USGS
science for a changing world

National Water-Quality Assessment Program

Predicted Nitrate and Arsenic Concentrations in Basin-Fill Aquifers of the Southwestern United States

Scientific Investigations Report 2012–5065

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

Predicted Nitrate and Arsenic Concentrations in Basin-Fill Aquifers of the Southwestern United States

By David W. Anning, Angela P. Paul, Tim S. McKinney, Jena M. Huntington,
Laura M. Bexfield, and Susan A. Thiros

National Water-Quality Assessment Program

Scientific Investigations Report 2012–5065

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

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2012

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

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

To order this and other USGS information products, visit http://store.usgs.gov

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Suggested citation:
Anning, D.W., Paul, A.P., McKinney, T.S., Huntington, J.M., Bexfield, L.M., and Thiros, S.A., 2012, Predicted nitrate and arsenic concentrations in basin-fill aquifers of the southwestern United States: U.S. Geological Survey Scientific Investigations Report 2012–5065, 78 p.

Contents

Executive Summary 1
Introduction 2
Basin-fill Aquifers 2
Regional Analysis 5
Motivation for Study 5
Purpose and Scope 6
Previous Investigations 7
Southwest Principal Aquifers Studies 7
State- and National-Scale Studies of Spatial Distribution and Vulnerability 8
Approach and Methods 9
Compilation and Selection of Groundwater Chemistry Data 11
Compilation and Processing of Explanatory Variables 14
Source Variables 23
Aquifer-Susceptibility Variables 25
Geochemical Variables 28
Random Forest Classifier 28
Strengths and Limitations 29
Algorithm Overview and Application 30
Variable Selection and Classifier Goodness-Of-Fit 32
Evaluation of Conceptual Model 34
Nitrate and Arsenic Classifiers and Predicted Concentrations 35
Nitrate 36
Nitrate Classifier Descriptions and Goodness-Of-Fit 37
Regional Distribution and Depth Variation of Predicted Nitrate Concentrations 41
Comparison of the Nitrate Classifiers and Conceptual Model 43
Effects of Selected Natural and Human-Related Factors on Predicted Nitrate Concentration 50
Nitrate Summary and Vulnerability Assessment 52
Arsenic 53
Arsenic Classifier Descriptions and Goodness-Of-Fit 53
Regional Distribution and Depth Variation of Predicted Arsenic Concentrations 58
Comparison of the Arsenic Classifiers and Conceptual Model 61
Effects of Selected Natural and Human-Related Factors on Predicted Arsenic Concentration 69
Arsenic Summary and Vulnerability Assessment 72
Summary and Conclusions 73
References Cited 73
Appendixes 1 and 2 .... A-1
Appendix 3 .... A-2
Appendix 4 .... A-7
Appendix 5 .... A-12
Appendix 6 .... A-13
Appendix 7 .... A-17

Appendix 8 ..... A-29
Appendix 9 ..... A-30
Appendix 10 ..... A-31
Appendix 11 ..... A-32
Appendix 12 ..... A-33
Appendix 13 ..... A-34
Appendix 14 ..... A-35
Appendix 15 ..... A-36
Appendix 16 ..... A-37
Appendix 17 ..... A-38

Figures

1. Map showing principal aquifers and locations of basins previously studied by the National Water-Quality Assessment Program in the Southwest Principal Aquifers study area 3
2. Map showing spatial distribution of observed nitrate and arsenic concentrations in groundwater samples from basin-fill aquifers in the Southwest Principal Aquifers study area used to train the random forest classifiers 14
3. Graph showing distribution of nitrate and arsenic concentrations in groundwater samples collected from each of the principal aquifers in the Southwest Principal Aquifers study area used to train the random forest classifiers 16
4. Graph showing distribution by aquifer-penetration depth of nitrate and arsenic concentrations in groundwater samples from basin-fill aquifers in the Southwest Principal Aquifers study area used to train the random forest classifiers, 1980–2009 17
5. Map showing biotic communities in the Southwest Principal Aquifers study area 24
6. Maps showing spatial distribution of flow-path variables in an example basin from the Southwest Principal Aquifers study area 26
7. Diagrams showing example of recursive binary partitioning 31
8. Graphs showing statistical distribution of misclassification errors for the random forest classifiers of nitrate concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area 40
9. Map showing predicted nitrate concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area 41
10. Chart showing distribution of predicted nitrate concentrations as a function of distance along generalized groundwater flow path, land development, presence of wetlands, and biotic community for basin-fill aquifers of the Southwest Principal Aquifers study area 51
11. Pie chart showing statistical distribution of the mean inorganic-nitrogen concentration for atmospheric-deposition monitoring sites in the Southwest Principal Aquifers study area 52
12. Graph showing statistical distribution of misclassification errors for random forest classifiers of arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area 57
13. Map showing predicted arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area 60

14. Chart showing distribution of predicted arsenic concentrations as a function of distance along generalized groundwater flow path, geologic setting, and recharge characteristics for basin-fill aquifers of the Southwest Principal Aquifers study area 71

Textbox 1 How is depth treated in the prediction classifiers? 12

Textbox 2 Have human activities in agricultural and urban areas affected arsenic concentrations across the Southwest? 64

Textbox 3 Arsenic accumulation underneath playas 66

Textbox 4 Arsenic and iron oxide interactions under different pH and redox conditions 68

Tables

1. Case-study basins in the Southwest Principal Aquifers study area included in analyses of this report 3
2. Potentially important sources and factors for inclusion in assessments and modeling of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area 8
3. Explanatory variables used in random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area 18
4. Estimated sampling error for random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area 33
5. Relation between observed nitrate and arsenic concentration classes and geochemical variables representing conditions in basin-fill aquifers of case-study basins in the Southwest Principal Aquifers study area 36
6. Standardized importance scores for the prediction and confirmatory random forest classifers of nitrate concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area 38
7. Training-observation classification summary for the prediction and confirmatory random forest classifiers of nitrate concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area 40
8. Statistical distribution of predicted nitrate concentrations, by aquifer-penetration depth, principal aquifer, and state, for basin-fill aquifers in the Southwest Principal Aquifers study area 42
9. Relation between predicted nitrate concentrations and explanatory variables representing conditions for basin-fill aquifers of the Southwest Principal Aquifers study area 45
10. Statistical distribution of predicted nitrate concentrations for relative background conditions, by selected biotic community, in basin-fill aquifers of the Southwest Principal Aquifers study area 47
11. Percentage of basin-fill aquifer model grid cells in the Southwest Principal Aquifers study area that are predicted to exceed the relative background nitrate concentration, by percentage of agricultural and urban land use in the model grid cell 48
12. Standardized importance scores for the prediction and confirmatory random forest classifiers of arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area 54

13. Training observation classification summary for the prediction and confirmatory random forest classifiers of arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area 56
14. Statistical distribution of predicted arsenic concentrations, by aquifer-penetration depth, principal aquifer, and state, for basin-fill aquifers in the Southwest Principal Aquifers study area 59
15. Relation between predicted arsenic concentrations and explanatory variables representing conditions for basin-fill aquifers of the Southwest Principal Aquifers study area 62
A1. Explanatory variable data and observed, predicted, and misclassification error data for observations in the training dataset for random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers in the Southwest Principal Aquifers study area A1
A2. Explanatory variable data and predicted concentration class data for observations in the prediction dataset for random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area A1
A3. Summary statistics for explanatory variables in the prediction and training datasets used in the random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area A2
A4. Count of training observations and average misclassification errors by basin for the prediction classifers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area A7
A5. Count of training observations and average misclassification errors by basin for the confirmatory classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area A12
A6. Average misclassification error and count of training observations by percentile range for explanatory variables used in the prediction classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifer study area A13

Conversion Factors

Inch/Pound to SI

Multiply	By	To obtain
	Length	
inch (in)	2.54	centimeter (cm)
inch (in)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Area	
acre	4,047	square meter (m^2)
acre	0.4047	hectare (ha)
acre	0.4047	square hectometer (hm^2)
acre	0.004047	square kilometer (km^2)
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
	Volume	
acre-foot (acre-ft)	1,233	cubic meter (m^3)
acre-foot (acre-ft)	0.001233	cubic hectometer (hm^3)
	Flow rate	
acre-foot per year (acre-ft/yr)	1,233	cubic meter per year (m^3/yr)
acre-foot per year (acre-ft/yr)	0.001233	cubic hectometer per year (hm^3/yr)
inch per year (in/yr)	25.4	millimeter per year (mm/yr)
	Mass	
pound (lb)	0.4536	kilogram (kg)
pound (lb)	453.6	gram (g)
	Application rate	
pounds per acre per year [(lb/acre)/yr]	1.121	kilograms per hectare per year [(kg/ha)/yr]

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

°F=(1.8×°C)+32

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

°C=(°F-32)/1.8

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

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

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

Acronyms

DO	dissolved oxygen
GIS	Geographic Information System
MRLs	minimum reporting levels
NED	National Elevation Dataset
REDOX	reduction-oxidation
STATSGO	State Soil Geographic

Organizations

NAWQA	National Water-Quality Assessment
NWIS	USGS National Water Information System
SWPA	Southwest Principal Aquifers
USEPA	U.S. Environmental Protection Agency
USGS	U.S. Geological Survey

Units of measurement

kg/yr	kilogram per year
km	kilometer
m	meter
mg/L	milligram per liter
mm/yr	millimeter per year
µg/g	microgram per gram
µg/L	microgram per liter

Predicted Nitrate and Arsenic Concentrations in Basin-Fill Aquifers of the Southwestern United States

By David W. Anning, Angela P. Paul, Tim S. McKinney, Jena M. Huntington, Laura M. Bexfield, and Susan A. Thiros

Executive Summary

Human-health concerns and economic considerations associated with meeting drinking-water standards motivated a study of the vulnerability of basin-fill aquifers to nitrate contamination and arsenic enrichment in the southwestern United States. Statistical models were developed by using the random forest classifier algorithm to predict concentrations of nitrate and arsenic across a model grid that represents about 190,600 square miles of basin-fill aquifers in parts of Arizona, California, Colorado, Nevada, New Mexico, and Utah. The statistical models, referred to as classifiers, reflect natural and human-related factors that affect aquifer vulnerability to contamination and relate nitrate and arsenic concentrations to explanatory variables representing local- and basin-scale measures of source, aquifer susceptibility, and geochemical conditions. The classifiers were unbiased and fit the observed data well, and misclassifications were primarily due to statistical sampling error in the training datasets.

The classifiers were designed to predict concentrations to be in one of six classes for nitrate, and one of seven classes for arsenic. Each classification scheme allowed for identification of areas with concentrations that were equal to or exceeding the U.S. Environmental Protection Agency drinking-water standard. Whereas 2.4 percent of the area underlain by basin-fill aquifers in the study area was predicted to equal or exceed this standard for nitrate (10 milligrams per liter as N; mg/L), 42.7 percent was predicted to equal or exceed the standard for arsenic (10 micrograms per liter; µg/L). Areas predicted to equal or exceed the drinking-water standard for nitrate include basins in central Arizona near Phoenix; the San Joaquin, Inland, and San Jacinto basins of California; and the San Luis Valley of Colorado. Much of the area predicted to equal or exceed the drinking-water standard for arsenic is within a belt of basins along the western portion of the Basin and Range Physiographic Province in Nevada, California, and Arizona. Predicted nitrate and arsenic concentrations are substantially lower than the drinking-water standards in much of the study area—about 93.0 percent of the area underlain by basin-fill aquifers was less than one-half the standard for nitrate (5.0 mg/L), and 50.2 percent was less than one-half the standard for arsenic (5.0 µg/L).

The classifiers for nitrate and for arsenic were consistent with previously published conceptual models of natural and human-related factors affecting each constituent. Prediction accuracy for both classifiers was most sensitive to variables representing geochemical conditions. While prediction accuracy also was sensitive to the variables representing source and aquifer susceptibility conditions, neither of these two general types of variables was more important than the other overall. Another finding for both constituents was that prediction accuracy was more sensitive to variables representing local conditions within the model grid cell than to comparable variables representing basin-scale conditions. For example, prediction accuracy was more sensitive to local agricultural land use than to agricultural land use in the whole basin.

Several conditions were found to increase the vulnerability of basin-fill aquifers to nitrate contamination. These conditions include fertilizer use, livestock manure production, development of land for agricultural or urban uses, presence of desert legumes, absence of hydric soils or soils with high organic-matter content, presence of soils with high infiltration rates, high rates of water-use for irrigation or public supply from groundwater or surface-water supplies, low natural recharge from precipitation, high mean air temperatures and potential evapotranspiration, and oxic geochemical conditions.

The distribution of predicted nitrate concentrations varied by biotic community, indicating the importance of natural sources and processes in the nitrogen cycle, including nitrogen fixation by desert legumes in the Sonoran Desert. Relative background concentrations determined from areas with minimal agricultural or urban land uses were less than 2.0 mg/L for most biotic communities, except for the Semidesert Grassland, Mojave Desertscrub, Sonoran Desertscrub-Arizona Uplands, and Sonoran Desertscrub-Lower Colorado River biotic communities, where relative background concentrations were determined to be less than 5.0 mg/L. Concentrations exceeding these relative background concentrations are, for the most part, only found in areas with agricultural or urban development. The likelihood of exceeding relative background concentrations increases where large amounts of land are developed for agricultural or urban land uses. Areas dominated by agricultural lands and areas dominated by urban lands have similar percentages of predicted concentrations that exceed relative background concentrations. Where lands are entirely developed for agricultural or urban land uses, about 48 percent of the basin-fill aquifers by area were predicted to exceed relative background nitrate concentrations. About 15 percent of the area with more than 50 percent of the land developed for agricultural and urban uses was predicted to have nitrate concentrations equal to or greater than 10 mg/L. Nearly all areas with wetlands, regardless of the presence of agricultural or urban developed lands, have predicted nitrate concentrations less than 0.50 mg/L.

The importance of human-related sources to nitrate contamination is a contrast to arsenic enrichment, where the sources are primarily natural—the basin-fill sediments and their parent bedrock. Conditions found to increase the vulnerability of basin-fill aquifers to arsenic enrichment include presence of volcanic bedrock in the surrounding mountains, low rates of natural recharge from precipitation, high potential evapotranspiration rates, minimal or absent groundwater outflow from the basin, and geochemical conditions.

An innovation developed in this investigation was the use of variables that allowed for definition of the approximate location of each model grid cell relative to likely groundwater flow paths within each of 422 defined alluvial basins. Such variables included distance to the basin margin, distance to selected geologic units, land-surface slope, and land-surface elevation as a percentile of elevations occurring within the basin. Use of these variables enhanced classifier accuracy and provided information about variation in nitrate and arsenic concentrations along flow paths from the basin margin to the basin lowlands, thereby adding information to the conceptual model for factors affecting those concentrations.

For undeveloped areas outside the Sonoran Desertscrub biotic communities, predicted nitrate concentrations along the upper basin margins, where mountain-front recharge occurs, are typically between 0.5 and 2.0 mg/L, which is comparable to concentrations observed in precipitation. Predicted concentrations in these areas generally decreased along the groundwater flow path to less than 0.50 mg/L in the basin lowlands. For undeveloped areas in Sonoran Desertscrub biotic communities, which have significant loading from nitrogen fixation by desert legumes, predicted concentrations are generally higher—between 1.0 and 5.0 mg/L—and do not vary much along groundwater flow paths from the basin margin to the basin lowlands.

Generally, arsenic concentrations were predicted to be higher in basins where the basin fill was derived from volcanic or crystalline bedrock and lower in basins where the basin fill was derived predominately from carbonate or clastic sedimentary bedrock. Each geologic setting was evaluated under low (less than 1.7 inches per year; in/yr) and high (equal to or greater than 1.7 in/yr) recharge conditions. In each geologic category and recharge condition, arsenic concentrations generally increased along a general flow path originating near the upper basin margins and extending to the basin lowlands. In areas with low recharge that were surrounded primarily by volcanic and crystalline bedrock, the percent basin-fill area where groundwater arsenic concentrations were predicted to exceed 10 µg/L generally increased from 50 near the upper basin margins to 65 percent in the basin lowlands. Under similar recharge conditions in basins surrounded by carbonate and clastic sedimentary bedrock, predicted arsenic concentrations exceeding 10 µg/L increased downgradient from 30 percent near the upper basin margins to 69 percent in the basin lowlands. High recharge conditions generally attenuated arsenic enrichment although a general increase in concentration still was predicted to occur from upper basin margins to the basin lowlands.

Introduction

The National Water-Quality Assessment (NAWQA) Program of the U.S. Geological Survey (USGS) is conducting a regional analysis of water quality in the principal aquifer systems across the United States (Lapham and others, 2005). The Southwest Principal Aquifers (SWPA) study is building a better understanding of the susceptibility and vulnerability of basin-fill aquifers in the region to groundwater contamination by synthesizing baseline knowledge of groundwater-quality conditions in 16 basins previously studied by the NAWQA Program (table 1; fig. 1). The improved understanding of aquifer susceptibility and vulnerability to contamination is assisting in the development of tools that water managers can use to assess and protect the quality of groundwater resources.

