collected between 1991 and 2005 from a network of 379 private wells, 41 public-supply wells, and 870 monitoring wells as part of the NAWQA Program. The correlations were used to develop statistical models of the probability of nitrate occurrence above 10 threshold concentrations: 1 to 10 mg/L as N. The models use logistic regression to estimate the probability that nitrate concentration in a well exceeds the threshold concentration. The general form of logistic regression is as follows:

$$P = exp\left(B_o + B_1V_1 + B_2V_2 + \ldots + B_nV_n\right)/$$
$$1 + \left(B_o + B_1V_1 + B_2V_2 + \ldots + B_nV_n\right) \quad , \quad (1)$$

where

P is the probability of exceeding the threshold concentration,

B_o is the intercept,

B_n is the fitted slope parameter, and

V_n is the explanatory variable.

Three well-construction variables and four spatial variables were included, in differing combinations, in the logistic-regression models. The variables used in each of the 10 logistic-regression models developed by Warner and Arnold (2010) are summarized in table 1 and figure 3A. Well-construction variables include open-interval diameter, length of the open interval (the part of the well casing open to the aquifer), and depth to the top of the open interval; a fourth variable is an interaction term between depth to top of the open interval and open-interval diameter. Spatial variables include three environmental variables derived from the U.S. Department of Agriculture State Soil Geographic (STATSGO) database (Wolock, 1997): silt content, alfisol content, and depth to water. Alfisols, which cover about 25 percent of the glacial aquifer system, are clayey soils formed from weathering processes in humid areas with deciduous forest or mixed vegetative cover (Warner and Arnold, 2010). An additional spatial variable was the 10-year mean nitrogen fertilizer-application rate (1992–2001) estimated from county data and applied to agricultural land delineated in the Enhanced National Land Cover Data 1992 (NLCDe92) database (Nakagaki and others, 2007). The slope coefficients for explanatory variables and intercepts are summarized in table 1. Larger slope coefficients indicate that a variable has a larger effect on predicted probabilities than variables with smaller coefficients for a given unit of measure; a positive coefficient indicates that the probability increases as the variable increases, and a negative slope indicates that the probability decreases as the variable increases. The effect of a spatial variable on the predicted probability is the product of the slope coefficient and the value of the variable.

Environmental variables—silt content, alfisol content, and depth to water—were based on STATSGO data and had the same spatial variability; seven value fields were represented within the Great Miami River Basin (fig. 4A). Silt content ranged from 48.55 to 57.20 percent; alfisol content, where present, from 69 to 94 percent; and depth to water from 2.48 to 4.88 ft. Application rates for nitrogen were estimated from county usage data applied to agricultural land delineated in the NCLDE92 database. The mean nitrogen-application rates (1992–2001) on agricultural land in the Great Miami River Basin ranged from 1,413 to 2,202 kg/yr (fig. 4B). The value of each spatial variable used in developing the models was determined by calculating the area-weighted mean for a circular area of radius 1,640 ft [500 meters (m)] around a well (fig. 3B); the circular buffer encompassed an area of about 0.3 mi^2.

The 7 variables included in the statistical models were a subset of 18 variables that were correlated with nitrate concentrations at a 95-percent confidence level (Warner and Arnold, 2010) by using Spearman's rho rank-correlation coefficient (Helsel and Hirsch, 1992). The 18 variables were, in turn, a subset of 46 variables for which visual inspection of scatter plots suggested a linear correlation with nitrate concentrations; the variables included categories related to land use, watershed characteristics, soil properties,

Table 1. Summary of statistical models of nitrate occurrence in the glacial aquifer system for threshold concentrations of 1 to 10 milligrams per liter as nitrogen (Warner and Arnold, 2010).

[mg/L, milligrams per liter; N, nitrogen; >, greater than; kg/yr, kilograms per year; --, variable not used in model; logistic regression models used variable values for each private well in the glacial aquifer system; from Warner and Arnold, 2010]

Variable type	Variable description	Variable abbreviation	Nitrate concentration, in mg/L as N									
			Model parameters, slope coefficients, and intercepts									
			>10	>9	>8	>7	>6	>5	>4	>3	>2	>1
Intercept			-2.1454	-1.0447	-1.396	-1.187	-4.7459	-6.8249	-5.3277	-4.6698	-5.1108	-3.9552
Source variables	10-year mean (1992–2001) nitrogen applied to farms, in kg/yr	NFARM_10YR	0.000658	0.000474	0.00047	0.000422	0.000555	0.00042	0.000211	--	--	--
Transport variables	Depth to top of open interval, in inches	DEPTH_TOI_	-0.0123	-0.0104	-0.00812	-0.00793	-0.00563	-0.004	-0.00187	-0.00165	-0.00152	-0.012
	Open interval diameter, in inches	OI_CASE_DI	-0.26	-0.2021	-0.1614	-0.1599	-0.0909	-0.05	0.0151	--	--	--
	Length of open interval, in inches	OILENGTH_I	--	--	--	--	--	--	--	0.00111	0.000985	0.00309
	Silt in soil, in percent	SILTAVE	--	--	--	--	--	0.039	0.0326	0.0481	0.0535	0.0356
	Depth to water, in feet	WTDEPAVE	--	--	--	--	0.7818	0.8839	0.7355	0.6468	0.8013	0.7219
Attenuation variables	Alfisols, in percent	ORDA	--	--	--	--	-0.0211	-0.019	-0.0127	-0.0156	-0.0188	-0.0164
	Interaction term between depth to top and diameter of open interval	OI_CASE_D*DEPTH_TOI_	0.00135	0.00103	0.000776	0.000772	0.000541	0.000367	--	--	--	--
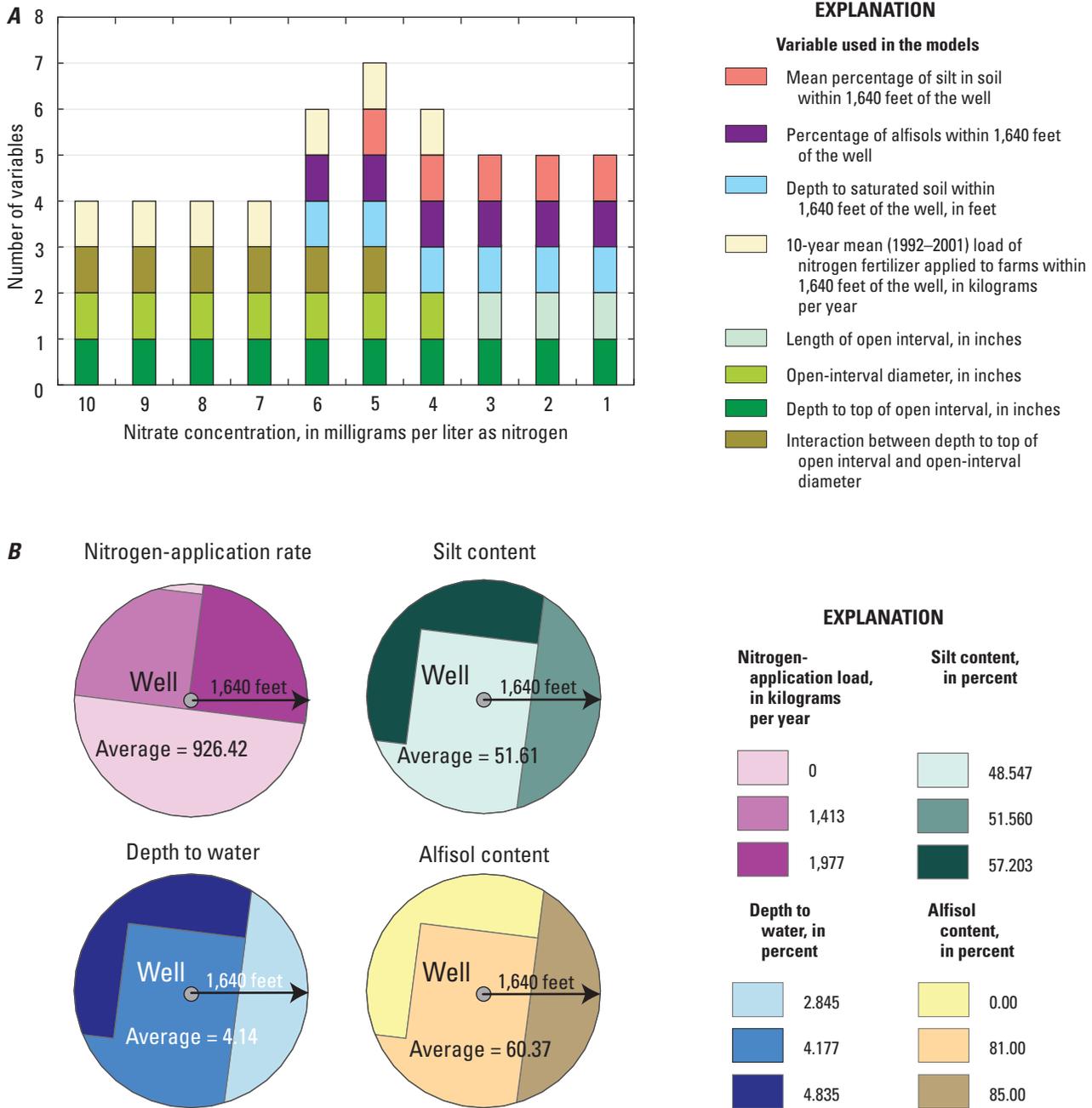
Figure 3. *A,* Sets of explanatory variables used in 10 logistic-regression models to describe nitrate occurrence in the glacial aquifer system (modified from Warner and Arnold, 2010) and *B,* example of spatial variables within a 1,640-ft radius of a hypothetical well in the Great Miami River Basin, southwestern Ohio.

Base from U.S. Geological Survey digital data, 1:24,000, 1983
Albers Equal Area projection
Standard parallels 37°50' and 41°10', central meridian −84°

EXPLANATION

STATSGO data

Depth to water, in feet	Silt content, in percent	Alfisol content, in percent
2.48	48.55	0.0
2.85	51.09	69.0
3.63	51.43	74.0
4.18	51.56	81.0
4.76	52.08	84.0
4.84	52.88	85.0
4.88	57.20	94.0

County data

Nitrogen-application load, in kilograms per year

	0
	1,413
	1,861
	1,977
	2,202

Figure 4. Spatial distribution of *A*, environmental variables—silt content, alfisol content, and depth to water—and *B*, nitrogen application rates in the modeled area of the Great Miami River Basin, southwestern Ohio.

population density, and waste disposal (Warner and Arnold, 2010). The 7 variables explained the variability in nitrate concentrations at a significance level exceeding 0.05 as determined by a stepwise-forward-selection method (Warner and Arnold, 2010).

There are a number of assumptions inherent in statistical models of nitrate occurrence. The most important assumption is that the source of water to a pumped well is within a 1,640-ft-radius circular buffer around the well. In natural systems, the areas contributing recharge to a well are a function of the hydrologic setting (the magnitude and direction of hydraulic gradients), the depth of the well, and the pumping rate. It is likely that explanatory variables estimated for the area within the circular buffers would differ from those estimated for a physically based representation of the areas contributing recharge to pumped wells.

Numerical Model of the Great Miami River Basin

Advection refers to the movement of a solute with average groundwater flow and generally is the dominant component of transport in unconsolidated surficial aquifers, particularly in areas with permeable sediments and high recharge rates. Groundwater-flow models that simulate hydraulic gradients are effective for representing the advective transport of chemically conservative solutes, such as nitrate in oxic environments. Numerical groundwater-flow models generally use finite-difference methods to simulate heads and flows in a spatially discretized representation of an aquifer system. Model input parameters include intrinsic aquifer properties, hydraulic boundary conditions, and natural and anthropogenic stresses. Particle-tracking methods use

simulated cell-by-cell flows to represent the movement of water within the aquifer, and can, in turn, be used to simulate contributing recharge areas and the advective transport of contaminants from surface sources to pumped wells.

A regional groundwater-flow model developed by Dumouchelle (1998) for part of the Great Miami River Basin was modified and recalibrated to simulate advective-transport paths and contributing recharge areas to public-supply wells (Sheets, 2007); detailed descriptions of the original and modified models are presented in Sheets (2007). The glacial aquifer system in the study area is a buried-valley aquifer in which glacial sediments fill bedrock valleys; the sediments, which are over 300 ft thick in some areas, generally are coarse grained with lenses of finer grained material (fig. 5) and can be productive sources of water (Sheets, 2007). Recharge from precipitation, which is about 40 in/yr in the basin, and flow from adjacent uplands—as streamflow or as shallow groundwater flow—are the sole sources of water to the aquifer; water leaves the system as streamflow or is withdrawn from wells (fig. 5). The part of the aquifer represented by the regional model also receives water from downvalley flow within the aquifer from the upgradient part of the aquifer and contributes water to downgradient parts of the aquifer. The aquifer is bounded laterally and below by relatively impermeable bedrock consisting of Ordovician and Silurian shale, limestone, and other types of carbonate rock (fig. 5).

The calibrated regional model of the Great Miami River Basin (Sheets, 2007) used MODFLOW–2000 (Harbaugh and others, 2000) to simulate hydrologic conditions in the glacial aquifer system and the inverse-calibration capabilities of MODFLOW–2000 (Hill and others, 2000) to calibrate the model to observed hydraulic conditions. In the current set of
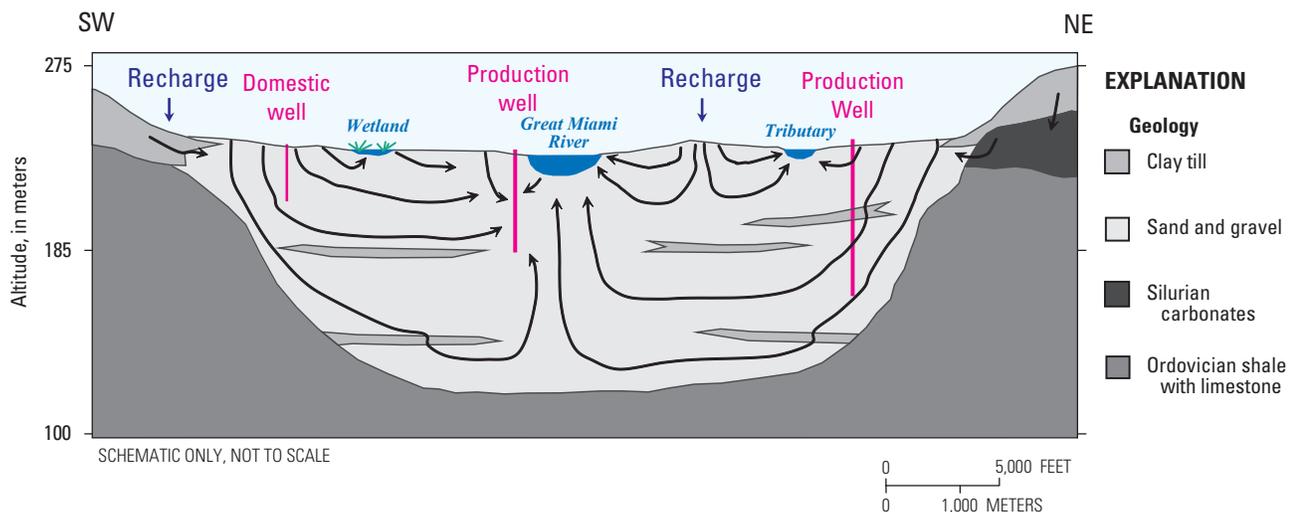
Figure 5.   Generalized geology and groundwater-flow patterns in a buried-valley glacial aquifer typical of the upper part of the Great Miami River Basin, southwestern, Ohio. (Modified from Sheets, 2007)

analyses, the particle-tracking program MODPATH (Pollock, 1994) was used to simulate the sources of water to pumped wells—existing and hypothetical—and the advective transport of nitrate for a number of hydrologic settings and pumping rates; the geoprocessing software MODTOOLS (Orzol, 1997) was used to convert model and particle-tracking output to georeferenced forms.

There are a number of assumptions inherent in steady-state numerical models of groundwater systems: (1) the hydraulic stresses (pumping and recharge) are constant over time, (2) aquifer properties are homogenous within model cells, and (3) calibrated aquifer properties and recharge reasonably match those of the real system. Although recharge varies over time, transport of contaminants often occurs over time scales long enough to make the assumption of steady-state conditions reasonable. Regional models, such as those of the Great Miami River Basin, address regional-scale questions that likely are not affected by variations of aquifer properties within model cells. The accuracy of calibrated aquifer properties can be assessed by comparing the match between measured and observed heads and flows, as discussed in Sheets (2007).

The regional model was originally developed to evaluate the source of water and transport of anthropogenic contaminants to public-supply wells, primarily in urbanized areas around Dayton, Ohio (Sheets, 2007). The analyses presented in this report represent a use of the model that differs from its original purpose; however, the results indicate that the model was appropriate for these analyses in most instances. Some model limitations of importance to these analyses, as discussed later in the report, likely are not of relevance to the analyses for which the model was originally designed.

Model Design and Calibration

The regional model finite-difference grid consists of 230 rows, 370 columns, and 3 layers; has a uniform horizontal discretization of 500 ft; and is rotated 25 degrees (°) east of north (fig. 6). The active model area encompasses about 254 mi^2 and incorporates the buried-valley aquifer system near Dayton, Ohio (fig. 2). The simulated aquifer generally is thickest near the axis of buried bedrock valleys and thinnest in areas near bedrock and glacial till uplands and areas of bedrock highs within the aquifer; the latter are overlain by glacial till and soil. Simulated water-table altitudes range from over 1,000 ft near bedrock and glacial till uplands and along valley margins to about 600 ft where the Great Miami River exits the aquifer system (fig. 6). About 377 public-supply wells operate within the basin, generally near urban areas.

Hydraulic conductivity (HK1) in the top layer of the model and recharge (RCH) are grouped into four parameter zones within the aquifer system: buried-valley sediments (BV), bedrock/till uplands (UP), sediments marginal to uplands (EDGE), and sediments in the Oakwood area in the south-central part of the aquifer (OAK) (fig. 7). Multipliers

were used to allow values in individual cells to vary within zones. Parameters also represent horizontal and vertical hydraulic conductivity in layers 2 and 3 and river leakances (RIV) in buried-valley and upland areas. A detailed discussion of the model parameterization is presented in Sheets (2007).

Formal sensitivity analyses and parameter estimation, the sensitivity analysis (SEN) and parameter estimation (PES) Processes in MODFLOW–2000 (Hill and others, 2000), respectively, were used to improve calibration of the model to observed hydrologic conditions (Sheets, 2007). Simulated equivalents at observation locations were sufficiently sensitive to allow for the estimation of six parameters: HK1_BV, HK2, RCH_BV, RCH_UP, RCH_EDGE, and RIV_BV; the remaining parameters used specified values. A detailed discussion of the inverse-calibration methods, sensitivity analyses, and model fit to observed hydrologic conditions is presented in Sheets (2007). Calibrated hydraulic conductivities in the top layer of the model ranged from less than 10 to more than 400 feet per day (ft/d) (fig. 8A), and recharge ranged from less than 6 to more than 12 inches per year (in/yr) (fig. 8B). Sediment porosity, which affects groundwater velocity and transport times, was uniform and specified as 0.20.

Modifications to Regional Model of the Great Miami River Basin

The current set of analyses used the regional model of the Great Miami River Basin documented in Sheets (2007). Model layering, parameter values, hydrologic boundaries, and natural recharge stresses were the same as in the documented model. Subregional models were developed for three areas in the southeastern part of the model (fig. 6) to allow for the representation of pumped wells with low withdrawal rates, such as are typical in domestic-supply wells. Spatial discretization in these areas was decreased from 500 to 10 ft, and the subregional models were linked to the regional model by using a specified-head boundary condition derived from regional-model-calculated heads; vertical discretization was the same as in the regional model. In each case, pumping stresses were consistent between the regional and subregional models, and simulated heads and hydrologic budgets for the subregional models were checked to make sure they were consistent with the coincident area within the regional model.

The steady-state regional model was temporally discretized to simulate contributing recharge areas resulting from a transient pumping stress. The length of the transient stress period was 50 years divided into 300 uniform time steps. The transient stress period used the same constant recharge stresses as the regional model and a constant pumping stress, which resulted in a solution at the end of the transient period that was similar to that from the steady-state regional model. Model-calculated steady-state heads were used as an initial condition. Specific yield and storativity of the aquifer sediments were uniform and specified as 0.23 and 0.0001, respectively.
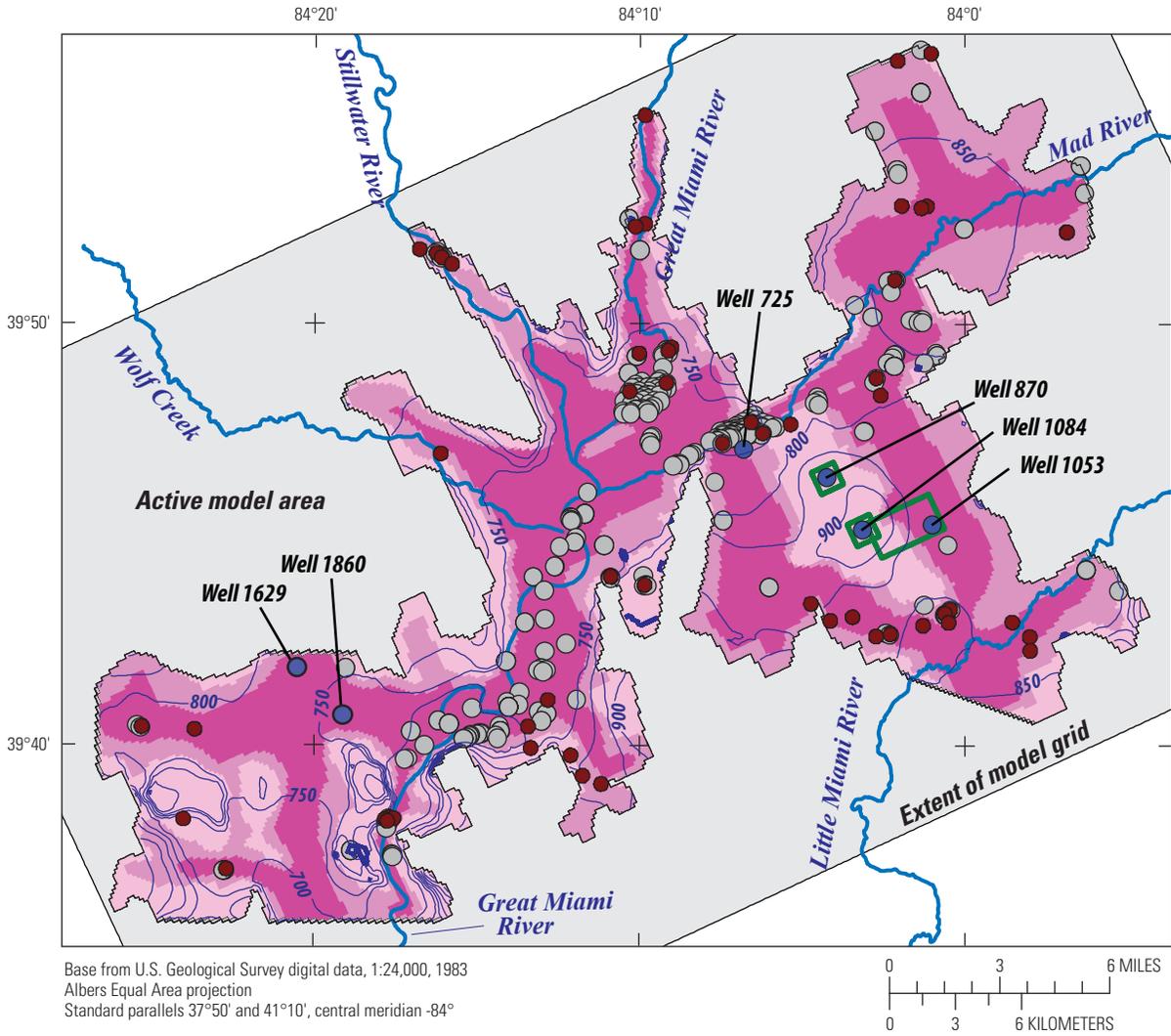
Figure 6. Regional model grid, active model layers, extents of subregional models, simulated water table, and locations of production wells, upper part of the Great Miami River Basin, southwestern Ohio.
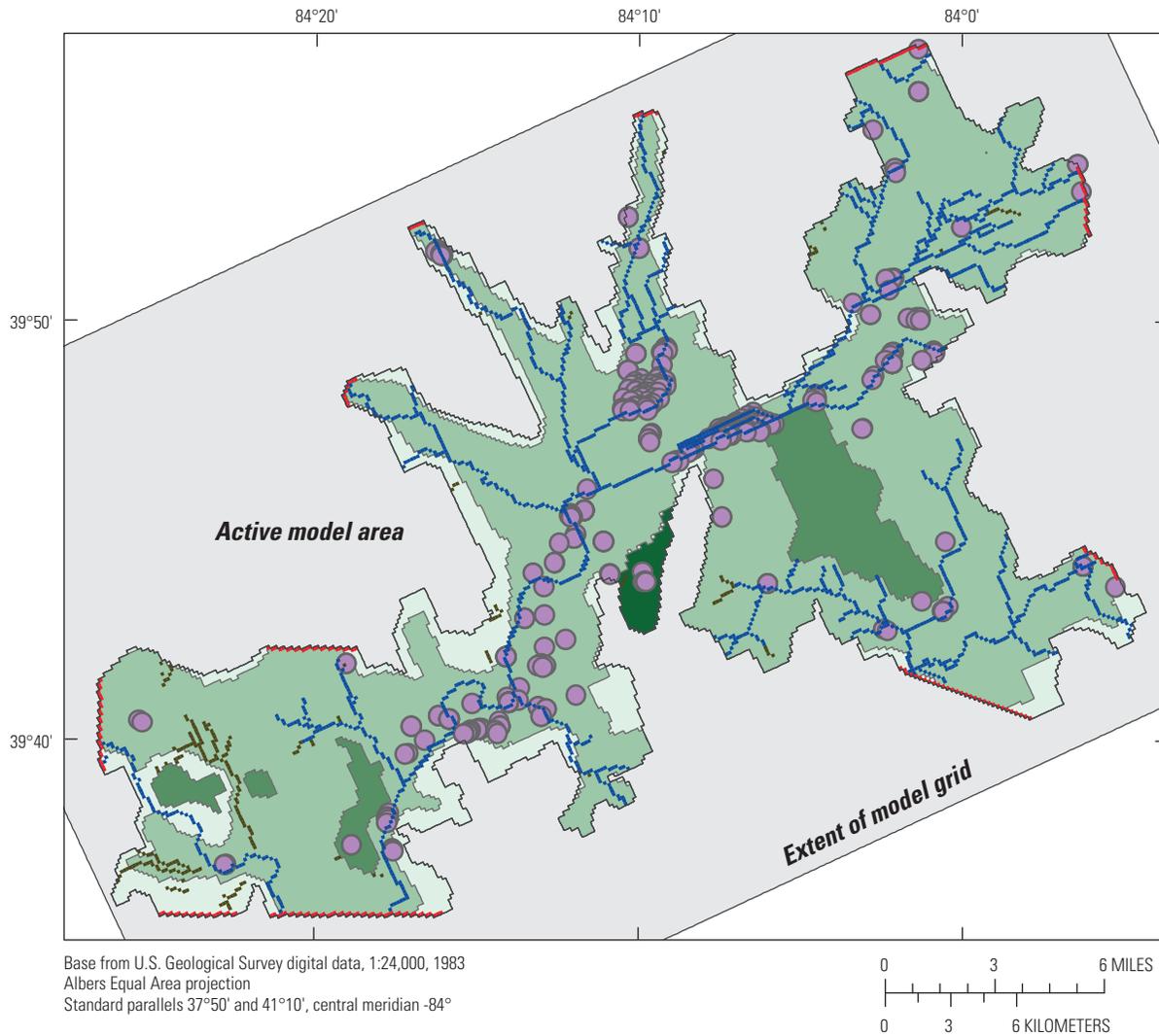
Figure 7. Hydrologic boundaries and hydraulic-conductivity and recharge parameter zones used to represent the regional aquifer system, upper part of the Great Miami River Basin, southwestern Ohio.

EXPLANATION

Hydraulic conductivity, in feet per day

> 0 to 10
> 10 to 20
> 20 to 30
> 30 to 50
> 50 to 100
> 100 to 150
> 150 to 200
> 200 to 300
> 300 to 400
> 400

Recharge, in inches per year

> 5.5 to 6.0
> 6.0 to 7.0
> 7.0 to 8.0
> 8.0 to 9.0
> 9.0 to 10.0
> 10.0 to 11.0
> 11.0 to 12.0
> 12.0 to 13.0

Base from U.S. Geological Survey digital data, 1:24,000, 1983
Albers Equal Area projection
Standard parallels 37°50' and 41°10', central meridian -84°
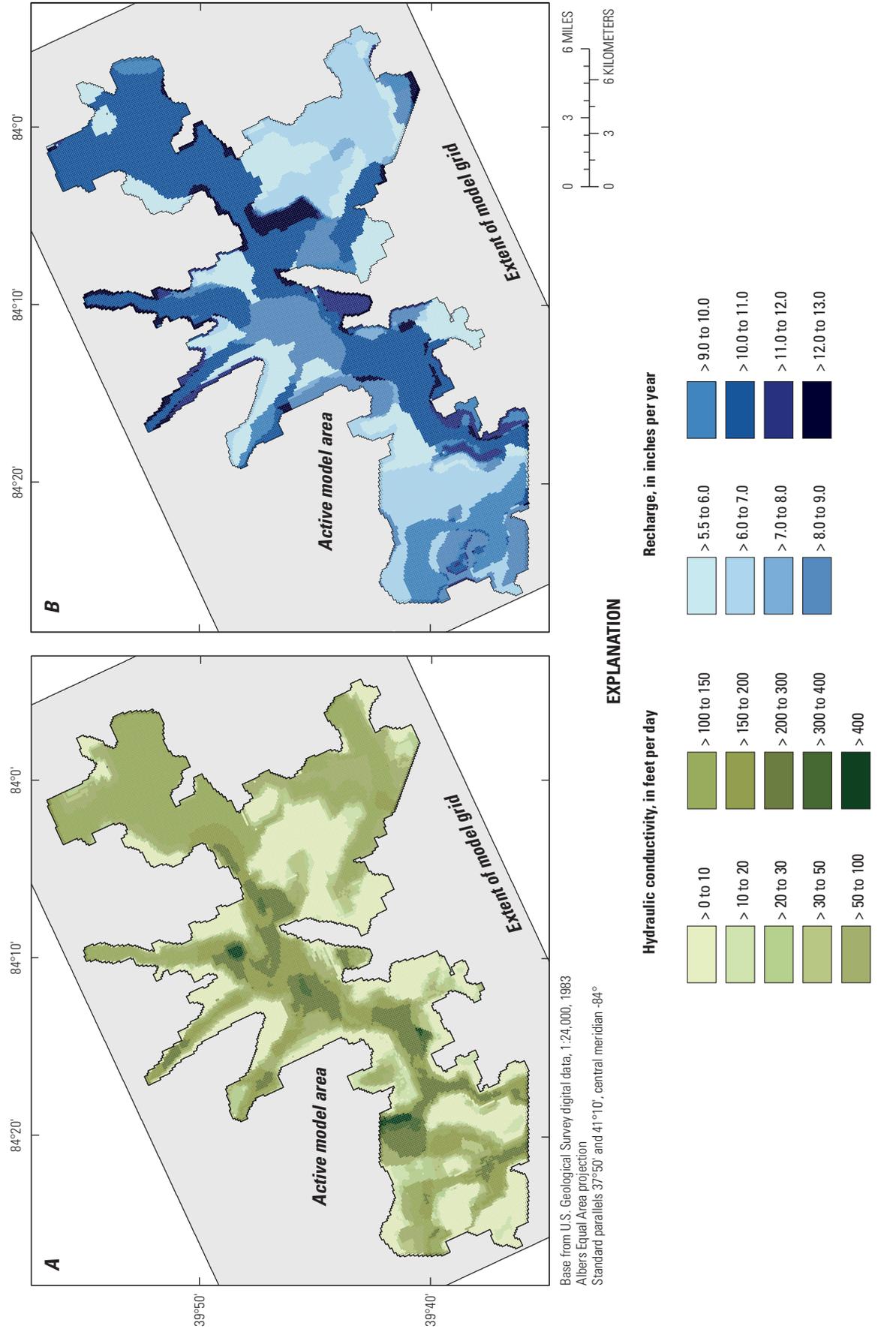
Figure 8.   Distribution of *A*, horizontal hydraulic conductivity values in layer 1 and *B*, recharge from precipitation in the calibrated regional model of the Great Miami River Basin, southwestern Ohio.

Sources of Water and Advective Transport of Nitrate to Pumped Wells

The contributing recharge area to a pumped well refers to the area at the water table that contributes water to the well. Under steady-state conditions, there is a mass balance between water entering through the contributing recharge area and withdrawal from the well. In cases where recharge is the sole source of water to a pumped well—a common situation for surficial aquifers—the contributing recharge area is proportional to the pumping rate of the well, and the proportionality constant is the recharge rate. In some cases, sources of water to a well can include other sources, such as induced infiltration from nearby surface waters or, under transient conditions, storage in the aquifer.

An assumption implicit in statistical models of nitrate occurrence (Warner and Arnold, 2010) is that the source of water to a pumped well is within a radius of 1,640 ft (500 m) of the well (fig. 3B) and that area-weighted mean spatial variables for the area within that circular buffer approximate the mean values for the area of the land surface overlying the contributing recharge area to the well. Differences between a true (or modeled) contributing recharge area and a circular buffer around the well can result in differences in the estimates of area-weighted spatial variables. Process models that simulate groundwater flow within an aquifer can be used to estimate contributing recharge areas to wells and account for hydrologic conditions, including aquifer properties, hydrologic boundaries, and recharge and pumping stresses. A contributing recharge area to a well pumped at a rate equivalent to the volume of recharge through a circular buffer would be the same as the area of the circular buffer but would account for groundwater processes and better represent the actual source area to the well.

Factors Affecting Steady-State Contributing Recharge Areas to Pumped Wells

The source of water to a pumped well is a function of the pumping rate, hydrologic setting, depth, and duration of pumping of the well and characteristics of the aquifer system, including aquifer geometry, hydrologic boundaries, natural stresses, and sediment characteristics (Reilly and Pollock, 1993; Barlow, 1994; Masterson and others, 1998; Walter and Whealan, 2003; Walter and others, 2004). The pumping rate determines the size of the contributing recharge area. The shape of the contributing recharge area is a function of hydraulic gradients, and the location of the contributing recharge area relative to the well is affected by the depth of the well screen. For a given pumping rate and screen elevation, local aquifer heterogeneities can result in additional complexities in the shape of the contributing recharge area.