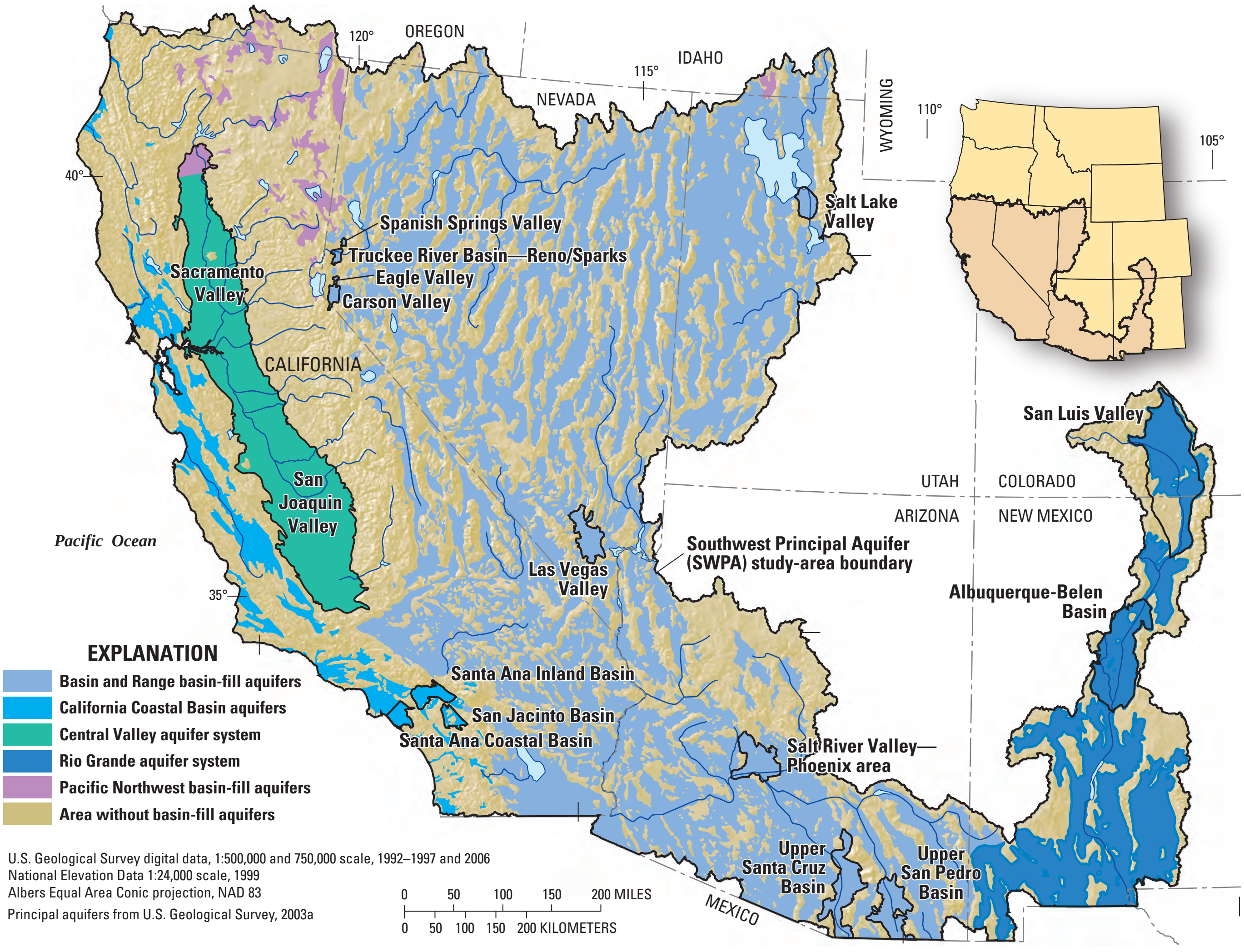
About 46.6 million people live in the SWPA study area (Oak Ridge National Laboratory, 2005), mostly in urban areas, but also in rural, agricultural communities that cultivate about 14.4 million acres of cropland (U.S. Geological Survey, 2003b). Other rural areas contain small communities with mining, retirement, or tourism/recreational-based economies. Because of the generally limited availability of surface-water supplies in the arid to semiarid climate, cultural and economic activities in the region are particularly dependent on good-quality groundwater supplies. In the year 2000, about 33.7 million acre-feet (acre-ft) of surface water was diverted from streams, and about 23.0 million acre-ft of groundwater was withdrawn from basin-fill aquifers in the SWPA study area (U.S. Geological Survey, 2004). Irrigation and public-supply withdrawals from basin-fill aquifers in the study area for 2000 were about 18.0 million acre-ft and 4.1 million acre-ft, respectively, and together account for about one quarter of the total withdrawals from all aquifers in the United States (Maupin and Barber, 2005). Although irrigation and public supply are the primary uses of basin-fill aquifer withdrawals in the study area, water use varies locally by basin, and withdrawals for industrial, mining, and electric power generation also are substantial in some areas.

Basin-fill Aquifers

Basin-fill aquifers underlie about half of the 409,000 square miles (mi^2) SWPA study area and are the primary groundwater supply for most cities and agricultural communities. In

Table 1. Case-study basins in the Southwest Principal Aquifers study area included in analyses of this report.

Case-study basin in this report	Geographically comparable basin from Thiros and others (2010) and Bexfield and others (2011)	State	Principal aquifer system
Albuquerque–Belen Basin	Middle Rio Grande Basin	New Mexico	Rio Grande aquifer system
Carson Valley	Carson Valley	Nevada	Basin and Range basin-fill aquifers
Eagle Valley	Eagle Valley	Nevada	Basin and Range basin-fill aquifers
Las Vegas Valley	Las Vegas Valley	Nevada	Basin and Range basin-fill aquifers
Sacramento Valley	Central Valley (northern part)	California	Central Valley aquifer system
Salt Lake Valley	Salt Lake Valley	Utah	Basin and Range basin-fill aquifers
Salt River Valley–Phoenix area	West Salt River Valley	Arizona	Basin and Range basin-fill aquifers
San Jacinto Basin	San Jacinto Basin of the Santa Ana Basin	California	California Coastal Basin aquifer system
San Joaquin Valley	Central Valley (southern part)	California	Central Valley aquifer system
San Luis Valley	San Luis Valley	Colorado and New Mexico	Rio Grande aquifer system
Santa Ana Coastal Basin	Coastal Basin of the Santa Ana Basin	California	California Coastal Basin aquifer system
Santa Ana Inland Basin	Inland Basin of the Santa Ana Basin	California	California Coastal Basin aquifer system
Spanish Springs Valley	Spanish Springs Valley	Nevada	Basin and Range basin-fill aquifers
Truckee River Basin–Reno/Sparks	Truckee Meadows	Nevada	Basin and Range basin-fill aquifers
Upper San Pedro Basin	Sierra Vista Subbasin	Arizona	Basin and Range basin-fill aquifers
Upper Santa Cruz Basin	Upper Santa Cruz Basin	Arizona	Basin and Range basin-fill aquifers



U.S. Geological Survey digital data, 1:500,000 and 750,000 scale, 1992–1997 and 2006
National Elevation Data 1:24,000 scale, 1999
Albers Equal Area Conic projection, NAD 83
Principal aquifers from U.S. Geological Survey, 2003a

Figure 1. Principal aquifers and locations of basins previously studied by the National Water-Quality Assessment Program in the Southwest Principal Aquifers study area.

several areas, these aquifers provide base flow to streams that support important aquatic and riparian habitats. When aggregated across the study area, the basin-fill aquifers compose five of the principal aquifers of the United States: the Basin and Range basin-fill aquifers in California, Nevada, Utah, and Arizona; the Rio Grande aquifer system in New Mexico and Colorado; the Coastal Basin aquifers and the Central Valley aquifer system in California; and the Pacific Northwest basin-fill aquifers in California, Oregon, and Nevada (fig. 1; U.S. Geological Survey, 2003a). About 55 percent of the area covered by SWPA basin-fill aquifers is located in California (52,450 mi^2) and Nevada (52,030 mi^2). The remaining 86,132 mi^2 of the basin-fill area primarily is located within the States of Arizona (19 percent), Colorado (2 percent), New Mexico (12 percent), and Utah (12 percent). Basin-fill areas in Idaho, Oregon and Texas, although modeled in this report, collectively represent less than 1 percent (1,689 mi^2) of the area covered by SWPA basin-fill aquifers (fig. 1). Numerous groundwater investigations have been conducted in individual basins within the five principal aquifers, several of which formed the basis for the principal-aquifer summaries found in the Groundwater Atlas of the United States (Miller, 1999).

Basin-fill aquifers primarily consist of sand and gravel deposits that partly fill structurally formed depressions and are bounded by consolidated rock mountains. In some areas, fine-grained deposits of silt and clay are interbedded with porous sand and gravel deposits and form confining layers that retard vertical movement of groundwater. Most basins contain thick sequences of basin-fill deposits, and the sediment becomes increasingly more compacted and less permeable with depth. Many basins are drained by a stream that flows through a gap in the consolidated rock; however, some basins are closed, and groundwater and surface water are removed naturally only by evapotranspiration. High-energy mountain streams form alluvial fans with coarse-grained sediment deposited along the mountain fronts. The unsaturated zones below alluvial-fan surfaces usually are several hundred-feet thick and underlain by an unconfined aquifer. Steep alluvial fans transition to a relatively flat valley floor where lacustrine and fluvial depositional environments often have created layers of fine-grained sediment interbedded with more permeable layers of sand and gravel. This usually results in confined conditions and upward vertical gradients in discharge areas in the central part of the basin. Somewhat continuous clay layers occur within about one hundred feet of the land surface in some basins, forming a shallow aquifer system above the uppermost clay layer that can be perched or that can contribute to or receive water from the underlying confined aquifer.

The primary sources of natural recharge to the deeper parts of the basin-fill aquifers are precipitation on the surrounding mountains and infiltration from streams. Runoff from the surrounding mountains seeps into the coarser-grained stream-channel and alluvial-fan deposits near the upper basin margins. Precipitation also can infiltrate the consolidated mountain rock where it is fractured or porous and can move into the basin-fill deposits. Low precipitation rates, combined with high evaporation rates, result in a relatively small contribution of groundwater recharge from precipitation occurring on the basin floor (generally less than 5 percent of annual precipitation). Much of that recharge is focused as infiltration through ephemeral stream channels. Some recharge occurs as subsurface inflow of groundwater from adjacent basins.

Since human development of water resources in alluvial basins, the balance of recharge and discharge has demonstrated increasing disequilibrium. Prior to this time, groundwater discharge typically was by evapotranspiration of shallow groundwater in wetlands or playas (in closed basins) and by discharge to streams flowing through the basin. The cities of Las Vegas, Nevada; Tucson, Arizona; and San Bernardino, California, owe their locations to the availability of groundwater that used to discharge to streams or springs throughout the year. In some basins, natural discharge occurs as subsurface outflow to adjacent basins; however, in other basins, faulting and constrictions in the bedrock that surrounds the aquifer restrict groundwater outflow.

With water development, some basin-fill aquifers have changed considerably as a result of an increase in the amount and number of mechanisms for recharge and discharge. Artificial recharge sources include seepage of irrigation water applied to crops and lawns; seepage from canals, water distribution pipes, sewer pipes and septic systems; infiltration of storm-water runoff from retention basins, recharge basins, and dry wells; seepage of treated wastewater through streambeds or irrigated fields; and infiltration in recharge ponds or well-injection of surface water or imported water. Recharge from these artificial sources introduces water to parts of the groundwater system that previously received little or no recharge from the land surface. For some basins, the increase in recharge and redistribution of water to areas that previously did not receive recharge has resulted in increased saturated thicknesses, increased flow velocities, or changes in flow directions.

For many basins, withdrawal from pumping wells has become the primary source of groundwater discharge and is greater than groundwater recharge. In some basins, the imbalance between recharge and discharge has led to large decreases in groundwater storage, decreases in groundwater discharge to streams and evapotranspiration, or both. Water-level declines and changes in flow directions and magnitudes occur where groundwater withdrawals are large. The increased rates of recharge and discharge associated with water development have increased flow through many basin-fill aquifers, especially from the land surface to the shallower parts of the aquifer. However, groundwater withdrawals from the deep wells typically used for public supply also have resulted in enhanced movement of groundwater from shallower to deeper parts of basin-fill aquifers. Water development, therefore, typically results in aquifers being more susceptible to water-quality degradation by human activities at the land surface and more vulnerable to contaminants where sources are present.

Regional Analysis

Similarities in the hydrogeology, land- and water-use practices, and water-quality issues allow for regional analysis of the vulnerability of basin-fill aquifers to contamination in the SWPA study area. The intrinsic susceptibility to contamination is a direct function of the ease with which water enters and moves through an aquifer, and is dependent on properties and characteristics such as recharge rate, the presence or absence of an overlying confining layer, groundwater travel time, thickness and characteristics of the unsaturated zone, and pumping (Focazio and others, 2002). Aquifers can be susceptible to contamination but are not vulnerable until a contaminant source is introduced. The vulnerability of groundwater to contamination is the probability for contaminants to reach a specific part of an aquifer after being introduced, usually at the land surface. Vulnerability is dependent on the properties of the groundwater system (susceptibility), the proximity of contaminant sources, and the contaminant's chemical characteristics. Long groundwater-residence times and slow rates of contaminant degradation in basin-fill aquifers of the SWPA study area can make the process of contaminating groundwater virtually irreversible and treatment prohibitively expensive or otherwise impractical. It is imperative, therefore, to understand the primary natural and human-related factors associated with the susceptibility and vulnerability of these aquifers to contamination to enable water managers to plan for their optimal protection and utilization.

The SWPA regional analysis of groundwater vulnerability to contamination ultimately began with the first data-collection and analysis phase from 1991 to 2001, when the NAWQA Program sampled wells and established baseline water-quality conditions for basin-fill aquifers in 16 basins across the study area (SWPA case-study basins; fig. 1 and table 1). Groundwater quality also was investigated relative to natural and human-related factors on the basis of a wide suite of constituents including major ions, nutrients, trace elements, pesticides, and volatile organic compounds (VOCs). These studies developed detailed knowledge of local conditions and factors affecting groundwater quality for each basin individually. The SWPA study is developing a regional understanding by synthesizing results from these basin studies into a common set of factors and themes found to affect water quality in basin-fill aquifers across the Southwest. In some cases, different names and slightly different boundaries were used in the different NAWQA basin-specific and synthesis studies. Table 1 shows the crossover for basin names used in this study compared to those used in other studies. The regional synthesis consists of three major components:

1. A review that summarizes current knowledge about the groundwater systems and the status of, trends in, and influential factors affecting groundwater quality of basin-fill aquifers in the 16 case-study basins previously studied by NAWQA (Thiros and others, 2010).

2. Development of conceptual models of the primary natural and human-related factors commonly affecting groundwater quality, leading to a regional understanding of the susceptibility and vulnerability of basin-fill aquifers to contamination (Bexfield and others, 2011).

3. Development of statistical models for prediction of specific constituent concentrations and for further expansion of the understanding of the vulnerability of basin-fill aquifers to contamination resulting from natural and human-related sources and the aquifer's intrinsic susceptibility (this report).

Resource managers and scientists will be able to use the results of the SWPA regional water-quality studies in assessing the vulnerability of groundwater to contamination in both well-studied and sparsely-studied basins across the SWPA study area. By identifying natural and human-related factors and processes affecting the occurrence and transport of selected contaminants, the assessments will allow managers and scientists to apply findings to broader classes of contaminants. Regional-scale models and other decision-support tools that integrate aquifer characteristics, land use, and water-quality monitoring data will help water managers to estimate water-quality conditions in unmonitored areas, assess the vulnerability of groundwater under different future basin-development scenarios, and develop cost-effective groundwater-monitoring programs.

Motivation for Study

The motivation for study of nitrate and arsenic concentrations in basin-fill aquifers in the SWPA study area arose from concerns about human-health issues and economic costs associated with the protection and treatment of drinking water for these constituents, as well as the potential for contaminant concentrations to increase over time and degrade the quality of groundwater in the aquifers as development progresses.

The U.S. Environmental Protection Agency (USEPA) primary drinking-water standard for nitrate is 10 mg/L because of the potential for elevated nitrate to restrict oxygen transport in the blood of infants in a condition known as methemoglobinemia (U.S. Environmental Protection Agency, 2009). This enforceable standard specifies the maximum allowable concentration of a contaminant in drinking water delivered to the consumer by a public-supply system. Recent concern also has arisen over transformation of nitrate within the human body into *N*-nitroso compounds, which are known carcinogens (Ward and others, 2005).

Arsenic has been recognized as a toxic element for centuries and is a human-health concern because elevated concentrations can contribute to a wide variety of adverse health effects, including skin damage and circulatory problems. In addition, arsenic in drinking water can lead to several types of cancers, including bladder, lung, skin, and, possibly, kidney and liver (National Research Council, 2001). The USEPA has set a primary drinking-water standard for arsenic of 10 µg/L (U.S. Environmental Protection Agency, 2009).

The drinking-water standards for nitrate, arsenic, and other compounds are implemented to reduce the risk of human-health problems associated with those compounds; however, from a regulatory standpoint, they do not apply to water from domestic wells. While most of the population in the SWPA study area receives water from public-supply systems regulated by the USEPA, about 1.6 million people in 2005, mostly in rural areas without access to public supply, were reliant on unregulated domestic-supply wells tapping basin-fill aquifers (Maupin and Arnold, 2010). An additional risk associated with domestic-supply wells is that they often are completed in shallower parts of the aquifer, where water quality tends to be adversely affected by activities at the land surface more quickly than in deeper parts of the aquifer. The presence of confining layers and upward hydraulic gradients naturally help to protect deeper parts of basin-fill aquifers from contamination in shallower parts (Bexfield and others, 2011). Although water-supply wells were historically designed to maximize quantity and not necessarily with regard to water-quality concerns, concentrations of nitrate above the drinking-water standard generally were not measured at depths typically used for public supply in the SWPA case-study basins (Bexfield and others, 2011, table 9B).