Pumping Rate

Under steady-state conditions, there is a condition of mass balance between recharge and pumping, and therefore, for a given recharge rate, the size of the contributing recharge areas is proportional to the pumping rate and, for a given pumping rate, the size of the contributing recharge areas is inversely proportional to the recharge rate. Simulated steady-state contributing recharge areas for three pumping rates from a hypothetical well (well 1053) in the southeastern part of the basin (fig. 6) are shown in figure 9. The well is in an area with steep hydraulic gradients, and the contributing recharge area extends upgradient from the well towards a regional water-table divide (fig. 9).

In this report, a high pumping rate refers to a withdrawal equal to the mean withdrawal for public-supply wells in the basin (about 0.25 million gallons per day (Mgal/d)); an intermediate pumping rate refers to a rate equal to the recharge through the corresponding 1,640-ft-radius circular buffer (a rate of about 0.08 Mgal/d at this location); and a low pumping rate refers to a rate similar to the withdrawal rate of a domestic-supply well [about 350 gallons per day (gal/d)]. About 19 and 33 percent of steady-state contributing recharge areas for well 1053 under high and intermediate pumping rates, respectively, are within the 1,640-ft-radius circular buffer. The contributing recharge area under a low pumping rate is completely within the 1,640-ft-radius circular buffer but extends to near the edge (fig. 9).

Hydrologic Setting

Contributing recharge areas to pumped wells are also a function of the depth of the well screen and regional hydraulic gradients around the well. Wells 870 and 1084 are near a bedrock/glacial till upland where the aquifer is thin; well 1084 is near a regional groundwater divide in an area with small, radial hydraulic gradients, and well 870 is in an area with steep hydraulic gradients (figs. 6 and 10). Well 1053 is in an area of steep hydraulic gradients within a buried valley where the aquifer is thicker. At each location, the intermediate pumping rate and resulting contributing recharge areas are equivalent to the rate of recharge through and size of the corresponding circular buffer, so the effect of regional gradients can be quantified by the overlap of the contributing recharge area and the circular buffer—defined as the fraction of the recharge area that is inside the circular buffer. Near the groundwater divide (well 1084), where hydraulic gradients are relatively flat, the contributing recharge area is similar to the area of the circular buffer; about 87 percent of the contributing recharge area is inside the circular buffer (fig. 10). In areas with steep hydraulic gradients, there is less overlap between contributing recharge areas and circular buffers—about 56 percent at well 870 and 33 percent at well 1053 (fig. 10).

Near the regional groundwater divide, where the contributing recharge areas are similar to the area of the circular buffer (well 1084), the entire area contributing
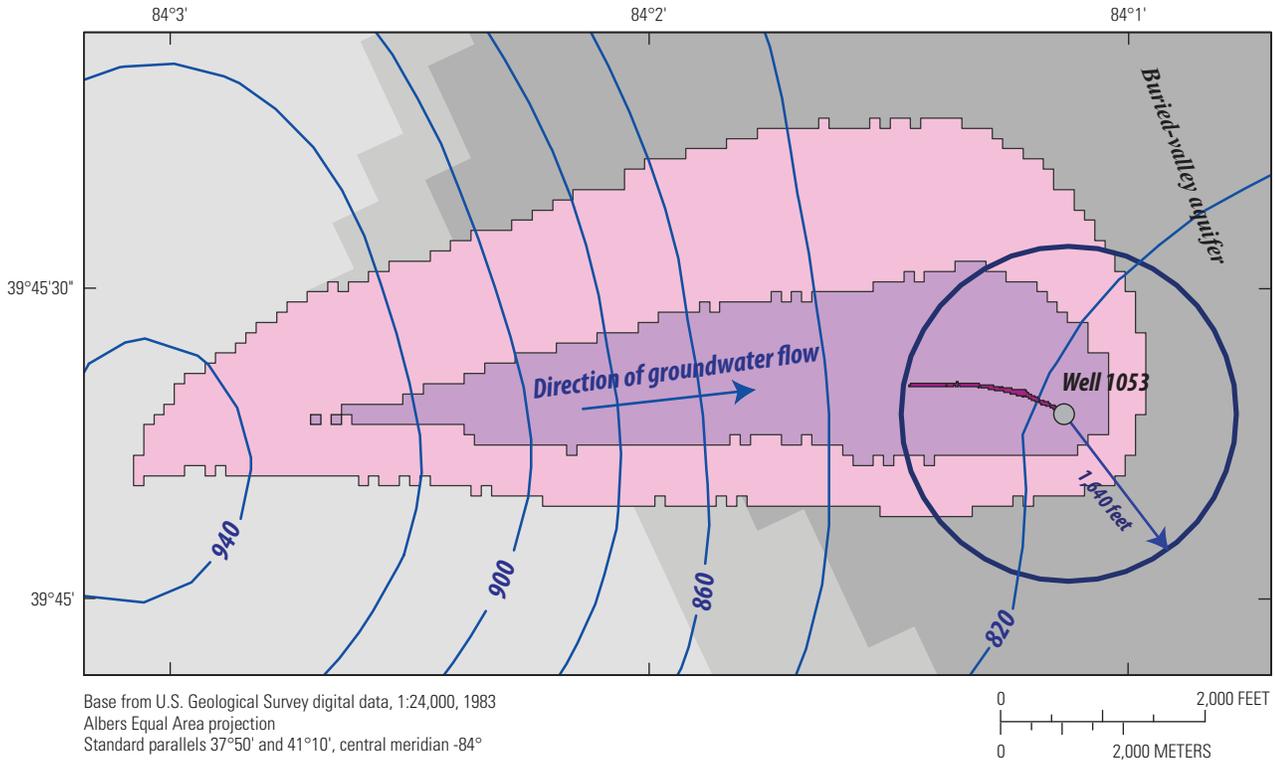
Figure 9. The 1,640-ft-radius circular buffer and steady-state contributing recharge areas to well 1053 (a hypothetical well) for three pumping rates, southeastern part of the Great Miami River Basin, southwestern Ohio.
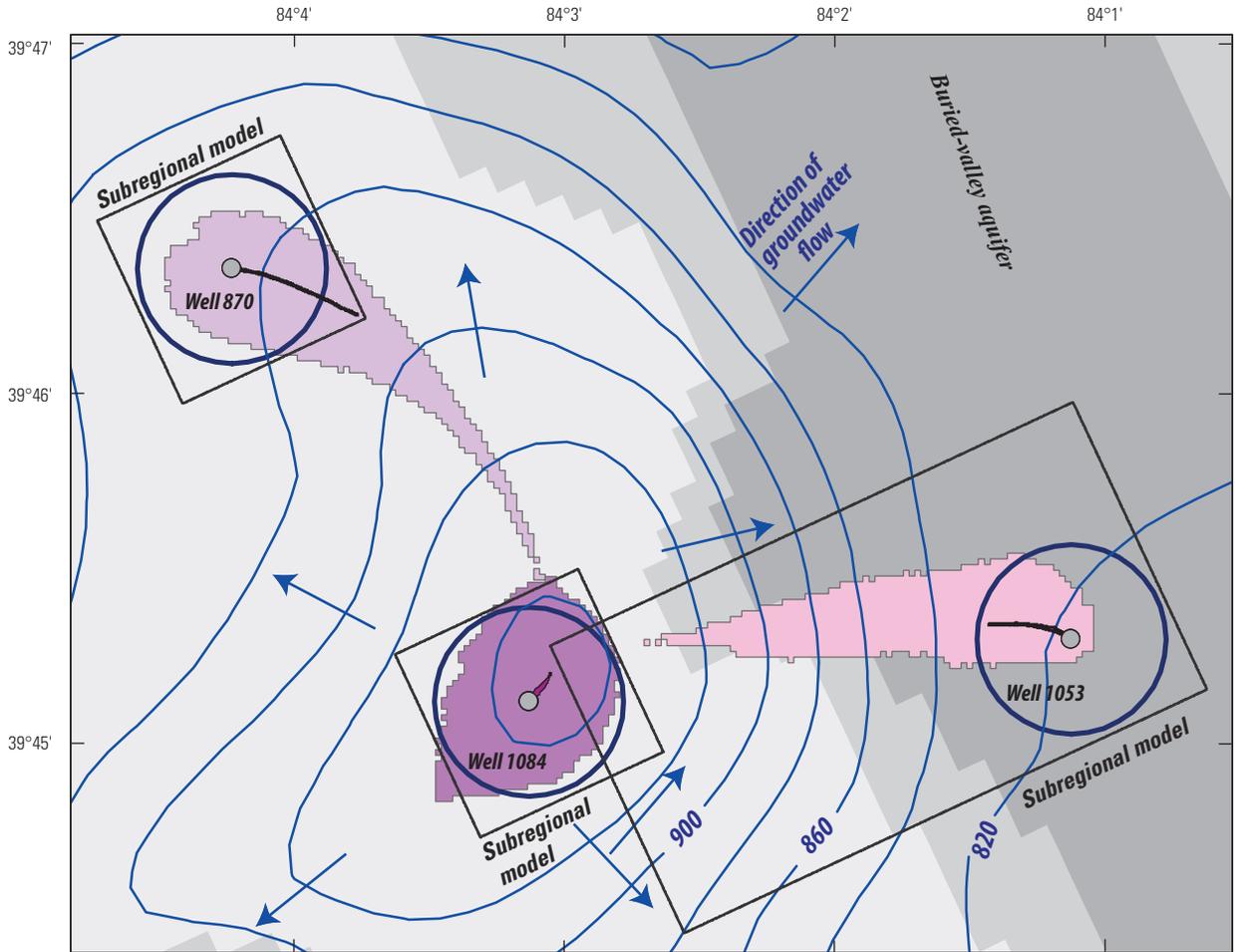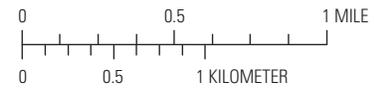
Figure 10. The 1,640-ft-radius circular buffer and steady-state contributing recharge areas for intermediate and low pumping rates in three hydrologic settings, southeastern part of the Great Miami River Basin, southwestern Ohio.

recharge under low-pumping conditions is within 500 ft of the well (fig. 10). Low-pumping contributing recharge areas in areas of steep hydraulic gradients at wells 870 and 1053 extend 2,430 and 1,560 ft, respectively, upgradient from the well. About 33 percent of the low-pumping contributing recharge area at well 870 is outside of the 1,640-ft-radius circular buffer (fig. 10).

Well Depth

Wells with screens deeper in the aquifer generally withdraw water with longer flow paths and have contributing recharge areas that extend farther upgradient than wells screened in shallower parts of the aquifer (figs. 11A and 11B). Under intermediate pumping conditions, in which the pumping rate is equivalent to the recharge rate through the corresponding buffer zone, and the contributing recharge area is of a similar size, the contributing recharge areas to pumped wells screened deeper in the aquifer (altitudes of the boundaries of layer 2 are 785 and 680 ft, and 680 and 638 ft for layer 3), corresponding to screen tops 34 and 139 ft below the water table, respectively) extend farther upgradient than the contributing recharge areas screened in the shallowest part of the aquifer (altitudes of the boundaries of layer 1 are 819 and 785 ft, corresponding to screen tops at the water table). The amounts of overlap between the contributing recharge areas for wells in layers 2 and 3 and the circular buffer were about 20 and 11 percent, respectively, compared to about 33 percent for a well screened in layer 1 (fig. 11A). For high pumping rates, contributing recharge areas for wells pumped in layers 1, 2, and 3, were essentially the same; the amounts of overlap with the circular buffer were 18.7, 17.6, and 16.6 percent, respectively (fig. 11A). At high rates, pumping changes regional gradients and creates strong gradients to the pumped well; under these conditions, well depth has a minimal effect.

The effect of well depth on a contributing recharge area is more pronounced for a low pumping rate (fig. 11B); wells pumped at low rates in deeper parts of the aquifer capture water from longer flow paths with recharge locations farther upgradient than for wells pumped in shallow parts of the aquifer because natural regional gradients are relatively unchanged as a result of pumping at low rates. The contributing recharge area to a pumped well near the water table (layer 1) extends about 1,420 ft upgradient from the well and is completely within the 1,640-ft-radius circular buffer. Contributing recharge areas to pumped wells in layer 2 and 3 extend 4,610 and 9,040 ft, respectively, from the well (fig. 11B). About 5.8 percent of the contributing recharge area to a well pumped in layer 2 is within the circular buffer for low pumping rates. For a well pumped in the deepest part of the aquifer (layer 3) with a screen altitude about 150 ft below the simulated water table, the contributing recharge area is outside the circular buffer; the nearest part of the contributing recharge area is about 2,910 ft from the edge of the circular buffer (fig. 11B). Because the mean depth of domestic-supply wells in the glacial aquifer system is 91 ft (Warner and Arnold, 2010), some

wells in the NAWQA monitoring network might have source areas outside of the 1,640-ft-radius circular buffer. Although all or most of the contributing recharge areas to wells pumped at a low pumping rate in layers 2 and 3 are outside of the circular buffer, they are within the contributing recharge areas to the wells pumped at an intermediate rate (that is, a rate equivalent to recharge through the circular buffer). This result indicates that the use of equivalent contributing recharge areas (and wells pumped at a rate equal to the volume of recharge through the circular buffer) is an improved representation of a spatial buffer for wells pumped at low rates.

Duration of Pumping

The assumptions for a steady-state contributing recharge area are that recharge and pumping are constant over time and that the two stresses are in a state of mass balance. However, at the start of pumping, the pumped well receives water primarily from storage within the aquifer framework. As the flow system adjusts to the pumping stress, and hydraulic conditions approach steady state, a mass balance between recharge and pumping is approached. Whereas the system may adjust hydraulically within a relatively short time (days or weeks), the contributing recharge area is defined by the process of advective transport over time scales of multiple decades. This transient effect causes the contributing recharge area to become a function of the duration of pumping. The particle methodology used to estimate transient contributing recharge areas involves a phased analysis in which particles are tracked in the direction of groundwater flow to a pumped well under transient flow conditions for a specified period of pumping; the resulting particle field is then tracked in reverse (opposite the direction of flow) to recharge locations under steady-state nonpumping conditions to estimate the transient contributing recharge area.

The contributing recharge area to well 1053 after 2 years of pumping generally is within, but about 56.1 percent the size of, the steady-state contributing recharge area (fig. 12A). This result indicates that, initially, storage in the aquifer is a significant source of water to the well. After 10 years of pumping, the contributing recharge area is about 87 percent of the size of the steady-state contributing recharge area, and after 20 years of pumping, the contributing recharge area is essentially the same size and generally within the steady-state contributing recharge area (fig. 12B). Although conditions are site specific, these results indicate that steady-state contributing recharge areas likely are reasonable approximations of the true contributing recharge areas for wells in operation for more than 10 to 20 years in similar hydrologic settings. This conclusion is consistent with the results of analyses of transient contributing recharge areas presented in Masterson and others (2002) and Masterson and others (2004). The results also indicate that contributing recharge areas to relatively new wells may not be sufficiently represented by steady-state conditions. It should be noted that the upgradient parts of the 2- and 20-year contributing

Figure 11. The 1,640-ft-radius circular buffer and steady-state contributing recharge areas to well 1053 for *A,* high and intermediate, and *B,* low and intermediate pumping rates for three well depths, southeastern part of the Great Miami River Basin, southwestern Ohio.
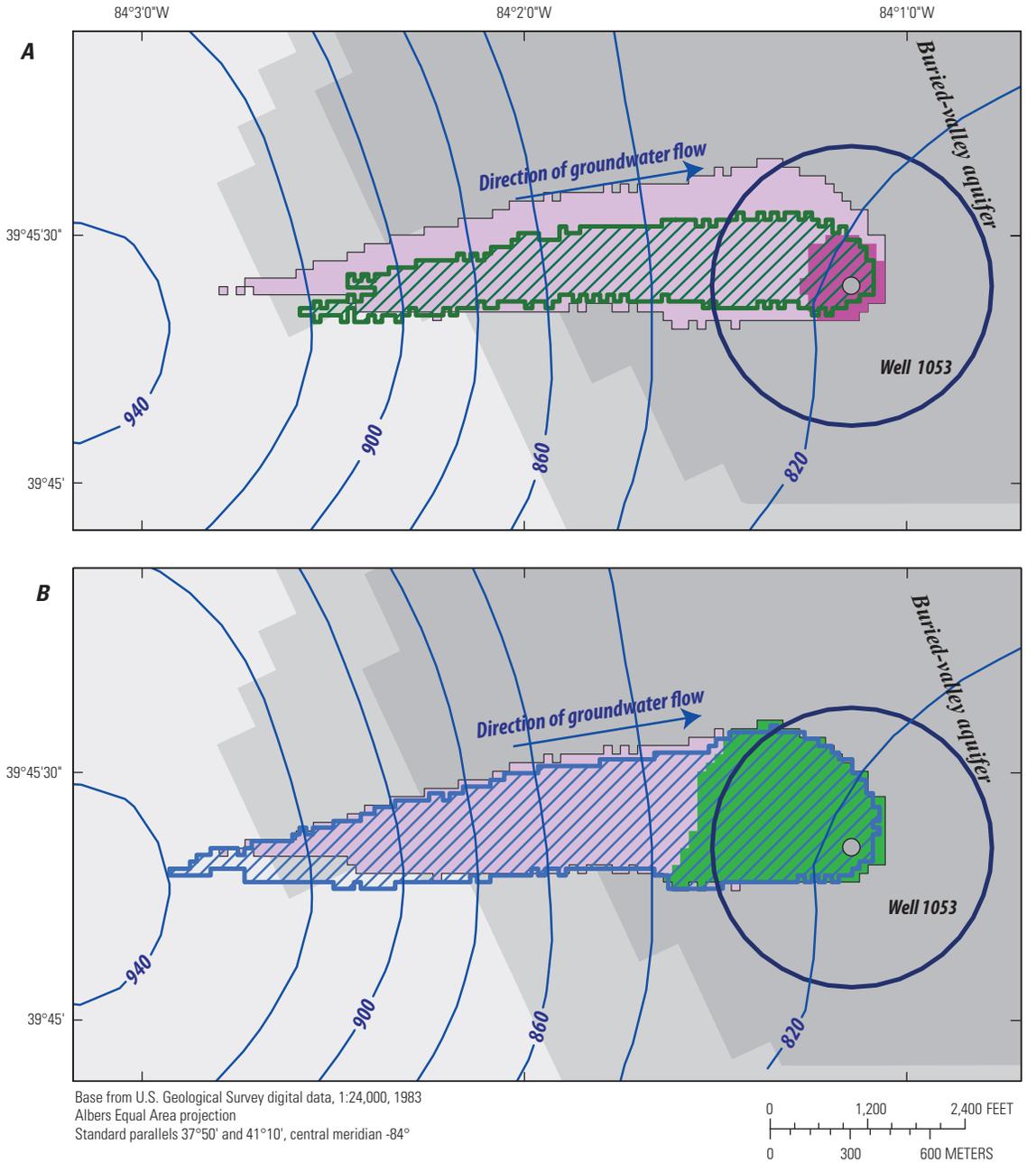
Figure 12.   Circular buffer of radius 1,640 feet, transient contributing recharge areas, steady-state recharge areas for intermediate pumping rates, and traveltimes under steady-state transport to well 1053 after *A,* 2 years and *B,* 20 years of pumping, southeastern part of the Great Miami River Basin, southwestern Ohio.

recharge areas are outside of the steady-state contributing recharge area; traveltimes within the contributing recharge area can exceed 100 years and, as a result, most of the advective transport to the well from the upper part of the contributing recharge area after 20 years of pumping is under steady-state nonpumping conditions, which produce hydraulic gradients that differ from those produced by pumping conditions.

Advective Transport of Nitrate

Statistical models can predict the probability of a contaminant exceeding a threshold concentration by utilizing relations between observed concentrations and a set of defined explanatory variables, whereas process models can be used to predict solute concentrations by simulating the physiochemical processes controlling the transport of a solute, such as nitrate, from surficial sources to receptors, such as pumped wells. The one-dimensional transport of a solute in the principal direction of groundwater flow can be described by the generalized equation:

$$\delta C/\delta t = -v\left(\delta C/\delta x\right) + D_L\left(\delta^2 C/\delta x^2\right) - \delta q/\delta t \qquad (2)$$

(Advection) (Dispersion) (Reactions),

where
   C is the concentration of a solute (moles per
               liter (M/L^3)),
   t is the time (t),
   x is the distance (L),
   v is the groundwater velocity (L/t),
   D_L is the longitudinal dispersivity (L), and
   Q is the concentration loss caused by
               reactions (M/L^3).

The first term refers to advection, which is the movement of a mass of solute with average groundwater velocity, and the second term—dispersion—refers to the spreading of mass resulting from scale-dependent aquifer heterogeneity. Advection and dispersion are physical processes, and the relative importance of the two is a function of hydrogeologic setting; advective transport generally is the dominant component of transport in permeable surficial aquifers such as the glacial aquifer system. The third term refers to the effect of chemical reactions on solute concentrations. An important biologically mediated chemical reaction affecting nitrate concentrations in surficial aquifers is denitrification, which is the transformation of nitrate to nitrogen gas under anaerobic conditions. Nitrate generally is conservative and mobile under the oxic conditions prevalent in the glacial aquifer system, including the Great Miami River Basin (Sheets, 2007).

Groundwater-flow models simulate cell-by-cell flows within an aquifer; particle-tracking methods use this information to simulate the movement of water through the aquifer. Solute-transport models use the cell-by-cell flows

to simulate physical transport processes—advection and dispersion—and can represent single-solute reactions such as sorption and decay. Reactive transport models use these same flows to represent advection and dispersion but can also simulate complex reactions among solutes, including chemical equilibrium and rate-dependent oxidation-reduction reactions like denitrification.

The advective transport of a conservative solute, such as nitrate in oxic conditions, to a well is a function of the spatial and temporal distribution of sources within the contributing recharge area to the well and traveltimes in the aquifer. Groundwater models can represent the advective transport of conservative solutes by using mass-weighted particle tracking—a method in which a mass of solute is assigned to particles, and the number and traveltimes of the particles are used to approximate time-varying loads or concentrations at critical receptors. Kauffman and others (2001) present a detailed analysis of the advective transport of nitrate in a coastal-plain aquifer in northern New Jersey by using mass-weighted particle tracking; the analysis presents the capabilities and limitations of using the method to represent the effect of transport processes and time-varying sources on nitrogen concentrations in pumped wells. Walter (2007) presents a comparison of mass-weighted particle tracking and solute-transport methods in estimating nitrogen loads and concentrations in wells and ecological receptors in a glacial aquifer on Cape Cod, Mass.

Even under the assumptions of no subsurface attenuation and limited dispersion, a number of considerations that relate to intrinsic aquifer properties and surface-source characteristics apply to the use of a groundwater-flow model to represent the advective transport of nitrate. Time-varying nitrate concentrations are affected by intrinsic aquifer properties, such as porosity and dispersivity. Groundwater velocities and therefore traveltimes are proportional to the porosity of aquifer sediments and, although differing porosities result in the same steady-state concentration, the temporal pattern of nitrate concentrations in a pumped well varies with porosity. Also, groundwater models and mass-weighted particle tracking do not account for the effects of dispersion arising from aquifer heterogeneities. Dispersion causes nitrate to arrive at pumping wells sooner than it would through advective transport alone. Steady-state concentrations arising from a continuous surface source are similar for simulations that both do and do not account for dispersion (Walter, 2007) because the largest component of dispersion is in the direction of groundwater flow (longitudinal). Surface-source characteristics include source history and biologic and chemical processes that affect nitrate concentrations in water recharging the aquifer at the water table. Changes in land use and associated changes in surface sources of nitrate affect the source term used in models simulating nitrate transport and thus the estimates of time-varying nitrate concentrations in a pumped well.

Processes near the land surface and in unsaturated sediments above the water table, such as uptake by plants

and denitrification, also affect the source terms used to estimate time-varying nitrate concentrations in pumped wells. Uptake of nitrogen from applied fertilizer varies by crop; for cereal crops, the average nitrogen uptake efficiency (NEU) is about 33 percent (Raun and Johnson, 1999) and can exceed 50 percent for corn (Bundy and Andraski, 2005). In addition to plant uptake, five additional processes—soil denitrification, volatilization, gaseous plant emissions, subsurface tile drains, and surface runoff—can cause additional near-surface losses of nitrogen (Raun and Johnson, 1999). The remaining nitrogen that is not lost through these processes leaches into the ground and potentially enters the aquifer at the water table with recharge. Although nitrate in a glacial aquifer may be transported conservatively under oxic conditions, substantial losses of nitrogen—exceeding 30 percent—can occur in association with reducing conditions in the unsaturated zone prior to recharge at the water table (Weiskel and Howes, 1992). Because these processes have an effect on the source term used to simulate the advective transport of nitrate to a pumped well, understanding those processes is critical. Although subsurface processes affecting nitrogen concentrations, such as dispersion and biochemical transformation, can be explicitly simulated by using solute-transport and reactive-transport models, surficial processes can be implicitly represented by altering source terms to reflect those processes. Some implicitly represented surficial processes are analogous to spatial variables of potential use in statistical models of nitrate occurrence.

Traveltimes

Groundwater age is an important consideration in evaluating the vulnerability of a well to contamination from surface sources of nitrogen; groundwater age has a control on the occurrence of nitrogen as well as the geochemical conditions that contribute to nitrogen contamination (Warner and Arnold, 2010). Traveltimes through an aquifer from recharge locations to pumped wells are a function of the depths of the flow paths; deeper flow paths generally are associated with longer traveltimes (fig. 13A). Steady-state traveltimes within a contributing recharge area, which are proportional to the porosity of aquifer sediments, can be defined by using particle tracking. Steady-state traveltimes in pumped wells in the Great Miami River Basin range from essentially instantaneous to more than 100 years; the longest traveltimes are for water recharging the upgradient parts of the contributing recharge areas, where flow paths are longer and deeper than for water recharging closer to the pumped well (fig. 13B). Maximum steady-state traveltimes to wells pumped in three hydrologic settings at a rate equivalent to the volume of recharge through the corresponding 1,640-ft-radius circular buffer exceeded 150 years, but median traveltimes to wells 870, 1084, and 1053 were about 17.6, 24.8 and 36.6 years, respectively. The three wells have a similar distribution of traveltimes (fig. 14A); the longest traveltimes are from upgradient parts of the contributing recharge areas.

Steady-state traveltimes are valid after a well has been in operation, and the resulting hydraulic gradients have been established for a time longer than the longest traveltime within the contributing recharge area, which can exceed 200 years. In most cases, pumped wells in the glacial aquifer system have been in operation for traveltimes that are shorter than the maximum steady-state traveltimes within the contributing recharge area. At well 1053, the transient contributing recharge area approximates the steady-state contributing recharge area in less than 20 years (fig. 12B); however, transient traveltimes are longer than those under steady-state pumping conditions (fig. 14B). Steady-state traveltimes reflect steeper hydraulic gradients around the well caused by pumping and therefore are faster than under less steep, natural gradients. The median traveltimes within the contributing recharge area after 2, 10, and 20 years of operation were about 23, 25, and 41 years, respectively (fig. 14B). As a result, the distributions of traveltimes after 30, 40, and 50 years are similar; the median is about 50 years compared to a median of 36.6 years under steady-state conditions. Transient contributing recharge areas for a given period of pumping are much larger than for the equivalent steady-state traveltime band; for 2 and 20 years of operation, the corresponding 2- and 20-year steady-state contributing areas were 15.9 and 40 percent of the area of the transient contributing recharge area (fig. 12).

Steady-state traveltimes are a function of the depth of the pumped well and the pumping rate. At an intermediate pumping rate (equivalent to the rate of recharge through the corresponding circular buffer), traveltimes are longer for wells screened deeper in the aquifer because the wells capture water from longer flow paths. Median traveltimes for a well pumped in layers 1, 2, and 3 are 36.6, 50.3, and 58.2 years, respectively (fig. 15). The increased gradient causes the well to capture water from flow paths associated with a large range of traveltimes—from essentially instantaneous to more than 150 years. The traveltimes for wells pumped in layers 1, 2, and 3 overlap over most of the ranges of traveltimes (fig. 15) because pumping alters regional hydraulic gradients and steepens gradients to the well; this result indicates that the recharge areas correspond to a similar range of traveltimes for wells pumped at different depths. Pumping has less effect on regional hydraulic gradients at low pumping rates, and traveltimes reflect natural regional age patterns. As a result, traveltimes have smaller ranges and less overlap between wells pumped from different depths (fig. 15); this result indicates that recharge areas for wells pumped at different depths correspond to different ranges of traveltimes. For a shallow pumped well (layer 1), traveltimes are less than 35.2 years with a median of 14.9 years, compared to maximum and median traveltimes of 188.3 and 36.6 years, respectively, for an intermediate pumping rate. Median traveltimes for pumped wells in layers 2 and 3 are 93.5 and 196.9 years, respectively, for low pumping rates (fig. 15), larger than those of 50.3 and 58.3 for the intermediate pumping rate.
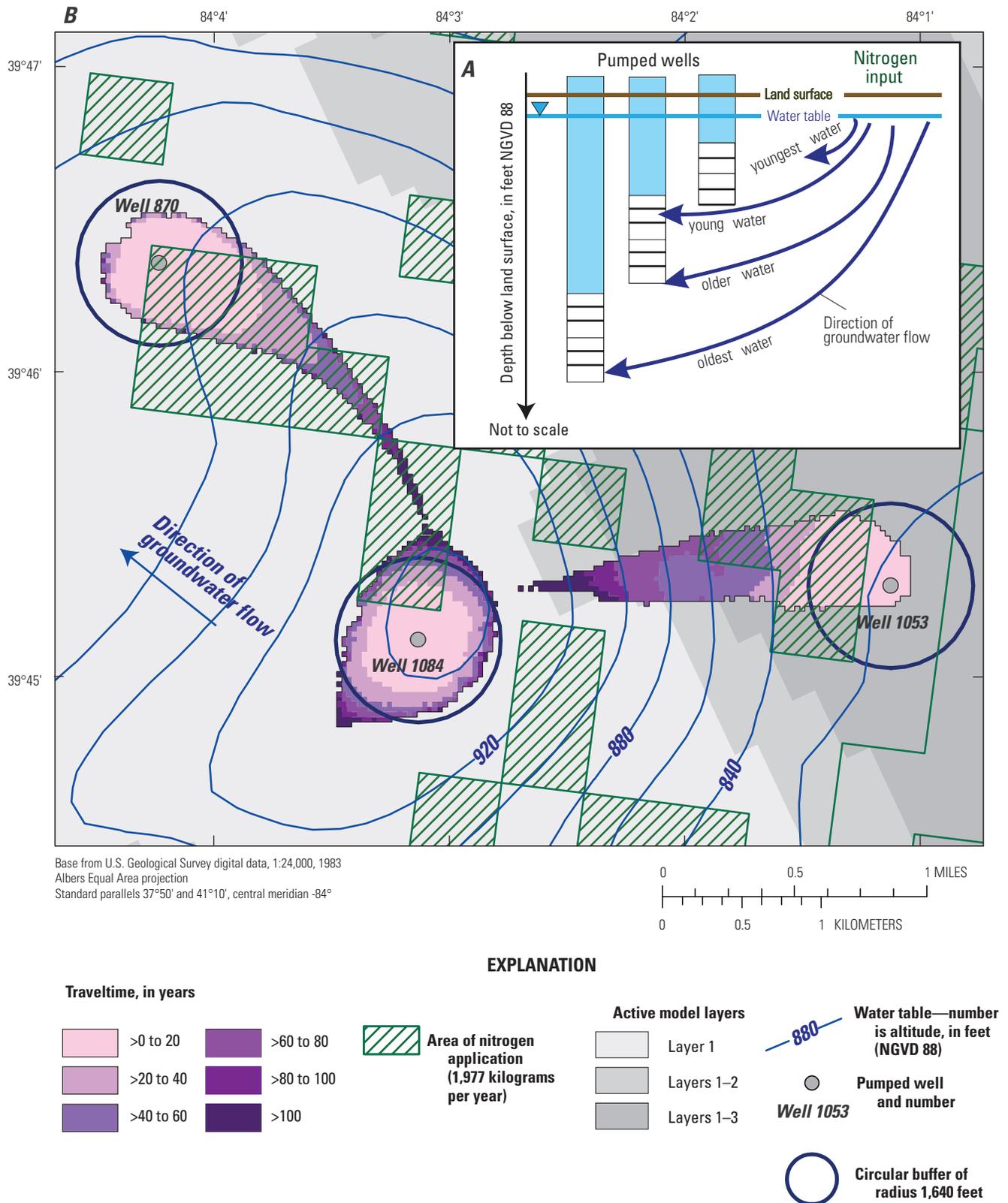
Figure 13. *A,* Relation between well depth and traveltime; and *B,* map showing 1,640-ft-radius circular buffers and steady-state traveltimes within contributing recharge areas to pumped wells in three hydrogeologic settings and areas of nitrogen application from fertilizer, southeastern part of the Great Miami River Basin, southwestern Ohio.
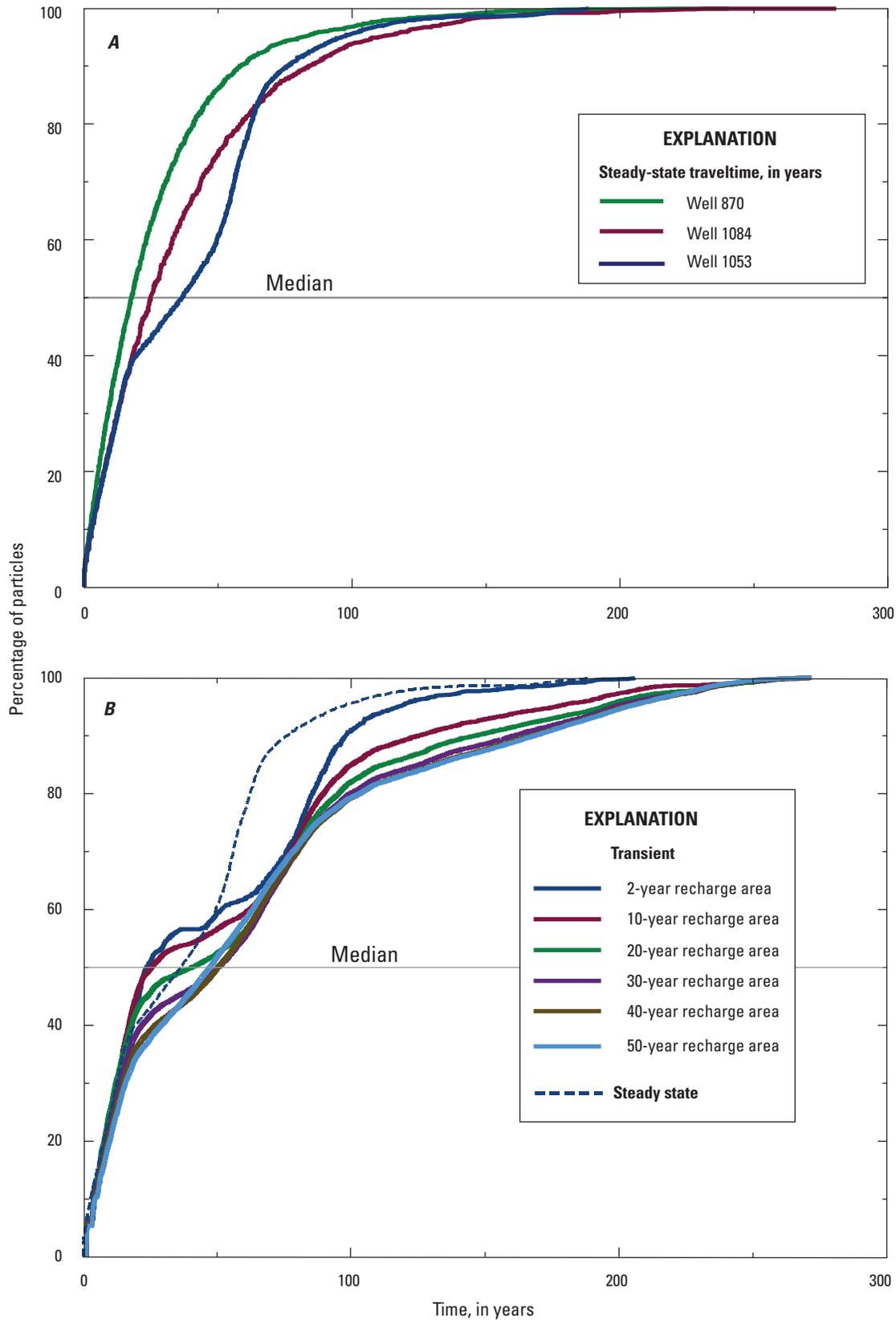
Figure 14.   *A,* Steady-state traveltimes to pumped wells in three hydrologic settings; and *B,* transient traveltimes after different periods of pumping by well 1053, southeastern part of the Great Miami River Basin, southwestern Ohio.
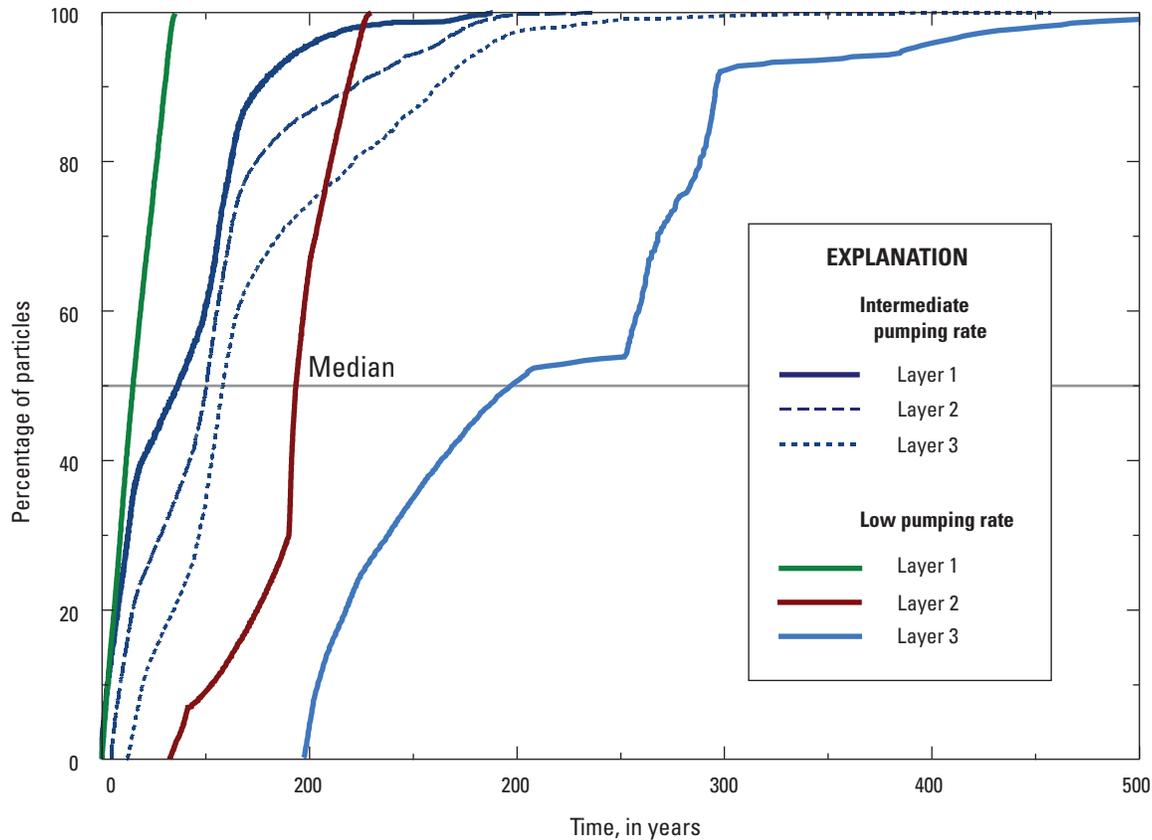
Figure 15. Steady-state traveltimes for well 1053 pumped at different depths and at different pumping rates, southeastern part of the Great Miami River Basin, southwestern Ohio.

Transport of Nitrate

If nitrogen is transported conservatively in the form of nitrate, as under oxic conditions in the glacial aquifer system, groundwater-flow models and mass-weighted particle tracking can be used to simulate the advective transport of nitrate and estimate time-varying concentrations at a pumped well. The use of mass-weighted particle tracking entails assigning a mass of nitrate to a set of particles and tracking those particles forward to pumped wells. Adding the associated masses for different particle traveltimes can allow approximation of the time-varying load of nitrate to a well; the pumping rate of the well can be used to convert the loads to concentrations. The method requires a uniform spatial distribution of particles at the water table and a defined source term to determine mass weights. Nitrogen concentrations in a pumped well are a function of several factors, including traveltimes between recharge locations at the water table and the well and the distribution of sources of nitrogen at the land surface.

Ten-year mean nitrogen-application rates (1992–2001) reported in Warner and Arnold (2010) were assembled from county usage data and applied to agricultural land identified in the NLCDe92 database. Application rates in the Great Miami River Basin, in kilograms per year (kg/yr)

(fig. 4B), are average annual rates applied 500-meter land-use parcels. The nitrogen-application rate on agricultural land in the area around hypothetical wells 870, 1084, and 1053 was 1,977 kg/yr (fig. 13B). It should be noted that statistical and process models are based on a steady-state application rate, which does not take into account complex, time-varying source terms. Agricultural land composed about 70.6, 20.3, and 47.0 percent of the total contributing recharge area to wells 870, 1084, and 1053, respectively (fig. 13B). Nitrate concentrations estimated under steady-state pumping conditions increase steadily until a steady-state concentration is reached after a transport time equal to the maximum traveltime within the contributing recharge area, which can exceed 100 years (fig. 13B). Under the assumptions of no attenuation of nitrogen prior to recharge and conservative transport, steady-state nitrate concentrations in wells 870, 1084, and 1053 were 34.6, 9.9, and 26.9 mg/L as N, respectively (fig. 16A). The concentrations are within the range reported for monitoring wells in the glacial aquifer system but exceed the 90th percentile (Warner and Arnold, 2010). The steady-state concentrations were reached after 65 to 70 years of transport and reflect the proportions of agricultural land and the distribution of traveltimes within the contributing recharge areas (fig. 13B).
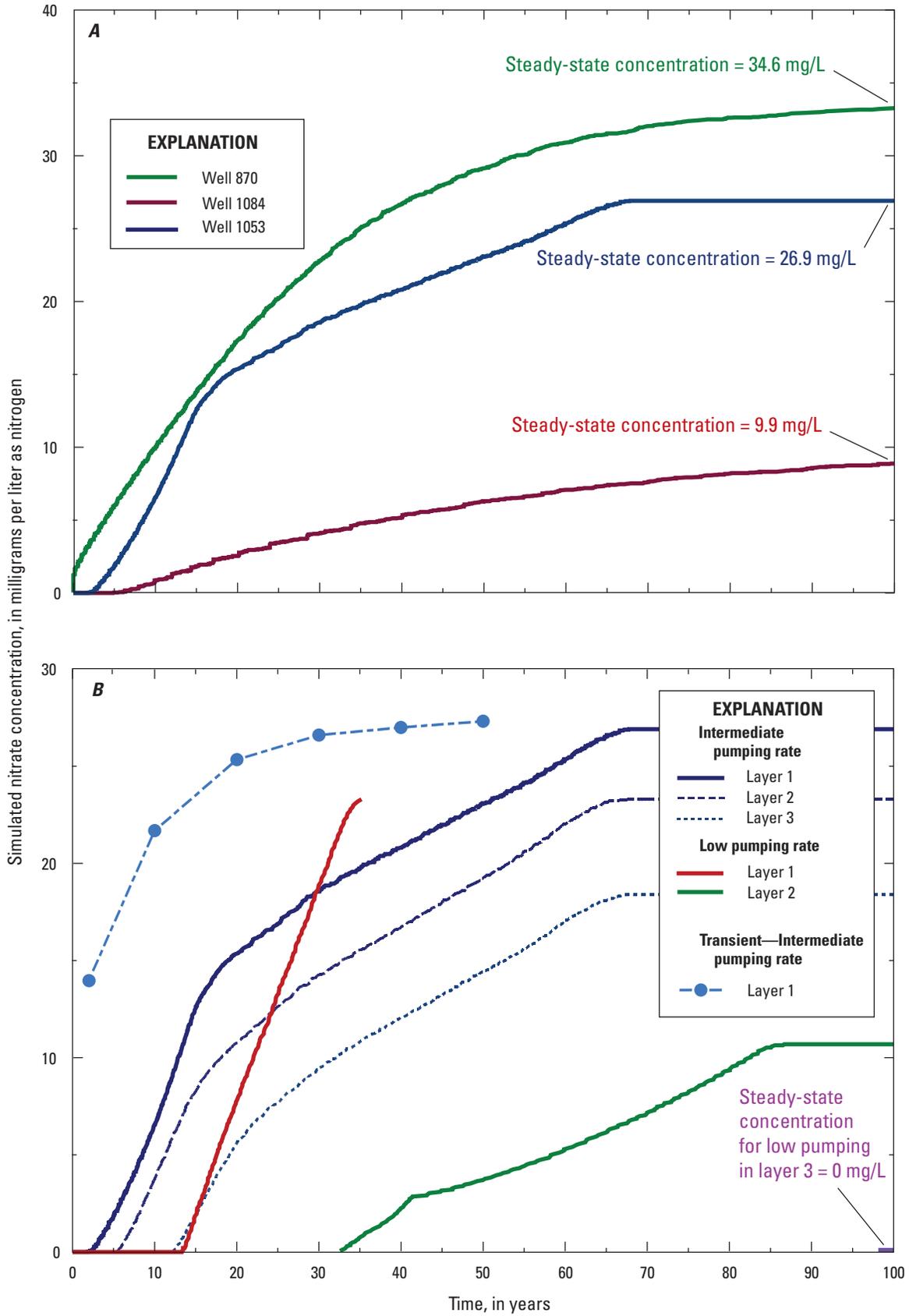
Figure 16. Time-varying nitrate concentrations for *A,* wells in three hydrologic settings and *B,* well 1053 for three well depths and low and intermediate pumping rates.

Estimated nitrate concentrations are affected by the same factors as the simulated contributing recharge areas and traveltimes, including well depth and pumping rates. At intermediate pumping rates, steady-state nitrate concentrations in well 1053 pumped in model layers 1, 2, and 3 were 26.9, 23.3, and 18.4 mg/L as N, respectively (fig. 16B). The contributing recharge areas to the well shift upgradient for pumping in successively deeper model layers (fig. 11), resulting in less agricultural source area within the contributing recharge areas. Steady-state concentrations under low pumping rates differ from those under higher pumping rates and are more affected by well depth (fig. 16B). Nitrate concentrations in a shallow well (layer 1) pumped at a low rate steadily increased to a concentration of 23.3 mg/L as N over the range of total traveltime—about 35 years—and likely may be near steady state (fig. 16B). Nitrate concentrations in a well pumped at a low rate deeper in the aquifer (layer 2) reached a steady-state concentration of 10.7 mg/L as N after about 86 years of transport (fig. 16B). Concentrations in a well pumped at low rate in the deepest part of the aquifer (layer 3) were zero over the range of transport times. Wells pumped at a low rate have less effect on regional gradients, and therefore regional hydraulic gradients and traveltimes have a larger effect on predictions of nitrate concentrations.

Most pumped wells are in operation for less time than the maximum steady-state traveltime within the contributing recharge area; as a result, concentrations from transient simulations of traveltimes (fig. 15) could differ from steady-state simulated concentrations. The method used to estimate transient contributing recharge areas results in a nonuniform particle field that makes assigning mass weights problematic. The transient contributing recharge areas after 2, 10, 20, 30, 40, and 50 years were used to implicitly estimate the concentration in the pumped well at those times on the basis of the average nitrogen load for particles in the contributing recharge area and the total volume of recharge. Using the same method for steady-state contributing recharge areas confirmed that the results do match steady-state concentrations simulated by using mass-weighted concentrations. Nitrate concentrations under transient transport initially are higher and increase faster than those estimated under assumptions of steady-state transport but approach the same steady-state concentration after sufficiently long transport times (fig. 16B). The nitrate concentration in well 1053, pumped at an intermediate rate for 2 years in a shallow part of the aquifer, was 13.9 mg/L as N compared to 0 for the nitrate concentration estimated after 2 years of steady-state transport. After 20 years, estimated concentrations under transient and steady-state flow conditions were 25.3 and 15.4 mg/L as N, respectively. The assumption of transient flow accounts for the fact that, initially, a pumped well may capture older water with nitrate transported under natural hydraulic gradients from recharge locations beneath existing sources, whereas the assumptions of steady-state flow are that the well has been in operation for a time equivalent to the maximum traveltime from the contributing recharge

area and that there initially is no mass of nitrate in the aquifer. Changes in land use or the rate of nitrogen application onto agricultural land would affect estimated nitrogen concentrations; the effect would be a function of the timing of the resultant changes in the source term and the traveltimes to the pumped well.

Estimates of time-varying concentrations from groundwater-flow models representing advective transport are affected by a number of factors, including aquifer porosity and surface-source characteristics. Groundwater velocities and therefore traveltimes are proportional to the effective porosity of aquifer sediments, which generally range from 0.2 to 0.4 for unconsolidated sediments (Freeze and Cherry, 1979). Increasing porosity causes slower velocities and longer traveltimes. For well 1053, increasing the porosity from 0.2 to 0.3 increases the amount of time required to reach the same steady-state concentration in a shallow pumped well from about 68 to about 100 years (fig. 17). The source term used to represent inputs of nitrate into the aquifer and the transport of nitrate to a pumped well also is affected by changes in the rate of application of nitrate at the land surface as well as the transformation and loss of nitrate between the land surface and the water table. If nitrogen in agricultural areas near well 1053 were assumed to be applied for only 20 years, the steady-state concentration of 15.3 mg/L as N would be reached in 20 years; this concentration is less than the steady-state concentration of 26.9 mg/L as N derived from an assumed continuous source over the entire period of steady-state pumping (fig. 17). Also, assuming a loss of 50 percent of nitrate prior to recharge to the water table from plant uptake and biochemical processes results in a steady-state concentration of 13.4 mg/L as N, which is half the concentration under the assumption of no loss (fig. 17). An assumed nitrogen loss of 50 percent or greater prior to recharge likely is more realistic than the assumption of conservative transport from land surface to the water table.

Comparison of Statistical and Process Models

Differences in predictions of nitrate occurrence by statistical and process models could arise from two sources. The first relates to differences in the assumed source of water to a pumped well—circular buffers (statistical models) or contributing recharge areas (process models)—that is used to estimate the values of the explanatory variables for which relations to nitrate occurrence have been defined. The second relates to the two different modeling approaches: predictions that utilize observed relations between nitrate occurrence and a set of explanatory variables (statistical models) and those that explicitly represent surface sources and transport processes (process models).
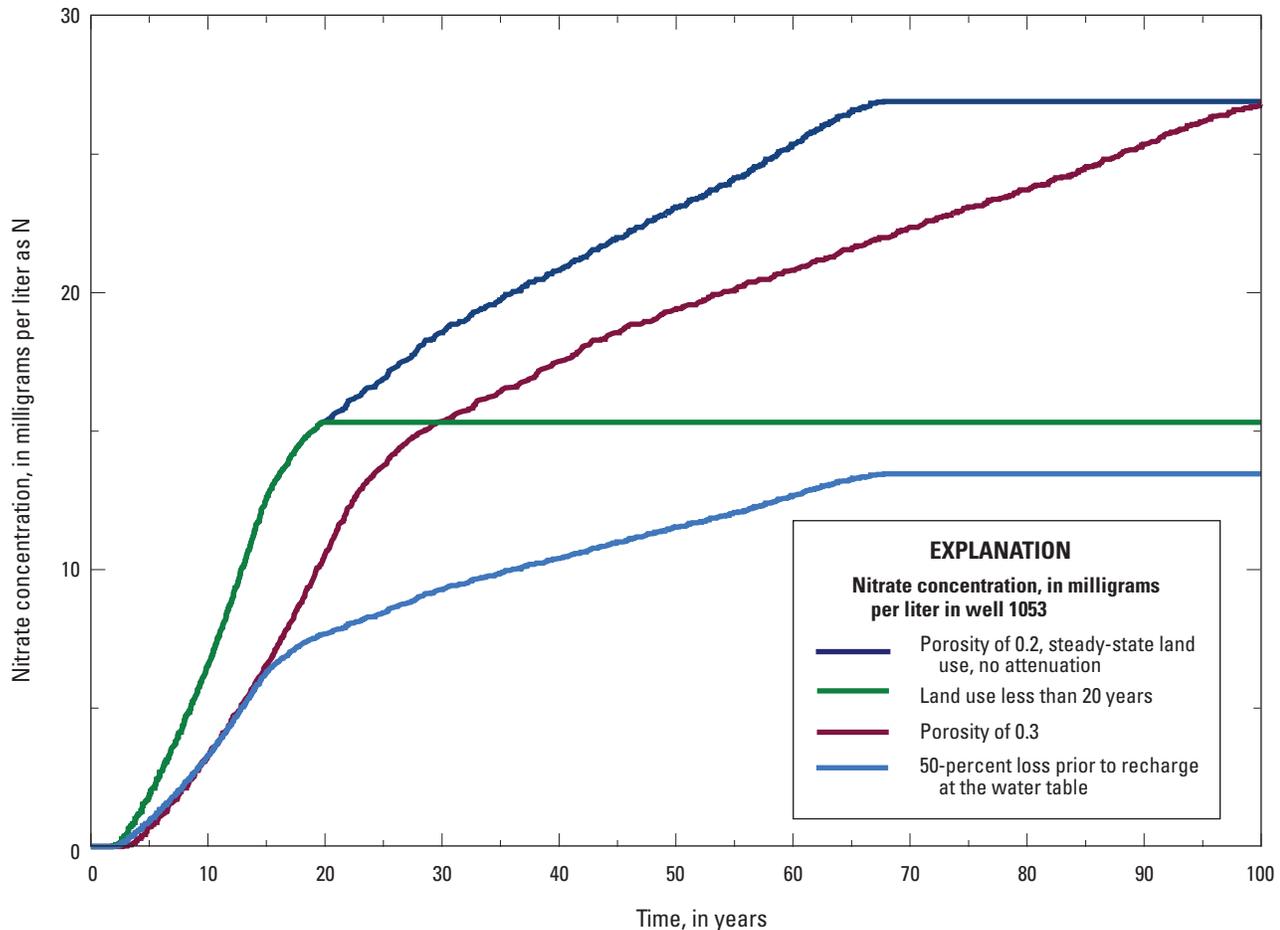
Figure 17.   Time-varying nitrate concentrations at well 1053 for different porosities and source-term characteristics under steady-state transport for intermediate pumping rates.

Predictions of Nitrate Occurrence by Statistical and Process Models at Selected Wells

The statistical models of nitrate occurrence presented in Warner and Arnold (2010) utilize relations between observed nitrate concentrations and a set of explanatory variables to predict the probability that nitrate concentrations at a well will exceed a given concentration. The set of variables includes environmental variables and a source variable defined as an area-weighted mean within 1,640 ft (500 m) of the pumped well. The assumption is that the circular buffer approximates the source of water to the well; this assumption does not account for local groundwater-flow conditions and can result in area-weighted explanatory variables and resultant model predictions that differ from those determined for the actual source area. The source of water to the well, or contributing recharge area, is a function of several factors, including pumping rate, screen depth, and hydrologic setting (figs. 10–12), and therefore area-weighted explanatory variables and model predictions would be affected by the same three factors.

The 1,640-ft-radius circular buffer around well 1053 differs from the contributing recharge area to the well pumped at a rate equivalent to the volumetric rate of recharge through the circular buffer (fig. 18). The volume-equivalent contributing recharge area is similar in size but has a different shape, extends upgradient from the circular buffer, and results in different estimates of explanatory variables. The distribution of mean 10-year nitrogen-application rates estimated from county data and environmental variables (silt content, alfisol content, and depth to water) estimated from STATSGO data (figs. 4A and 4B) differ between the circular buffer and volume-equivalent contributing recharge area (fig. 18). As an example, the area-weighted mean nitrogen-application rate within the circular buffer and contributing recharge area were 928 and 779 kg/yr, respectively. The area-weighted silt content, alfisol content, and mean depth to water were 49.4 percent, 82.1 percent, and 4.64 ft, respectively, within the circular buffer and 51.1 percent, 84.5 percent, and 4.27 ft respectively, within the volume-equivalent contributing recharge area. Differences in

EXPLANATION

 Area of nitrogen-fertilizer application
(1,977 kilograms per year)

 Alfisol content of 85 percent, silt content of 51.56 percent,
and depth to water of 4.18 feet

 Area of no nitrogen fertilizer application—Alfisol content
of 81 percent, silt content of 48.55 percent, and
depth to water of 4.84 feet

 Circular buffer of radius of 1,640 feet

Well 1053

 Pumped well and number

Nitrogen-application rate, in kilograms per year;
Alfisol and silt content, in percent; and depth to
water, in feet

Recharge area

 0 kg/yr, 85%, 51.56 %, 4.18 ft

 1,997 kg/yr, 85%, 51.56 %, 4.18 ft

 0 kg/yr, 81%, 48.55%, 4.84 ft

Circular buffer

 0 kg/yr, 81%, 48.55%, 4.84 ft

 1,977 kg/yr, 81%, 48.55%, 4.84 ft

 1,997 kg/yr, 85%, 51.56%, 4.18 ft

Figure 18.   Circular buffer of radius 1,640-ft and equivalent contributing recharge area to well 1053 and spatial distribution
of explanatory variables used to estimate nitrate occurrence, Great Miami River Basin, southwestern Ohio. Ten-year mean
nitrogen-application rate derived from county data, and environmental variables—alfisol content, silt content, and depth to
water—derived from STATSGO data.

estimated source and environmental variables between circular buffers and contributing recharge areas are affected by the same factors that affect the simulated contributing recharge area; these factors include pumping rate, screen depth, and hydrologic setting.

Role of Pumping Rate and Well Depth

The contributing recharge area to a well is affected by the rate at which the well is pumped and the depth of the well screen; the size of the contributing recharge area is proportional to the pumping rate, and wells screened deeper in the aquifer have contributing recharge areas that extend farther upgradient than shallow-screened wells. Generally, the effect of screen depth on the contributing recharge area is greater for wells pumped at low rates because natural regional hydraulic gradients are a more important control on the contributing recharge area (fig. 13). At a given location, the distributions of spatial variables and the area-weighted means of these variables in the contributing recharge areas differ for each combination of pumping rate and screen depth (fig. 19).

Estimates of Area-Weighted Spatial Variables

Area-weighted mean variables at well 1053 vary for different simulated pumping rates from three screen depths (fig. 20A) corresponding to screen midpoints of about 17, 86, and 162 ft below the water table. For a high pumping rate—equivalent to the mean pumping rate of public-supply wells in the basin—values of environmental variables (alfisol and silt content and depth to water) are similar for shallow, intermediate, and deep well screens (fig. 20A). The ranges for all three area-weighted environmental variables are less than 0.5 percent of the averages for the three well-screen depths. These results are consistent with the similarity in simulated contributing recharge areas for a well pumped at a high rate at different depths (fig. 19). For an intermediate pumping rate—equivalent to the volume of recharge through the corresponding 1,640-ft-radius circular buffer—the ranges of the three area-weighted environmental variables for different well-screen depths were between 0.6 and 2 percent of the average. With an assumed low pumping rate of 350 gal/d, which is similar to the withdrawal rate from a domestic-supply well, ranges in area-weighted environmental variables for the three well-screen depths were between 2 and 7 percent of the average. The coefficients of variation—defined as the ratios of the variances to the means—for silt content, alfisol content, and depth to water were 0.007, 0.07, and 0.004, respectively, for the different combinations of pumping rates and screen depths. Estimates of the 10-year mean nitrogen-application rate (1992–2001) within contributing recharge areas varied more (fig. 20A) with a coefficient of variation of 321.8 for different combinations of pumping rates and screen depths. The ranges of area-weighted nitrogen-application rates for three different well-screen depths and high and intermediate pumping rates were about 1.6 and 33.7 percent, respectively,

of the average value. Screen depth had a greater effect on area-weighted estimates of nitrogen-application rates at low pumping rates because natural gradients have a greater effect on flow patterns, and contributing recharge areas vary more with depth (fig. 19). The area-weighted mean nitrogen-application rates for wells pumped in shallow, intermediate, and deep parts of the aquifer were 1,111; 525; and 0 kg/yr, respectively (fig. 20A).

Area-weighted spatial variables for contributing recharge areas and different combinations of pumping rates and screen depths all differed from those for the corresponding 1,640-ft-radius circular buffer around well 1053 (fig. 20A). Area-weighted alfisol and silt content were lower and depth to water higher within the circular buffer than in the contributing recharge areas for all combinations of pumping rates and screen depths. The mean nitrogen-application rate within the circular buffer was in the range of variation of values from different combinations of pumping rates and screen depths (fig. 20A).

Predictions of Nitrate-Exceedance Probabilities

The effects of variations in area-weighted spatial variables on predictions of nitrate occurrence by logistic regression are functions of the explanatory variables used in a model, the scale of variability of the area-weighted explanatory variables, and the weight of the explanatory variables in the logistic regression models indicated by the slope coefficient (B_n) in equation 1. In addition to the explanatory variables estimated from spatial data, the statistical models include well-construction variables (fig. 3A): depth to and length of the well screen and the diameter of the well. In this analysis, a well diameter of 7 in., a screen length of 7.8 ft—corresponding to the medians for private-supply wells used to develop the statistical models (Warner and Arnold, 2010)—and a well depth of 10 ft were assumed; the well depth of 10 ft is similar to the median of the simulated depth to water of 7 ft in the basin and consistent with the depths of wells screened in shallow parts of the aquifer (layer 1). For purposes of comparison, consistent well-construction variables are used so that differences in predicted probabilities are a function of only differences in area-weighted explanatory variables for simulated source areas. The statistical models (Warner and Arnold, 2010) were developed based on an assumed 1,640-ft-radius circular buffer as the source area to the well; these models include spatial variables estimated in the same manner. Although spatial variables estimated from contributing recharge areas were used in this analysis, the results provide insight into the degree to which improving representation of the source area would affect the statistical model development and performance.

The probabilities of exceeding threshold concentrations of 1, 2, 3, and 4 mg/L as N in well 1053 were predicted to be 49, 41, 33, and 21 percent, respectively, by using median values of the area-weighted spatial variables for the circular buffer in the logistic regression models presented in Warner

Figure 19.   Circular buffer of radius 1,640-ft and equivalent contributing recharge areas to well 1053 for different combinations of well depths and pumping rates, and spatial distribution of explanatory variables used to estimate nitrate occurrence, Great Miami River Basin, southwestern Ohio. Ten-year mean nitrogen-application rate derived from county data, and environmental variables—alfisol content, silt content, and depth to water—derived from STATSGO data.

Figure 20.   *A,* Estimates of area-weighted explanatory variables and *B,* the probability of nitrate occurrence above threshold concentrations of 1 to 4 milligrams per liter (mg/L) as nitrogen by statistical models for the 1,640-ft circular buffer and different combinations of well depth and pumping rate at well 1053.

and Arnold (2010) (fig. 20B). Logistic regression models for predicting exceedances of 1, 2, and 3 mg/L as N do not include the 10-year nitrogen-application rate as an explanatory variable. The predicted probabilities based on the circular buffer around the well were higher than predicted probabilities based on area-weighted statistical variables estimated from contributing recharge areas for different combinations of pumping rates and screen depths (fig. 20B). The highest variability in probabilities for different screen depths was at low pumping rates. The coefficients of variation for the probabilities of nitrate exceeding 1, 2, 3, and 4 mg/L as N were less than 0.003. The exceedance probabilities varied less than the values of the area-weighted spatial variables (figs. 20A and B) because the probabilities are functions of both spatial and well-construction variables, which only partially control the variability in predictions of nitrate occurrence.

Role of Hydrologic Setting

The importance of well location on estimates of area-weighted spatial variables for contributing recharge areas arises from two sources: (1) the hydrologic setting of the pumped well, and (2) the areal distribution of spatial characteristics upgradient from the well. The hydrologic setting of a well affects the shape of the contributing recharge area and therefore the areal distribution of characteristics defined by land-use or soil data layers upgradient from the well. The interaction of hydrologic setting and the distribution of spatial variables results in unique estimates of area-weighted spatial variables for a given set of well characteristics (location, pumping rate, and depth).

Wells in three hydrologic settings—shallow and deep saturated sediments in areas of relatively steep hydraulic gradients (well 870 and 1053) and shallow saturated sediments near a groundwater divide (well 1084)—are associated with differing areal distributions of spatial characteristics within circular buffers and contributing recharge areas (fig. 21). When a well near a groundwater divide is pumped at a rate equivalent to recharge through the corresponding circular buffer, the recharge area approximates the circular buffer, and the two likely would have similar spatial characteristics. Contributing recharge areas to similar pumped wells in areas of steep hydraulic gradients extend upgradient from the well, beyond the 1,640-ft-radius circular buffer (fig. 21), and likely have spatial characteristics that differ from the corresponding circular buffers.

Estimates of Area-Weighted Spatial Variables

Estimates of area-weighted spatial variables differ for wells pumped in the three different locations and hydrologic settings at intermediate and low pumping rates (fig. 22A). The circular buffers and contributing recharge areas for wells 870 and 1084 are within the same areas of defined environmental characteristics from the STATSGO database (silt content,

alfisol content, and depth to water) (fig. 21), so there is no variability among estimates of area-weighted mean values for these two areas (fig. 22A). Contributing recharge areas and the circular buffer for well 1053 encompass two different regions of defined environmental characteristics resulting in some variability in estimates of area-weighted mean values (fig. 22A).

The circular buffers and contributing recharge areas of all three wells encompass two regions of defined 10-year mean nitrogen-application rates. The coefficients of variation of estimated area-weighted mean nitrogen-application rates for wells 870, 1084, and 1053 were 292.5, 243.3, and 29.5, respectively. Well 870 had the most variable and highest mean area-weighted nitrogen-application rate (1,363.0 kg/yr) (fig. 22A). The area-weighted mean nitrogen-application rate ranged from 715.6 kg/yr for the circular buffer to 1,977 kg/yr for the contributing recharge area at low pumping rates. About 36 percent of the circular buffer overlaps an area with a mean nitrogen-application rate of 1,977 kg/yr—the remaining part of the circular buffer overlaps an area with no nitrogen application—whereas the entire contributing recharge area to the well for low pumping rates is within the area of nitrogen application (fig. 21). Conversely, the contributing recharge area to well 1084 for low pumping rates is almost completely within an area with no nitrogen application, and the circular buffer overlaps an area with nitrogen application (fig. 21).

Predictions of Nitrate-Exceedance Probabilities

Probabilities of exceeding nitrate concentrations of 1, 2, 3, and 4 mg/L as N at the three wells by using circular buffers and contributing recharge areas for low and intermediate pumping rates are a function of the estimated spatial variables (fig. 22B). Probability predictions were the same for the circular buffers and the contributing recharge areas for low and intermediate pumping rates for wells 870 and 1084 (fig. 22B) because the three logistic regression models for nitrate exceedances of 1, 2, and 3 mg/L as N used the same estimated values of the three environmental spatial variables for both the circular buffers and the contributing recharge areas (fig. 22A). Estimated spatial variables for the contributing recharge areas and the circular buffer for well 1053 and the resulting predictions of nitrate exceedances as N are more variable (fig. 22B), but the coefficients of variation of the predictions for the three models are less than 0.003.

The logistic regression model for nitrate exceedance of 4 mg/L as N includes the mean 10-year nitrogen-application rate as an explanatory variable, which is the variable with the largest coefficient of variation for circular buffers and contributing recharge areas for the three wells (fig. 22A). The predicted probabilities of exceeding 4 mg/L as N range from 14 to 21 percent (fig. 22B); the coefficient of variation of all the predictions of exceeding 4 mg/L as N was about 0.004. The three well locations include a range of spatial variables, particularly nitrogen application rate (0 to 1,977 kg/yr). Probabilities of nitrate exceeding a given

Figure 21. Circular buffer of radius 1,640-ft and steady-state contributing recharge areas to wells in three hydrologic settings for intermediate and low pumping rates, and spatial distribution of explanatory variables used to estimate nitrate occurrence, southeastern part of Great Miami River Basin, southwestern Ohio.