Potential exists for concentrations of nitrate and arsenic to increase over time and degrade the quality of groundwater in the aquifer where human activities add contaminant sources or increase the susceptibility of the aquifer through water-resources development. Nitrate can be added to a basin-fill aquifer from agricultural sources, such as infiltrating excess irrigation water containing fertilizer and wastewater from animal feedlots, as well as from urban sources, such as wastewater infiltrating from septic tanks and leaking sewer systems and urban runoff (Bexfield and others, 2011, table 9B). In addition, artificial recharge in areas where little natural recharge occurred previously can transport naturally accumulated nitrate from the soil zone to the water table. Deeper parts of basin-fill aquifers can be vulnerable to increasing concentrations of nitrate in areas where there are discontinuous or no confining layers and the presence of downward hydraulic gradients, or where downward hydraulic gradients have been enhanced by modifications to groundwater flow systems caused by artificial recharge and groundwater withdrawals. Also, the long screen intervals of many public-supply and irrigation wells can short circuit groundwater flow through an aquifer by hydraulically connecting different depths, providing preferential pathways for groundwater, and enabling contaminant movement across confining layers (Landon and others, 2009).

The oxidation-reduction (redox) conditions in an aquifer are a primary control on the persistence or fate of nitrate and arsenic in water sampled in the SWPA case-study basins (Bexfield and others, 2011). Changes in the groundwater flow system, such as new sources of recharge and water-level declines caused by pumping, can alter redox conditions. Therefore, water development can increase the vulnerability of basin-fill aquifers in the Southwest to contamination from nitrate, arsenic, or both by mobilizing these constituents to groundwater used for water supply.

The recharge of oxidized irrigation water and groundwater pumping usually results in a fluctuating water table, which can mobilize naturally occurring arsenic sorbed to basin-fill sediment in the groundwater (Jurgens and others, 2009). Arsenic that was naturally concentrated in the unsaturated zone by evapotranspiration in areas that previously received little recharge also can be transported by excess irrigation water to the water table in parts of some basins (Busbee and others, 2009).

Traditional wellhead-protection approaches generally are designed to prevent groundwater contamination through reduction of human-related contaminant sources, such as nitrate. These programs, however, generally do not protect against natural sources of contaminants already present in the aquifer and, therefore, are unlikely to be effective for arsenic. Possible actions available to address elevated concentrations of nitrate or arsenic in groundwater used for drinking include treating the water, blending it with other available sources, importing new sources of water, or abandoning existing wells and drilling new ones. These actions can be costly or strategically difficult to implement compared to taking measures that reduce groundwater vulnerability by mitigating the effects of human alteration of the landscape and aquifers.

Purpose and Scope

This report documents statistical models that relate concentrations of nitrate and arsenic in basin-fill aquifers of the SWPA study area to selected natural and human-related factors representing contaminant sources, aquifer susceptibility, and geochemical conditions. Specifically, this report presents the following:

- The spatial and statistical distribution of nitrate and arsenic concentrations in basin-fill aquifers across the SWPA study area, as determined by using predictions from statistical models.
- An evaluation of previously published conceptual models for the effects of natural and human-related factors on nitrate and arsenic in groundwater. The evaluation provides insight to factors leading to the vulnerability of basin-fill aquifers to nitrate contamination or arsenic enrichment and is achieved through comparison of the conceptual models to diagnostics and predictions from the statistical models.

This report builds upon individual studies of factors that affect water quality, as summarized in Thiros and others (2010), and synthesis of that information into generalized conceptual models for selected contaminants as described by Bexfield and others (2011). Several natural and human-related factors that affect groundwater quality are represented by datasets compiled by McKinney and Anning (2009) as part of the SWPA study.

Previous Investigations

Publications on groundwater quality and aquifer vulnerability to contamination in individual basins are far too numerous to list here; however, only a limited number of studies have provided a regional analysis for basin-fill aquifers in the SWPA study area. While some of those regional investigations were conducted by the SWPA study and lay the foundation for this report, other noteworthy investigations assess groundwater vulnerability by development of predictive models for specific contaminants in select states within the Southwest or for the entire Nation. This section provides a summary of those studies.

Southwest Principal Aquifers Studies

Initial investigations by the SWPA study documented and modeled natural and human-related effects on selected constituents in basin-fill aquifers across the SWPA study area and focused on the shallow part of several aquifers (Paul and others, 2007) or on dissolved-solids concentrations (Anning and others, 2007). These investigations were followed by the two companion reports that lay the foundation for this report—Thiros and others (2010) and Bexfield and others (2011).

Paul and others (2007) used NAWQA data collected for 1993–2004 to investigate water quality of the shallow, upper parts of several basin-fill aquifers in the SWPA study area and found them vulnerable to high nitrate concentrations (greater than 10 mg/L) where fertilizer is used, land is irrigated, and oxidizing conditions are present in the groundwater. Similarly, Paul and others (2007) found that occurrence of selected pesticides is affected by oxidation/reduction conditions, soil permeability, groundwater temperature, and depth to the well's screen interval. Occurrence of selected VOCs was found to be affected by oxidation/reduction conditions, pH, and industrial land use. Anning and others (2007) investigated salinity in many basin-fill aquifers of the SWPA study area and found that dissolved-solids concentrations typically increase along flow paths as a result of geochemical reactions with the aquifer matrix, dissolution of disseminated salts and massive evaporite deposits, and evapotranspiration of shallow groundwater by natural vegetation or by agricultural crops. Mixing with higher concentration inflows of groundwater, stream seepage, or irrigation seepage also causes increases in dissolved-solids concentrations along flow paths.

Thiros and others (2010) summarizes current knowledge about the groundwater systems and the status of, trends in, and influential factors affecting groundwater quality of basin-fill aquifers in the 16 individual basins (fig. 1) previously studied by NAWQA. Each basin description provides information about various influential factors believed to affect the groundwater quality in that basin, including population, land use, water use, recharge and discharge mechanisms, and flow directions. Data for several of the natural and human-related factors presented in Thiros and others (2010) and Bexfield and others (2011), particularly those factors relating to physiography, population, land use, and water use, are compiled in McKinney and Anning (2009) for all 422 basins within the SWPA study area. Several of the variables compiled by McKinney and Anning (2009) also are used in this report.

Summaries of the factors that affect groundwater quality in individual basins by Thiros and others (2010) were synthesized by Bexfield and others (2011) into conceptual models of the primary natural and human-related factors commonly affecting groundwater quality with respect to selected contaminants, thereby helping to build a regional understanding of the susceptibility and vulnerability of basin-fill aquifers to those contaminants. The conceptual models were intended to provide a general understanding of major factors that should be considered in broad-scale characterization of the vulnerability of basin-fill aquifers of the SWPA study area to contamination, and to help guide future efforts at statistical modeling of contaminant occurrence (this investigation; table 2).

Synthesis of information from the 16 SWPA case-study basins indicates that although many commonalities exist with respect to hydrogeology, climate, and other characteristics, multiple factors with the potential to substantially affect groundwater quality exhibit a broad range of conditions among basins (Bexfield and others, 2011). These factors include characteristics related to potential contaminant sources, such as the geologic composition of bedrock in recharge areas, land use within the alluvial basins, and population density. The general distribution, quantity, and mechanisms of groundwater recharge and discharge, which can be very important to aquifer vulnerability, also are quite variable. The variability is related to such factors as how much water (and associated contaminant) is transported to the water table and at what locations, where groundwater flows from or to, and how fast groundwater travels. General aquifer characteristics that affect recharge, groundwater flow, and contaminant persistence (such as the thickness of the overlying unsaturated zone, the presence or absence of effective confining layers, and redox conditions) also exhibit substantial variability among case-study basins.

Bexfield and others (2011) found that differences in important natural and human-related characteristics among the 16 case-study basins are reflected in observed differences in the areal and vertical extent of individual contaminants in the basin-fill aquifers above levels of concern, and in the sources and hydrogeologic controls that have been documented to affect those contaminants. Six relatively common contaminants (dissolved solids, nitrate, arsenic, uranium) or contaminant classes (VOCs, pesticide compounds) were investigated for sources and controls affecting their occurrence and distribution above specified levels of concern in groundwater of the case-study basins, and conceptual models of factors that are important to aquifer vulnerability with respect to those contaminants were subsequently formed. Conceptual models for nitrate and arsenic are summarized in the following paragraphs and in table 2; see Bexfield and others (2011) for summaries of conceptual models for dissolved solids, uranium, VOCs, and pesticide compounds.

Table 2. Potentially important sources and factors for inclusion in assessments and modeling of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.

[Source: Bexfield and others (2011). Factors that previous investigations in some alluvial basins have shown as likely to be important, but that could not be adequately assessed for multiple case-study basins with the currently available information, are shown with a "*." **Abbreviations**: Redox conditions, reduction-oxidation conditions; —, not available]

Natural sources	Anthropogenic sources	Natural hydrogeologic factors	Anthropogenic factors
Nitrate			
Where known, presence of soil-zone accumulations of nitrate (resulting from natural physical and biological processes) in areas where recharge could occur periodically	Presence of agricultural sources as a whole or individually	Rate of evapotranspiration	Depth to water in areas of artificial recharge
—	Presence of urban sources as a whole or individually	Redox conditions	Contribution of artificial recharge to overall basin groundwater budget
—	—	Presence of confining units or naturally upward hydraulic gradients	Magnitude of pumping stresses
—	—	—	Well depth
—	—	—	Presence of urban recharge in areas of previous agriculture
Arsenic			
Presence of high-arsenic rocks/sediments within the recharge area	None	Redox conditions *	None
Presence of high-arsenic rocks/sediments within the aquifer	—	pH values *	—
Presence of saline (often geothermal) water sources within or adjacent to basin *	—	Groundwater residence time	—

Information synthesized by Bexfield and others (2011) indicates that nitrate concentrations exceeding 5.0 mg/L were common at shallow aquifer depths in either localized or broad areas of many basin-fill aquifers. Although natural sources involving soil-zone accumulations of nitrate have been documented in a few basins, human-related sources are the primary contributors—in particular, excess irrigation water infiltrating through both agricultural fields and urban turf areas where fertilizer has been applied, and seepage of water from sewer and septic systems. Other common contributors are agricultural wastewater, urban wastewater applied to crops or urban turf, and diffuse urban runoff. Hydrogeologic and geochemical factors most commonly affecting nitrate concentrations include redox conditions in the aquifer, substantial human modification of aquifer recharge or discharge processes, the presence of a shallow water table in areas of high artificial recharge, evapotranspiration of recharge (both natural and artificial), and the presence of confining layers or upward hydraulic gradients that help to protect the deeper aquifer. Another factor of importance in several case-study basins is the occurrence of urban recharge in areas of previous agricultural activity.

Arsenic concentrations exceeding 5.0 µg/L were found by Bexfield and others (2011) to be common to many basin-fill aquifers across varying areal extents. Where data are sufficient to assess the vertical extent of elevated arsenic concentrations, they are observed at most or all aquifer depths. Elevated arsenic concentrations in groundwater are attributable primarily to high-arsenic rocks or sediments within the aquifers and their recharge areas. Although further investigation of the hydrogeologic and geochemical factors resulting in release of arsenic from rocks and sediments is needed in the SWPA study area, available studies suggest that important factors include redox conditions, pH, and the presence of groundwater with long residence times.

State- and National-Scale Studies of Spatial Distribution and Vulnerability

A limited number of state and national-scale studies have presented the spatial distribution of nitrate and arsenic concentrations in groundwater or aquifer vulnerability to nitrate contamination or arsenic enrichment. The nitrate studies focused on the vulnerability to contamination as determined from logistic regression or nonlinear regression models, whereas studies of arsenic did not develop such models and only presented the spatial distribution of observed concentrations.

The NAWQA Program has supported several national-scale investigations of nitrate and other nutrients in groundwater and surface water of the Nation and results of those studies are available online (U.S. Geological Survey, 2010a). Included in those studies was an assessment of the vulnerability of the Nation's shallow groundwater and drinking water to nitrate contamination (Nolan and Hitt, 2006). In that assessment, two nonlinear regression models were developed at the national scale to (1) predict nitrate concentration in recently recharged groundwater in the shallow uppermost part of the aquifer, about 25 ft below land surface, and (2) predict ambient nitrate concentration in deeper supplies used for drinking, about 160 ft below land surface. The models have a mechanistic structure that segregates the effects on concentration from different nitrogen sources and different physical factors that enhance or restrict nitrate transport and accumulation in groundwater. Factors contributing to high nitrate concentrations in the models include high nitrogen application rate, high water input, well-drained soils, fractured rocks or those with high porosity, and lack of denitrification or dilution by natural or artificial recharge waters (Nolan and Hitt, 2006). Notable areas, with predicted concentrations of nitrate greater than 10 mg/L in shallow groundwater within basin-fill aquifers of the SWPA study area, include parts of the Central Valley, Palo Verde Valley, Salinas Valley, Salton Sea, and an assortment of areas in southern coastal parts of California; the Yuma Valley in Arizona; and the San Luis Valley of Colorado (Nolan and Hitt, 2006). These areas generally also were predicted to have nitrate concentrations greater than 10 mg/L in deeper parts of the aquifer used for drinking water.

Rahman and Uhlman (2009) developed nitrate contamination vulnerability maps of groundwater in basin-fill aquifers and other aquifers in Arizona. The study predicted the probability of exceeding nitrate concentrations greater than 3.0 mg/L, 5.0 mg/L, and 10 mg/L from logistic regression models. The explanatory variables found to be significant in the three exceedance models had little variation and reflected land-surface slope, water-use type, well density, precipitation, geology, land cover, sewage-treatment type, population density, occurrence within a groundwater-management area or in an irrigation district, and proximity to streams, lakes, or point sources. Reported overall misclassification rates were 27 percent for the 3.0 mg/L exceedance model, 23 percent for the 5.0 mg/L exceedance model, and 10 percent for the 10 mg/L model. On the basis of model predictions, the study found that 13 percent of the groundwater in Arizona had an 80 percent probability of exceeding 3.0 mg/L, 9 percent had an 80 percent probability of exceeding 5.0 mg/L, and less than 1 percent had an 80 percent probability of exceeding 10 mg/L. Identified locations of concern included agricultural lands surrounding the towns of Buckeye, Casa Grande, Chandler, and northwestern Phoenix.

Lopes (2006) investigated the quality of Nevada's aquifers (1990–2004) and their vulnerability to contamination. The study found nitrate concentrations greater than 10 mg/L in groundwater most frequently occurred in urban areas. In attempts to develop logistic regression equations that related the probability of nitrate exceeding 2.0 mg/L and other thresholds to explanatory variables, such as precipitation, land-surface slope, soil properties, land use and geology, Lopes (2006) found relations to be weak and determined that the models performed little better than random prediction. Lopes (2006) found susceptibility conditions to be unique to each basin, which demonstrates the importance of understanding physical and chemical variables that control contaminant transport through basin-fill aquifers, especially in populated basins where large amounts of chemicals are used.

The vulnerability of aquifers in Colorado to nitrate contamination was assessed by Rupert (2003), who used logistic regression models to predict the probability of detecting concentrations greater than 5.0 mg/L. Factors found to influence the vulnerability to nitrate contamination of groundwater were localized within a 2-kilometer (km) buffer (6,560 ft) and included fertilizer use, soil properties (available water capacity, clay content, and organic matter content), and land cover (shrubland, percent row crops, and small grains crops). Inspection of mapped predictions indicates a significant portion (authors estimate one quarter) of the San Luis Valley has greater than 50 percent probability of exceeding 5.0 mg/L.

As part of the NAWQA Program, USGS data were used to supplement existing data available from the National Arsenic Occurrence Survey (Frey and Edwards, 1997) to evaluate the distribution of arsenic throughout the United States. In this study, data were analyzed on a county-scale basis by using data from 18,850 wells, representative of 76 percent of large, and 61 percent of small, public-water supply systems in the U.S. (Focazio and others, 2002). Water from about 14 percent of the public-supply wells analyzed for exceedance of target arsenic concentrations throughout the country exceeded 5.0 µg/L, and only about 8 percent exceeded 10 µg/L. The national median arsenic concentration was less than or equal to 1.0 µg/L. Welch and others (1999 and 2000) found that groundwater arsenic concentrations were generally higher in western states (including Alaska and Hawaii) than in eastern states, which was attributed to differences in general climatic and geologic characteristics. Welch and others (2000) found that arsenic concentrations exceeding the current drinking-water standard of 10 µg/L are more frequently found in groundwater samples collected from the western United States, where sources include geothermal water, release from felsic volcanic rocks under alkaline conditions, and enrichment by evaporative concentration.

Approach and Methods

A statistical modeling approach was taken to meet the two main objectives of (1) assessing the vulnerability of the basin-fill aquifers across the SWPA to nitrate contamination and arsenic enrichment, and (2) evaluating the conceptual models developed by Bexfield and others (2011) that summarize the current understanding of the effects of natural and human-related factors on nitrate and arsenic concentrations in the 16 case-study basins. The statistical models reflect natural and human-related factors affecting aquifer vulnerability to contamination by relating constituent concentration to explanatory variables representing local- and basin-scale measures of source, aquifer susceptibility, and geochemical conditions. Source variables reflect the presence of, and in some cases the flux of, natural or human-related contaminants. Aquifer susceptibility variables reflect mechanisms and the ease through which water enters and moves through an aquifer. Geochemical variables reflect chemical processes that affect the fate of contaminants in the groundwater or on the aquifer substrate.

The statistical models were constructed by using the random forest classifier algorithm (Breiman, 2001), and are hereafter called 'classifiers.' In short, random forest classifiers learn the relations between known object classes and known object characteristics. These relations take the form of decision trees, which once constructed, are used to identify the otherwise unknown class for a new set of objects based on their known characteristics. The fundamental components of the random forest classifier include (1) the training dataset, which contains a tabulated set of known object classes and known object characteristics; (2) the random forest, which consists of the decision trees that define the relations between object class and object characteristics; and (3) the prediction dataset, which contains known object characteristics and the predicted object class. Whereas the prediction dataset represents the complete set of objects that are of interest, the training dataset usually represents a subset of the complete set. Typically, financial or other limitations preclude collecting object class data for every object of interest and, as a result, classifiers (or other statistical models) are used to predict object classes where they are desired but are unknown. While this approach is less costly than collecting the complete set of data that is desired, the penalty is a larger uncertainty for predicted object classes than for known object classes.

In this study, the classifiers were constructed to identify a concentration class based on source, aquifer susceptibility, and geochemical characteristics. The vulnerability of the basin-fill aquifers to contamination was assessed by examining classifier predictions of nitrate and arsenic concentrations for basin-fill aquifers across the Southwest. The conceptual models of the effects of natural and human-related factors on aquifer vulnerability to contamination were evaluated through analysis of the predicted concentrations and of the diagnostic information available from the classifiers. An overview of the random forest classifier algorithm, as well as the compilation, selection, and processing of constituent concentration data and explanatory variable data are discussed in detail later in this section.

Separate classifiers were developed for nitrate and arsenic because each constituent was expected to be affected by a different set of factors, and each factor could have a different magnitude or directional influence (increase/decrease) on concentration. For each constituent, two different classifiers were developed—a prediction classifier and a confirmatory classifier. The prediction classifiers were developed specifically to predict nitrate and arsenic concentrations in basin-fill aquifers across the SWPA study area and were based on explanatory variables representing source and susceptibility conditions. These explanatory variables were available throughout the entire SWPA study area and, therefore, did not pose a limitation for using the classifiers to predict concentrations across the study area.

The confirmatory classifiers were developed to supplement the prediction classifiers in the evaluation of the conceptual model. The name, "confirmatory," reflects the classifier's purpose for evaluation of *a-priori* hypotheses and contrasts other general types of statistical models, such as those used for prediction or exploratory purposes. The confirmatory classifiers included the explanatory variables used in the prediction classifiers, as well as additional variables representing geochemical conditions and basin groundwater budget components. The inclusion of the geochemical and basin groundwater budget variables in the confirmatory classifiers allowed for further evaluation of the conceptual models, which was not possible with the prediction classifiers alone. The geochemical data, however, were only available at specific well locations, and consistent water-budget data were not available for every basin in the study area. The limited availability of the data for these variables constrained the confirmatory classifiers to observations from the 16 case-study basins (fig. 1, table 1) and precluded use of the confirmatory classifier for predicting concentrations across the SWPA study area. To contrast the scope of the two classifiers, the confirmatory classifiers were developed by using all available explanatory variables but with observations restricted to the 16 case-study basins, whereas the prediction classifiers were unrestricted with respect to spatial coverage because these were developed by using a subset of the explanatory variables that were available throughout the study area.

The classifiers were spatially referenced to a 3-km by 3-km model grid so that nitrate and arsenic concentrations were spatially tied to explanatory variables associated with a given location. Another important spatial reference was the hydrogeologic area. Each hydrogeologic area consists of (1) an alluvial basin, which contains a basin-fill aquifer, and (2) a contributing area, which consists of consolidated bedrock that drains surface runoff to that alluvial basin. Hydrogeologic areas were used for computing values for some of the explanatory variables and for spatially summarizing classifier results. The 422 hydrogeologic areas form a contiguous set of basins across the SWPA study area; most (344) were delineated by Anning and Konieczki (2005), and the remaining 78 were delineated by McKinney and Anning (2009).

The classifier training dataset and prediction datasets are fundamental components of the classifiers. The first step in developing the classifier training and prediction datasets was to classify each grid cell as representative of basin-fill aquifers or consolidated rocks on the basis of whether the cell was located within the alluvial basin of the hydrogeologic area or within the contributing area. For each grid cell representative of basin-fill aquifers, the values for each explanatory variable were determined and put in a master dataset. In many cases, the explanatory variables were populated on the basis of data from national-scale digital datasets by using geospatial interpretation techniques. Values of the explanatory variables represented conditions for either (1) the grid cell of interest or (2) all grid cells within the alluvial basin and contributing area of the hydrogeologic area containing the grid cell of interest—usually a basin average value or a total value. More detail of the model grid and the explanatory variables is presented in the section, "Compilation and Processing of Explanatory Variables."

The next step in developing the classifier training and prediction datasets was to assign nitrate and arsenic concentration data from groundwater samples to each model grid cell. A maximum of two concentration observations of a given constituent per model grid cell was enforced to avoid biasing the classifiers to conditions in cells for which multiple samples were collected. Each observation represented one sample from a single well. More detail on the concentration data is presented in the section, "Compilation and Selection of Groundwater Chemistry Data." A training dataset for each classifier was constructed by the following:

1. Selecting all concentration observations for the constituent of interest that occur within the spatial extent of the classifier, recalling that the confirmatory classifiers were limited in extent to the 16 case-study basins.
2. For each concentration observation, creating a row in a tabular dataset that has columns populated with the concentration observation and explanatory variables associated with the concentration observation. Note that in some cases there are two concentration observations in a given grid cell, in which case there are two rows in the training dataset that have nearly identical explanatory variable data, with well depth, depth to water, and aquifer-penetration depth values being different.

A total of four training datasets were made: one for each classifier. For documentation purposes these training sets were condensed into a single dataset (appendix 1). Each of these datasets was used to train its intended classifier. More detail on training the classifiers is presented later in the section "Random Forest Classifier."

For the two prediction classifiers, prediction datasets were prepared and consist of a tabular dataset with each row representing a given grid cell and the columns representing the explanatory variable data. Concentration predictions were obtained by using the random forest algorithm, which routes the explanatory data for each grid cell through the trained classifier and appends the predicted concentration into the prediction file. For documentation purposes, the prediction datasets for the nitrate and the arsenic prediction classifiers were condensed into a single dataset (appendix 2).

Compilation and Selection of Groundwater Chemistry Data

Available well-construction and water-quality data from the USGS National Water Information System (NWIS; U.S. Geological Survey, 2010b) databases from each of the six states that have considerable area within the SWPA study area (Arizona, California, Colorado, Nevada, New Mexico, and Utah) were compiled into one dataset. The minimum data requirement for each well was that it have at least one measurement of arsenic or nitrate. In NWIS, nitrate data were available from analyses where nitrate or nitrate plus nitrite were analyzed. Where available, nitrite data were subtracted from the nitrate plus nitrite data. If nitrite concentrations were unavailable, nitrate plus nitrite data were still used because nitrite concentrations were often negligible, less than one-tenth or even one-hundredth the concentration of nitrate. Where both nitrate and nitrate plus nitrite data were available, the preference was to select the nitrate-only analysis.

In order to ensure that each well was represented once and to maintain consistency between the arsenic and nitrate analyses, a single sample was selected to represent conditions for wells sampled more than once. The representative sample was chosen as the one that had both arsenic and nitrate, if available. In the event that multiple samples from a well had both arsenic and nitrate, the sample selected was the one that had the most data for the following constituents: dissolved oxygen (DO), pH, iron, manganese, and sulfate. In the event of a tie, the sample chosen to represent the well was randomly chosen among those tied for the most constituents.

After selecting samples to represent each well, the next step in building the classifier training datasets was to select wells to represent each model grid cell. In most cases only one well was available to represent a given model grid cell; however, there were several cells where two or more wells were available. To avoid over-representation of any given cell in the classifier, a restriction was implemented to include a maximum of two wells per model grid cell into the training dataset. Where possible, one was selected to represent the "shallow" part of the aquifer and another was selected to represent the "deep" part of the aquifer. To facilitate this selection, "aquifer-penetration depth" was calculated for each well as the well depth minus the water-level depth below land surface, and thereby represents the part of the aquifer from which the well is likely to be primarily drawing groundwater (see text box "How is depth treated in the prediction classifiers?"). As part of the well selection process, each well was categorized as having a shallow (less than 150 feet), deep (equal to or greater than 150 feet), or uncategorized aquifer-penetration depth (for instances where the well depth or water-level depth data were not available). In the event that multiple wells in the same grid cell were of the same penetration category (shallow or deep), the well with the most available water-quality data, as described previously, was chosen to represent that cell for that penetration category. In the event of a tie within a cell and penetration category, a well was chosen randomly to represent the specified penetration category.

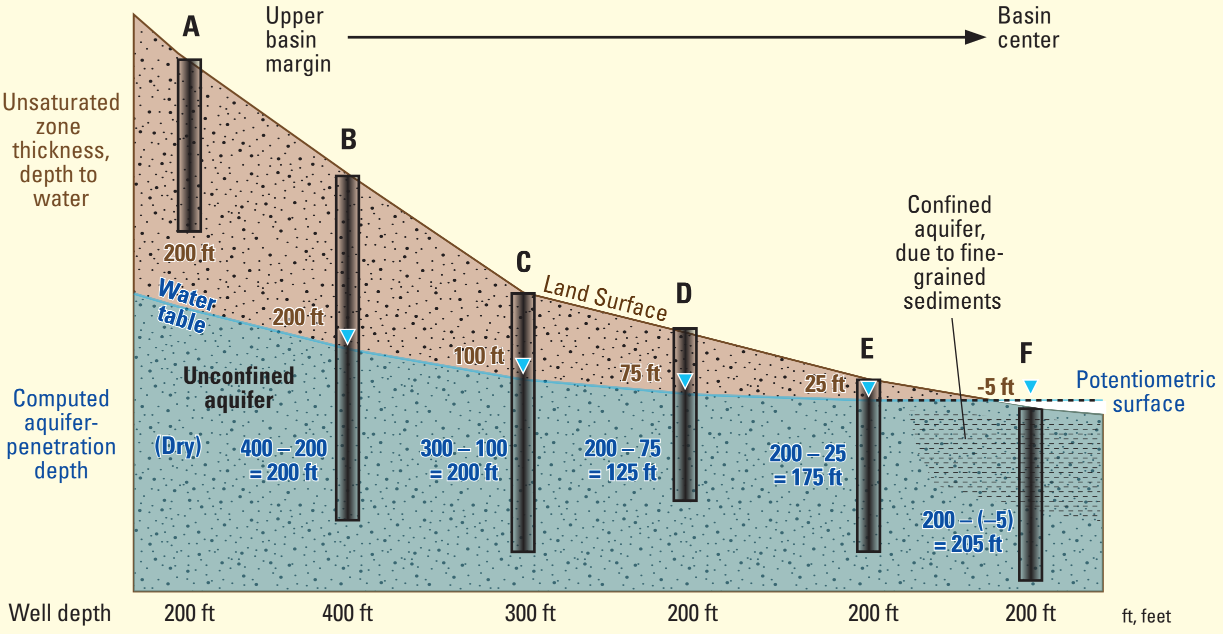
The final result of the sample and well selection process was a set of nitrate and arsenic concentration observations for the classifier training datasets that consist of 6,234 different wells placed among 4,634 model grid cells representing basin-fill aquifers. Well-depth data were available for about 83 percent of the wells, and water-level depth data were available for about 73 percent of the wells, which allowed for aquifer-penetration depth to be computed in about 68 percent of the wells. The aquifer-penetration depth was shallow for 2,002 wells, deep for 2,246 wells, and uncategorized for 1,985 wells. The median values of selected aquifer-penetration depth-related characteristics for the wells in the training dataset were 290-ft well depth, 70-ft depth to water, and 163-ft aquifer-penetration depth (appendix 3).

How is depth treated in the prediction classifiers?

Concentrations of nitrate and arsenic can vary with depth; however, accounting for this variation in a regional-scale model is problematic. The prediction classifiers were constructed to account for concentration variation with depth by using the variable ‘aquifer-penetration depth.’ This variable is defined as the vertical distance from the top of the aquifer to a specific elevation within the aquifer. For the classifier training datasets, aquifer-penetration depth for a specific well was calculated as the depth of the well minus the depth to water in the well, with both measurements relative to the land-surface elevation. For example, the aquifer-penetration depth for well C in the illustration below is 300−100=200 ft. In this calculation, it is assumed that the water level in the well is the same as that in the aquifer. This assumption is valid in most parts of the aquifer except where confined conditions occur. In these areas, as shown for well F, the aquifer-penetration depth is overestimated because the water level in the well is higher than the elevation of the groundwater in the aquifer.

The influence of aquifer-penetration depth on arsenic and nitrate concentrations was evaluated by iteration using the training classifiers and prescribing aquifer-penetration depths at designated intervals: 50, 100, 150, 200, 250, 500, 750, and 1,000 ft. It is important to note that predicted concentrations for a given aquifer-penetration depth represent conditions for an upper region of the aquifer rather than a single point in the aquifer. For example, predicted concentrations for an aquifer-penetration depth of 200 ft represent conditions from the water table to 200 ft below that elevation, not conditions at exactly 200 ft below the water table.

An alternative strategy to account for depth in the classifiers would have been to use a variable representing the depth of a well below the land surface instead of the aquifer-penetration depth. While this would be simpler because it would only require one measurement and not two, such an approach for the basin-fill aquifers would be problematic because the thickness of unsaturated zone in the basin fill varies across the basin. Along the basin margins, it can be a few hundred feet to water, but in the basin center, it can be just a few feet. Consequently, a 200 ft deep well along the margins can be dry (well A), but a 200 ft deep well in the basin center could penetrate the aquifer for most of its depth (wells D and E). Without knowing depth to water throughout the basin-fill aquifers in the SWPA study area, prescribing a well depth for each cell for model predictions is not possible. For aquifer-penetration depth, values from 50 to 1,000 ft are reasonable for most model grid cells, except for those along the basin margin, where the aquifer thickness might not exceed 750 or 1,000 ft, but probably exceeds the other six prescribed depths.



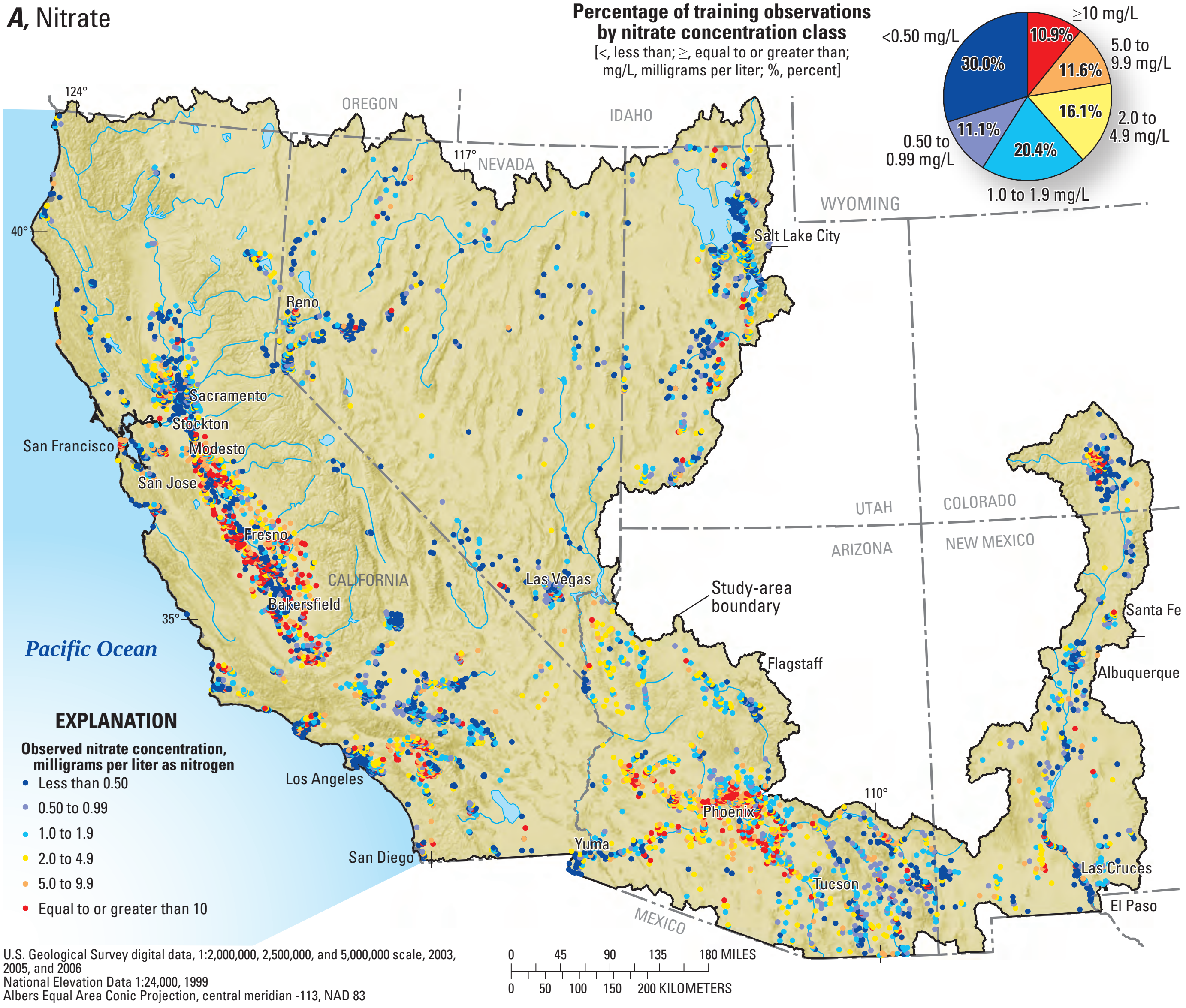
Nitrate and arsenic concentrations range widely, and for each constituent a significant portion of the data were censored below the minimum reporting levels (MRLs) of the laboratory analyses. For each constituent, there were multiple MRLs, and, consequently, the most strategic approach for developing the classifiers was to treat the concentration data as categorical. Classes for the concentration data were developed in consideration of the objectives to spatially define variations in concentrations and to provide water-resource managers information relevant to drinking-water standards. The boundary points between concentration classes were 0.50, 1.0, 2.0, 5.0, and 10 mg/L for nitrate and 1.0, 2.0, 3.0, 5.0, 10, and 25 µg/L for arsenic. Note that the boundary point of 0.50 mg/L between the first two nitrate classes is higher than the highest MRL (0.10 mg/L) for nitrate, which ensured that all of the censored data were incorporated into the lowest concentration class. The boundary point of 1.0 µg/L between the first two arsenic classes likewise was selected such that all of the censored values would be incorporated within the lowest arsenic concentration class. The boundary points of 10 mg/L nitrate and 10 µg/L arsenic were specifically selected so that classifier predictions would identify areas likely to have concentrations equal to or greater than the primary drinking-water standards for these constituents.