Figure 22. *A,* Estimates of area-weighted explanatory variables and *B,* predictions of nitrate occurrence above threshold concentrations from 1 to 4 milligrams per liter as nitrogen by using statistical models for wells in three hydrologic settings for intermediate and low pumping rates.

threshold concentration generally showed less variability than the estimated spatial variables because the predicted probabilities are functions of the same sets of well-construction variables.

Predictions of Nitrogen-Occurrence Probabilities for Public-Supply and Network Wells

A total of 309 pumping centers—each with 1 or more wells—are represented in the regional model of the Great Miami River Basin (Sheets, 2007); these wells withdraw about 18.4 Mgal/d from the glacial aquifer, primarily around the city of Dayton, Ohio (figs. 2 and 6). Spatial variables estimated for 1,640-ft-radius circular buffers for these pumping centers differed to varying degrees from estimates for corresponding steady-state contributing recharge areas to these centers; adjusted R^2 values ranged from 0.27 to 0.44. The highest correlations were for mean nitrogen-application rate (R^2=0.41) and silt content (R^2=0.44) (fig. 23A). Spatial variables for circular buffers estimated from the STATSGO database—silt content, alfisol content, and depth to water— were statistically different than those for contributing recharge areas. Comparisons of estimated variables for circular buffers and contributing recharge areas by a paired t-test (Helsel and Hirsch, 1992) showed that silt content was lower (p was less than 0.001), and alfisol content and depth to water were higher (p was less than 0.001), within circular buffers (fig. 23A), Nitrogen-application rates estimated for circular buffers and contributing recharge areas were not statistically different.

The logistic-regression models presented in Warner and Arnold (2010) and spatial variables estimated for circular buffers and corresponding contributing recharge areas were used to predict the probability of exceeding nitrate concentrations of 1 to 4 mg/L as N; well-construction data were used to determine nonspatial explanatory variables. Predictions of the probability of exceeding 1 to 4 mg/L as N generally were correlated for the two sets of spatial-variable estimates (fig. 23B). Adjusted R^2 values for probabilities of exceeding 1 to 4 mg/L as N were 82 and 94 percent, respectively. Models of nitrate exceedance of 1 to 4 mg/L as N are useful for comparison because (1) the 1 mg/L as N includes only environmental variables from the STATSGO database, and (2) 4 mg/L as N is the lowest threshold concentration that includes all spatial variables—all of the environmental variables and nitrogen-application rate. Predictions of exceeding 1 mg/L as N generally were greater for the set of spatial variables estimated from the circular buffers than those estimated from the contributing recharge areas (fig. 23B). This result is consistent with differences observed in area-weighted mean values estimated from the STATSGO database. The slope coefficients of silt content, alfisol content, and depth to water for the prediction model for 1 mg/L as N are 0.0356, -0.0164, and 0.7219, respectively (table 1), indicating that lower silt content and higher depth to water contribute to higher probabilities of exceedance. The

large slope coefficient for depth to water indicates that the probabilities predicted for the circular buffers (fig. 23B) could be higher because of the higher area-weighted depth to water within the circular buffers.

A network of 1,290 wells was used to evaluate water quality in the glacial aquifer system (DeSimone and others, 2009); the network included public- and private-supply wells and monitoring wells used to develop statistical models of nitrate occurrence (Warner and Arnold, 2010). A total of 55 network wells were within the modeled area of the Great Miami River Basin (fig. 6A). Spatial data for 1,640-ft-radius circular buffers and corresponding contributing recharge areas for network wells—at a pumping rate equivalent to the volume of recharge through the circular buffer—show different trends than do spatial data from public-supply wells: silt content generally is higher (p is less than 0.001) inside circular buffers than in the corresponding contributing recharge areas (fig. 24A), and alfisol content and depth to water are lower. The probabilities of exceeding 1 and 4 mg/L as N in water from wells inside circular buffers and corresponding contributing recharge areas are highly correlated with adjusted R^2 values of 0.80 and 0.88, respectively (fig. 24B), similar to predictions for public-supply wells. However, predictions of exceeding 1 and 4 mg/L as N generally are higher (p less than 0.001) by using spatial variables for circular buffers than for contributing recharge areas (fig. 24B). The results show that, although depth to water generally is lower within circular buffers, predicted probabilities of exceeding 1 mg/L as N are higher. The results from individual wells show that large differences in alfisol content contributed to the higher predicted probabilities. The fact that different trends in estimated values of spatial variables but similar trends in predicted probabilities were observed in the two sets of wells underscores the hypothesis that physical processes likely do not underlie the observed trends. One possible reason could be the chosen locations of the public-supply wells, such as the siting of wells in areas of sandy soils or in areas near rivers, where hydraulic gradients are steeper. The results show that representing source areas as circular buffers or contributing recharge areas can greatly affect estimates of spatial variables and probability predictions, but that the effects are site specific.

Basin-Scale Comparisons of Statistical and Process Models

Differences in estimates of area-weighted spatial variables and the resultant predictions of nitrate occurrence by using circular buffers and contributing recharge areas are functions of several factors, including (1) hydrologic setting, which in turn is a function of aquifer geometry and properties and hydraulic boundaries and gradients; and (2) areal differences in the values of explanatory variables derived from spatial data (fig. 4), which can affect the resultant predictions
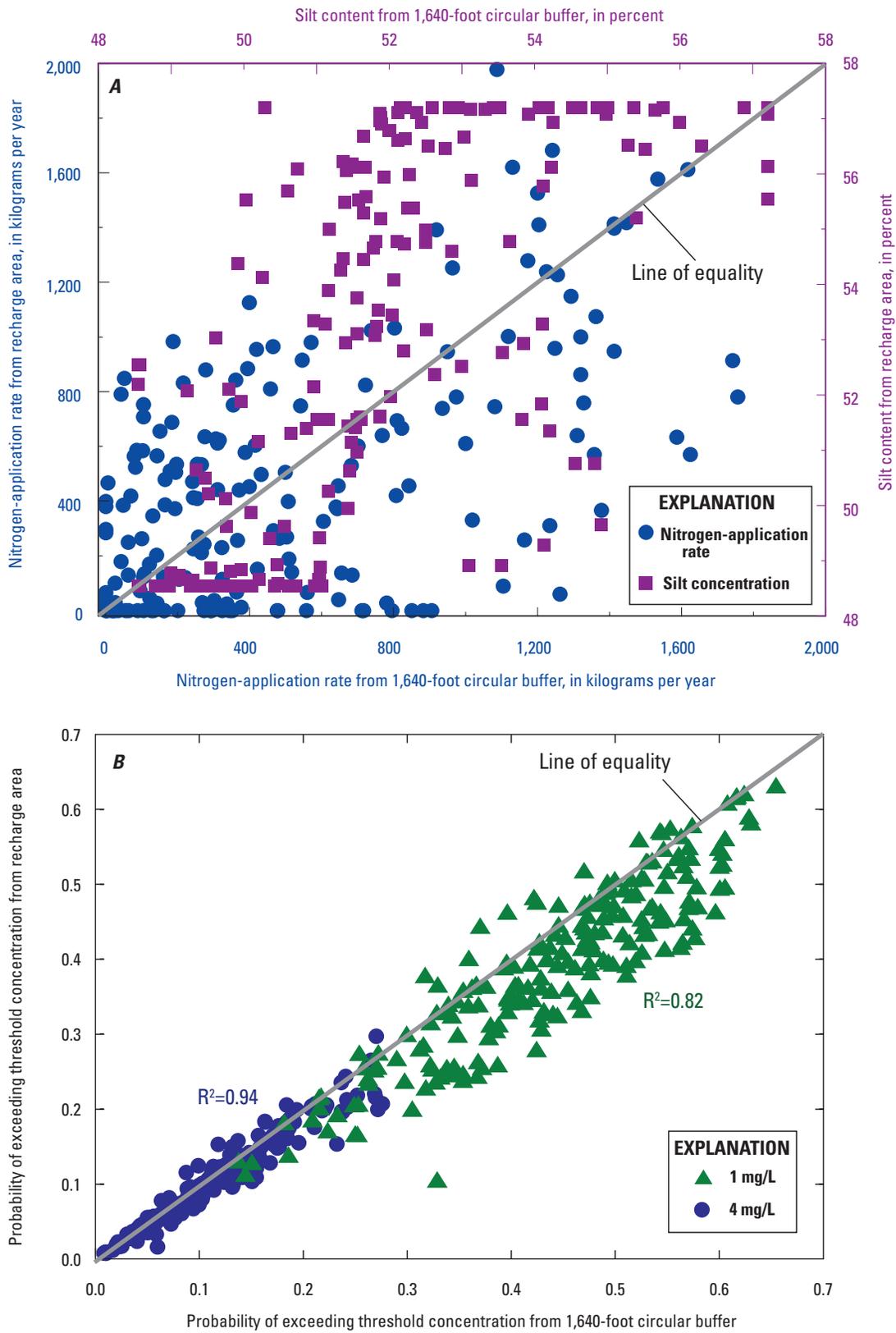
Figure 23. *A,* Estimates of area-weighted explanatory variables and *B,* predictions of nitrate occurrence above threshold concentrations of 1 to 4 milligrams per liter as nitrogen determined by using contributing recharge areas and circular buffers for production wells, Great Miami River Basin, southwestern Ohio.
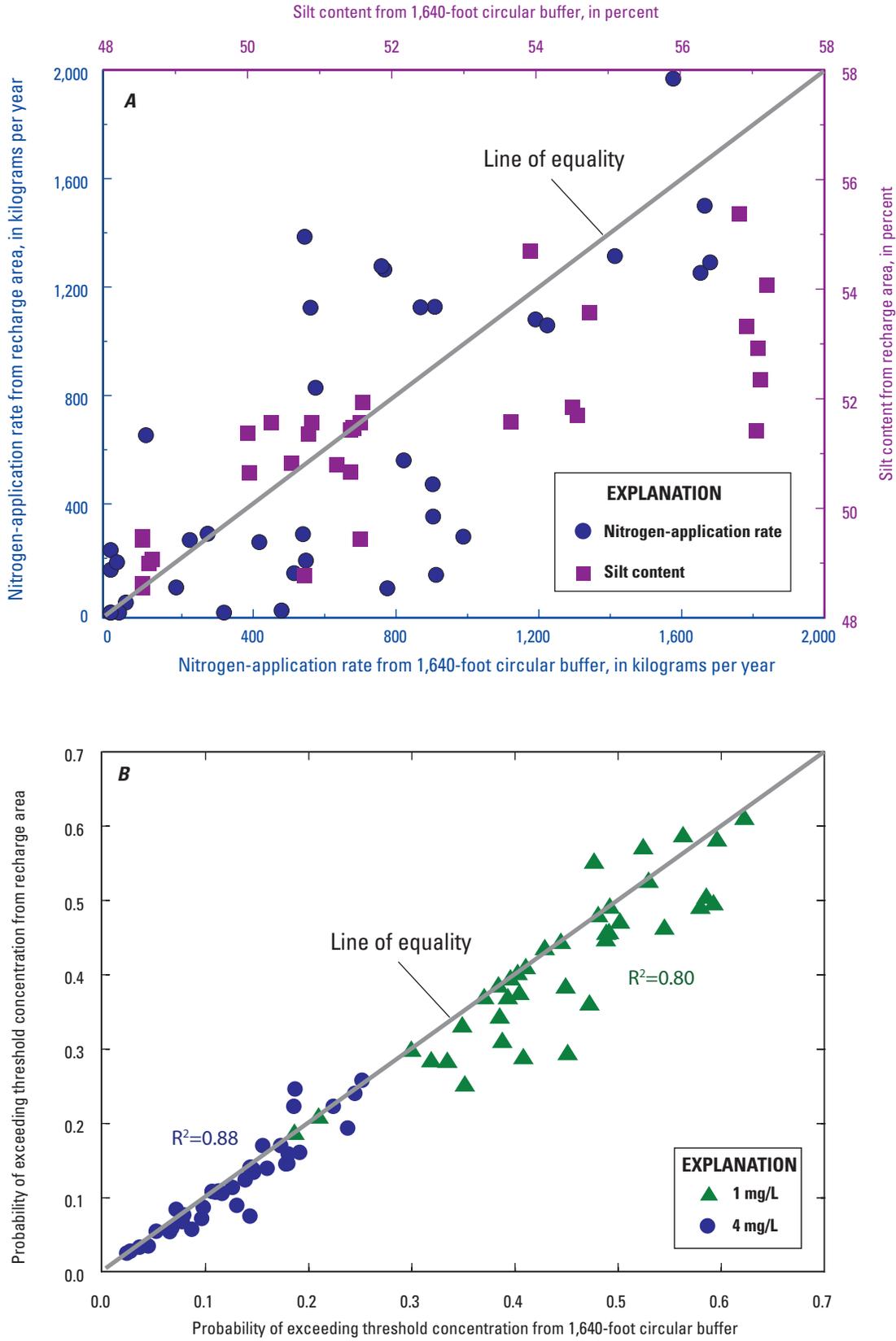
Figure 24.   *A,* Estimates of area-weighted explanatory variables and *B,* predictions of nitrate occurrence above threshold concentrations of 1 and 4 milligrams per liter as nitrogen determined by using recharge areas and circular buffers for network wells in the Great Miami River Basin, southwestern Ohio.

of nitrate occurrence. Contributing recharge areas simulated by process models account for regional groundwater-flow conditions and better represent spatial characteristics in the source area to a well than do circular buffers.

The effect of hydrologic setting on the correlations between estimates of spatial variables and predictions of nitrogen occurrence in contributing recharge areas and circular buffers were evaluated at the basin scale within the Great Miami River Basin. Estimates of spatial variables and predictions of nitrate occurrence were made for 21,774 hypothetical wells spaced at regular intervals of 500 ft. Wells within 1,640 ft of the edge of the active model area were excluded from the analysis so that the spatial variables for the circular buffers, like the simulated contributing recharge areas, would represent only the active modeled area (fig. 4).

At each location, the area-weighted means of explanatory spatial variables—mean 10-year nitrogen application, silt content, alfisol content, depth to water, and volumetric rate of recharge —were estimated for a 1,640-ft-radius circular buffer around the well. Volumetric rates of recharge through the 1,640-ft-radius circular buffers ranged from 10,550 to 23,589 ft^3/d (about 0.08 to 0.18 Mgal/d) (fig. 25) and were spatially correlated with modeled recharge to the aquifer (fig. 8B). A well was then pumped in layer 1 of the model at a rate equivalent to the volume of recharge through the circular buffer. In cases where the model went dry in layer 1, the well was pumped in successive layers until the layer around the well remained saturated. In cases where no solution could be obtained, the well was discarded. A total of 854 simulations either went dry or did not yield a solution; this outcome was common where saturated sediments were thin, such as near bedrock uplands. Only the hypothetical well of interest was pumped during each simulation so that natural gradients were represented to the extent possible; pumping at rates typical of the recharge rates through corresponding circular buffers generally did not alter natural gradients. The relative areal overlap—defined as the ratio of the portion of the contributing recharge area within the circular buffer to the total area of the contributing recharge area—was determined for each well site.

For each of the 20,920 successful model runs, the contributing recharge area to the well and area-weighted means of the spatial explanatory variables were determined. The sets of estimated spatial variables from contributing recharge areas and circular buffers were used to predict probabilities of nitrate exceeding threshold concentrations of 1 to 10 mg/L as N by using the logistic regression models in Warner and Arnold (2010). Well-construction variables were the same in all predictions, so that differences in the probabilities were functions only of differences in spatial variables which, in turn, are functions of the differences between circular buffers and volume-equivalent contributing recharge areas. A well diameter of 7 in. and screen length of 7.8 ft were used in the statistical models, consistent with

the mean values for private-supply wells in the NAWQA monitoring network (Warner and Arnold, 2010); a screen depth of 10 ft—consistent with the screen depth for a shallow pumped well, but less than the mean for private-supply wells (about 91 ft)—was also used. Assuming a shallower well depth would result in larger probabilities of nitrate exceedance for a given set of spatial variables, because well depth has a negative slope coefficient and is inversely related to predicted probabilities (table 1). However, on the basis of results given earlier in this report, trends in differences between estimates of spatial variables and resultant predictions for probabilities estimated for circular buffers and contributing recharge areas would be similar. In addition to estimates of area-weighted spatial variables and predictions of nitrate occurrence, mass-weighted particle tracking was used to estimate time-varying nitrate concentrations arising from fertilizer application on agricultural areas within the contributing recharge area to each of the 20,920 successfully modeled wells.

Comparison of Circular Buffers and Contributing Recharge Areas as the Assumed Sources of Water to Wells

The potential effects of hydrologic setting on estimates of spatial variables for a source area to a pumped well can be quantified by using the overlap between a circular buffer around a well and the simulated contributing recharge area to the well pumped at a rate equivalent to the volumetric rate of recharge through the circular buffer. In areas where hydraulic gradients are nearly flat, such as near groundwater divides, contributing recharge areas are nearly circular and generally overlap the circular buffer, whereas wells pumped in regions with steeper regional hydraulic gradients have elongated contributing recharge areas that overlap less of the circular buffer (figs. 21, 26). At three locations in the southeastern part of the basin, the amount of overlap ranged from 85.6 percent in well 1084, which was near a regional groundwater divide, to 32.7 percent in well 1053, which was in an area of steep hydraulic gradients (fig. 26).

The amount of overlap between circular buffers and corresponding contributing recharge areas in the basin ranged from 2 to 99 percent (fig. 27). Amounts of overlap generally were larger near upland margins, near groundwater divides and away from discharge areas, although basin-scale spatial trends were variable. The mean and median overlaps were 28 and 32 percent, respectively (fig. 28). The results indicate that, at a basin scale, contributing recharge areas differ substantially from circular buffers, and large differences in estimated spatial variables and resulting predictions are likely.

Base from U.S. Geological Survey digital data, 1:24,000, 1983
Albers Equal Area projection
Standard parallels 37°50' and 41°10', central meridian -84°

EXPLANATION

Recharge through 1,640-foot circular buffers, in cubic feet per day

24,000

10,000

Area not modeled
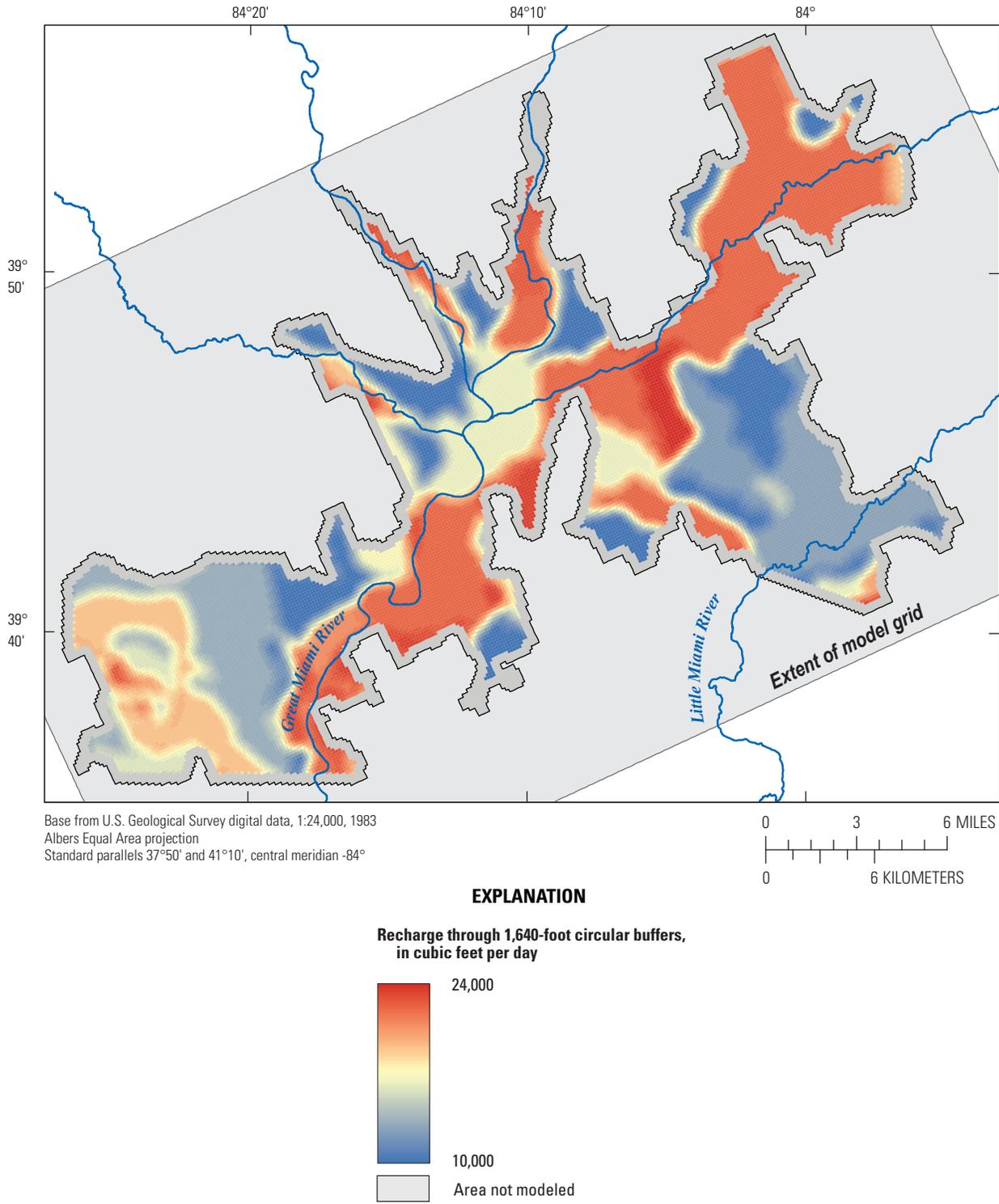
Figure 25.   Volumetric recharge rates through circular buffers for hypothetical wells uniformly spaced at 500-foot intervals across the Great Miami River Basin, southwestern Ohio.
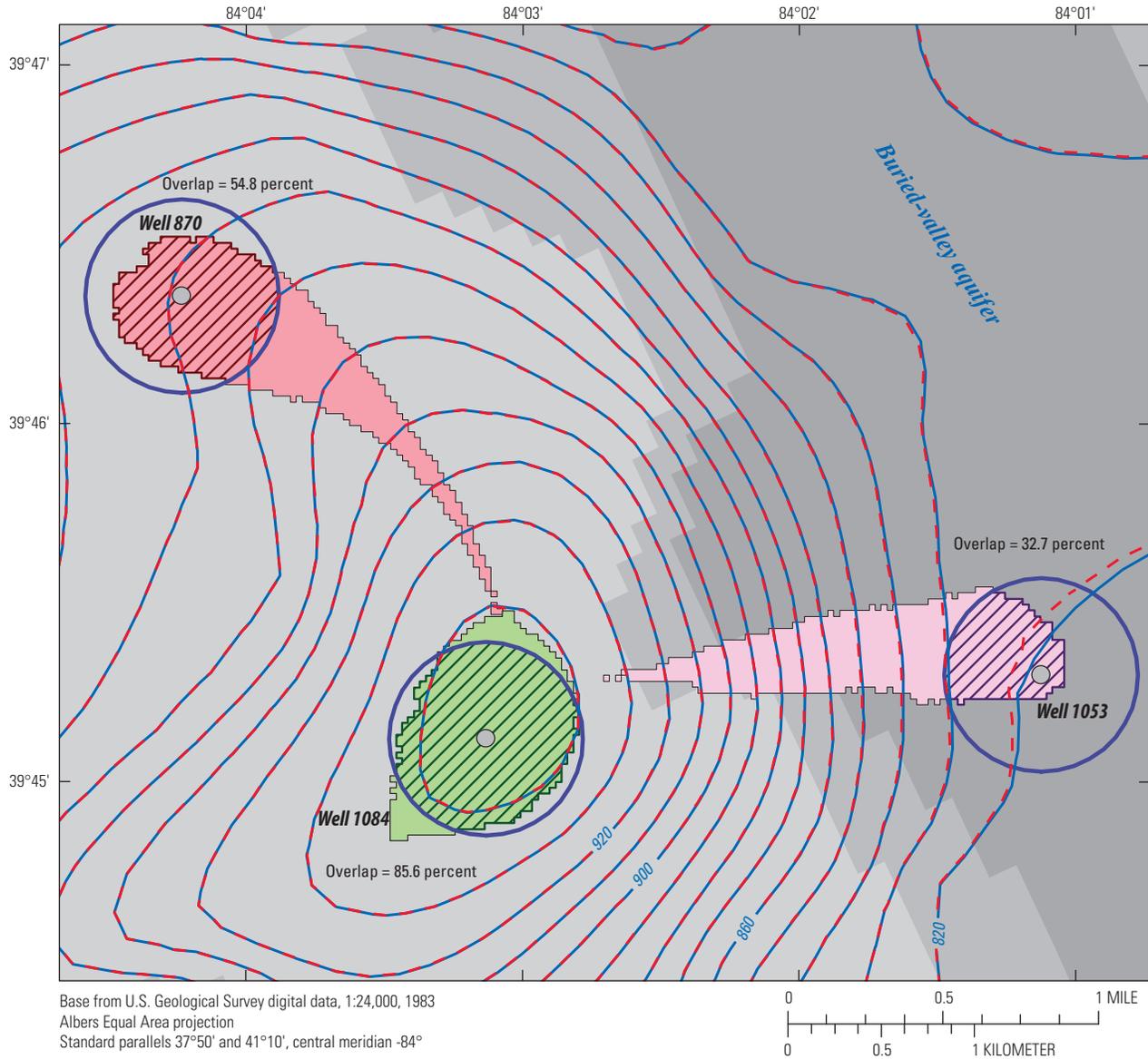
Figure 26. Overlap between contributing recharge areas and circular buffers in three hydrologic settings, southeastern part of the Great Miami River Basin, southwestern Ohio.
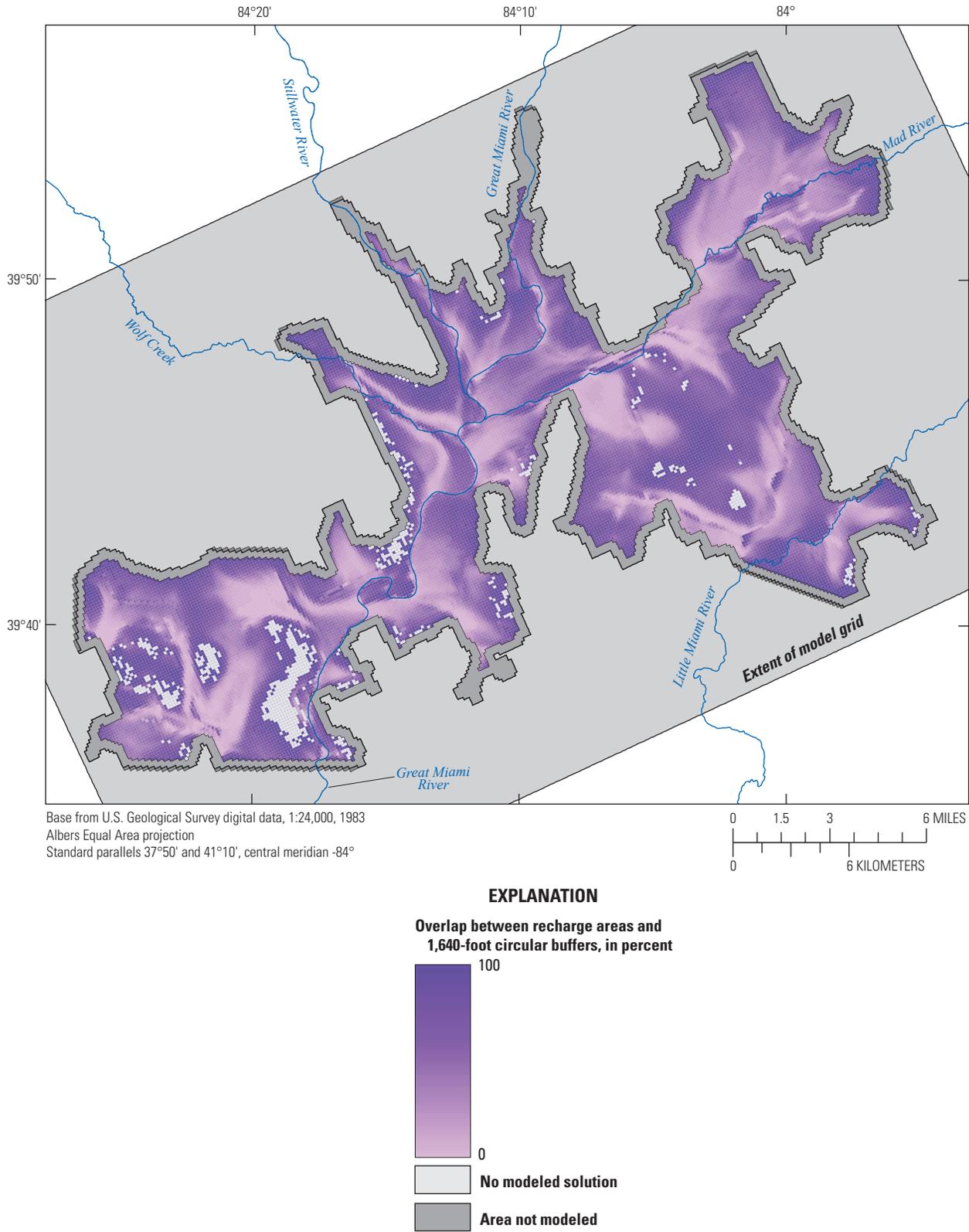
Figure 27.   Overlap between contributing recharge areas and circular buffers for 20,920 hypothetical wells uniformly spaced at 500-foot intervals across the Great Miami River Basin, southwestern Ohio.
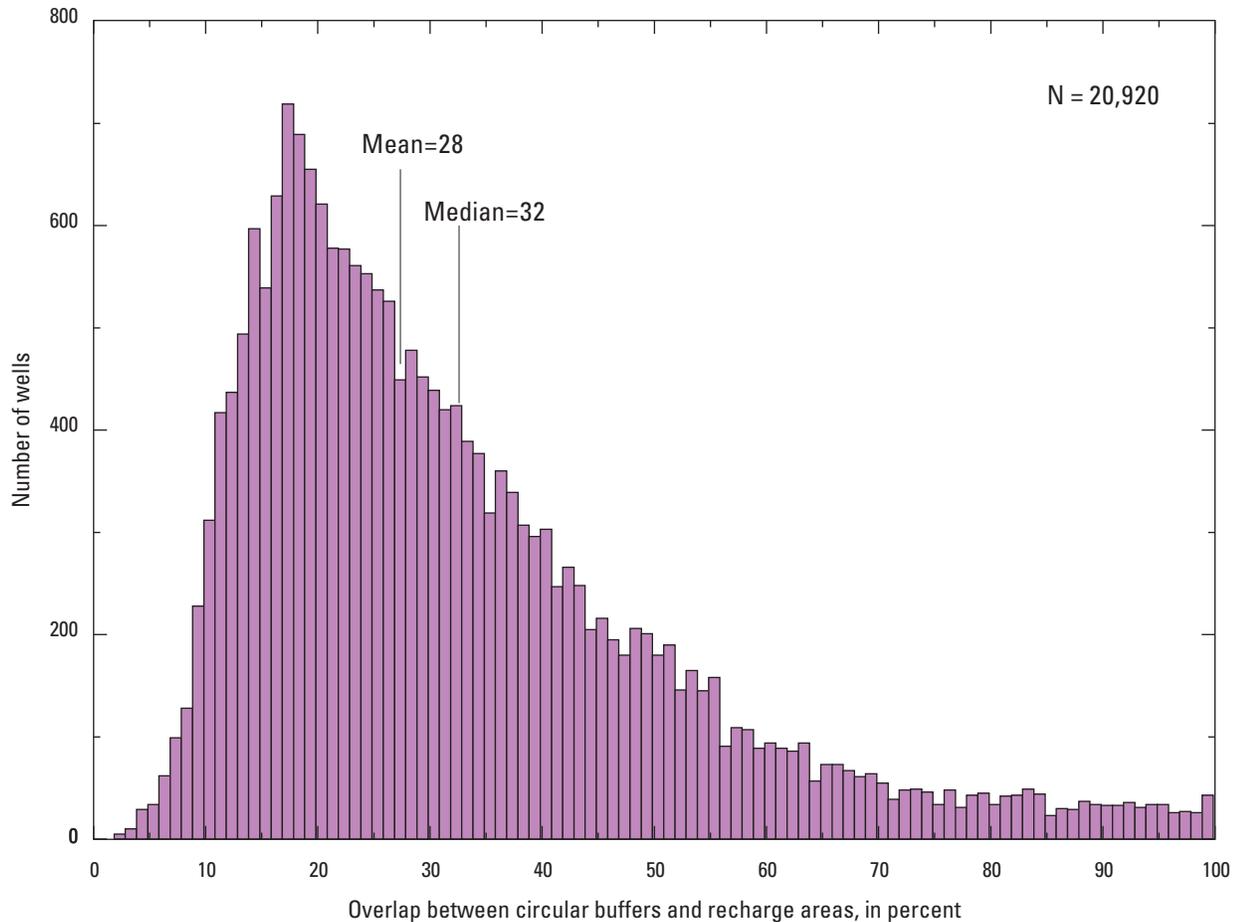
Figure 28. Distribution of overlaps between contributing recharge areas and circular buffers for 20,920 hypothetical wells in the Great Miami River Basin, southwestern Ohio.

Predictions of Nitrate Occurrence by Using Statistical Models

Values of spatial variables were obtained from two sources: soil characteristics from the STATSGO database (Wolock, 1997) and nitrogen-application rates from county data mapped to areas of agricultural data in the NLCDe92 database. Differences in the area-weighted means of spatial variables between circular buffers and contributing recharge areas, as indicated by the generally small overlap between the two at the basin scale (figs. 27 and 28), would cause differences in the resultant predictions of nitrate occurrence by statistical models.

Estimates of Area-Weighted Spatial Variables

Area-weighted mean nitrogen-application rates inside circular buffers and contributing recharge areas generally are correlated ($R^2=0.71$, p less than 0.01) with no statistical difference between the two sets of estimates (fig. 29A).