Examination of the training data indicates about 11 percent of the observations exceed the 10 mg/L drinking-water standard for nitrate for all Southwest Principal aquifers, and about 25 percent exceed the 10 µg/L drinking-water standard for arsenic (fig. 2). About 41 percent of the nitrate concentration observations are less than one-tenth of the drinking-water standard, 1.0 mg/L, which is the same as the concentration estimated for areas with minimal effects from human activities, or “relative background” conditions across the Nation (Dubrovsky and others, 2010). In contrast, only 15 percent of the observed arsenic concentrations are less than one-tenth of the drinking-water standard (1.0 µg/L).

The training data show that concentrations are predominantly high in some areas but predominantly low in others (fig. 2). As an example of this contrast in concentration distributions, about 50 percent of the nitrate observations from the Central Valley aquifer system are greater than 2.0 mg/L, whereas, in the Pacific Northwest aquifers, observed concentrations are much lower—only 21 percent exceed 2.0 mg/L, and 79 percent are less than 2.0 mg/L. As another example, about 68 percent of the arsenic observations are greater than 2.0 µg/L in the Basin and Range basin-fill aquifers, whereas concentrations in the California Coastal Basin aquifers are much lower: about 79 percent of the arsenic concentrations are less than 2.0 µg/L. Within each principal aquifer, there are also areas with generally high concentrations and other areas with generally low concentrations. For example, in the Basin and Range basin-fill aquifers, many of the nitrate observations in central Arizona are greater than 2.0 mg/L, and many of the arsenic observations in western Arizona, southeastern California, and western Nevada are equal to or greater than 10 µg/L (figs. 2 and 3).

In consideration of the study objectives, to adequately characterize spatial variations in concentrations throughout large parts of the study area that have predominantly low concentrations, at least a few concentration classes are needed to represent these lower concentration ranges. Likewise, at least a few classes are needed to represent higher concentration ranges to adequately characterize spatial variations in areas that have predominantly high concentrations. While using fewer concentration classes would increase the accuracy of the random forest classifier predictions, six classes of nitrate concentrations and seven classes of arsenic concentrations were used so that spatial variations would be elucidated throughout the SWPA study area.

At the regional scale, principal aquifer nitrate and arsenic training observations (fig. 3) show larger differences in the concentration distributions spatially than with aquifer-penetration depth (fig 4). While vertical stratification of concentrations can occur at local to basin spatial scales, as in Rosen (2003) and Burow and others (2008), for example, neither observed nitrate nor arsenic concentrations show clear and strong vertical stratification in a spatially consistent manner for the basin-fill aquifers as a whole across the Southwest (fig. 4). While plots, such as figure 4, can show large and prevalent trends in concentrations, small trends in concentrations can be hard to distinguish because many environmental conditions besides aquifer-penetration depth can affect concentrations. An alternate approach was used whereby observed concentrations from model-grid cells that have two observations were compared in a pair-wise manner while accounting for aquifer-penetration depth and sample collection date. This approach held many of the environmental conditions constant between the two concentrations being compared, unless those conditions change with time, with depth in the aquifer, or spatially in the grid cell. By using this approach, differences in concentration class were determined for paired nitrate observations in a given cell and then regressed against difference in aquifer-penetration depth and difference in sample collection date (time). Results from the analysis showed no trend in nitrate concentrations over time; however, a negative trend was found with aquifer-penetration depth at the p-value less than 0.01 confidence level. The regression coefficients indicated that the trend was very small—nitrate would increase only one concentration class for each 770 ft increase in aquifer-penetration depth. The same analysis was performed for arsenic, but there were no trends detected over time or with increased aquifer-penetration depth. On the basis of these regression results, aquifer-penetration depth was included as a variable in the classifiers; however, the sample collection date for the observed concentrations was not included.



U.S. Geological Survey digital data, 1:2,000,000, 2,500,000, and 5,000,000 scale, 2003, 2005, and 2006
National Elevation Data 1:24,000, 1999
Albers Equal Area Conic Projection, central meridian -113, NAD 83

Figure 2. Spatial distribution of observed nitrate and arsenic concentrations in groundwater samples from basin-fill aquifers in the Southwest Principal Aquifers study area used to train the random forest classifiers, 1980–2009: *A*, Nitrate; *B*, Arsenic.

Compilation and Processing of Explanatory Variables

The training and prediction datasets for the classifiers used to predict nitrate and arsenic concentrations were founded on a set of explanatory variables representing source, aquifer susceptibility, and geochemical conditions, which in most cases were developed by using a Geographic Information System (GIS). Development of the variables relied on three major steps:

1. Acquisition of spatial data layers and tabular attribute datasets.
2. Processing and conversion of each data layer to a common raster format and analysis environment, such that they were scaled and aligned with the model grid.
3. Computation of statistics from the raster format data to represent the explanatory variables for each grid cell.

The remainder of this section describes the procedures used to assign the systematic set of source, aquifer susceptibility, and geochemical explanatory variables to each model grid cell throughout the SWPA study area. A summary for each variable has been provided, including the general variable type (source, susceptibility, or geochemical), the representative area, the classifiers in which it was used, and the original source of the data (table 3). In addition, summary statistics for each explanatory variable in the training dataset (appendix 1) and the prediction dataset (appendix 2) are provided in appendix 3 and include percent of training observations with data, the minimum and maximum value, and values for the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles.

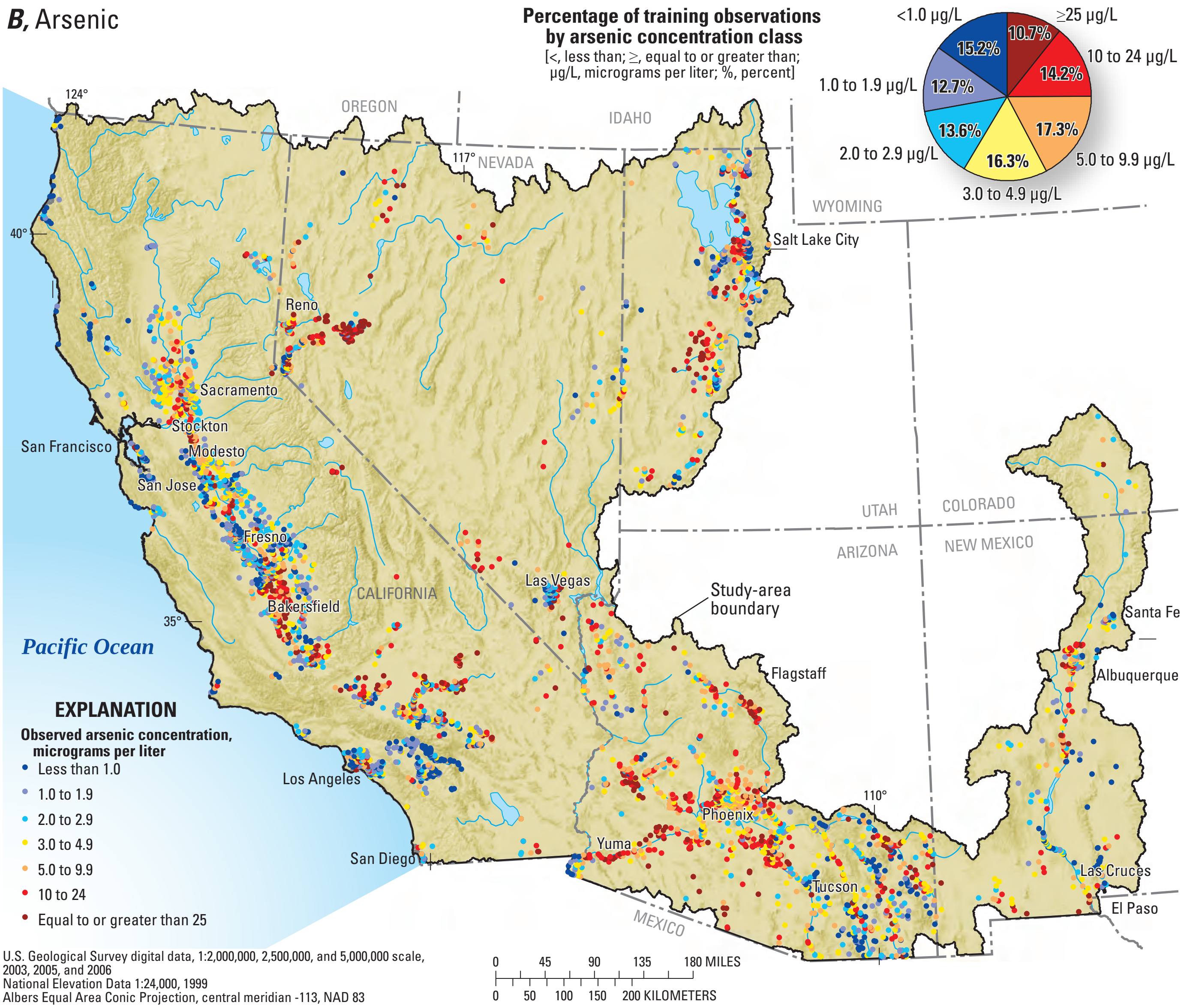


Figure 2. Spatial distribution of observed nitrate and arsenic concentrations in groundwater samples from basin-fill aquifers in the Southwest Principal Aquifers study area used to train the random forest classifiers, 1980–2009: *A*, Nitrate; *B*, Arsenic.—Continued

The classifiers are spatially referenced to a model grid for the entire study area that has a 3-km cell resolution and comprises 119,981 cells, including those in the alluvial basins and those in the contributing areas of the hydrogeologic areas. Each grid cell was assigned a unique identifier that was used as the key field in GIS analysis and relational database transactions. For the purposes of this study, only those cells that coincided with basin fill were used in the classifier prediction or training datasets, which totaled 54,854 cells.

Most of the explanatory variables were developed by using GIS analysis. Datasets originally acquired in vector format were transformed to raster format, which represents a given variable as a matrix of cells in a continuous space. To ensure accurate cell-by-cell combination of the raster datasets, common analysis environment parameters were configured in ArcGIS Spatial Analyst (McCoy and others, 2001). The appropriate cell sizes (100-meter), output extent, snap raster and mask were configured before processing each dataset. An output extent defines the maximum and minimum spatial extent for cell-level computations, the snap raster processes data to common cell boundaries, and the mask allows select cells within the analysis window to be used for data processing. The GIS vector (point, line, or polygon feature) data layers described in this report were converted to raster data, or were “rasterized” using the common analysis environment parameters.

For many of the explanatory variables, the value for each grid cell was determined by using zonal-statistic computations. Calculation of a zonal statistic involves combining a zone layer that defines a specific area in space (for example, an alluvial basin or a grid cell) with one or more value layers for an explanatory variable (for example, geology, elevation, land

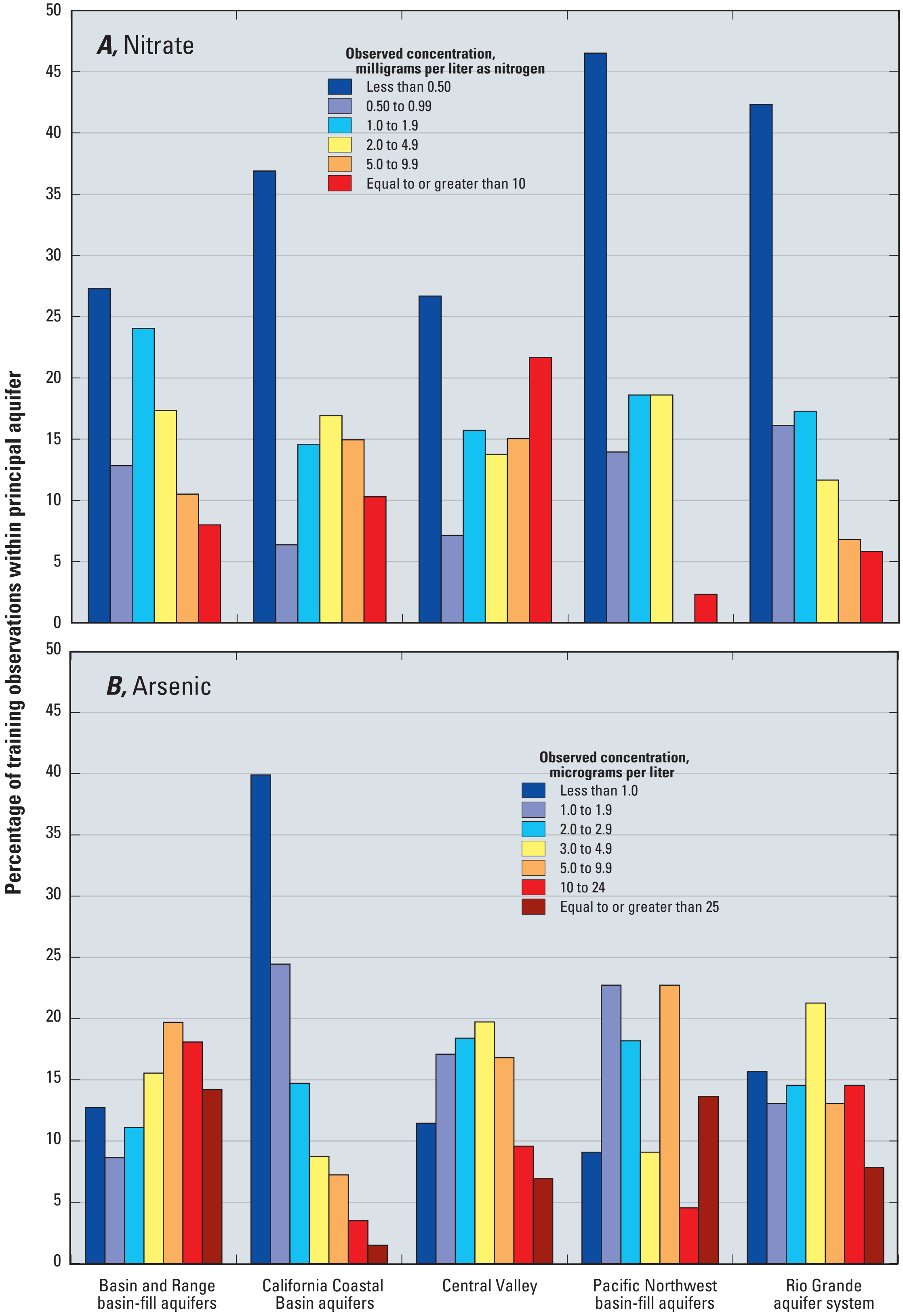


Figure 3. Distribution of nitrate and arsenic concentrations in groundwater samples collected from each of the principal aquifers in the Southwest Principal Aquifers study area used to train the random forest classifiers, 1980–2009: *A*, Nitrate; *B*, Arsenic.