Area-weighted means of STATSGO data—silt content, alfisol content, and depth to water—also generally were correlated (R^2 values between 0.58 and 0.62, p less than 0.01) (fig. 29B). Silt content was statistically higher, and alfisol content and depth to water were statistically lower (p less than 0.001), within circular buffers than in contributing recharge areas; however, there likely are no physical processes underlying those differences. Differences in area-weighted mean spatial variables between circular buffers and contributing recharge areas are functions of the overlap between the buffers and contributing recharge areas (fig. 27) and the variability of the two sets of spatial data from which the area-weighted means were derived (fig. 4). Differences in area-weighted mean nitrogen-application rates between circular buffers and contributing recharge areas, determined by subtracting the value estimated for the contributing recharge area from the value estimated for the circular buffer, ranged from about -1,560 kg/yr (indicating that the value for the circular buffer was less than the value for the contributing recharge area) to 1,880 kg/yr. The mean and median differences were 7.8 and 0.0 kg/yr, respectively (fig. 30A). Area-weighted means
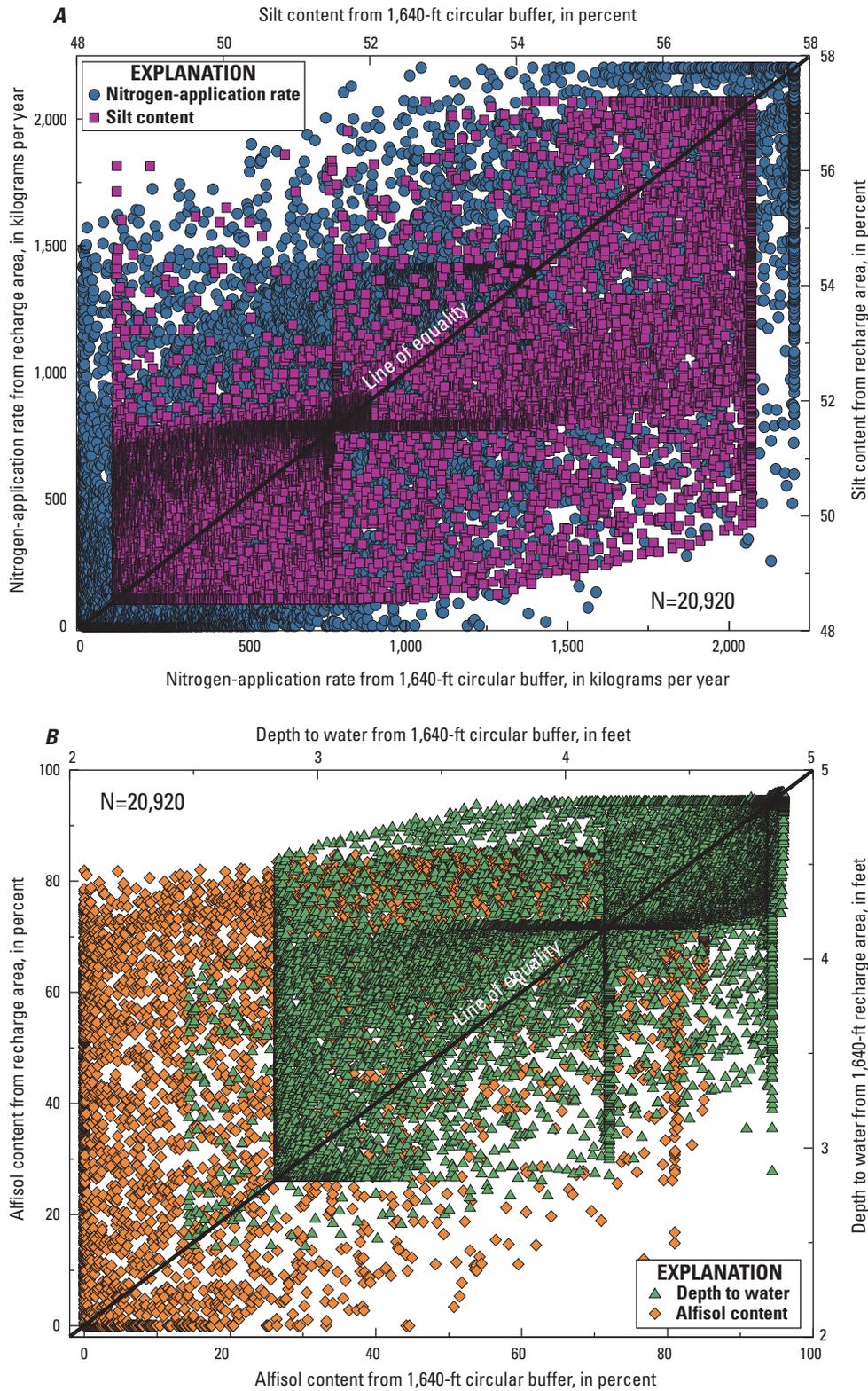
Figure 29.   Relation between contributing recharge areas and circular buffers for *A,* nitrogen-application rate and silt content, and *B,* alfisol content and depth to water.
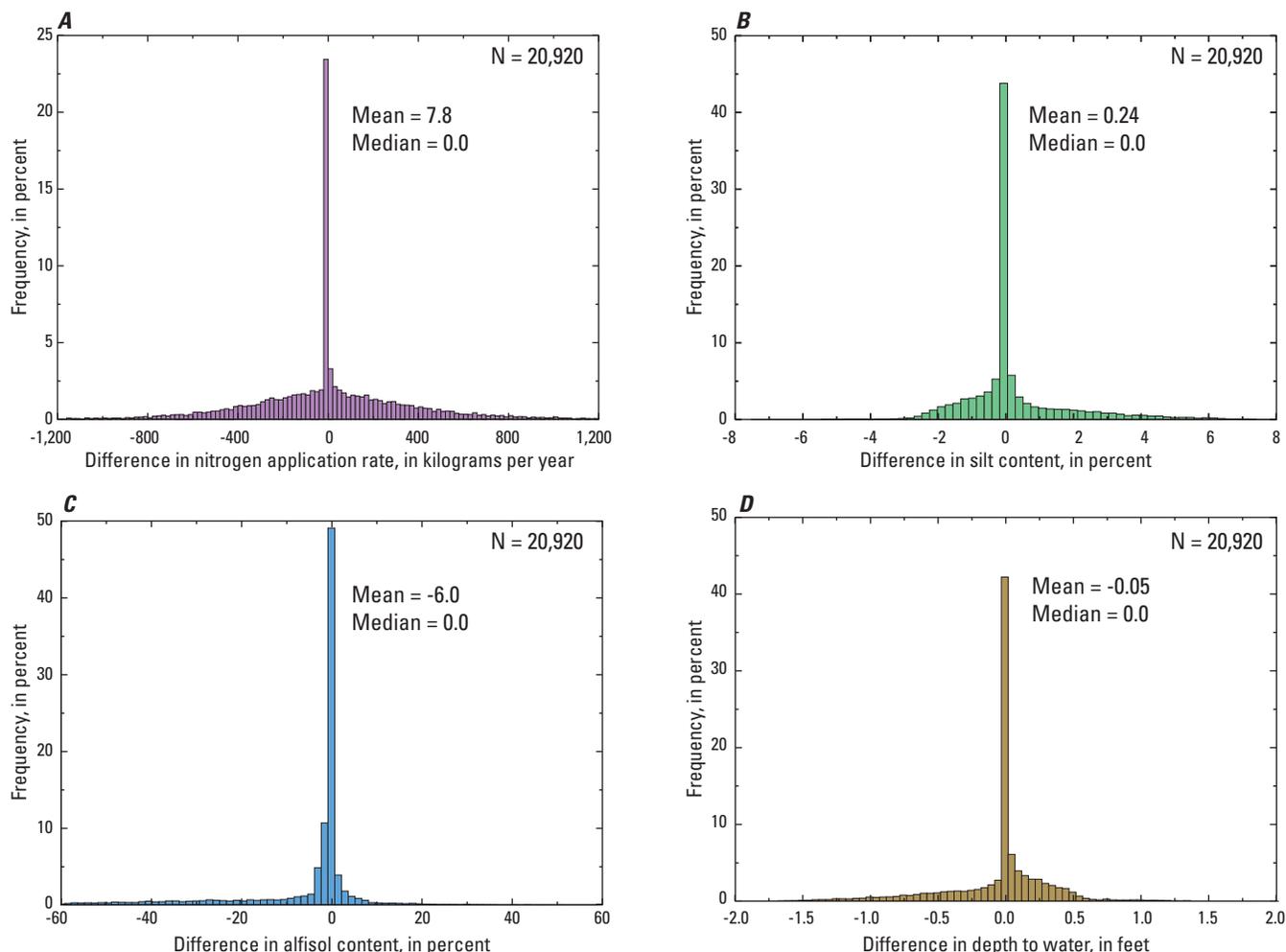
Figure 30. Differences in estimated area-weighted means for *A,* nitrogen-application rate, *B,* silt content, *C,* alfisol content, and *D,* depth to water for contributing recharge areas and circular buffers. A positive difference indicates that the prediction for circular buffers was larger.

from circular buffers and contributing recharge areas were not statistically different. The estimated mean nitrogen-application rates were equal at 4,209 locations, representing about 20.4 percent of the total.

Area-weighted means for spatial variables estimated from the STATSGO database—silt content, alfisol content, and depth to water—were the same for 6,915 locations, about 33.6 percent of the total. Differences ranged from -7.5 to 48.9 percent for silt content, from -81.8 to 81.4 percent for alfisol content, and from -1.68 to 4.77 ft for depth to water. The mean differences, defined as the value for the recharge area subtracted from the value for the circular buffer, for silt content, alfisol content, and depth to water were 0.24 percent, -6.0 percent, and -0.05 ft, respectively (figs. 30B–30D). Silt content was statistically higher, and alfisol content and depth to water statistically lower, inside circular buffers than inside contributing recharge areas (p less than 0.001). The spatial distribution of differences in estimates of spatial variables

for circular buffers and contributing recharge areas (fig. 30) reflects trends in the two sets of underlying spatial data (fig. 4). Differences in nitrogen-application rates were largest and more variable in the northeastern and southeastern parts of the basin (fig. 31A), generally corresponding to agricultural areas with the largest rates of fertilizer application (fig. 4B). Differences in silt content between circular buffers and contributing recharge areas (fig. 31B) generally were largest and more variable in areas where the STATSGO data are more variable, such as near major streams in the center of the basin (fig. 4A).

Predictions of Nitrate-Exceedance Probabilities

Nitrate exceedance probabilities are functions of the scales of the differences and the weights of the variables in the logistic regression model, as indicated by the slope

Base from U.S. Geological Survey digital data, 1:24,000, 1983
Albers Equal Area projection
Standard parallels 37°50' and 41°10', central meridian -84°

EXPLANATION

**Difference in nitrogen-application rate,
in kilograms per year, from county data**

1,900

-1,600

**Difference in silt content,
in percent, from STATSGO data**

49

-8

No difference

No modeled solution

Area not modeled
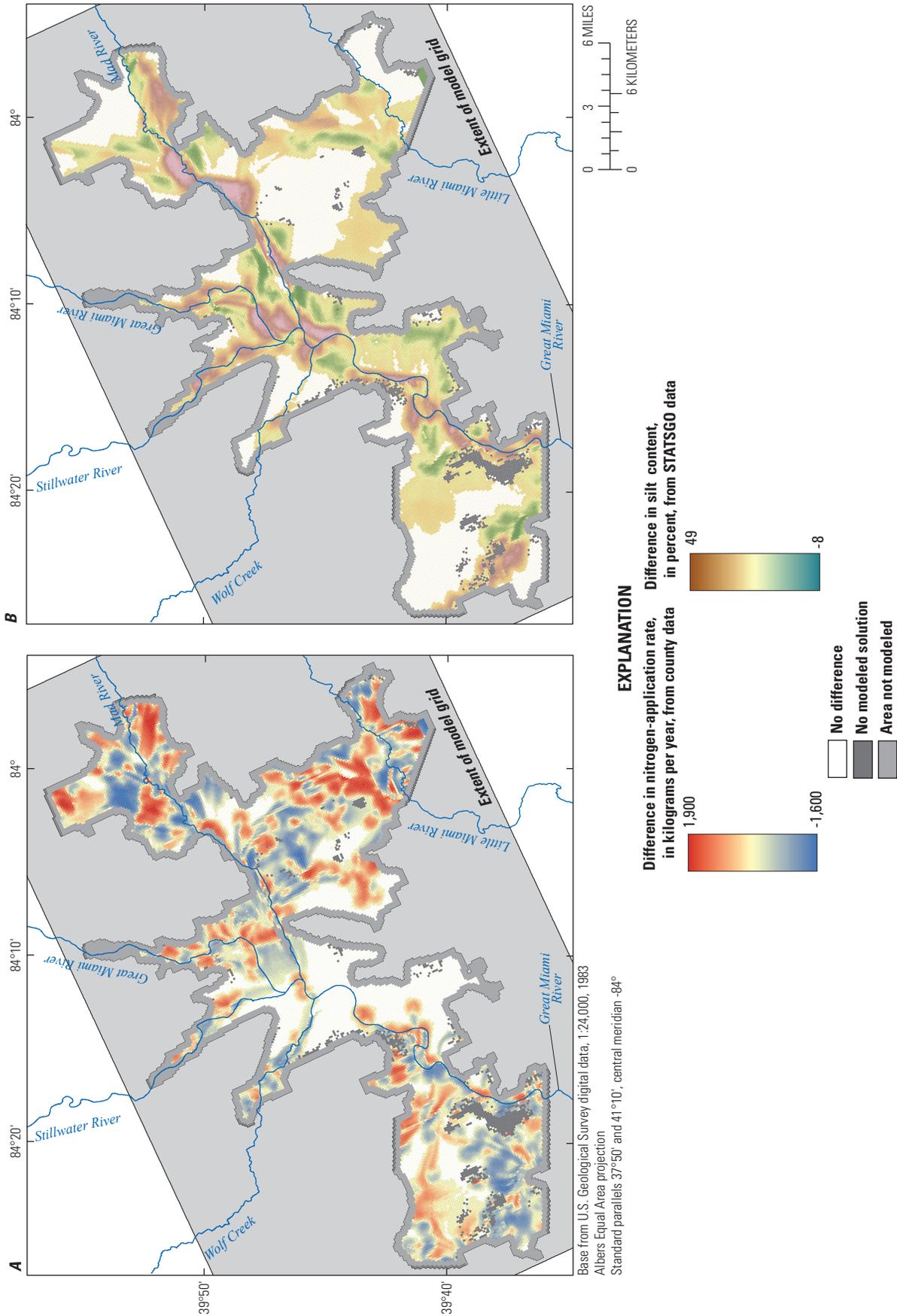
Figure 31.   Differences in the area-weighted means of the *A*, nitrogen-application rate and *B*, silt content estimated for contributing recharge areas and circular buffers, Great Miami River Basin, southwestern Ohio.

coefficients (table 1). When calculated for circular buffers, the probability of exceeding a nitrate concentration of 1 mg/L as N ranged from 20.1 to 57.0 percent, and the mean and median were both 47.0. The highest probabilities generally were near major streams in the center of the basin and in the southeastern part of the basin (fig. 32A). Data for water in some wells indicated that high predicted probabilities for the center of the basin were related to the absence of alfisol soils, whereas high probabilities for upland areas were related to large depths to water (fig. 4A). The predicted probability of exceeding 4 mg/L as N ranged from 5.4 to 30.4 percent, and the mean and median were 19.5 and 18.7, respectively. The largest areas of high probability were in the southeastern and northeastern parts of the basin (fig. 32B); the logistic regression model for predicting nitrate exceeding 4 mg/L as N included the nitrogen-application rate as an explanatory variable, and the high predicted probabilities for the northeastern part of the basin were related to areas of high rates of fertilizer application (fig. 4B). Probabilities predicted on the basis of spatial variables estimated for contributing recharge areas (figs. 33A and 33B) showed trends that generally were similar to trends in probabilities estimated for circular buffers (figs. 33A and 33B). The mean and median of probabilities of nitrate exceeding 1 mg/L as N predicted on the basis of variables estimated from contributing recharge areas were 45.2 and 42.7 percent, respectively; the mean and median of probabilities of exceeding 4 mg/L as N were 18.7 and 18.3 percent, respectively.

Probabilities predicted on the basis of spatial variables estimated for circular buffers and contributing recharge areas generally were well correlated with adjusted R^2 values of 0.69 and 0.68 for the predictions of concentrations exceeding 1 and 4 mg/L as N, respectively (fig. 34). Predictions based on variables estimated for circular buffers were statistically higher than were predictions for contributing recharge areas (p less than 0.001) for both 1 and 4 mg/L as N; this result reflects statistical differences in the estimated spatial variables. Statistical differences were likely not caused by any underlying physical processes and thus could differ in other basins. Differences in predicted probabilities of exceeding 1 mg/L as N on the basis of spatial variables from circular buffers and contributing recharge areas ranged from -17.6 to 48.7 percent; a positive value indicated higher predicted probabilities for circular buffers. The mean and median values were 1.8 and 0.2 percent, respectively (fig. 35A). For predicted probabilities of exceeding 4 mg/L as N, differences ranged from -9.2 to 21.8 percent and had a mean and median of 0.8 and 0.4 percent, respectively (fig. 35B). Predictions for exceeding concentrations of 1 and 4 mg/L as N were identical for circular buffers and contributing recharge areas in 35.4 and 10.2 percent of the hypothetical wells, respectively. The number of locations for the 4 mg/L threshold was lower than the number of locations for the 1 mg/L threshold because the use of nitrogen-application rate as an explanatory variable in the logistic regression model for the 4 mg/L threshold resulted in more potential for differences in predictions from

contributing recharge areas and circular buffers (table 1). Although predictions generally were correlated, differences for some areas were large (figs. 32 and 33), and using spatial variables for contributing recharge areas can significantly affect statistical model predictions.

Spatial trends in differences among predictions generally reflect spatial patterns in the differences among the spatial variables estimated for circular buffers and contributing recharge areas. The largest differences among predictions of exceeding concentrations of 1 mg/L as N generally were near major streams near the center of the basin (fig. 36A), for which differences in environmental variables from the STATSGO database were large (fig. 31B). Areas for which predictions for circular buffers and contributing recharge areas were the same—about 35 percent of the modeled basin—generally corresponded to large contiguous areas for which environmental variables from the STATSGO database were similar (fig. 4A). Differences in predictions of exceeding 4 mg/L as N were more widespread; predictions were the same for circular buffers and contributing recharge areas in only about 10 percent of the modeled basin (fig. 36B). Areas where differences for exceeding 4 mg/L as N were larger than differences for exceeding 1 mg/L as N generally corresponded to areas where differences in the nitrogen-application rate were larger, such as agricultural areas in the southeastern and northeastern parts of the basin (fig. 4B). Despite the large differences in predictions caused by differences in spatial variables, the spatial patterns were similar. Areas where the probability of nitrate concentrations exceeding the threshold was higher than the median probability predicted for circular buffers generally coincided with the areas similarly predicted for contributing recharge areas (figs. 37A and 37B). The areas where predicted probabilities of exceeding 1 mg/L as N were greater than the median values—about 47 percent for circular buffers and about 43 percent for contributing recharge areas—overlapped by 93.0 percent (fig. 37A). For predicted probabilities of exceeding 4 mg/L as N, areas greater than the medians of about 19 and 18 percent for circular buffers and contributing recharge areas overlapped by 88.6 percent (fig. 37B). These differences do not reflect an underlying nonrandom process, however, and patterns in differences among spatial variables for this area would not be transferable to other basins.

Role of Well Depth

Contributing recharge areas to wells pumped near the water table partially overlap the circular buffers to those wells; however, in many cases, circular buffers and contributing recharge areas to deeper wells do not overlap because contributing recharge areas to wells pumped deeper in the aquifer are likely to extend farther upgradient than contributing recharge areas to shallower wells (fig. 11). As an example, the contributing recharge area to well 725, pumped in the shallow part of the aquifer, is partially inside (about 13 percent) and upgradient from the corresponding

Base from U.S. Geological Survey digital data, 1:24,000, 1983
Albers Equal Area projection
Standard parallels 37°50' and 41°10', central meridian –84°

EXPLANATION

Probability of exceeding nitrate concentrations, in percent

1 mg/L
57
20

4 mg/L
30
5

Area not modeled
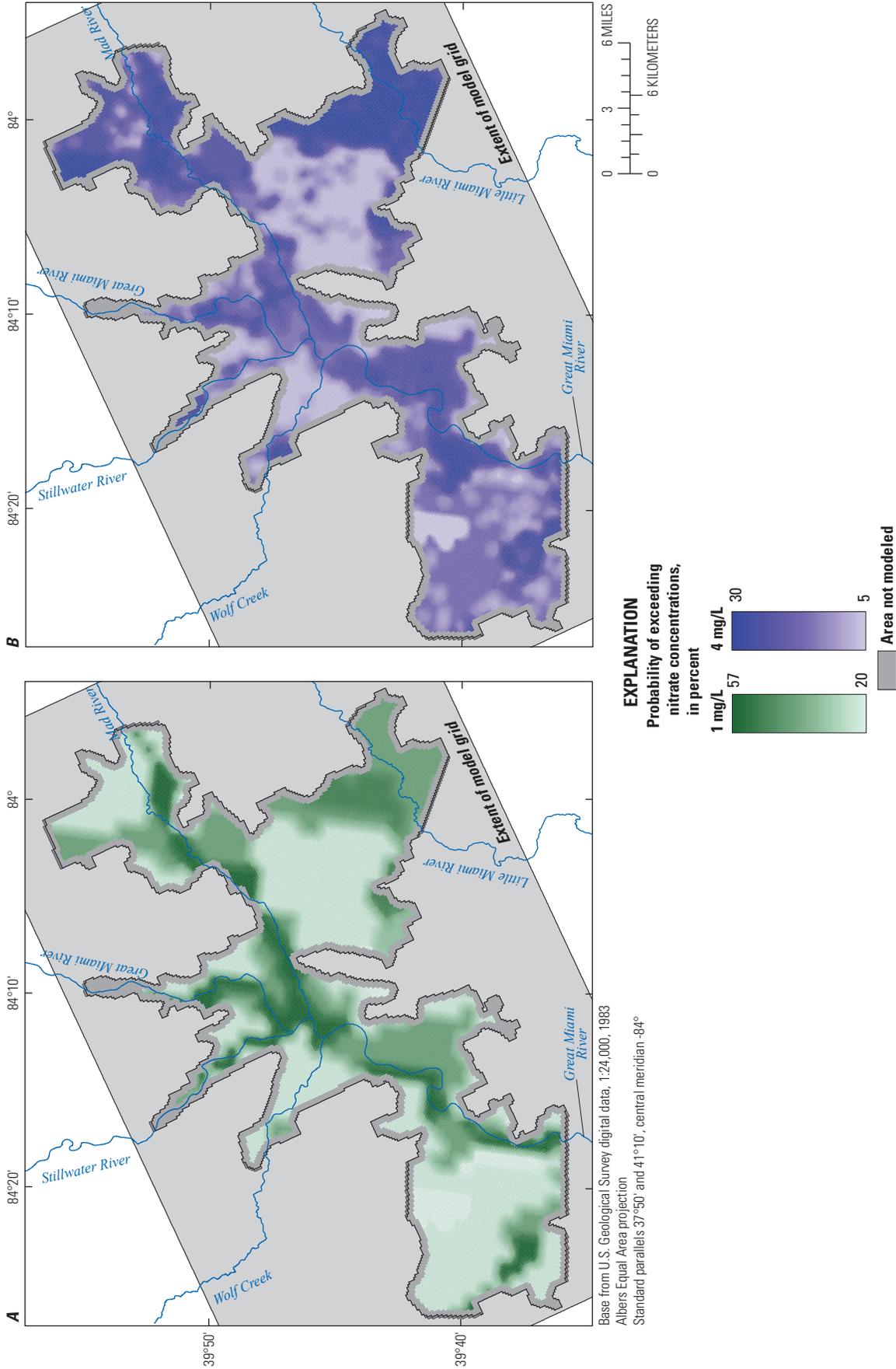
Figure 32.   Probabilities of exceeding nitrate concentrations of *A*, 1 milligram per liter as nitrogen, and *B*, 4 milligrams per liter as nitrogen predicted from logistic regression models on the basis of spatial variables for circular buffers, Great Miami River Basin, southwestern Ohio.

Base from U.S. Geological Survey digital data, 1:24,000, 1983
Albers Equal Area projection
Standard parallels 37°50' and 41°10', central meridian -84°
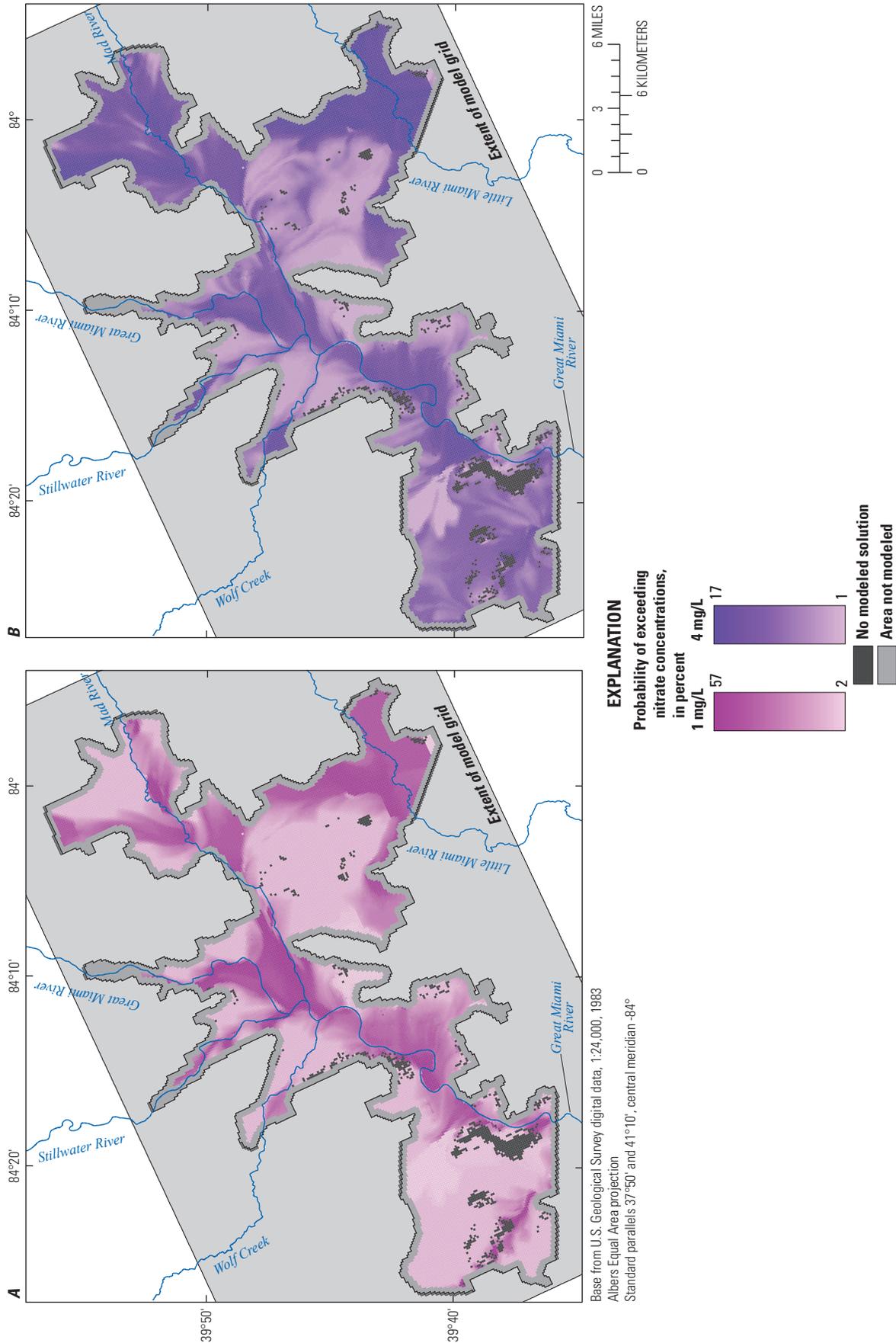
Figure 33.   Probabilities of exceeding nitrate concentrations of *A*, 1 milligram per liter as nitrogen, and *B*, 4 milligrams per liter as nitrogen predicted by logistic regression models on the basis of spatial variables for contributing recharge areas, Great Miami River Basin, southwestern Ohio.
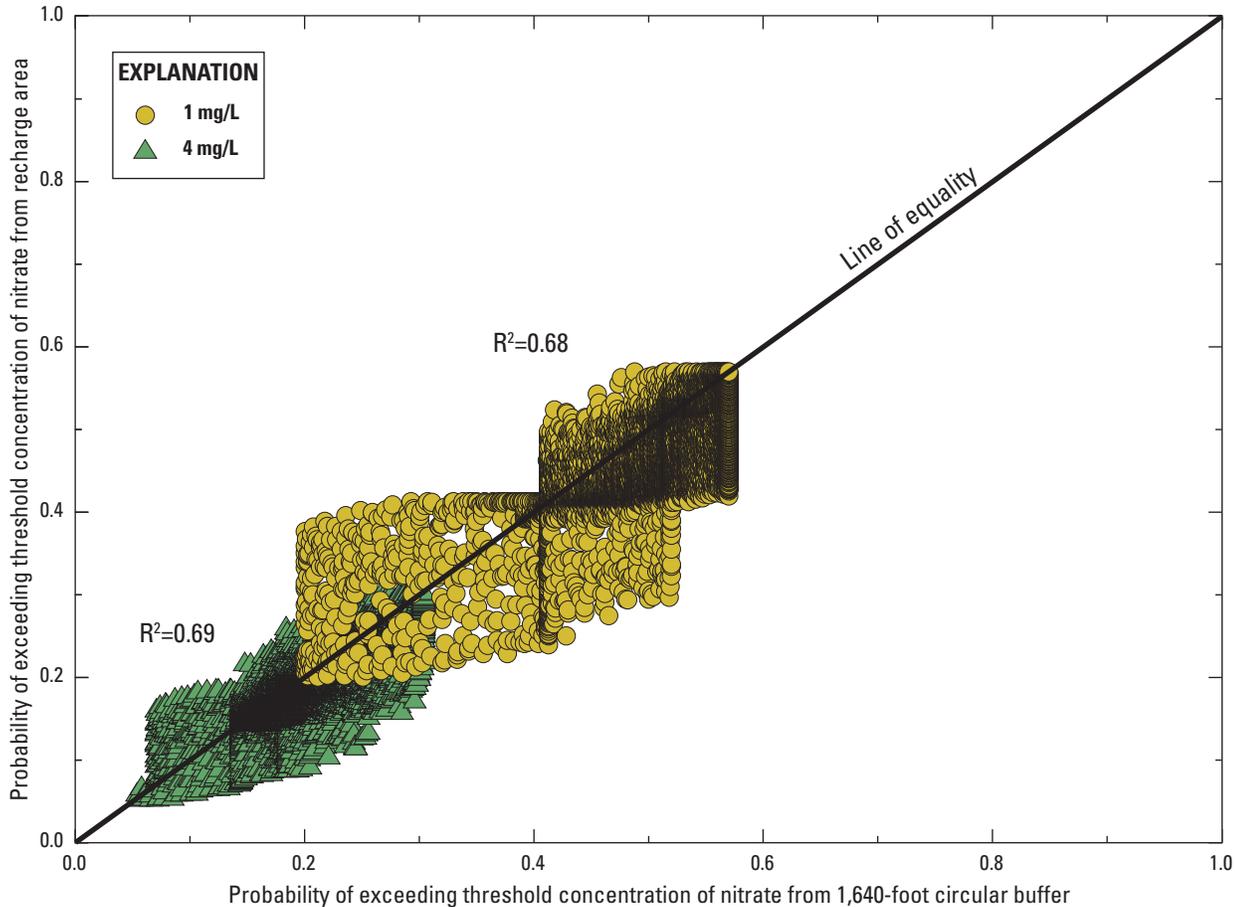
Figure 34.   Relation between probabilities of exceeding nitrate concentrations of 1 and 4 milligrams per liter as nitrogen on the basis of area-weighted mean explanatory variables for contributing recharge areas and circular buffers.

circular buffer, whereas the contributing recharge area to a well pumped deep in the aquifer (layer 3) at the same location is far upgradient (fig. 38). The closest edge of the contributing recharge area to the deep well is about 15,000 ft from the edge of the 1,640-ft radius circular buffer (fig. 38). The fraction of overlap between contributing recharge areas and circular buffers is generally lower for deep wells than for shallow wells (fig. 39). The mean and median of the fraction of overlap between circular buffers and contributing recharge areas to deep wells were 0.24 and 0.27, respectively; the contributing recharge areas to about 63.1 percent of wells pumped from deep parts of the aquifer did not overlap any of the corresponding circular buffers. For deep wells where the contributing recharge area partly overlapped the circular buffer, the fractions of overlap for shallow and deep wells were positively correlated (fig. 39).

Differences among contributing recharge areas to wells pumped at different depths indicate that differences in estimates of spatial variables and resultant predictions of nitrate occurrence that arise from differences in circular buffers and contributing recharge areas are greater for wells

pumped in deep than for those pumped in shallow parts of the aquifer, and that estimated spatial variables for contributing recharge areas are likely to be different for wells pumped at different depths. Estimates of nitrogen-application rates for contributing recharge areas to shallow wells (layer 1) and deep wells (layer 3) generally were correlated (R^2=0.64) (fig. 40A), and the two sets of estimated variables were not statistically different; however, there is a large amount of scatter in the data arising from large differences in contributing recharge areas from shallow and deep wells at individual locations (fig. 40A). The correlation is less for estimates of spatial variables such as silt content from STATSGO data between contributing recharge areas from shallow and deep pumped wells (fig. 40A). Correlations of silt and alfisol content were statistically higher, and depth to water statistically lower, in contributing recharge areas to deep wells than contributing recharge areas to shallow wells (p less than 0.001). No underlying physical processes likely caused these differences.

Probabilities of nitrate concentrations exceeding 1 and 4 mg/L as N in shallow and deep wells at a consistent depth of 10 ft (to identify trends in the predictions arising only
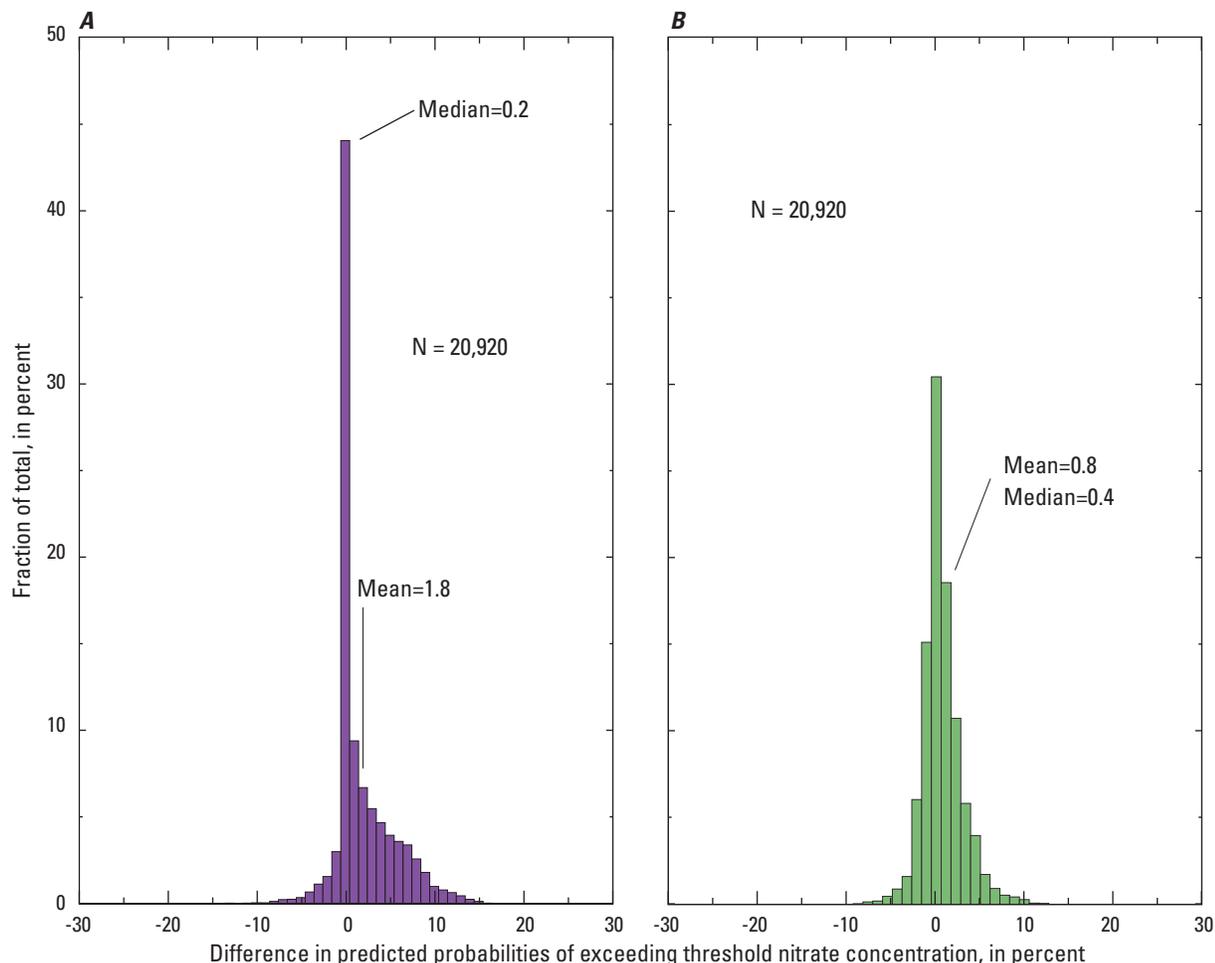
Figure 35. Differences in predicted probabilities of exceeding concentrations of *A,* 1 milligram per liter as nitrogen; and *B,* 4 milligrams per liter as nitrogen on the basis of area-weighted mean explanatory variables for contributing recharge areas and circular buffers. A positive difference indicates that the prediction for circular buffers was larger.

from differences in estimated spatial variables) generally were correlated (fig. 40B). The probabilities of exceeding 1 and 4 mg/L as N predicted for contributing recharge areas to shallow wells (layer 1) are statistically higher than those predicted for contributing recharge areas to deep wells (layer 3), likely reflecting statistical differences in spatial variables estimated from the STATSGO database. Based on a well depth of 91 ft, which is the mean depth of private wells in the NAWQA monitoring network and is common for wells pumped in layer 3 of the model, predicted probabilities of exceeding 1 and 4 mg/L as N are much lower than those predicted on the basis of an assumed well depth of 10 ft. The negative slope coefficients of the well-depth explanatory variables in the logistic regression models (table 1) indicate that predicted probabilities decrease with increasing depth to the top of the well screen. The variability in predicted probabilities arising from differences in estimated spatial variables for contributing recharge areas to shallow and deep wells is substantially smaller for larger assumed well-screen

depths (fig. 40B). This result indicates that although differences in estimated spatial variables are larger for deep wells, the use of well-screen depths consistent with deeper wells lessens the effects of those differences on predictions of nitrate occurrence.

Estimates of Depth to Water

Values of the spatial variables included as explanatory variables in the logistic regression models are based on national databases and generally are coarse at the local scale. Nitrogen application rates are calculated by applying total fertilizer use, based on county-specific data, to agricultural parcels in the NLCDe92 database; these data can provide a source term for statistical and process models of nitrate occurrence or transport. Soil characteristics that are based on STATSGO data—silt content and alfisol content—can inform process and statistical models regarding the potential

Base from U.S. Geological Survey digital data, 1:24,000, 1983
Albers Equal Area projection
Standard parallels 37°50' and 41°10', central meridian –84°

EXPLANATION

Difference in probability of exceeding nitrate concentration, in percent, from circular buffers and recharge areas

No difference

No modeled solution

Area not modeled
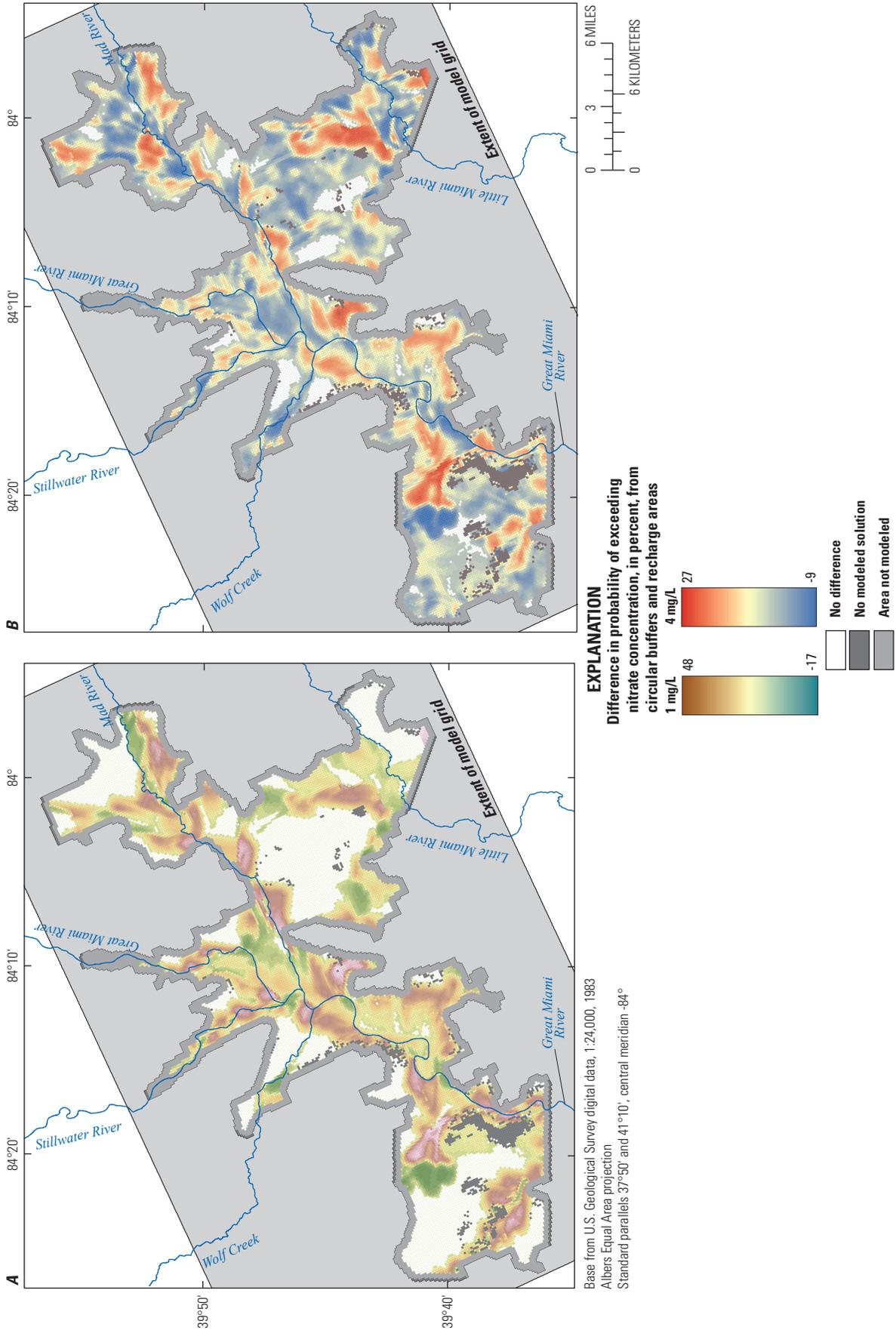
Figure 36.   Differences in the probabilities of nitrate concentrations exceeding *A*, 1 milligram per liter as nitrogen, and *B*, 4 milligrams per liter as nitrogen predicted by logistic regression models on the basis of spatial variables for contributing recharge areas and circular buffers, Great Miami River Basin, southwestern Ohio.

EXPLANATION

Area where probability of nitrate concentrations exceeding 1 mg/L is greater than median predicted probability

Circular buffer, median of 47 percent

Recharge area, median of 43 percent

Area where probability of nitrate concentrations exceeding 4 mg/L is greater than median predicted probability

Circular buffer, median of 19 percent

Recharge area, median of 18 percent

Area not modeled

No model solution

Base from U.S. Geological Survey digital data, 1:24,000, 1983
Albers Equal Area projection
Standard parallels 37°50' and 41°10', central meridian -84°
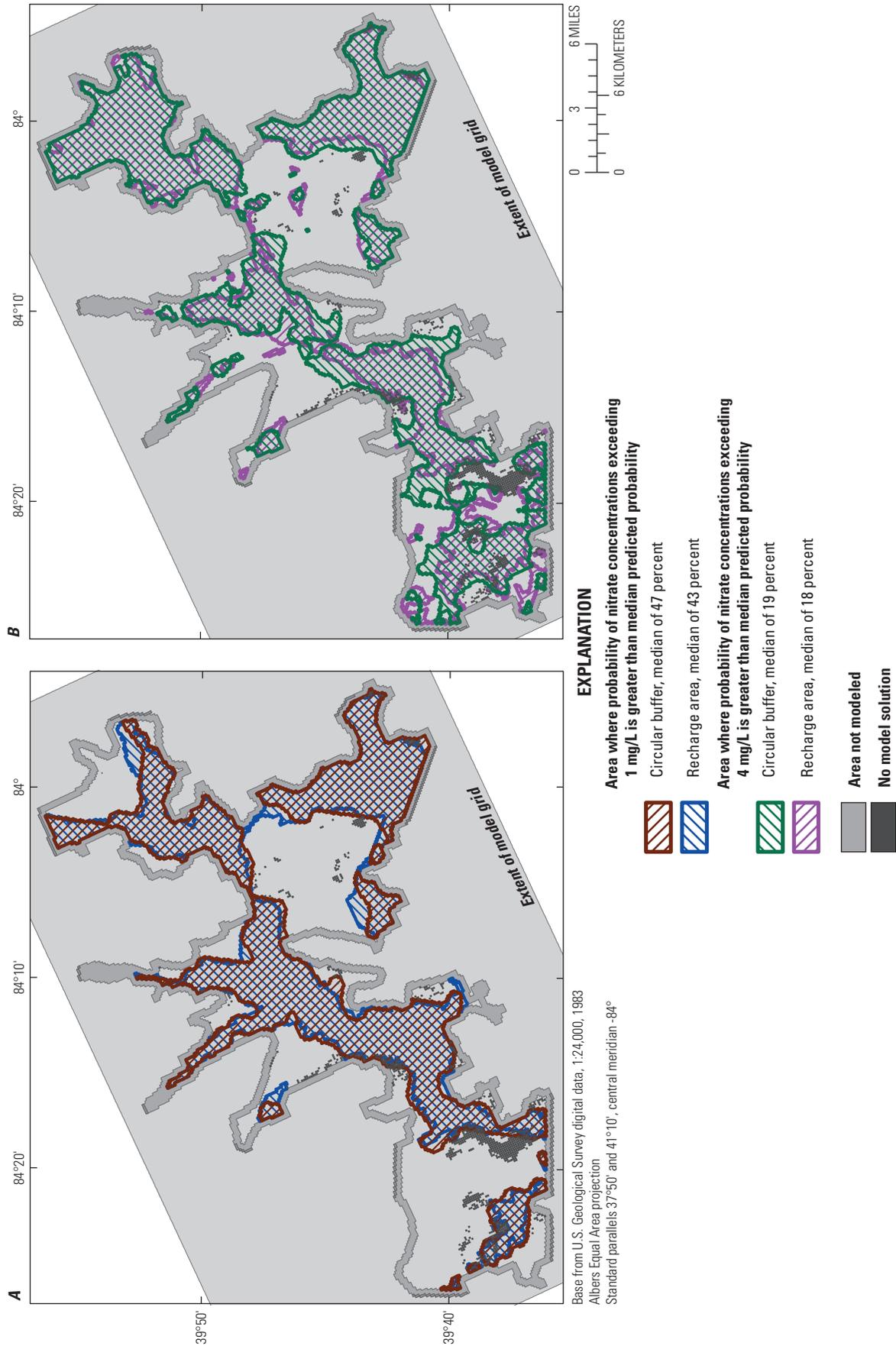
Figure 37.   Areas where probabilities of nitrate concentrations exceeding *A*, 1 milligram per liter as nitrogen, and *B*, 4 milligrams per liter as nitrogen were higher than the median probabilities predicted by logistic regression models for contributing recharge areas and circular buffers, Great Miami River Basin, southwestern Ohio.
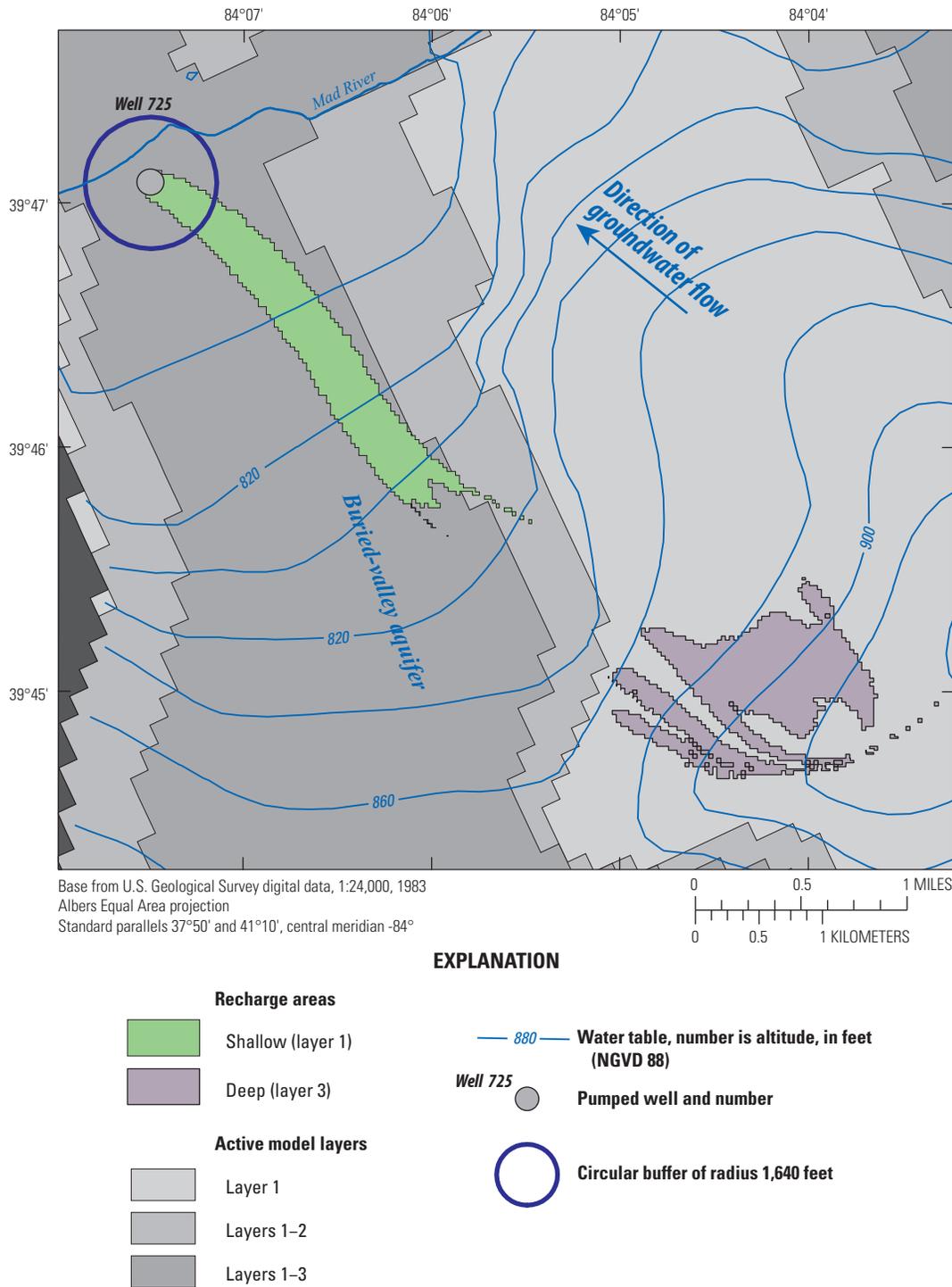
Figure 38.   Contributing recharge areas and the circular buffer to well 725 (a hypothetical well) for shallow and deep well screens, southeastern part of the Great Miami River Basin, southwestern Ohio.
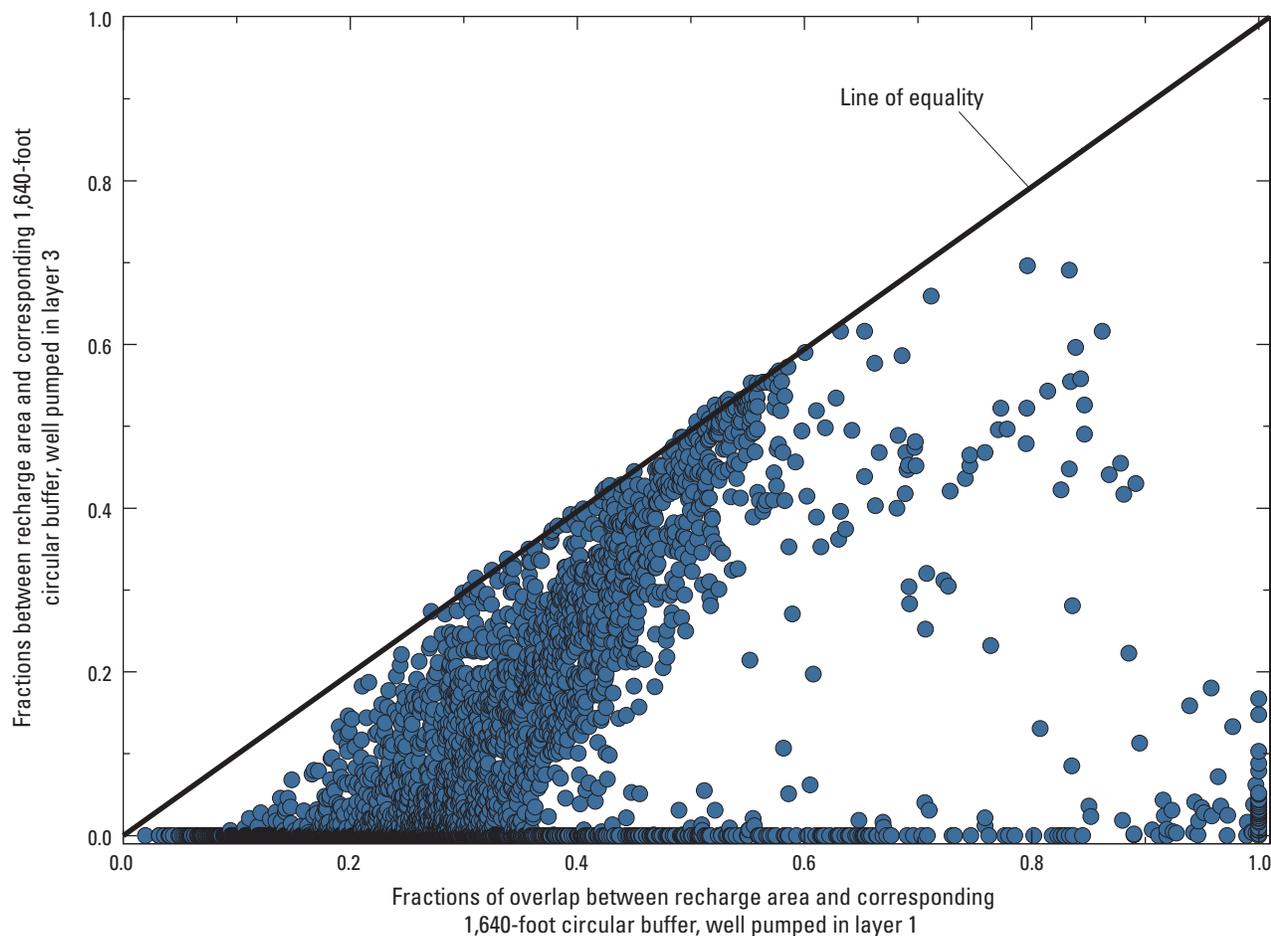
Figure 39. Relation between fractions of overlap between contributing recharge areas and corresponding circular buffers for wells pumped in shallow and deep parts of the glacial aquifer system, Great Miami River Basin, southwestern Ohio.

for nitrogen attenuation, but are external to the models. Unlike other spatial variables used in the logistic-regression models, depth to water (saturates soil), derived from the STATSGO database, represents the depth from land surface to the water table and is a function of the local flow system; the variable, therefore, can be calculated from a process model that simulates hydrologic processes. Depth to water has large slope coefficients in the statistical models of nitrate occurrence above six of the ten threshold concentrations (1 to 10 mg/L as N) (table 1) (Warner and Arnold, 2010) and therefore has an important control on predictions of nitrate occurrence.

Depth to water in the STATSGO database ranged from 2.48 to 4.88 ft (fig. 4A); area-weighted mean values for the 1,640-ft-radius circular buffers to the 21,774 hypothetical wells had the same range (fig. 41A). Area-weighted mean estimates of depth to water determined from the groundwater-flow model, representing the distance between the land surface and the calculated water-table altitude, ranged from 0.02 to 78 ft—a much larger range than area-weighted means determined from the STATSGO database (fig. 41A). Model-calculated depths to water better represent local hydrologic

conditions (determined from aquifer properties, hydrologic boundaries, and recharge conditions) and, if the model is well posed, likely represent more accurate and detailed estimates of the explanatory variable used in statistical models.

For the range of depths to water in the STATSGO database (2.48 to 4.88 ft), the probabilities of exceeding 1 and 4 mg/L as N predicted by using the subset of model-calculated depths to water and depths to water from the STATSGO database differed substantially (fig. 14B). The two sets of predictions were not correlated and statistically different (p less than 0.001). As an example, for a prediction of 0.5 for an exceedance of 1 mg/L as N on the basis of STATSGO data, the predicted model-calculated depths to water ranged from 0.18 to 0.73 (fig. 41B). The large difference between probabilities predicted on the basis of depths to water estimated from the STATSGO database and probabilities calculated by the groundwater-flow model indicates that the more detailed, site-specific estimates from the process (groundwater-flow) model could inform or improve statistical models of nitrate occurrence.
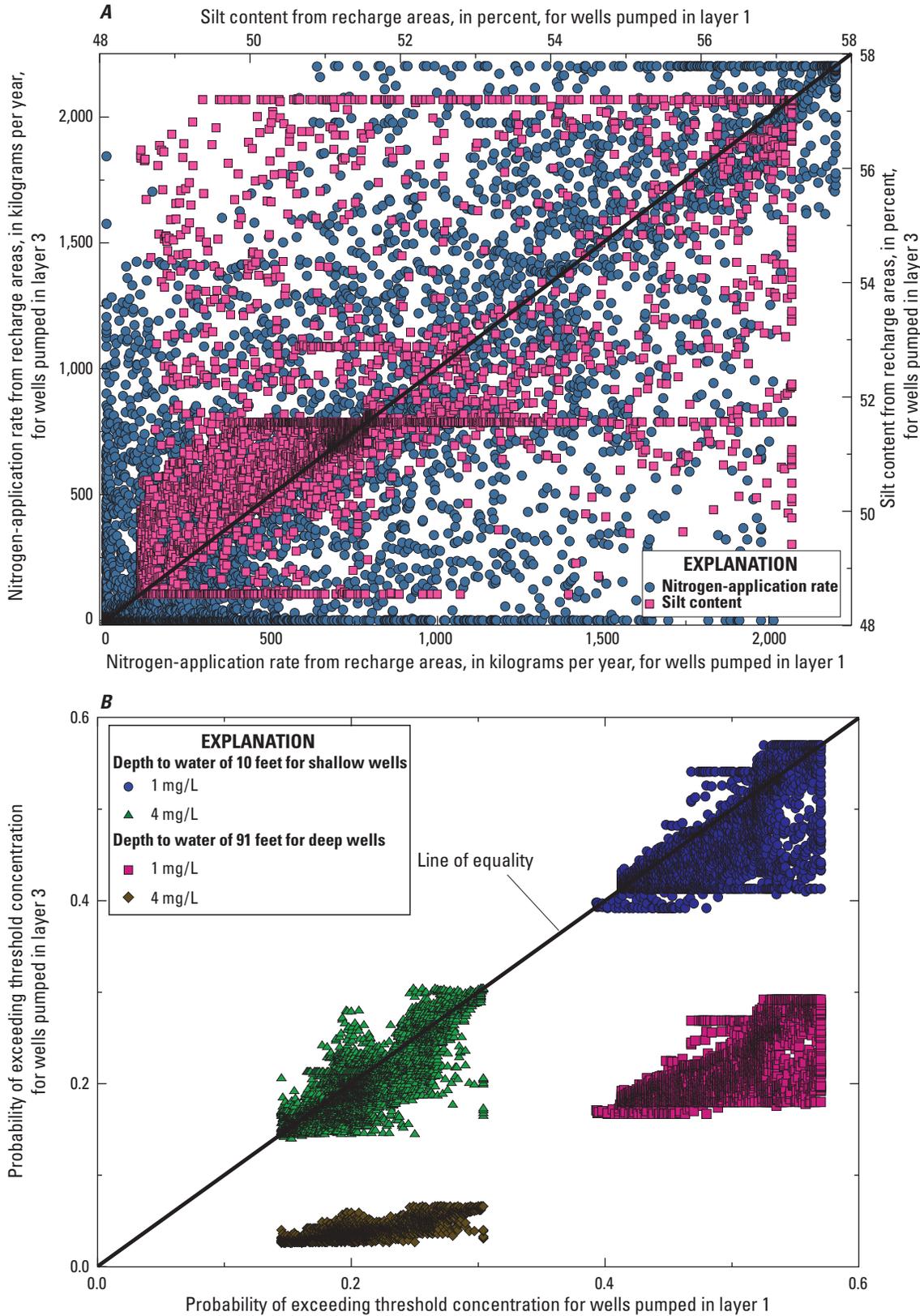
Figure 40.   *A,* Estimates of nitrogen-application rate and silt content, and *B,* probabilities of nitrate occurrence above threshold concentrations of 1 and 4 milligrams per liter as N determined for contributing recharge areas and circular buffers for shallow and deep pumped wells, Great Miami River Basin, southwestern Ohio.
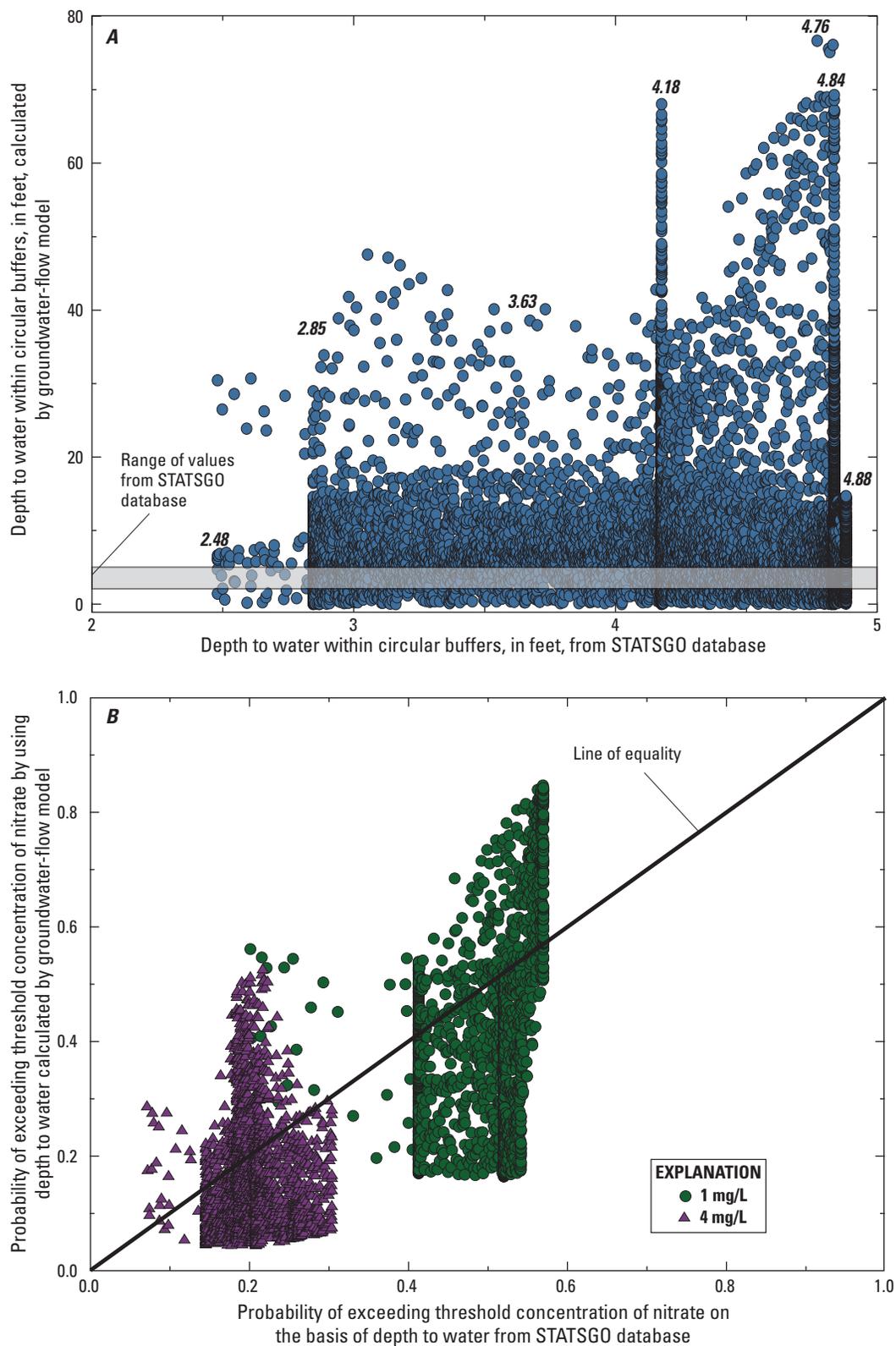
Figure 41. Relations between *A,* depths to water within circular buffers estimated from STATSGO data and calculated by a groundwater-flow model, and *B,* probability of nitrate concentrations exceeding 1 and 4 milligrams per liter as nitrogen on the basis of the two sets of estimated depths to water.

The use of a process model of saturated flow to calculate depths to water has two important limitations: (1) the inability to represent perched water tables, which could cause lower depths to water, particularly for areas with clayey soils (such as alfisols), and (2) uncertainties associated with the flow model that can result in simulated water-table altitudes above the land surface and thus negative depths to water. Calculations of negative depths to water were excluded from the analysis.

Predictions of Nitrate Concentrations by Using Nitrogen Loading Rates

Process models that simulate sources, flow, and transport can be used to estimate concentrations of a solute from surface sources in an aquifer system. For conservative species, such as nitrate in oxic environments within the glacial aquifer system, process models that simulate only groundwater flow can use mass-weighted particle tracking to implicitly represent the advective transport of nitrate from surface sources, such as areas of fertilizer application, to pumped wells. The method predicts nitrogen concentrations by tracking water—represented as a series of particles—from recharge locations at the water table to a pumped well, analogous to the recharge area, and assigning to each particle a mass of nitrogen estimated from the associated source area at that location. Accountings of cumulative mass and transport times are used to estimate the time-varying concentration in the well. The analyses are presented as a way of comparing process-model predictions that explicitly represent physical transport processes with statistical-model predictions of the probability of exceeding a threshold concentration from the set of estimated explanatory variables. The analyses are based on the spatial distribution of nitrogen application that was used to develop the statistical models (fig. 4); the distributions of nitrogen application at the local scale and nitrogen from other sources are not represented. Also, assumptions regarding nitrate attenuation in the unsaturated zones and in the aquifer are inherent in the analyses.

Under the assumed nitrogen-application rates from county data (fig. 4B), a nitrogen uptake of 50 percent (Bundy and Andraski, 2005), and losses of 30 percent to other biologic and geochemical processes near land surface and in the unsaturated zone (Weiskel and others, 1992), nitrate concentrations determined from mass-weighted particle tracking to 21,774 hypothetical pumped wells ranged from 0 to 12 mg/L as N after 50 years of transport through shallow parts of the Great Miami River Basin (fig. 42A). The median value was 2 mg/L as N, consistent with the median for monitoring wells in the NAWQA well network (Warner and Arnold, 2010). Nitrate concentrations arising from the application of fertilizer were spatially correlated with the distribution of nitrogen application in the basin; nitrate concentrations exceeding the median value of 2 mg/L as N (higher than the 50th percentile) were in the northeastern,

southeastern, and southwestern parts of the aquifer (fig. 42B) beneath the three areas of highest fertilizer-application rates on agricultural land (fig. 4B). It should be noted that the concentrations account only for fertilizer application and not other sources such as septic systems in residential areas.