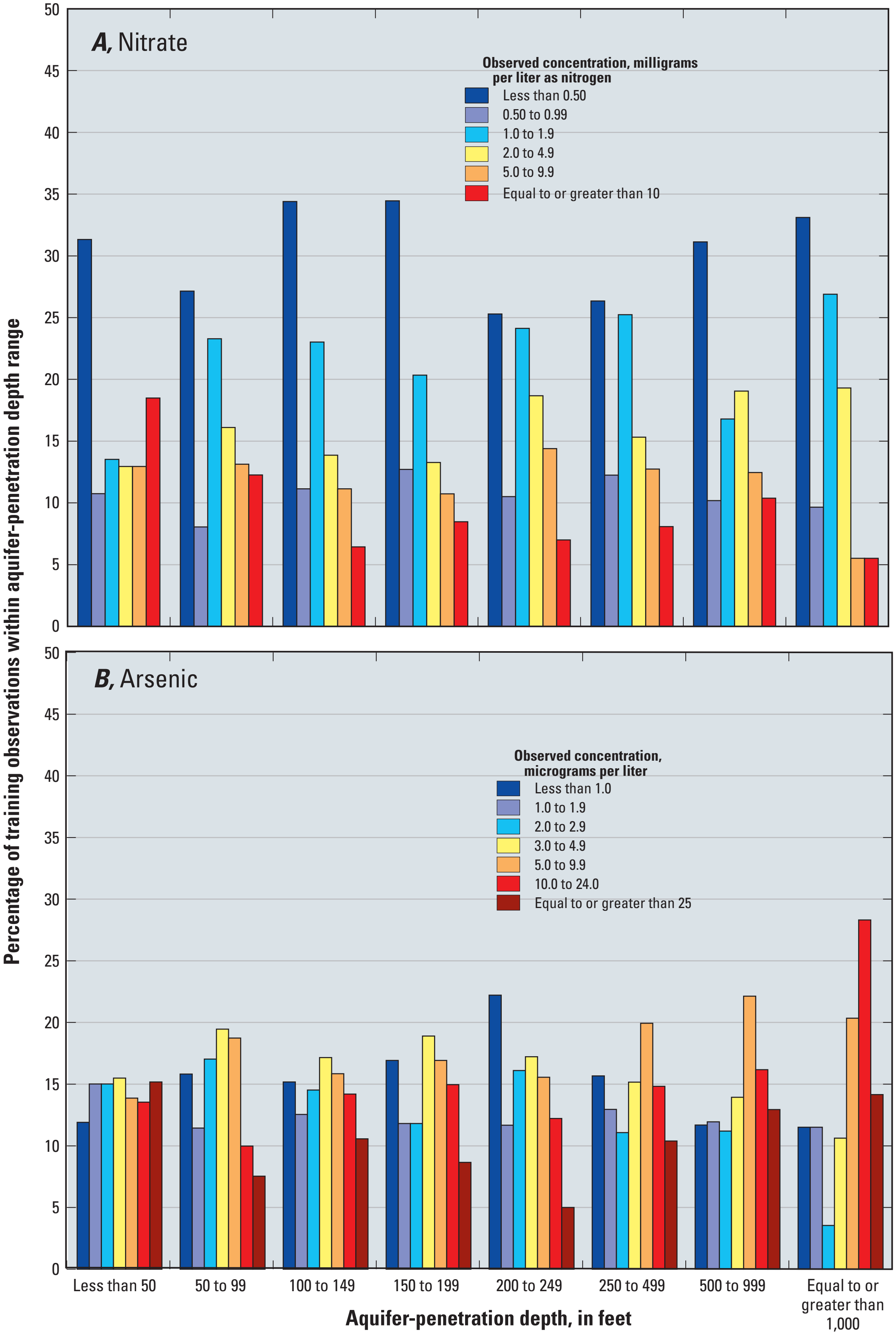


Figure 4. Distribution by aquifer-penetration depth of nitrate and arsenic concentrations in groundwater samples from basin-fill aquifers in the Southwest Principal Aquifers study area used to train the random forest classifiers, 1980–2009: *A*, Nitrate; *B*, Arsenic.

Table 3. Explanatory variables used in random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.

[Grid cells are 3 km by 3 km. The represented area for an explanatory variable is restricted to the grid cell itself, the alluvial basin that contains the grid cell, or the contributing area surrounding the alluvial basin and within the hydrogeologic area that contains the cell. For variables representing the alluvial basin or contributing area, the area of interest was defined on the basis of 1:500,000 scale state geologic data.[1] **Abbreviations**: acre-ft/yr, acre-feet per year; As, arsenic; ft, feet; in/yr, inches per year; kg/yr, kilograms per year; km, kilometers; km^2, square kilometers; m, meters; µg/L, micrograms per liter; mg/L, milligrams per liter; mm, millimeters; NO_3, nitrate; ppm, parts per million; X, constituent tested in classifier; —, constituent not tested in classifier]

Variable group	Explanatory variable	Description	Represented area	Tested in classifier:				Data source
				Prediction		Confirma-tory		
				NO_3	As	NO_3	As	
Source variables								
Nitrogen loading	Nitrogen, atmospheric	Nitrogen loading from atmospheric deposition, 1987–2002, in total kg/yr	Grid cell	X	—	X	—	Ruddy and others (2006)
	Nitrogen, farm fertilizer	Nitrogen loading from farm fertilizer, 1987–2002, in total kg/yr	Grid cell	X	—	X	—	Ruddy and others (2006), revised with a correction for a processing error in the non-farm versus farm allocation performed by J.M. Gronberg and N.E. Spar in 2008
	Nitrogen, non-farm fertilizer	Nitrogen loading from non-farm fertilizer, 1987–2002, in total kg/yr	Grid cell	X	—	X	—	Ditto
	Nitrogen, confined manure	Nitrogen loading from manure in spatially confined areas (dairies, feedlots, etc.), 1987–2002, in total kg/yr	Grid cell	X	—	X	—	Ruddy and others (2006)
	Nitrogen, unconfined manure	Nitrogen loading from manure in spatially unconfined areas, 1987–2002, in total kg/yr	Grid cell	X	—	X	—	Ditto
	Nitrogen, total	Nitrogen loading from above sources, 1987–2002, in total kg/yr	Grid cell	X	—	X	—	Calculated from above nitrogen sources
Agricultural, urban, and biotic sources	Biotic community	Index number representing biotic community of North America	Grid cell	X	X	X	X	Brown and others (2007)
	Septic/sewer ratio	Ratio of 1990 US Census of housing units on septic relative to those on sewer	Grid cell	X	X	X	X	Hitt (1997)
	Local population	Human population, count	Grid cell	X	X	X	X	Oak Ridge National Laboratory (2005)
	Local population density	Human population density, in persons/km^2	Grid cell	X	X	X	X	Ditto
	Basin population	Human population, count	Alluvial basin	X	X	X	X	Ditto
	Basin population density	Human population density, in persons/km^2	Alluvial basin	X	X	X	X	Ditto
	Local urban land	Urban land, percentage	Grid cell	X	X	X	X	U.S. Geological Survey (2008)
	Local agricultural land	Agricultural land, percentage	Grid cell	X	X	X	X	Ditto
	Basin urban land	Urban land, percentage	Alluvial basin	X	X	X	X	Ditto
	Basin agricultural land	Agricultural land, percentage	Alluvial basin	X	X	X	X	Ditto
	Basin rangeland	Rangeland, percentage	Alluvial basin	X	X	X	X	Ditto
	Basin other land cover	All other land cover, excluding agricultural, urban, and rangeland, percentage	Alluvial basin	X	X	X	X	Ditto
Geologic sources	Geology, carbonate rocks	Area of carbonate rocks in contributing area, percentage	Contributing area	X	X	X	X	1:500,000 scale state geology[1]
	Geology, crystalline rocks	Area of crystalline rocks, mostly granitic and metamorphic, in contributing area, percentage	Contributing area	X	X	X	X	Ditto
	Geology, clastic sedimentary rocks	Area of clastic sedimentary rocks in contributing area, percentage	Contributing area	X	X	X	X	Ditto
	Geology, mafic volcanic rocks	Area of mafic volcanic rocks in contributing area, percentage	Contributing area	X	X	X	X	Ditto
	Geology, felsic and silicic volcanic rocks	Area of felsic and silicic rocks in contributing area, percentage	Contributing area	X	X	X	X	Ditto
	Geology, intermediate composition volcanic rocks	Area of intermediate composition volcanic rocks in contributing area, percentage	Contributing area	X	X	X	X	Ditto

Table 3. Explanatory variables used in random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Grid cells are 3 km by 3 km. The represented area for an explanatory variable is restricted to the grid cell itself, the alluvial basin that contains the grid cell, or the contributing area surrounding the alluvial basin and within the hydrogeologic area that contains the cell. For variables representing the alluvial basin or contributing area, the area of interest was defined on the basis of 1:500,000 scale state geologic data.[1] **Abbreviations**: acre-ft/yr, acre-feet per year; As, arsenic; ft, feet; in/yr, inches per year; kg/yr, kilograms per year; km, kilometers; km^2, square kilometers; m, meters; µg/L, micrograms per liter; mg/L, milligrams per liter; mm, millimeters; NO_3, nitrate; ppm, parts per million; X, constituent tested in classifier; —, constituent not tested in classifier]

Variable group	Explanatory variable	Description	Represented area	Tested in classifier:				Data source
				Prediction		Confirmatory		
				NO_3	As	NO_3	As	
Source variables								
Geologic sources—Continued	Geology, undifferentiated volcanic rocks	Area of undifferentiated volcanic rocks in contributing area, percentage	Contributing area	X	X	X	X	1:500,000 scale state geology[1]
	Geology, distance to carbonate rocks	Distance from cell center to nearest carbonate rock outcrop, in m	Grid cell	X	X	X	X	Ditto
	Geology, distance to crystalline rocks	Distance from cell center to nearest crystalline rock outcrop, in m	Grid cell	X	X	X	X	Ditto
	Geology, distance to clastic sedimentary rocks	Distance from cell center to nearest clastic sedimentary rock outcrop, in m	Grid cell	X	X	X	X	Ditto
	Geology, distance to mafic volcanic rocks	Distance from cell center to nearest mafic volcanic rock outcrop, in m	Grid cell	X	X	X	X	Ditto
	Geology, distance to felsic and silicic volcanic rocks	Distance from cell center to nearest felsic and silicic volcanic rock outcrop, in m	Grid cell	X	X	X	X	Ditto
	Geology, distance to intermediate composition volcanic rocks	Distance from cell center to nearest intermediate composition volcanic rock outcrop, in m	Grid cell	X	X	X	X	Ditto
	Geology, distance to undifferentiated volcanic rocks	Distance from cell center to nearest undifferentiated volcanic rock outcrop, in m	Grid cell	X	X	X	X	Ditto
	Soil and rock equivalent uranium-238 concentration	Equivalent uranium-238 concentration in parts per million as calculated from the counts received by a gamma-ray detector in the energy window corresponding to bismuth-214. Collected using aerial gamma-ray surveys	Grid cell	—	X	—	X	Kucks (2005)
Aquifer susceptibility variables								
Flow path	Aquifer-penetration depth	Depth into aquifer that well penetrates; computed as well depth minus water-level depth, in ft	Grid cell	X	X	X	X	Groundwater site information from U.S. Geological Survey (2010a)
	Well depth	Depth from land surface to bottom of well, in ft	Grid cell	—	—	X	X	Ditto
	Water-level depth	Depth from land surface to water level, in ft	Grid cell	—	—	X	X	Ditto
	Land-surface slope	Mean slope from 30-m grid, in degrees	Grid cell	X	X	X	X	U.S. Geological Survey (2005)
	Land-surface elevation	Mean elevation from 30-m grid, in m	Grid cell	X	X	X	X	Ditto
	Land-surface elevation percentile	Land-surface elevation of grid cell, expressed as a percentile of elevations for all grid cells in alluvial basin	Grid cell	X	X	X	X	Ditto
	Basin elevation	Average land-surface elevation in alluvial basin, in m	Alluvial basin	X	X	X	X	Ditto
	Distance to basin margin	Distance from cell center to margin of basin-fill at contact with consolidated rocks, in km	Grid cell	X	X	X	X	1:500,000 scale state geology[1]

Table 3. Explanatory variables used in random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Grid cells are 3 km by 3 km. The represented area for an explanatory variable is restricted to the grid cell itself, the alluvial basin that contains the grid cell, or the contributing area surrounding the alluvial basin and within the hydrogeologic area that contains the cell. For variables representing the alluvial basin or contributing area, the area of interest was defined on the basis of 1:500,000 scale state geologic data.[1] **Abbreviations**: acre-ft/yr, acre-feet per year; As, arsenic; ft, feet; in/yr, inches per year; kg/yr, kilograms per year; km, kilometers; km^2, square kilometers; m, meters; µg/L, micrograms per liter; mg/L, milligrams per liter; mm, millimeters; NO_3, nitrate; ppm, parts per million; X, constituent tested in classifier; —, constituent not tested in classifier]

Variable group	Explanatory variable	Description	Represented area	Tested in classifier:				Data source
				Prediction		Confirmatory		
				NO_3	As	NO_3	As	
Aquifer susceptibility variables—Continued								
Soil properties	Soil, seasonally high water depth	Spatial average depth to seasonally high water table, in ft	Grid cell	X	X	X	X	Unpublished version of Wolock (1997) that has 100-m spatial resolution rather than 1 km
	Soil, hydric	Area in which hydric (water-saturated) soils were identified, percentage	Grid cell	X	X	X	X	Ditto
	Soil, hydrologic group A	Area with soil hydrologic group A, percentage. Group A is sand, loamy sand or sandy loam types of soils. It has low runoff potential and high infiltration rates even when thoroughly wetted. It consists chiefly of deep, well to excessively drained sands or gravels and have a high rate of water transmission	Grid cell	X	X	X	X	Ditto
	Soil, hydrologic group B	Area with soil hydrologic group B, percentage. Group B is silt loam or loam. It has a moderate infiltration rate when thoroughly wetted and consists chiefly or moderately deep to deep, moderately well to well drained soils with moderately fine to moderately coarse textures	Grid cell	X	X	X	X	Ditto
	Soil, hydrologic group C	Area with soil hydrologic group C, percentage. Group C soils are sandy clay loam. They have low infiltration rates when thoroughly wetted and consist chiefly of soils with a layer that impedes downward movement of water and soils with moderately fine to fine structure	Grid cell	X	X	X	X	Ditto
	Soil, hydrologic group D	Area with soil hydrologic group D, percentage. Group D soils are clay loam, silty clay loam, sandy clay, silty clay or clay. This hydrologic soil group has the highest runoff potential. It has very low infiltration rates when thoroughly wetted and consists chiefly of clay soils with a high swelling potential, soils with a permanent high water table, soils with a claypan or clay layer at or near the surface, and shallow soils over nearly impervious material	Grid cell	X	X	X	X	Ditto
	Soil, permeability	Permeability, in inches per hour	Grid cell	X	X	X	X	Ditto
	Soil, organic material	Organic material content, percentage	Grid cell	X	X	X	X	Ditto
	Soil, clay	Clay content, percentage	Grid cell	X	X	X	X	Ditto
	Soil, silt	Silt content, percentage	Grid cell	X	X	X	X	Ditto
	Soil, sand	Sand content, percentage	Grid cell	X	X	X	X	Ditto
Water use and hydroclimatic	Water-resources development index	Water-resources development index, a log-10 based measure for the annual surface water and groundwater use in an alluvial basin, and in some areas, parts of the contributing area too	Alluvial basin	X	X	X	X	Recomputed as part of this study based on method used by Anning and Konieczki (2005)
	Groundwater use, irrigated agriculture	Estimated irrigated agricultural groundwater use, in gallons per year	Grid cell	X	X	X	X	Water-use data from U.S. Geological Survey (2004), and land-cover data from U.S. Geological Survey (2008)

Table 3. Explanatory variables used in random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Grid cells are 3 km by 3 km. The represented area for an explanatory variable is restricted to the grid cell itself, the alluvial basin that contains the grid cell, or the contributing area surrounding the alluvial basin and within the hydrogeologic area that contains the cell. For variables representing the alluvial basin or contributing area, the area of interest was defined on the basis of 1:500,000 scale state geologic data.[1] **Abbreviations**: acre-ft/yr, acre-feet per year; As, arsenic; ft, feet; in/yr, inches per year; kg/yr, kilograms per year; km, kilometers; km^2, square kilometers; m, meters; µg/L, micrograms per liter; mg/L, milligrams per liter; mm, millimeters; NO_3, nitrate; ppm, parts per million; X, constituent tested in classifier; —, constituent not tested in classifier]

Variable group	Explanatory variable	Description	Represented area	Tested in classifier:				Data source
				Prediction		Confirmatory		
				NO_3	As	NO_3	As	
Aquifer susceptibility variables—Continued								
Water use and hydroclimatic—Continued	Surface-water use, irrigated agriculture	Estimated irrigated agricultural surface-water use, in gallons per year	Grid cell	X	X	X	X	Ditto
	Groundwater use, public water supply	Estimated public-supply groundwater use, in gallons per year	Grid cell	X	X	X	X	Ditto
	Surface-water use, public water supply	Estimated public-supply surface-water use, in gallons per year	Grid cell	X	X	X	X	Ditto
	Recharge, contributing area	Recharge for contributing area between alluvial basin boundary and hydrogeologic area boundary, computed using Maxey-Eakon method (1949), in in/yr	Contributing area	X	X	X	X	Precipitation data from PRISM Group (2004a)
	Recharge, basin	Recharge for alluvial basin, computed using Maxey-Eakon method (1949), in in/yr	Alluvial basin	X	X	X	X	Ditto
	Potential evapotranspiration	Potential evapotranspiration, 1970–2006, in mm	Grid cell	X	X	X	X	Flint and Flint (2007)
	Mean air temperature	Mean air temperature, 1971–2000, in degrees Fahrenheit	Grid cell	X	X	X	X	PRISM group (2004b)
Basin groundwater budget	Recharge, subsurface inflow	Annual subsurface inflow from adjacent basins, percentage of the basin groundwater budget	Alluvial basin	—	—	X	X	Bexfield and others (2011)
	Recharge, mountain front	Annual mountain-front recharge, percentage of the basin groundwater budget	Alluvial basin	—	—	X	X	Ditto
	Recharge, precipitation	Annual recharge from precipitation on alluvial basin, percentage of the basin groundwater budget	Alluvial basin	—	—	X	X	Ditto
	Recharge, stream infiltration	Annual recharge from streamflow infiltration, percentage of the basin groundwater budget	Alluvial basin	—	—	X	X	Ditto
	Recharge, irrigation	Annual recharge from infiltration of excess irrigation water, percentage of the basin groundwater budget	Alluvial basin	—	—	X	X	Ditto
	Recharge, artificial	Annual artificial recharge, percentage of the basin groundwater budget	Alluvial basin	—	—	X	X	Ditto
	Recharge, change	Change in annual recharge from predevelopment to modern (circa 2000) conditions, percentage	Alluvial basin	—	—	X	X	Ditto
	Storage, change	Change in annual storage change from predevelopment to modern (circa 2000) conditions, percentage	Alluvial basin	—	—	X	X	Ditto
	Recharge, total	Total recharge from components of the basin groundwater budget, in acre-ft/yr	Alluvial basin	—	—	X	X	Ditto
	Discharge, total	Total discharge from components of the basin groundwater budget, in acre-ft/yr	Alluvial basin	—	—	X	X	Ditto
	Discharge, change	Change in annual discharge from predevelopment to modern (circa 2000) conditions, percentage	Alluvial basin	—	—	X	X	Ditto
	Discharge, subsurface outflow	Annual subsurface outflow to adjacent basins, percentage of the basin groundwater budget	Alluvial basin	—	—	X	X	Ditto

Table 3. Explanatory variables used in random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Grid cells are 3 km by 3 km. The represented area for an explanatory variable is restricted to the grid cell itself, the alluvial basin that contains the grid cell, or the contributing area surrounding the alluvial basin and within the hydrogeologic area that contains the cell. For variables representing the alluvial basin or contributing area, the area of interest was defined on the basis of 1:500,000 scale state geologic data.[1] **Abbreviations**: acre-ft/yr, acre-feet per year; As, arsenic; ft, feet; in/yr, inches per year; kg/yr, kilograms per year; km, kilometers; km^2, square kilometers; m, meters; µg/L, micrograms per liter; mg/L, milligrams per liter; mm, millimeters; NO_3, nitrate; ppm, parts per million; X, constituent tested in classifier; —, constituent not tested in classifier]

Variable group	Explanatory variable	Description	Represented area	Tested in classifier:				Data source
				Prediction		Confirmatory		
				NO_3	As	NO_3	As	
Aquifer susceptibility variables—Continued								
Basin groundwater budget—Continued	Discharge, evapotranspiration	Annual evapotranspiration of shallow groundwater, percentage of the basin groundwater budget	Alluvial basin	—	—	X	X	Ditto
	Discharge, to streams	Annual groundwater discharge to streams, percentage of the basin groundwater budget	Alluvial basin	—	—	X	X	Ditto
	Discharge, to springs and drains	Annual groundwater discharge to springs and drains, percentage of the basin groundwater budget	Alluvial basin	—	—	X	X	Ditto
	Discharge, well withdrawals	Annual groundwater discharge to pumping or flowing wells, percentage of the basin groundwater budget	Alluvial basin	—	—	X	X	Ditto
	Residence time	Coarse estimate of residence time for groundwater in upper 1,000 ft of basin-fill aquifer, in years	Alluvial basin	—	—	X	X	Ditto
Geochemical variables								
Geochemical	Groundwater, pH	pH of groundwater sample, in standard units	Grid cell	—	—	X	X	Groundwater chemistry information from U.S. Geological Survey (2010a)
	Groundwater, dissolved oxygen	Dissolved-oxygen concentration of groundwater sample, in mg/L	Grid cell	—	—	X	X	Ditto
	Groundwater, dissolved solids	Dissolved-solids concentration of groundwater sample, in mg/L, computed as sum of individual ions	Grid cell	—	—	X	X	Ditto
	Groundwater, nitrate	Dissolved-nitrate concentration of groundwater sample, in mg/L as N	Grid cell	—	—	—	X	Ditto
	Groundwater, sulfate	Dissolved-sulfate concentration of groundwater sample, in mg/L as SO_4^{-2}	Grid cell	—	—	X	X	Ditto
	Groundwater, iron	Dissolved-iron concentration of groundwater sample, in µg/L	Grid cell	—	—	X	X	Ditto
	Groundwater, manganese	Dissolved-manganese concentration of groundwater sample, in µg/L	Grid cell	—	—	X	X	Ditto
	Groundwater, alkalinity	Alkalinity of groundwater sample, in mg/L	Grid cell	—	—	X	X	Ditto
	Groundwater, bicarbonate	Dissolved bicarbonate of groundwater sample, in mg/L	Grid cell	—	—	X	X	Ditto
	Groundwater, orthophosphate	Dissolved orthophosphate of groundwater sample, in mg/L as P	Grid cell	—	—	X	X	Ditto
	Groundwater, chloride	Dissolved chloride of groundwater sample, in mg/L	Grid cell	—	—	X	X	Ditto
	Groundwater, molybdenum	Dissolved molybdenum of groundwater sample, in µg/L	Grid cell	—	—	X	X	Ditto
	Groundwater, selenium	Dissolved selenium of groundwater sample, in µg/L	Grid cell	—	—	X	X	Ditto

[1] The 1:500,000 scale state geology was compiled from Green (1992), Green and Jones (1997), Hirschberg and Pitts (2000), Johnson and Raines (1996), Ramsey (1996), Saucedo and others (2000), Turner and Bawic (1996), and Walker and others (2003).

cover, or population) to calculate a statistic of the explanatory variable for an individual zone. In most cases the statistic computed was an average, total, or percentage of the total. The computed zonal statistic for all model grid cells within the alluvial basins of the study area forms a data layer for an explanatory variable, and all of these data layers were stored in a relational database-management system by grid cell. Spatial referencing of each explanatory variable was maintained in the relational database management system by including the model grid cell identifier that ties the data to a specific grid cell and by including latitude-longitude location coordinates for each cell as an additional data layer. For development of training and prediction datasets, explanatory variable data were retrieved in tabular format with observations forming rows and explanatory variables forming columns.

Source Variables

Variables representing sources of nitrate and arsenic were derived from existing datasets of different spatial scales and were re-calculated to represent the 3-km model grid cells used for the random forest classifiers. Nitrate sources were determined for agricultural, urban, and undeveloped landscapes, and included loads originating from fertilizer use, atmospheric deposition, and nitrogen fixing plants. The source of arsenic was predominately from various geologic units.

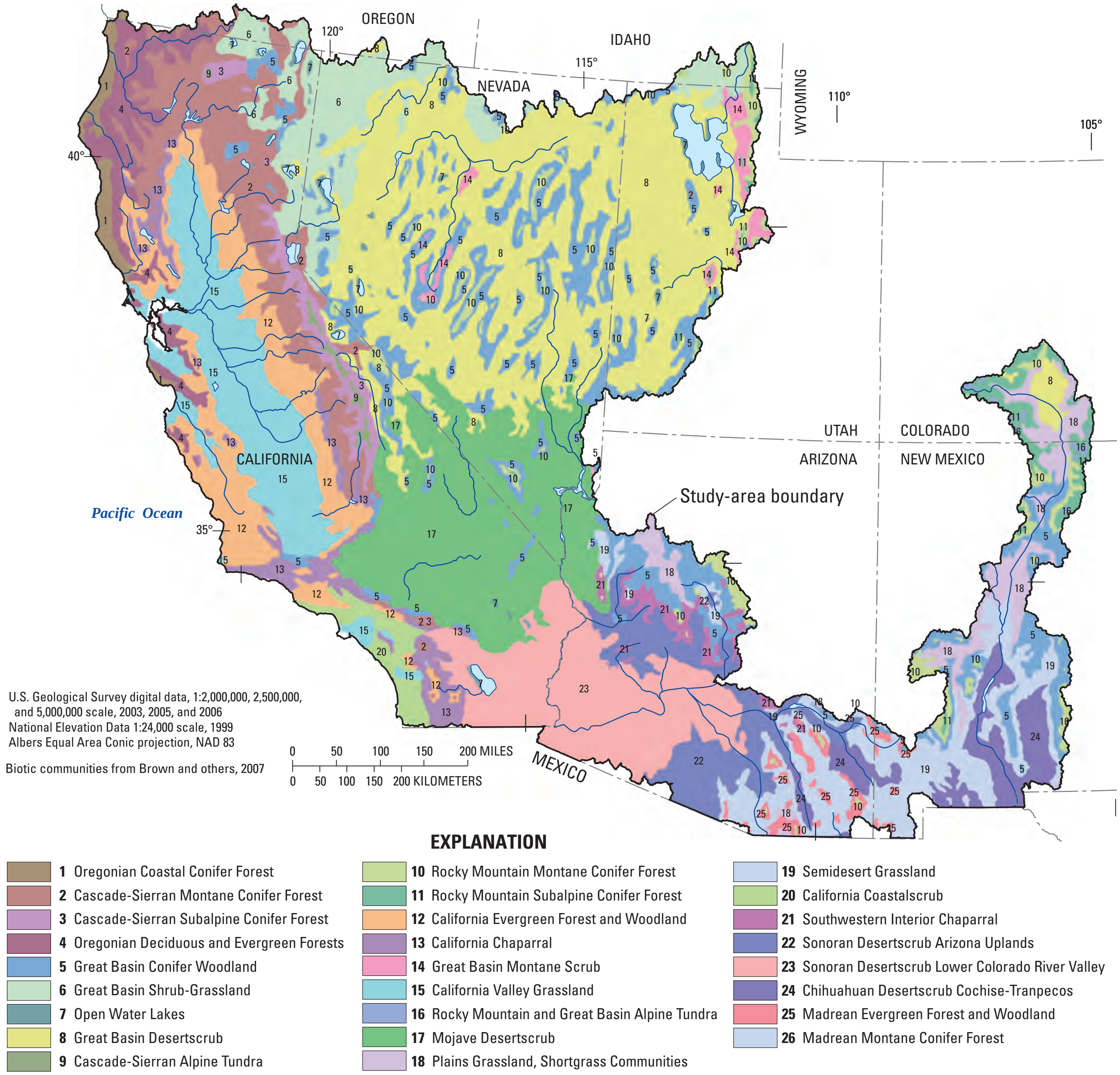
Nitrogen loading variables represent quantified nitrogen fluxes from atmospheric deposition, farm and non-farm fertilizer, and manure from confined (mostly dairies or feedlots) or from unconfined (mostly rangeland grazing) livestock operations. Note that nitrogen loads from unconfined livestock, in part, represent nitrogen cycling from atmospheric deposition into grasses and then through livestock. The original data for these variables were provided by the NAWQA Program’s National Nutrient Synthesis team and are 1-km cell resolution raster datasets of estimates of average annual nitrogen inputs (kilograms per year; kg/yr) for the years 1982 through 2001. The nitrogen input estimates for the fertilizer and livestock grids are based on county-level data that are apportioned to each grid cell on the basis of agricultural and urban land use data. For example, annual farm fertilizer inputs for a county are apportioned to only agricultural lands in that county. Although the nitrogen input grids are not available to the public, tabular county estimates are available in the report, “County-Level Estimates of Nutrient Inputs to the Land Surface of the Conterminous United States, 1982–2001” (Ruddy and others, 2006). The total nitrogen input from each source was computed for each model grid cell from the 1-km raster data by using zonal-average statistics. The spatial distribution of selected nitrogen-loading variables is shown in appendix 10.

Agricultural, urban, and biotic source variables include local and basin land use, septic/sewer ratio, local and basin population and population densities, and biotic community. These variables provide surrogates for nitrogen loading, but also can serve as surrogates for aquifer susceptibility to nitrate contamination or arsenic enrichment because certain transport mechanisms are associated with these variables and their intensities. The spatial distribution of selected agricultural, urban, and biotic source variables is shown in appendix 11.

Biotic community (fig. 5) is a categorical variable representing different biological sources of nitrogen. Some biotic communities can produce more nitrogen available for transport to the aquifer than others as a result of the different plant communities and soil microbes in them as well as the different climates in which they reside. The original dataset, biotic communities of North America, is a small scale (1:10,000,000) vector dataset of plant ecology classifications (Brown and others, 2007). The data for the SWPA study area represent 26 different classes of forests, woodlands, scrublands, grasslands, and deserts that coincide with the basin-fill aquifers. The biotic-community vector data were rasterized to a 100-meter (m) grid, and zonal majority statistics were computed for the 3-km model grid.

The septic/sewer ratio data provides a means to distinguish areas served by sewer systems from areas that predominantly use septic systems for waste disposal. The variable can serve to distinguish different source loading rates for nitrate contamination, but also can represent susceptibility conditions for nitrate and arsenic due to infiltration of septic leakage. The septic/sewer ratio originates from a 100-m cell resolution raster that represents the percentage of housing units with a septic-system disposal method for a given census block group. The data were derived from the 1990 Census of Population and Housing Summary Tape File 3A (U.S. Bureau of the Census, 1992) that was developed by the NAWQA Program on the basis of three fields in the table: public sewer (H0240001), septic tank or cesspool (H0240002), and other means of disposal (H0240003). The number of septic tanks was divided by the sum of the three sewage disposal methods for a given block group. The calculations were joined to a block-group raster on the basis of the block-group identifier. Zonal-average statistics for the 100-m raster were calculated for the 3-km model grid.

Population data represent a generalized surrogate variable for nitrate sources associated with human activities, as well as aquifer susceptibility to nitrate contamination or arsenic enrichment due to recharge processes associated with human activities. Population and population density data originate from a 1-km cell resolution raster of LandScan global population data for 2005. LandScan data are produced at the U.S. Census Block level and modified by using photographic interpretation, image analysis, and population modeling (Oak Ridge National Laboratory, 2005). The population data were used to represent model grid-cell population and population density, and basin population and population density. The population data were also used to disaggregate county-level public-supply water use (discussed below). Population density estimates were calculated as a zonal average for the 3-km model grid and for the alluvial basins. Total population estimates for the 3-km model grid and alluvial basins were calculated as the zonal average multiplied by the area of the zone.


OREGON
IDAHO
NEVADA
WYOMING
UTAH
COLORADO
ARIZONA
NEW MEXICO
CALIFORNIA
MEXICO
Pacific Ocean
Study-area boundary
120°
115°
110°
105°
40°
35°
0 50 100 150 200 MILES
0 50 100 150 200 KILOMETERS
EXPLANATION
1 Oregonian Coastal Conifer Forest
2 Cascade-Sierran Montane Conifer Forest
3 Cascade-Sierran Subalpine Conifer Forest
4 Oregonian Deciduous and Evergreen Forests
5 Great Basin Conifer Woodland
6 Great Basin Shrub-Grassland
7 Open Water Lakes
8 Great Basin Desertscrub
9 Cascade-Sierran Alpine Tundra
10 Rocky Mountain Montane Conifer Forest
11 Rocky Mountain Subalpine Conifer Forest
12 California Evergreen Forest and Woodland
13 California Chaparral
14 Great Basin Montane Scrub
15 California Valley Grassland
16 Rocky Mountain and Great Basin Alpine Tundra
17 Mojave Desertscrub
18 Plains Grassland, Shortgrass Communities
19 Semidesert Grassland
20 California Coastalscrub
21 Southwestern Interior Chaparral
22 Sonoran Desertscrub Arizona Uplands
23 Sonoran Desertscrub Lower Colorado River Valley
24 Chihuahuan Desertscrub Cochise-Tranpecos
25 Madrean Evergreen Forest and Woodland
26 Madrean Montane Conifer Forest

U.S. Geological Survey digital data, 1:2,000,000, 2,500,000, and 5,000,000 scale, 2003, 2005, and 2006
National Elevation Data 1:24,000 scale, 1999
Albers Equal Area Conic projection, NAD 83

Biotic communities from Brown and others, 2007

Figure 5. Biotic communities in the Southwest Principal Aquifers study area.