Simulations of nitrate concentrations by a groundwater-flow model and mass-weighted particle tracking use the nitrogen-application rate (fig. 4B) as the only source input and, as a result, the steady-state concentrations of nitrate in pumped wells are highly correlated with area-weighted mean nitrogen-application rates within contributing recharge areas (R^2=0.90) (fig. 43). The proportionality constant between the area-weighted mean nitrogen-application rate and steady-state nitrogen concentration is a function of the volume of recharge through the contributing recharge area; the scatter in the data arises from the spatially variable recharge volumes in the basin (fig. 25). The nitrate concentration also is correlated with the area-weighted mean nitrogen-application rate within circular buffers (R^2=0.65) (fig. 43) because area-weighted means from circular buffers and contributing recharge areas are correlated (R^2=0.71) (fig. 29A).

Nitrogen concentrations predicted by a groundwater-flow model are a function of simulated transport times because concentrations change over time until a steady-state concentration is reached (fig. 16A). After an assumed transport time of 50 years, simulated concentrations exceed 1 and 4 mg/L in about 57.6 and 28.4 percent of the basin (fig. 44A). After an advective transport time of 10 years, simulated concentrations exceed 1 mg/L as N in about 44.7 percent of the basin and 4 mg/L as N in about 5.6 percent of the basin (fig. 44B). Simulated concentrations also are a function of depth consistent with observations of the relations between nitrate concentrations, groundwater age, and well depth (Warner and Arnold, 2010). Simulated nitrate concentrations in water from the same source generally are lower for wells pumped from deep parts of the aquifer (layer 3) than those pumped from shallow parts (layer 1) (fig. 45); longer transport times for water withdrawn from greater depths in the aquifer result in lower concentrations than for water withdrawn from shallower depths after 50 years of transport. The large amount of scatter results from large differences among sources of water and associated nitrogen inputs to wells pumped at different depths (fig. 38).

Comparison of Predictions of Nitrate Occurrences and Concentrations

Logistic regression models use a set of explanatory variables that include combinations of spatial variables describing soil characteristics affecting nitrogen attenuation and sources of nitrogen in agricultural areas; the result is a prediction of the probability that concentrations at a location will exceed a specific threshold concentration. The use of a process model to simulate nitrogen transport, such as can be done by a groundwater-flow model coupled with

Base from U.S. Geological Survey digital data, 1:24,000, 1983
Albers Equal Area projection
Standard parallels 37°50' and 41°10', central meridian -84°

EXPLANATION

**Nitrate concentration,
in milligrams per liter as N**

12

0

Percentiles of nitrate concentration

0 to 20   >50 to 60

>20 to 30   >60 to 70

>30 to 40   >70 to 80

>40 to 50   >80 to 90

>90 to 100

No modeled solution

Area not modeled
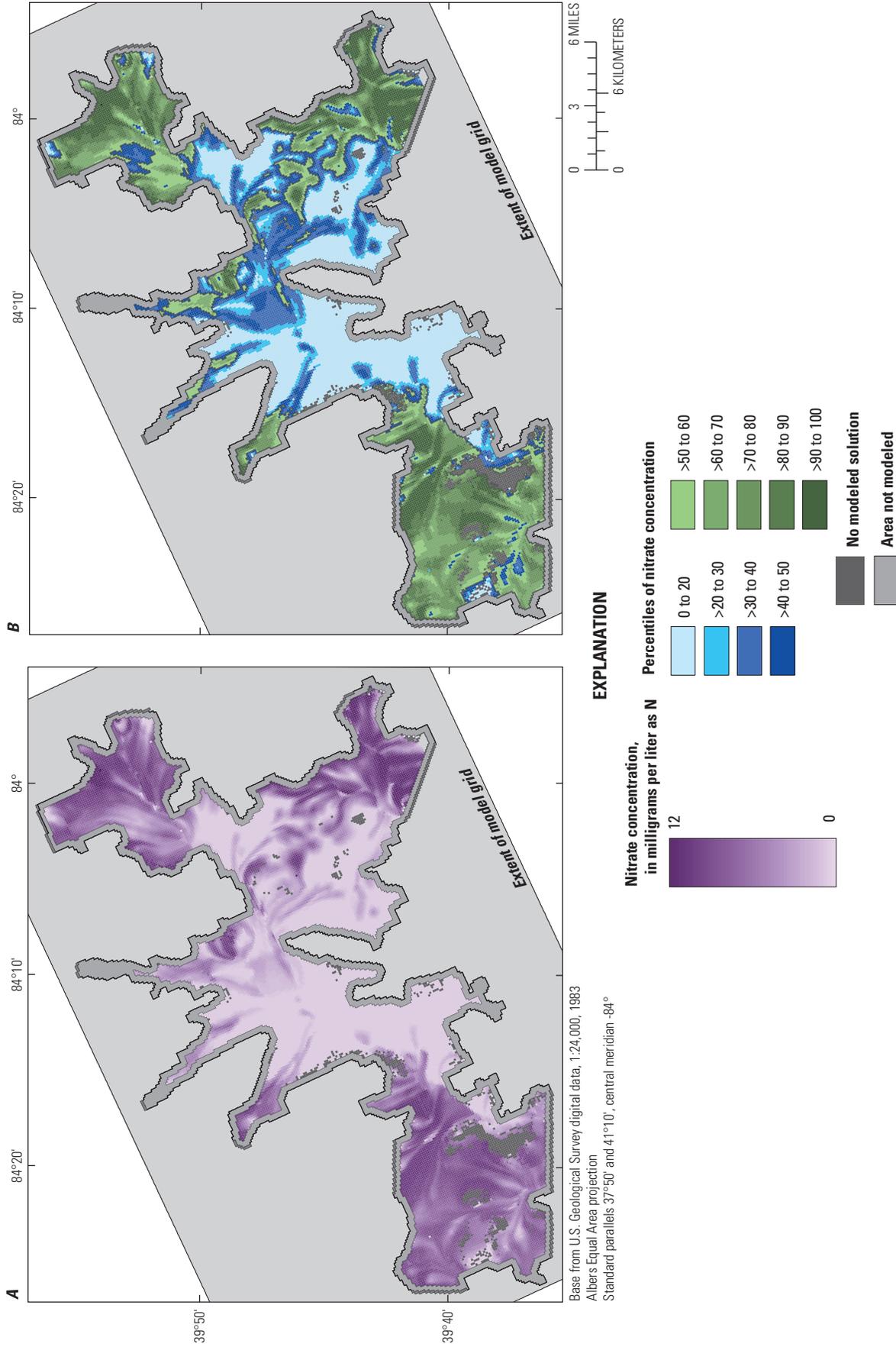
Figure 42. *A*, Nitrate concentrations (as N) and *B*, concentration percentiles after 50 years of steady-state transport from areas of nitrogen-fertilizer application to shallow pumped wells, Great Miami River Basin, southwestern Ohio.
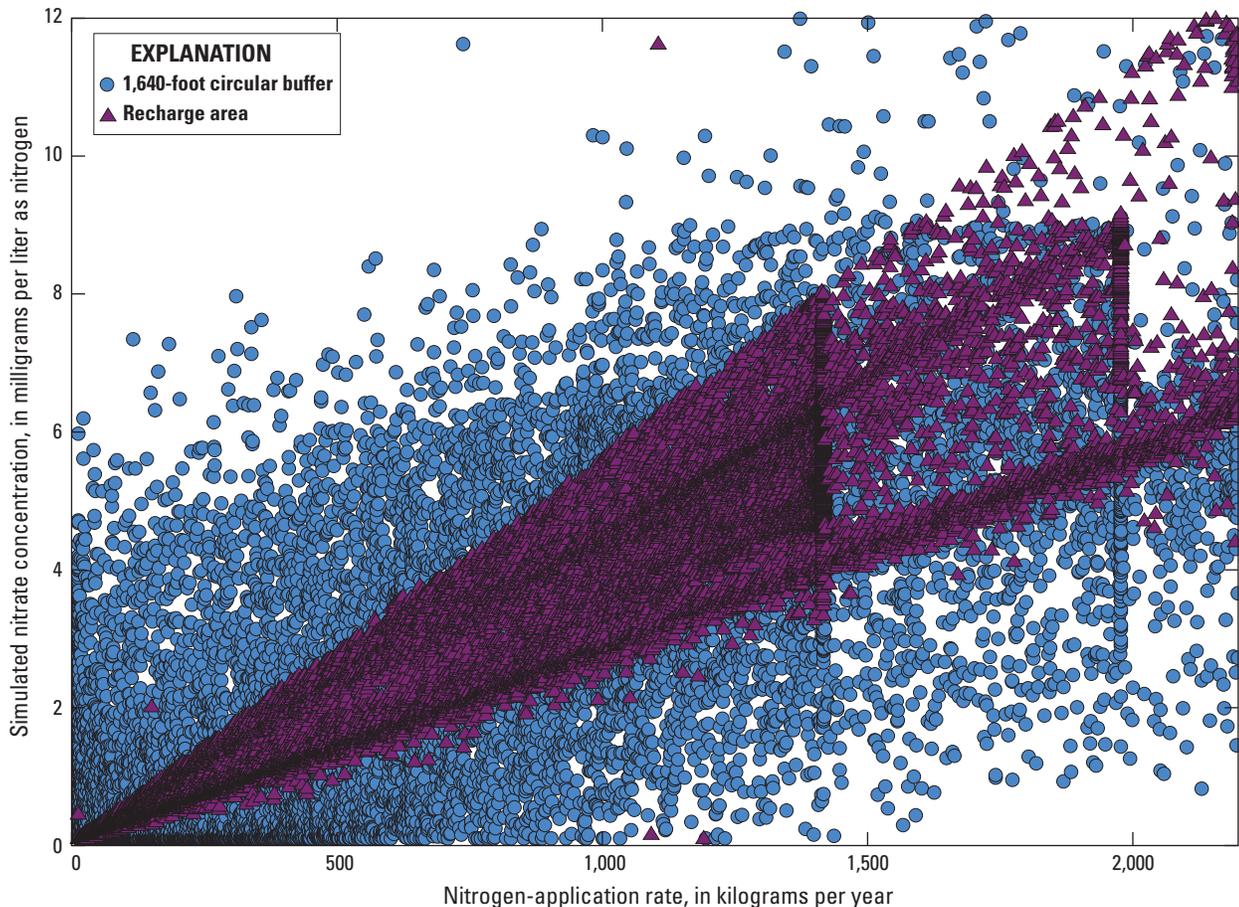
Figure 43.   Relations between simulated nitrate concentrations and area-weighted mean nitrogen-application rates within circular buffers and contributing recharge areas, Great Miami River Basin, southwestern Ohio.

mass-weighted particle tracking, produces an estimate of a concentration in a pumped well that incorporates source terms and advective-transport processes; attenuation variables affecting nitrogen prior to recharge to the aquifer are not explicitly included in the process model but can be included implicitly as correction factors to the source term. The two differing methodologies and sets of included information complicate a direct comparison between the two.

One possible comparison is to use a process model to develop a prediction more analogous to statistical-model predictions. A process model can identify areas where the predicted concentration exceeds a given threshold concentration. These areas can be compared to areas where the statistical models predict that concentrations are likely to exceed that same threshold concentration (corresponding to predicted probabilities exceeding 0.5). However, uncertainties associated with nitrogen sources and attenuation processes explicitly represented in the process model would limit the utility of this kind of comparison.

The basin-scale patterns of model-prediction values can be used to show how differing sets of variables can affect the

agreement between statistical and process models; this more limited comparison can show if the two methods qualitatively agree on where predictions of nitrogen concentration and occurrence are greater than a common measure. The patterns can be related by identifying areas where a given prediction is greater or less than the basin-wide median for that type of prediction; in this approach, attenuation processes do not affect results under the reasonable assumption that those processes, though unknown, are spatially consistent. The parts of the basin where simulated nitrate concentrations (from a process model) after 50 years of transport exceed the median value of 2 mg/L as N for the basin are in agricultural areas in the northeastern, southeastern, and southwestern parts of the basin (fig. 42B). Areas where probabilities of exceeding 1 mg/L as N, estimated on the basis of contributing recharge areas, are greater than the median value of 0.427 are primarily near the center of the basin and overlap about 42 percent of the area where nitrate concentrations (predicted by a process model) exceed the median for the basin (fig. 46A); the two have the same total areas, because areas greater and less than median values are equal in size. The two areas did not
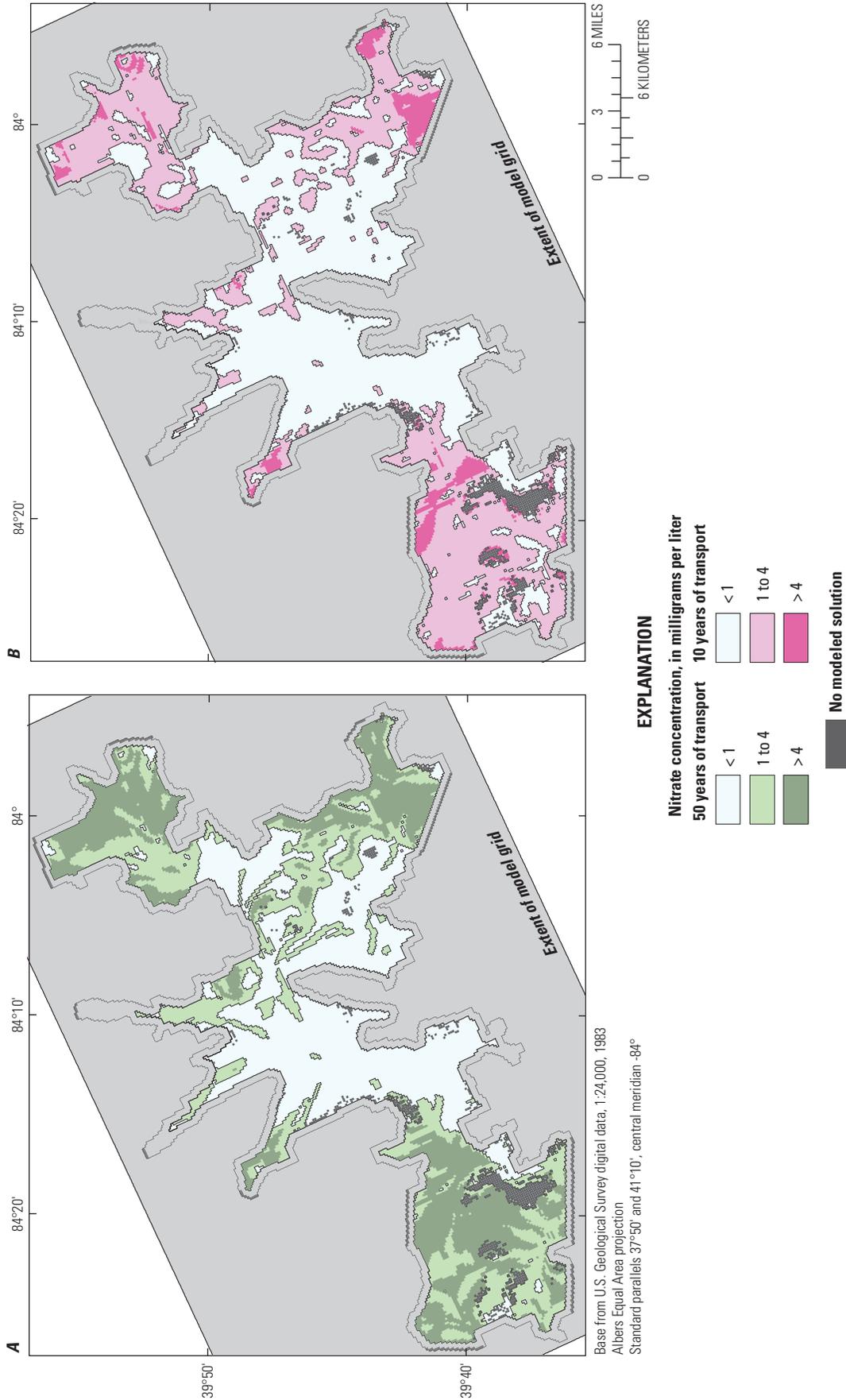
Figure 44.   Areas where nitrate concentrations are predicted to exceed 1 and 4 milligrams per liter as nitrogen by a process model based on transport times of *A*, 50, and *B*, 10 years, Great Miami River Basin, southwestern Ohio.
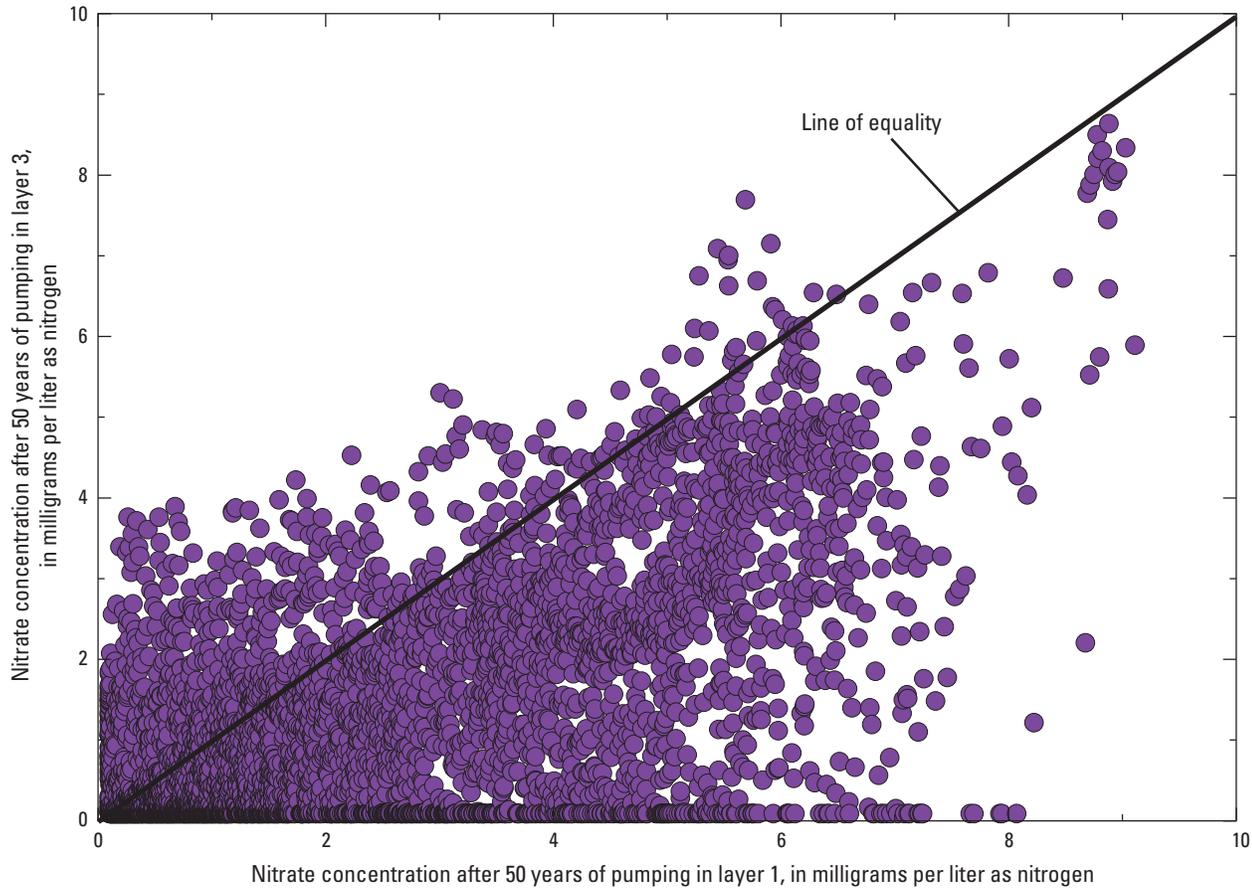
Figure 45.   Relationship between simulated nitrate concentrations for wells pumped in shallow and deep parts of the glacial aquifer, Great Miami River Basin, southwestern Ohio.

generally overlap each other because the explanatory variables for the logistic regression model used to predict probabilities of exceeding 1 mg/L as N did not include a nitrogen-source variable, whereas the process model uses only a nitrogen-source term to estimate concentrations.

Predictions of nitrate occurrence outside of agricultural areas could be more consistent with the process model if the latter incorporated nitrogen inputs from other sources, such as septic systems. A nitrogen-source variable is not included in models of threshold concentrations less than 4 mg/L as N because observed concentrations in network wells could not be correlated with defined sources to an extent that would warrant inclusion in the statistical models (Warner and Arnold, 2010). Areas where probabilities of nitrate concentrations exceeding 4 mg/L as N are greater than the median probability for the basin are more in agreement with areas where concentrations simulated by a process model also exceed the median for the basin—the two overlap by about 62 percent (fig. 46B). The percentage of overlap is higher because both the logistic regression and process models incorporate a nitrogen-source variable. Areas where the agreement improves by inclusion of

the nitrogen-application rate as an explanatory variable are the southwestern and northeastern parts of the basin (fig. 46B).

Both methods used in this study are associated with uncertainties. Statistical models have uncertainties associated with the degree of fit and an inability to represent potential variables and processes that may affect predictions. In this analysis, the process models had uncertainties associated with model parameters as well as with sensitivity to factors for which there is little information, such as attenuation processes prior to recharge. As an example, the lack of agreement between statistical and process models in estimating the extent of concentrations exceeding 4 mg/L as N in the basin could be the result of either the inaccuracy of the statistical models or an incorrect estimate of the amount of attenuation taking place prior to recharge.

Limitations of Analysis

An assumption in this steady-state modeling analysis is that there is a mass balance between recharge and pumping rates so that the volume of water entering the aquifer through

EXPLANATION

Area in which prediction exceeds basinwide median

Nitrate concentration

Probability of exceeding
1 milligram per liter, from recharge area

Probability of exceeding
4 milligrams per liter, from recharge area

No modeled solution

Area excluded
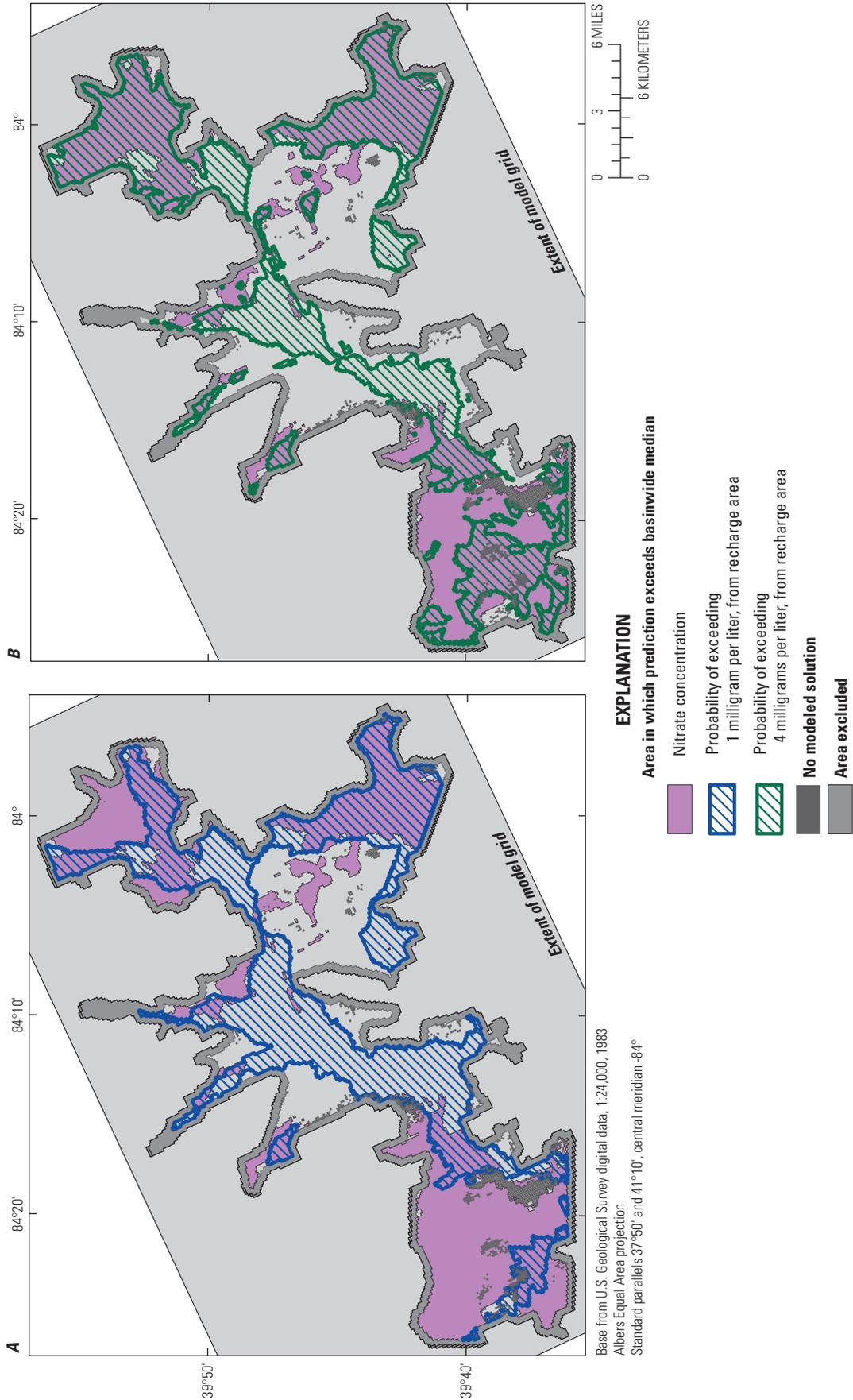
Figure 46. Areas where the probability of exceeding *A*, 1 milligram per liter as nitrogen, and *B*, 4 milligrams per liter as nitrogen exceeds the basinwide median probability from statistical models and areas where simulated nitrate concentrations from a process model exceed the median for the basin, Great Miami River Basin, southwestern Ohio.

the contributing recharge area is the same as the volume being withdrawn from the pumped well. Factors that affect the mass balance between recharge and pumping could affect comparisons of estimates of area-weighted mean variables and resultant predictions of nitrate-exceedance probabilities. Two factors that can affect this balance are hydraulic gradient effects and boundary effects.

Gradient effects are largest in areas of steep hydraulic gradients and are caused by the use of a finite-difference model in the analysis; if model cells are sufficiently large, so that flow through the cell containing the simulated well is larger than the pumping rate, the well is a weak sink and the contributing recharge area—which is the contributing recharge area to the model cell—will be larger than the actual contributing recharge area to the well. Although areas of steep gradients are areas where possible weak sinks need to be considered, these areas also are where a process model most likely can improve estimates of spatial variables over estimates based on a circular buffer. As an example, the contributing recharge area to well 1860 is 2.17 times the size of the circular buffer around the well—under equilibrium conditions, the buffer and the contributing recharge area should be the same size because the well is pumped at a rate equivalent to recharge through the buffer (fig. 47). This well is upgradient from a hydrologic boundary representing a major stream in an area of steep hydraulic gradients.

A boundary effect refers to the influence of a nearby hydrologic boundary on a contributing recharge area and may represent a real hydrologic condition. The result of a boundary effect is that the model cell containing the pumped well receives water from other sources in addition to recharge, usually represented in a model as a specified-head or head-dependent flux boundary; the result is a contributing recharge area smaller than the actual contributing recharge area. As an example, the contributing recharge area to well 1629, which is near a specified-head boundary along the edge of the active model domain, is about 30 percent the size of the circular buffer (fig. 47), indicating that most of the flow through the cell is from the specified-head boundary.

Gradient and Boundary Effects on the Mass Balance of Recharge and Pumping

The degree to which gradient and boundary effects could affect comparisons of estimated spatial variables and resultant predictions is site specific and would be expected to vary spatially. The ratio of the volumetric rate of pumping to the volumetric rate of flow through the contributing recharge area can be used to quantify the degree to which gradient and boundary effects could affect the balance between pumping and recharge. The mean and median of the ratio of pumping to recharge for shallow wells in the basin were 0.98 and 0.96, respectively (fig. 48A); pumping and recharge were within 10 percent for pumped wells in about 50 percent of the basin and within 20 percent in about 65 percent of the

basin, indicating that, generally, there was a reasonable mass balance between recharge and pumping across most of the modeled basin. Groundwater fluxes in porous, unconsolidated aquifers, such as glacial aquifers, increase exponentially with depth, and therefore the relative importance of gradient and boundary effects are a function of well depth. At locations in which wells were pumped in shallow (layer 1) and deep (layer 3) parts of the aquifer, the ratios of pumping to recharge were statistically higher in deep wells (p less than 0.001) with means of 0.97 (shallow) and 1.04 (deep), respectively. The comparison indicates that deeper wells generally were more affected by hydrologic boundaries (indicated by a ratio of pumping to estimated recharge greater than 1), but the relative importance of gradient and boundary effects was variable (fig. 48B).

Large gradient effects, defined as a condition where pumping is less than 50 percent of recharge, occur in pumped wells in about 5 percent of the basin, and large boundary effects, defined as a condition where recharge is less than 50 percent of pumping, occur in about 14 percent of the basin (fig. 49). Wells with large gradient effects generally are near discharge boundaries representing major surface-water features. Wells with large boundary effects generally are near specified-head boundaries along the edge of the active model domain or downgradient from head-dependent flux boundaries where simulated surface-water features are contributing water to the simulated well (fig. 49).

Gradient and Boundary Effects on Estimated Spatial Variables and Statistical-Model Predictions

Steep hydraulic gradients and hydrologic boundaries can affect the mass balance between pumping and recharge and, therefore, estimates of spatial variables and resultant predictions of the probability of exceeding threshold concentrations. The amount of overlap between contributing recharge areas and circular buffers was statistically higher (p less than 0.001) for the population of wells with mass balances within 10 percent than for the population of all wells in the basin; gradient and boundary effects likely would cause less agreement between contributing recharge areas and buffers. The mean and median values of the amount of overlap for the two populations were similar: median overlaps of 0.32 and 0.28 and mean overlaps of 0.34 and 0.33 for all wells and wells with mass balances within 10 percent, respectively (fig. 50).

The nitrogen-application rates within contributing recharge areas of the population of wells with mass balances within 10 percent and the population of all wells in the basin were not statistically different; the relation between area-weighted mean nitrogen-application rates within contributing recharge areas and circular buffers shows a similar degree of variability for both populations (fig. 51A). There was no statistical difference in silt content—estimated from

Figure 47. Contributing recharge areas that are examples of a gradient effect (well 1860) and a boundary effect (well 1629) in the northeastern part of the Great Miami River Basin, southwestern Ohio.
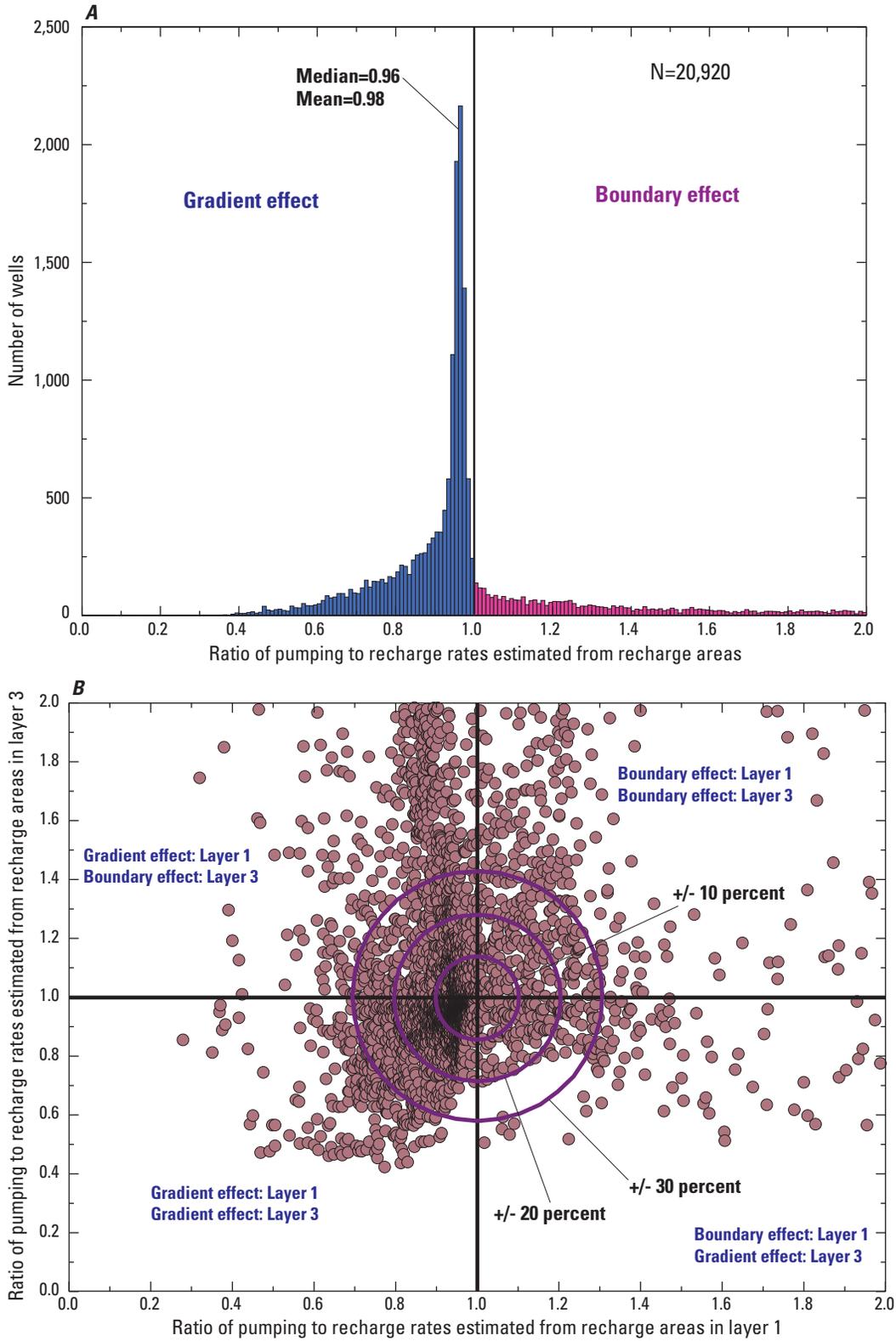
Figure 48.   *A,* Simulated ratio of pumping rates in shallow pumped wells to recharge rates estimated from contributing recharge areas and *B,* relative importance of gradient and boundary effects for shallow (layer 1) and deep (layer 3) pumped wells.

Base from U.S. Geological Survey digital data, 1:24,000, 1983
Albers Equal Area projection
Standard parallels 37°50' and 41°10', central meridian -84°

EXPLANATION
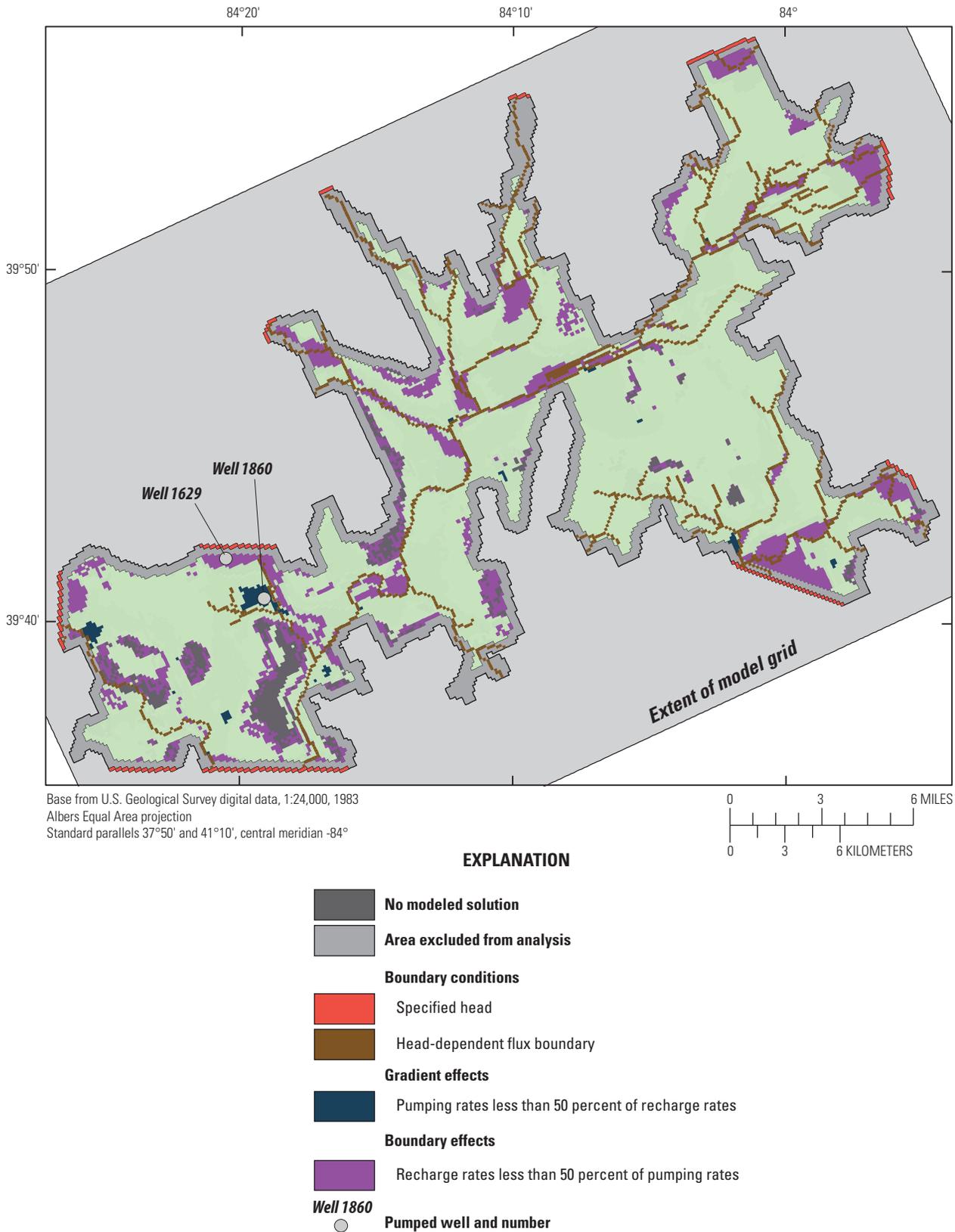
Figure 49.   Distribution of large gradient and boundary effects, defined as differences of more than 50 percent between recharge and pumping rates, Great Miami River Basin, southwestern Ohio.
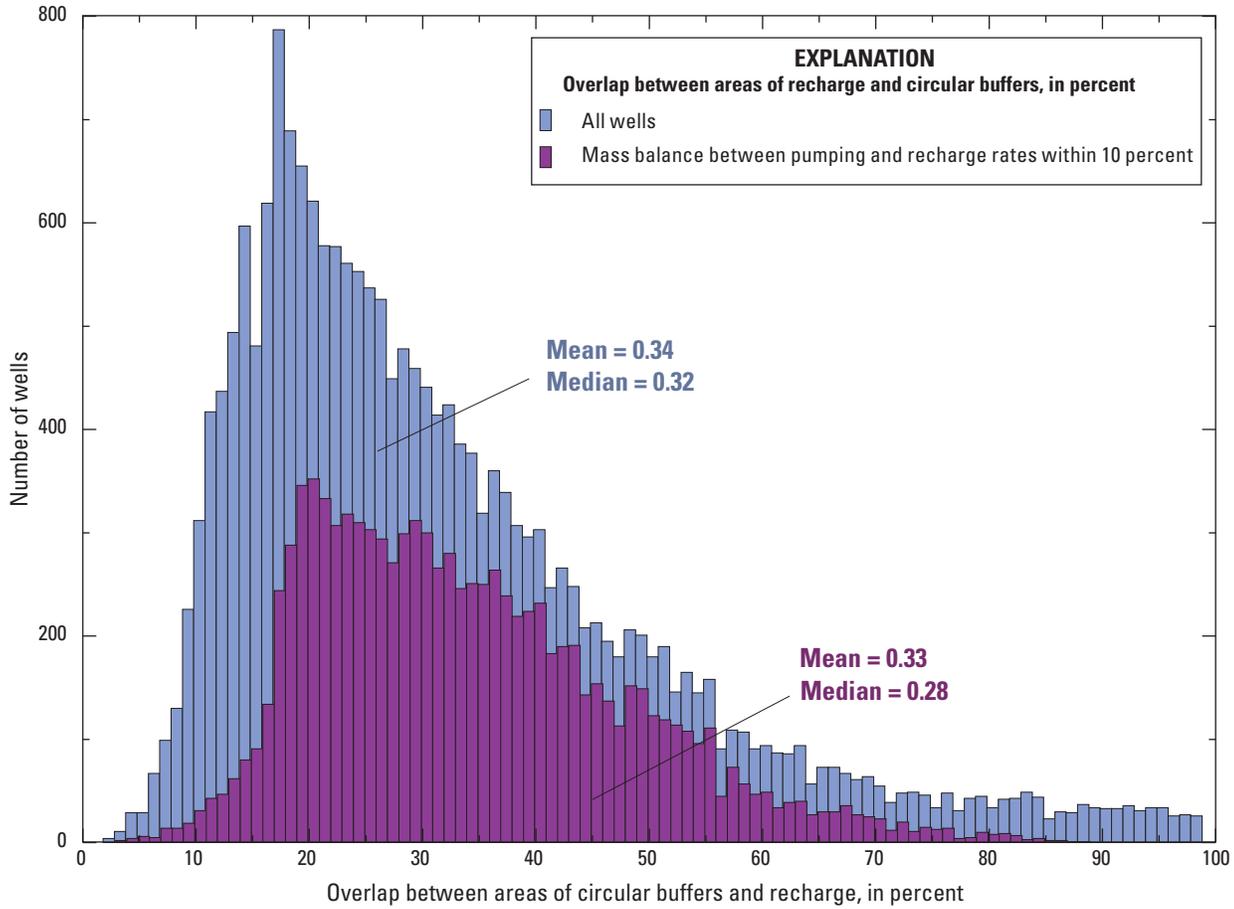
Figure 50.   Overlap between areas of circular buffers and contributing recharge areas for all wells in the basin and those wells with a mass balance between recharge and pumping rates within 10 percent, Great Miami River Basin, southwestern Ohio.
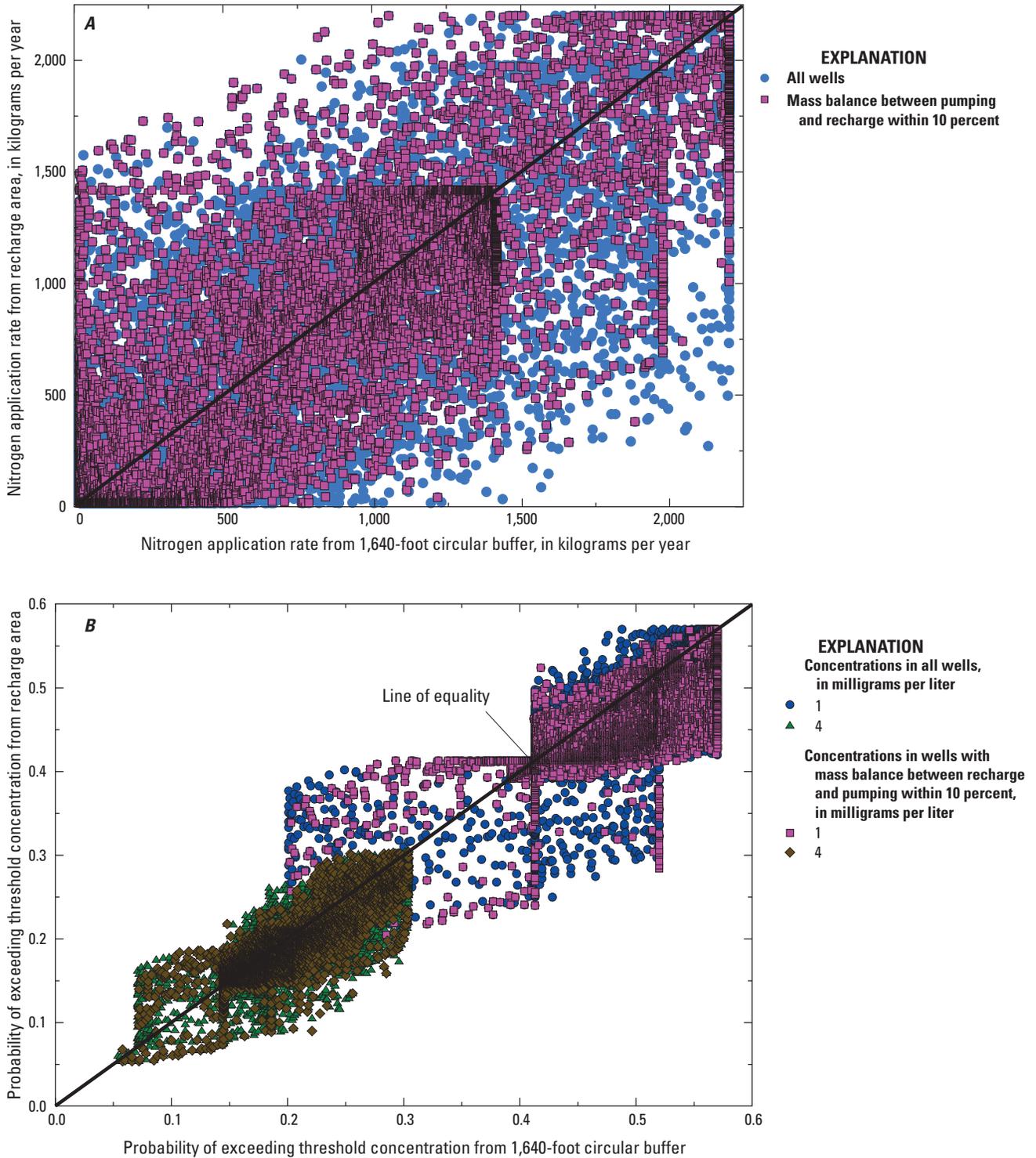
Figure 51. Differences in *A,* estimates of area-weighted spatial variables, and *B,* probabilities of exceeding 1 and 4 milligrams per liter as nitrogen determined for contributing recharge areas and circular buffers for all hypothetical wells in the basin and a subset of wells with a mass balance between pumping and recharge rates within 10 percent, Great Miami River Basin, southwestern Ohio.

STATSGO data—between the two populations. Alfisol content and depth to water were statistically lower (*p* less than 0.001) in wells with mass balances between recharge and pumping within 10 percent than in the population of all wells. Predictions of the probabilities of exceeding 1 and 4 mg/L as N show similar variability between the population of wells with mass balances between recharge and pumping within 10 percent and the population of all wells (fig. 51*B*); however, predicted probabilities for wells with a mass balance within 10 percent were statistically higher (*p* less than 0.001) than for the population of all wells; these differences likely reflect differences in estimates of the spatial data.

The results show that gradient and boundary effects can affect comparisons of statistical and process models arising from different representations of the source of water to a pumped well; however, the differences are small, and general trends and conclusions are not affected. In the regional model of the Great Miami River Basin, gradient and boundary effects (greater than 10 percent) occur in about half the model cells, but effects on estimates of spatial variables and predictions of nitrate occurrence at the basin scale are small. Gradient effects can be minimized by using models with sufficiently small discretization to allow simulated pumping rates to be essentially the only sink within the cell. Boundary effects can be minimized by using, to the extent possible, boundary conditions that do not contribute water to the aquifer; the utility of this approach depends on the system being simulated. Boundary effects may represent real processes, and if there are additional sources of water, the use of statistical models that are based on circular buffers or contributing recharge areas may not be appropriate.

Summary

Unconsolidated glacial sediments consisting of gravel, sand, silt, and clay in the northern conterminous United States compose an important regional aquifer system that, in 2000, supplied water to about 41 million people. About 40 percent of the water was pumped from private wells, primarily in rural areas where agriculture is the predominant land use. Wells in these areas are vulnerable to contamination from nitrogen, primarily in the form of nitrate, from fertilizer application on agricultural land. The USGS, through the NAWQA Program, has characterized the distribution of nitrogen in the glacial aquifer system and used logistic regression based on observed relations between nitrogen concentrations and sets of well-construction and spatial characteristics to develop statistical models of nitrate occurrence. A set of 10 models describes the probability of exceeding concentrations of 1 to 10 milligrams per liter as nitrogen (mg/L as N) by using a set of explanatory variables—well-construction variables (well depth, diameter, and screen length) and spatial variables. The latter set consists of soil characteristics (silt content and alfisol content) and depth to water from the U.S. Department of Agriculture

State Soil Geographic (STATSGO) database and nitrogen-application rates from county data mapped to NLCDe92 agricultural land use. The four explanatory variables used in the statistical models represent area-weighted means computed for a 1,640-foot-radius circular buffer around the well.

Statistical models may implicitly represent transport and attenuation process through inclusion of environmental characteristics but do not explicitly account for complexities in natural systems. By contrast, process models can explicitly represent physical and chemical processes by using numerical methods and have the capacity to represent, to an extent, those complexities. However, the use of process models to represent nitrogen transport requires detailed knowledge of nitrogen sources, attenuation processes from land surface to the point of interest in the aquifer, and the parameters that control transport. Process models can inform and improve statistical models primarily by providing for physically based source areas to pumped wells to account for advective-transport processes; enhanced information on nitrogen sources and nitrogen-attenuation processes that can affect source terms would allow further improvement of statistical models of nitrogen occurrence. To evaluate the potential use of process models to inform statistical models, the two methods were compared at existing and hypothetical wells at a local scale in different hydrologic settings in the Great Miami River Basin, southwestern Ohio, and at a basin scale by using a network of hypothetical wells uniformly spaced at 500-foot intervals across the basin.

Contributing recharge areas, which are a function of local hydrologic setting, were used to improve estimates of explanatory variables over estimates based on the simplified circular buffers used in statistical models. Differences in estimated explanatory variables were used to evaluate how incorporating an enhanced understanding of physical transport could improve statistical models. The larger the difference, the more likely that statistical models would be improved if contributing recharge areas were used to estimate explanatory variables. Comparing model predictions from the two sets of variables would provide an evaluation of the degree to which incorporation of physically based estimates of explanatory variables would affect the statistical models. The set of explanatory variables based on contributing recharge areas, however, was not used to develop the models; a direct comparison would require the use of predictions from separate sets of statistical models developed from the two sets of explanatory variables.

Steady-state contributing recharge areas, which represent a mass balance between recharge and pumping rates, are more accurate representations than circular buffers of the areas at the water table that contribute recharge to a pumped well. If a well is pumped at a rate equivalent to the volume of recharge through the circular buffer, the contributing recharge area to the well—referred to as a volume-equivalent contributing recharge area—is the same size as the circular buffer (about 0.3 square miles for a radius of 1,640 feet). A number of factors affect the size and shapes of contributing recharge

areas and the extent to which they differ from circular buffers, including pumping rate, well depth, hydrologic setting, and length of pumping. Source areas represented as circular buffers differ substantially from source areas represented as contributing recharge areas of equivalent size, and therefore spatial variables estimated for the assumed source areas and the resultant predictions of nitrate would also differ. The differences generally are larger in areas of steep hydraulic gradients where differences in circular buffers and contributing recharge areas are larger. Incorporating an improved understanding of groundwater processes through process models can inform statistical models of nitrate occurrence in glacial environments.

- Calculations of steady-state contributing recharge areas were based on the assumptions that hydraulic stresses are constant over time and that there is a mass balance between pumping and recharge rates; these assumptions are likely valid only after the well has been in operation longer than the maximum steady-state traveltime within the contributing recharge area, in some cases longer than 100 years. Most wells typically are in operation for much less time than is necessary to establish a true steady-state transport condition. Comparisons of steady-state contributing recharge areas with transient contributing recharge areas to a well after different durations of pumping showed that hydrologic conditions simulated for the two types of contributing recharge areas were in reasonable agreement after about 10 to 20 years of simulated pumping, and that the steady-state assumption underlying recharge delineation was valid in most wells.

- Near groundwater divides, where hydraulic gradients are small, volume-equivalent steady-state contributing recharge areas were similar to circular buffers, whereas contributing recharge areas extended upgradient from wells in areas with steep hydraulic gradients and differed greatly in shape from circular buffers. At three hypothetical wells in the same area but in different hydrologic (gradient) settings in the southeastern part of the Great Miami River Basin, the amount of overlap between contributing recharge areas and circular buffers ranged from about 87 percent near a groundwater divide to about 33 percent where hydraulic gradients were steeper.

- Contributing recharge areas to wells pumped in shallow parts of an aquifer were more similar to the areas of circular buffers than the contributing recharge areas to wells pumped in deeper parts of the aquifer; the importance of well depth on the agreement between contributing recharge areas and circular buffers is a function of pumping rate. In an area of steep hydraulic gradients, contributing recharge areas to shallow and deep wells pumped at a high rate (equal to the mean

pumping rate of 0.25 million gallons per day for public-supply wells in the basin), were similar; the amount of overlap differed by about 2 percent. For a low pumping rate of 350 gallons per day (generally consistent with a domestic well), there was a large difference in contributing recharge areas to shallow and deep wells. The contributing recharge area to a shallow well pumped near the water table in an area of steep hydraulic gradients was completely within the circular buffer, whereas the contributing recharge area to a deeper well was more than 2,000 feet outside of the circular buffer. This difference reflects the strong effect of the natural hydraulic gradients on the contributing recharge areas at low pumping rates.

- The mean depth of domestic wells in the NAWQA sampling network is 91 feet, similar to the deep well simulated in the model. This well depth coupled with the low pumping rates typical of domestic wells suggests that contributing recharge areas to these wells could be outside of the circular buffer. However, the contributing recharge areas to wells pumped at different depths at a low pumping rate are within the corresponding volume-equivalent contributing recharge areas. These results suggest that the use of volume-equivalent contributing recharge areas not only better represents groundwater-flow processes, but also more likely includes the actual source areas to wells pumped at low rates regardless of depth.

Differences in contributing recharge areas result in differences in area-weighted means for the four explanatory spatial variables—STATSGO data (silt content, alfisol content, and depth to water) and nitrogen-application rate. These differences, in turn, affect predicted probabilities of nitrate occurrence. Estimates of area-weighted mean spatial variables depend on the same factors that affect the contributing recharge areas to a well—hydrologic setting, well depth, and pumping rate.

- Differences in estimates of area-weighted means for spatial variables varied among three hypothetical wells in different hydrologic settings in the southeastern part of the basin; area-weighted means based on STATSGO data were the same for a well near a groundwater divide, where the contributing recharge area is similar to the circular buffer, but differed substantially in an area with steep hydraulic gradients. The largest variability was for the nitrogen-application rate, which was determined from the Enhanced National Land Cover Data 1992 (NLCDE92) data.

- For wells pumped at low rates in an area of steep hydraulic gradients, well depth had the largest effect on estimated spatial variables, reflecting the large differences in contributing recharge areas. As an example, area-weighted mean nitrogen-application rates in contributing recharge areas to shallow and

deep wells pumped at a high rate were 731 and 743 kilograms per year (kg/yr), respectively, but 1,111 and 0 kg/yr, respectively, for wells pumped at a low rate.

- The differences in estimated spatial variables resulted in differences in the predicted probabilities of nitrate occurrence, which varied less than the estimated variables because variability in the spatial variable only partly explained variability in the predictions with the remainder being a function of well-constriction variables. Differences also were a function of the set of explanatory variables included in a given statistical model. As an example, predictions of nitrate exceeding 1, 2, and 3 mg/L as N were the same for three hypothetical wells in different hydrologic settings because the models included only explanatory variables from STATSGO data, whereas predictions of exceeding 4 mg/L as N differed because nitrogen-application rate, which varied the most, was included as an explanatory variable in the model.

- Estimates of area-weighted means for circular buffers and contributing recharge areas differed substantially between public-supply and NAWQA network wells in the Great Miami River Basin. Spatial variables for the buffers and the contributing areas were weakly correlated with large differences between the two sets of estimates. Predictions of nitrate exceeding 1 and 4 mg/L as N were more correlated, but there were large differences among individual wells.

The overlap between contributing recharge areas and circular buffers is highly site specific and varies greatly across the basin with hydrologic setting. Differences in estimates of area-weighted means are a function of not only hydrologic setting but also spatial variability in the data used to estimate the area-weighted means. These differences also result in basin-scale differences in the predicted probabilities of nitrate exceeding threshold concentrations. The differences in estimated explanatory variables and resultant statistical-model predictions between contributing recharge areas and circular buffers indicate that the incorporation of physically based contributing recharge areas likely would result in a different set of explanatory variables and an improved set of statistical models. The greatest improvement in the statistical models likely would be in areas with steep hydraulic gradients.

- In the Great Miami River Basin, the amount of overlap between circular buffers and volume-equivalent contributing recharge areas ranged from less than 5 to more than 95 percent and generally varied with location relative to groundwater divides and discharge locations. The mean and median amounts of overlap between contributing recharge areas and

circular buffers for the basin were 34 and 32 percent, respectively, indicating that there are substantial differences in the two representations of source areas to wells.

- The general lack of agreement between estimates of area-weighted mean explanatory variables for circular buffers and volume-equivalent contributing recharge areas resulted in large differences in these estimates across the basin. Differences in estimated variables derived from STATSGO data (silt content, alfisol content, and depth to water) were correlated with the spatial variability of that dataset; the largest differences were in the central part of the basin. The largest variability was for estimates of nitrogen-application rates. The differences ranged from 1,880 kg/yr (a higher mean within circular buffers) to -1,560 kg/yr and generally were highest in the northeastern and southeastern parts of the basin—generally correlated with the NLCDe92 data from which the estimates were derived.

- The predicted probabilities of nitrate exceeding 1 and 4 mg/L as N on the basis of spatial variables estimated for circular buffers and contributing recharge areas differed in about 65 and 90 percent of the basin; spatial patterns in predictions were related to spatial differences in estimates of spatial variables and were significant in some areas (exceeding 0.45 and 0.25, respectively).

- Soil characteristics (silt content and alfisol content) and nitrogen-application rate were external to the models, and information about these variables was limited to the underlying spatial data. However, depth to water is a function of local groundwater conditions, and a groundwater-flow model was considered able to inform statistical models by providing for more physically based estimates of depth to water. Depth to water from the STATSGO database ranged from 2.48 to 4.88 feet in the Great Miami River Basin, whereas depth to water calculated by the model ranged from about 0 to 78 feet. Based on area-weighted means for circular buffers and only model-calculated depths to water inside the circular buffer that were within the range of the STATSGO data, predictions of concentrations exceeding 1 and 4 mg/L as N differed greatly. This result suggests that use of a physically based depth to water could greatly change the development of statistical models.

- Contributing recharge areas to wells pumped in deeper parts of the aquifer overlapped less with circular buffers than did contributing recharge areas to shallow

wells. The mean and median amounts of overlap for deep wells were 0.24 and 0.27. Contributing recharge areas to most (about 65 percent) deep wells did not overlap any part of the corresponding circular buffers and, depending on hydrologic setting, were far from the buffers. The greater difference between circular buffers and contributing recharge areas to deep wells resulted in larger differences in estimates of spatial variables and resulting model predictions.

Groundwater-flow models coupled with mass-weighted particle tracking can simulate nitrate concentrations in groundwater transported by advection from surface sources. With an assumed a crop uptake of 50 percent and additional losses of 30 percent(consistent with published values) and the nitrogen-application rates used in the statistical models, the median simulated nitrate concentration in the Great Miami River Basin was similar to that measured in the nationwide glacial aquifer system—about 2 mg/L as N. Direct comparisons between results from statistical and process models cannot be made because statistical models predict the probability of exceeding a given concentration and not concentration values. The two methods generally are in poor agreement, but the respective uncertainties and limitations of the two methods would make the comparison difficult to analyze.

- The parts of the basin where probabilities of exceeding 1 mg/L as N exceeded the basinwide median prediction (about 0.43) overlapped about 42 percent of the parts of the basin where simulated nitrate concentrations exceeded the median nitrate concentration. For a concentration of 4 mg/L as N, the amount of overlap between areas where both statistical-model predictions and process-model predictions exceeded the basinwide median prediction, and simulated nitrate concentrations exceeded the median simulated nitrate concentration, was about 62 percent.

- The improved agreement resulted from the statistical model describing nitrate occurrence above 1 mg/L as N not including nitrogen-application rate as a variable, which was the only control on the distribution of simulated concentrations from the process model. The statistical model describing nitrate occurrence above 4 mg/L as N did include the nitrogen-source variable, resulting in more agreement in agricultural areas.

A model-calculated contributing recharge area better represents the source area to a well and is more likely to include the actual source area to a well with a low pumping rate; however, the use of a process model has limitations. An implicit assumption in using a contributing recharge area to define spatial variables that can affect water quality in a well is that pumping and recharge rates are in a state of mass balance.

Two possible factors that can affect the utility of a finite-difference model are gradient effects and boundary effects, both of which can affect simulated contributing recharge areas and, in turn, estimates of spatial variables and resultant model predictions.

- Gradient effects refer to a condition where pumping within a model cell is less than total flow through the cell and generally occur in areas with large hydraulic gradients. If the pumping rate is equivalent to the volume of recharge through the circular buffer, the result is a contributing recharge area larger than the circular buffer. Gradient effects are a function of model discretization and, in the model of the Great Miami River Basin, pumping was less than 50 percent of flow through the contributing recharge area in about 5 percent of the basin; these areas were generally upgradient from discharge locations. Gradient effects can be minimized by using a finer model discretization.

- Boundary effects refer to a condition where there are other sources of water to a cell containing a pumped well, resulting in a contributing recharge area that is smaller than it would be if pumping and recharge were in mass balance. Boundary effects occur near simulated hydrologic features and may represent real hydrologic processes, in which case the chosen model representation of the source area either as a circular buffer or contributing recharge area may not be appropriate. In about 14 percent of the area of the Great Miami River Basin, recharge was less than 50 percent of the total rate of inflow to wells; these boundary effects were in areas generally near the margins of the simulated aquifer and near simulated surface waters. Wells pumped in deep parts of the aquifer generally were more likely to be associated with boundary effects.

- Mass balances between pumping and recharge in the Great Miami River Basin are within 20 percent in about 65 percent and within 10 percent in about half of the model cells. The amount of overlap for the population of wells with mass balances within 10 percent were similar to the rates for the population of all wells. Estimates of spatial variables and predictions of nitrate occurrence exceeding 1 and 4 mg/L as N for the population of wells with a mass balance within 10 percent and the population of all wells generally were similar, indicating that boundary and gradient effects occur within the modeled area but generally do not affect analytical results. The importance of gradient and boundary effects is model specific and needs to be evaluated to see if changes in the model are warranted.

References Cited

Barlow, P.M., 1997, Particle-tracking analysis of contributing areas of public-supply wells in simple and complex flow systems, Cape Cod, Massachusetts: U.S. Geological Survey Water Supply Paper 2434, 66 p.

Bundy, L.G., and Andraski, T.W., 2005, Recovery of fertilizer nitrogen in crop residues and cover crops on an irrigated sandy soil: Soil Science Society of America Journal, v. 69, p. 640–648.

DeSimone, L.A., Hamilton, P.A., and Gilliom, R.J., 2009, Quality of water from domestic wells in principal aquifers of the United States, 1991–2004—Overview of major findings: U.S. Geological Survey Circular 1332, 48 p.

Dumouchelle, D.H., 1998, Simulation of ground-water flow, Dayton area, southwestern Ohio: U.S. Geological Survey Water-Resources Investigations Report 98–4048, 57 p.

Freeze, R.A., and Cherry, J.A., 1979, Groundwater: Englewood Cliffs, N.J., Prentice Hall, Inc., 604 p.

Harbaugh, A.W., Banta, E.R., Hill, M.C., and McDonald, M.G., 2000, MODFLOW-2000, the U.S. Geological Survey modular ground-water model—User guide to modularization concepts and the ground-water flow process: U.S. Geological Survey Open-File Report 00–92, 121 p.

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

Hem, J.D., 1985, Study and interpretation of the chemical characteristics of natural water (3d ed.): U.S. Geological Survey Water Supply Paper 2254, 263 p.

Hill, M.C., Banta, E.R., Harbaugh, A.W., and Anderman, E.R., 2000, MODFLOW-2000, the U.S. Geological Survey modular ground-water model—User guide to the observation, sensitivity, and parameter-estimation processes and three post-processing programs: U.S. Geological Survey Open-File Report 00–184, 210 p.

Kauffman, L.J., Baehr, A.L., Ayers, M.A., and Stackelberg, P.E., 2001, Effects of land use and traveltime on the distribution of nitrate in the Kirkwood-Cohansey Aquifer System in southern New Jersey: U.S. Geological Survey Water-Resources Investigations Report 01–4117, 49 p.

Masterson, J.P., Hess, K.M., Walter, D.A., and LeBlanc, D.R., 2002, Simulated changes in the sources of ground water for public-supply wells, ponds, streams, and coastal waters on western Cape Cod, Massachusetts: U.S. Geological Survey Water-Resources Investigations Report 2002–4143, 12 p.

Masterson, J.P., Walter, D.A., and LeBlanc, D.R.,1998, Delineation of contributing areas to selected public-supply wells, western Cape Cod, Massachusetts: U.S. Geological Survey Water-Resources Investigations Report 98–4237, 41 p.

Masterson, J.P., Walter, D.A., and LeBlanc, D.R., 2004, Transient analysis of the source of water to wells—Cape Cod, Massachusetts: Ground Water, v. 42, no. 1, p. 126–134.

Nakagaki, Naomi, Price, C.V., Falcone, J.A., Hitt, K.J., and Ruddy, B.C., 2007, Enhanced National Land Cover Data 1992 (NLCDe 92): Sacramento, Calif., U.S. Geological Survey, accessed April 4, 2011, at http://water.usgs.gov/lookup/getspatial?nlcde92.

Nolan, B.T., and Hitt, K.J., 2003, Nutrients in shallow ground waters beneath relatively undeveloped areas in the conterminous United States: U.S. Geological Survey Water-Resources Investigations Report 02–4289, 17 p.

Orzol, L.L., 1997, User's guide for MODTOOLS—Computer programs for translating data of MODFLOW and MODPATH into Geographic Information Systems files: U.S. Geological Survey Open-File Report 94–464.

Pollock, D.W., 1994, User's guide for MODPATH/MODPATH-PLOT, version 3—A particle tracking post-processing package for MODFLOW, the U.S. Geological Survey finite-difference ground-water flow model: U.S. Geological Survey Open-File Report 94–464, 248 p.

Raun, W.R., and Johnson, G.V., 1999, Improving nitrogen use efficiency for cereal production: Agronomy Journal, v. 91, no. 3, p. 357–363.

Reilly, T.E., and Pollock, D.W., 1993, Factors affecting areas contributing recharge to wells in shallow aquifers: U.S. Geological Survey Water Supply Paper 2412, 21 p.

Sheets, R.A., 2007, Hydrologic setting and ground-water flow simulations of the Great Miami River Basin regional study area, Ohio, section 7 *in* Paschke, S.S., ed., Hydrologic settings and ground-water flow simulations for regional studies of the transport of anthropogenic and natural contaminants to public-supply wells—Studies begun in 2001: U.S. Geological Survey Professional Paper 1737–A, p. 7-1–7-24.

Walter, D.A., 2007, Use of numerical models to simulate transport of sewage-derived nitrate in a coastal aquifer, central and western Cape Cod, Massachusetts: U.S. Geological Survey Scientific Investigations Report 2007–5259, 41 p.

Walter, D.A., Masterson, J.P., and Hess, K.M., 2004, Ground-water recharge areas and traveltimes to pumped wells, ponds, streams, and coastal water bodies, Cape Cod, Massachusetts: U.S. Geological Survey Scientific Investigations Map I–2857, 1 sheet.

Walter, D.A., and Whealan, A.T., 2005, Simulated water sources and effects of pumping on surface and ground water, Sagamore and Monomoy flow lenses, Cape Cod, Massachusetts: U.S. Geological Survey Scientific Investigations Report 2004–5181, 85 p.

Warner, K.L., and Arnold, T.L., 2006, Framework for regional synthesis of water-quality data for the glacial system in the United States: U.S. Geological Survey Scientific Investigations Report 2005–5223, revised May 2006, 6 p.

Warner, K.L., and Arnold, T.L., 2010, Relations that affect the probability and prediction of nitrate concentration in private wells in the glacial aquifer system in the United States: U.S. Geological Survey Scientific Investigations Report 2010–5100, 73 p.

Weiskel, P.K., and Howes, B.L., 1992, Differential transport of nitrogen and phosphorus from septic systems through a coastal watershed: Environmental Science and Technology, v. 26, no. 2, p. 352–360.

Wolock, D.M., 1997, STATSGO soil characteristics for the conterminous United States: U.S. Geological Survey Open-File Report 97–656, 1:250,000-scale digital data, accessed December 3, 2007, at http://water.usgs.giv/GIS/metadata/usgswrd/XML/rech48grd.xml.

THIS PAGE INTENTIONALLY LEFT BLANK

White-Miami study unit

National Water-Quality Assessment study units