Land-use variables include grid-cell scale and basin-scale agricultural land and urban land, and basin-scale rangeland and other (uncharacterized) land cover. Like population variables, these represent generalized surrogate variables for nitrate sources associated with human activities, as well as aquifer susceptibility to nitrate contamination or arsenic enrichment due to recharge processes associated with human activities. The original data were obtained from the National Land Cover Database (NLCD) dataset for 2001 (U.S. Geological Survey, 2003b), which was coordinated and produced by the Multi-Resolution Land Characteristics Consortium of nine federal agencies. The NLCD is a nationally consistent, 30-m resolution raster representing natural land cover and human-related land use for the United States (Homer and others, 2004). The data were generated from Landsat 5 Thematic Mapper and Landsat 7 Enhanced Thematic Mapper Plus satellite imagery in 2001 and depict 29 classes of land-cover data. Data from the 30-m raster and the 3-km model grid were resampled to 100-m cells. Area weighted averages were calculated for the NLCD land-use variables by using a raster combine (McCoy and others, 2001) and a pivot table. A “combine” is a geospatial analysis in which two or more rasters are merged together, and new values and cell counts are assigned for the coincident cells. Attribute tables from the original data are joined to the combine raster table on the basis of the original values. The new attribute table was exported and cross tabulated by using a pivot-table function. From the pivot table, summary statistics of weighted area for land use were calculated and assigned to a model grid cell. The NLCD data also were used to distribute county-level irrigated agricultural water use to individual raster cells within the study area, as described in the "Aquifer Susceptibility Variables" section.

Geologic sources consist of variables that represent the extent of seven different types of bedrock occurring in the contributing area of the hydrogeologic area that contains the grid cell for which the characterization was made. For each rock type, there is one variable to represent the percentage of that rock type within the contributing area, and another that represents the distance from the model grid cell to the nearest outcrop of that rock type. The rock types include carbonate, such as limestone or dolomitic rocks; crystalline, such as plutonic and metamorphic rocks; clastic sedimentary rocks, such as sandstones or siltstones; mafic volcanic rocks, such as basalt; felsic and silicic volcanic rocks, such as rhyolite; intermediate volcanic rocks, such as andesite; and undifferentiated volcanic rocks, which are mixed or undetermined regarding their magnesium and iron content. The geologic variables representing the distance to the rock type serve as source variables, but also can serve to represent geochemical conditions because geochemical conditions likely vary by rock type and with distance along a flow path away from the surrounding mountains. The geologic variables were derived from a 100-m raster dataset consisting of bedrock geology for California, parts of Nevada, Utah, Colorado, New Mexico, and Arizona, and is a modified version of existing state geologic maps as described in McKinney and Anning (2009). The area of each rock type was computed and expressed as a percentage for the contributing part of the hydrogeologic areas and then assigned to the model grid cells within the hydrogeologic area. Distances from the model grid cell of interest to the geologic units were calculated by using the Euclidean distance function (McCoy and others, 2001), which is the shortest, straight-line distance between the two. The spatial distribution of selected geologic source variables is shown in appendix 12.

Aquifer-Susceptibility Variables

Variables representing the susceptibility of the basin-fill aquifers to nitrate contamination or arsenic enrichment were derived from existing datasets of various scales and were re-calculated to represent the model grid cells used for the random forest classifiers. The aquifer susceptibility variables represent general position along a groundwater flow path, soil properties, water use and hydroclimatic conditions, and basin groundwater budget terms (table 3).

Numerous geochemical modeling studies have shown groundwater to evolve along flow paths from recharge areas to discharge areas (for example, Robertson, 1991, and Thomas and others, 1996). An innovation of this study was the development of "flow-path variables" that provide a general spatial reference for the position of a model grid cell within an aquifer of a given alluvial basin. The flow-path variables include aquifer-penetration depth, well depth, water-level depth, land-surface slope, land-surface elevation, land-surface elevation percentile, basin elevation, and distance to the basin margin. These variables provide a means for the nitrate and arsenic classifiers to reflect the effects of susceptibility factors associated with (1) upper basin margins as opposed to basin lowlands and (2) shallow as opposed to deep parts of the aquifer. These variables serve as surrogates for processes that tend to take place in different locations of the basin, such as recharge along the upper basin margin, discharge in the basin lowlands, and geochemical evolution of groundwater along flow paths between recharge and discharge areas. The variables also can represent aquifer matrix conditions, such as highly-permeable coarse-grained deposits near the margins and poorly-permeable clay deposits toward the central parts of the basin. In addition, depth-related variables allow for attenuation of contaminant transport through unsaturated and saturated basin fill, as well as geochemical conditions associated with shallow or deep groundwater. The spatial distribution of selected flow-path variables is shown in appendix 13.

Distance to the basin margin was small for cells adjacent to bedrock surrounding the basin-fill aquifers, and increased along flow paths toward the center of the basin (fig. 6*A*). Distance to basin margin from the model grid cell to the contact between the basin alluvium and bedrock was calculated by using the Euclidean distance function. Land-surface elevation percentile was calculated as the grid cell's average elevation (fig. 6*B*) as a percentile of the elevation data for all model grid cells within an alluvial basin, and was developed with an underlying assumption that land-surface elevation can serve as a proxy for groundwater elevations, with flow in the basin-fill aquifers moving from areas of higher land-surface elevations to areas of lower land-surface elevations. Land-surface elevation percentiles ranged from 100 for the highest model grid cell in the basin to 0 for the lowest and, therefore, represent a general measure for location in the basin-fill aquifer and along flow paths (fig. 6*C*). For topographically closed basins, the lowest point was typically in the central parts of the basin, whereas for topographically open basins, the lowest point was on the basin margin adjacent to the next down-gradient basin. Land-surface slope for a cell was calculated by using the rate of change in elevation for each of the eight neighboring cells. Land-surface slopes were generally steeper along the upper basin margins next to the bedrock mountain front and shallower in the lowland parts of the basin (fig. 6*D*).

Land-surface elevation, percentile, and slope for the model grid cells were computed on the basis of digital elevation model (DEM) raster data from the USGS 30-m National Elevation Dataset (NED; U.S. Geological Survey, 1999). First, NED data were hydraulically conditioned; artificial sinks and peaks were filled or leveled to remove inaccuracies resulting from errors in creating the elevation data (McCoy and others,

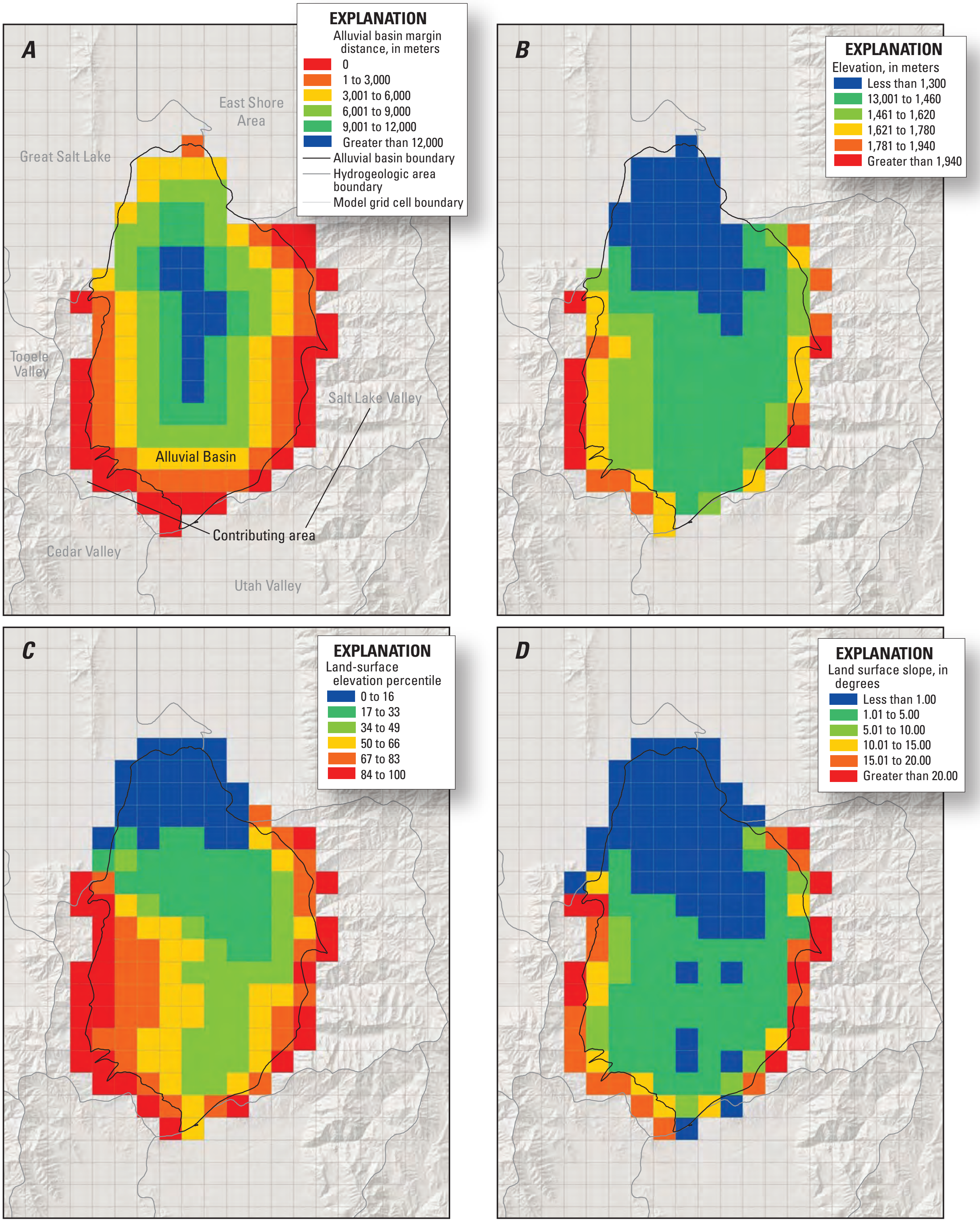


Figure 6. Spatial distribution of flow-path variables in an example basin from the Southwest Principal Aquifers study area: *A*, distance to basin margin; *B*, land-surface elevation; *C*, land-surface elevation percentile; *D*, land-surface slope.

2001). Next, the regional 2-degrees-latitude by 6-degrees-longitude NED data were merged into a single raster for the entire study area. Finally, slope, elevation, and elevation percentile statistics were calculated by using the merged NED data for each model grid cell and alluvial basin.

Water-level and well-depth data provide information about the vertical location of a groundwater sample within the aquifer and its position relative to the land surface, which is important especially where sources are on the land surface and contaminants are transported through the unsaturated zone or where certain geochemical conditions are likely to differ with depth. Water-level and well-depth data were obtained from the NWIS database. Aquifer-penetration depth was computed as the well depth minus the water-level depth and, thereby, represents that part of the aquifer from which the well could be drawing water. While water-level depth, well-depth, and aquifer-penetration depth were all used in the confirmatory classifier, only aquifer-penetration depth was used in the prediction classifier. This occurred because water-level and well-depth data were available only for the wells included in the training datasets and were not available universally for all model grid cells across the SWPA study area, which is a necessity for using the model to predict concentrations. For the prediction classifier, a prescribed aquifer-penetration depth was used to obtain model predictions. For more information on the aquifer-penetration depth variable, see the text box "How is depth treated in the prediction classifier?" In some cases, water-level depth, well depth, or aquifer-penetration depth were unavailable for training set observations. For the purpose of training the classifiers, these observations were filled in with the random forest classifier's algorithm to estimate missing data, which is discussed later in the section "Random Forest Classifier Overview." More precise estimation regarding the part of the aquifer tapped by a well could be determined by using top and bottom of well-screen data; however, such data were far less commonly available than water-level and well-depth data.

Soil properties can influence certain processes that affect contaminant transport, such as recharge. Likewise, in areas of shallow groundwater, soil properties can affect reduction/oxidation processes, which affect contaminant fate. Soil properties are represented by several explanatory variables: seasonally high water table, extent of hydric soils, organic material content, clay content, silt content, and sand content. Soil property data were derived from the State Soil Geographic (STATSGO) database soil maps, which were available at scales 1:12,000 to 1:63,360 (Natural Resources Conservation Service, 2010). The soils data used in this study were converted to a raster from STATSGO polygon data as described in Wolock (1997). As part of this conversion, associated soil attribute data in the STATSGO database first were processed into a weighted average by combining multiple soil layers and components in the database on the basis of thickness and percentage of composition in the areal extent. Area weighted averages of the STATSGO 1-km raster soil properties were calculated for the 3-km model grid by using the combine method described previously. The spatial distribution of selected soil-property variables is shown in appendix 14.

Water-use and hydroclimatic variables provide information about water fluxes that can affect nitrate contamination or arsenic enrichment. This group includes basin-scale variables, such as the water-resources development index, recharge for the contributing area, and recharge for the alluvial basin, as well as model-grid-cell scale variables, such as groundwater use for agricultural purposes and for public supply, surface-water use for agricultural purposes and for public supply, mean air temperature, and average annual potential evapotranspiration. The spatial distribution of selected water use and hydroclimatic variables is shown in appendix 15.

Large water withdrawals from wells and human influences stemming from urban and agricultural land uses increase the potential for the degradation of groundwater quality in basin-fill aquifers (Bexfield and others, 2011). Estimated use of groundwater and surface-water resources for irrigated agriculture and for public-supply were available from McKinney and Anning (2009) as 100-m cell resolution raster data of disaggregated county-level water-use data for the study area. The data are based on tabular data for the estimated use of water in the United States in year 2000 (U.S. Geological Survey, 2004). Methods for the development of the public-supply and irrigated agriculture water-use data are described in McKinney and Anning (2009). Zonal sum statistics for the model grid cells were calculated by using the 100-m water-use data.

The water-resources development index represents the variety and magnitude of human-related recharge and discharge processes (Anning and Konieczki, 2005). The index was computed as the log (base 10) of the sum of the annual public-supply and agricultural water use (described previously) for model grid cells within a hydrogeologic area. Hydrogeologic areas with greater water-resource development indexes are characterized by greater ground- and surface-water development. Each model grid cell was assigned the water-resources development index for the hydrogeologic area in which the cell resides.

Potential evapotranspiration and mean air temperature are strongly correlated with elevation within a basin and can provide information about how climate processes affect the susceptibility of basin-fill aquifers to nitrate contamination or arsenic enrichment. Potential evapotranspiration represents the evaporation and transpiration potential of a plant-covered land surface given an unlimited supply of water. Potential evapotranspiration was an average annual value computed from monthly data for 1970–2006 that were estimated by Flint and Flint (2007) from a 270-m raster-based model that utilized computed solar radiation, vegetation cover, and the Priestley-Taylor equation (Priestley and Taylor, 1972). Estimated mean air temperature data for the study area are based on 30-year (1971–2000) average annual temperature data available from the PRISM Group at Oregon State University (2004b). The PRISM approach is an iterative process that models monthly and annual temperature data by using weighted weather-station climate data and linear elevation regressions. The available station data do not necessarily represent conditions for the surrounding locations; therefore, the linear regression is modified for each model cell to reflect changes in climate and elevation. The PRISM raster data are available at 800-m cell resolution, and zonal average statistics were calculated for the potential evapotranspiration and mean air temperature rasters for the 3-km model grid.

Estimates of recharge from precipitation in the alluvial basin and in its contributing area were calculated by using the Maxey and Eakin (1949) empirical precipitation-recharge relationship. The method uses a fraction of precipitation (0–25 percent) that is based on the following five precipitation zones (Maxey and Eakin, 1949):

Precipitation zone	Fraction applied to precipitation
Less than 8 in/yr	0 percent
8 to 12 in/yr	3 percent
12 to 15 in/yr	7 percent
15 to 20 in/yr	15 percent
Greater than 20 in/yr	25 percent

The data layers used for this analysis were PRISM average annual precipitation data (PRISM Group, 2004a), hydrogeologic-area boundaries (McKinney and Anning, 2009) for delineation of contributing areas, and alluvial basins derived from geology. PRISM precipitation data were converted (reclassified) to recharge by multiplying each of the five precipitation zones by the appropriate fraction of precipitation tabulated as shown previously. Contributing areas for recharge were created by subtracting the alluvial basin from the hydrogeologic area. Zonal-sum statistics were calculated for recharge in the contributing area and alluvial basin of each hydrogeologic area. The recharge values were assigned to the model grid cells in a basin or contributing area.

Basin groundwater-budget variables were available from Bexfield and others (2011) for the 16 case-study basins (fig. 1) and included several variables that represent annual rates of recharge and discharge in the basin-fill aquifer for different hydrologic processes (table 3). Rates for individual components of recharge are expressed as percentages of total recharge, and rates were similarly expressed for discharge components. Basin groundwater-budget variables also included change in rates of recharge, discharge, and storage from predevelopment conditions to modern (circa year 2000) conditions, as well as a rough estimate of the groundwater residence time in the upper 1,000 feet of the aquifer. Methods for determining the basin groundwater-budget variables vary somewhat by basin and are discussed in more detail in Bexfield and others (2011) as well as Thiros and others (2010). Values for the basin groundwater-budget variables were assigned to each cell within a case-study basin. Because data for these variables were not available universally throughout the SWPA study area, the variables were only used in the confirmatory classifiers.

Geochemical Variables

Geochemical variables were included in the confirmatory classifiers to elucidate the relation of selected constituents and properties to nitrate and arsenic concentrations (table 3). As described previously, samples and wells were selected to represent each model grid cell for which nitrate or arsenic concentration data were available. For the samples selected to represent nitrate and arsenic concentrations in the training dataset, additional constituent concentration data were retrieved from the NWIS database. These constituents and properties include alkalinity, dissolved oxygen, pH, dissolved solids, and dissolved bicarbonate, chloride, iron, manganese, molybdenum, orthophosphate, selenium, and sulfate. Not all samples had data available for every constituent; the random forest classifier algorithm, however, has a means to handle missing data in the training and prediction datasets and is discussed in the "Random Forest Classifier" section. For the arsenic classifiers, nitrate concentrations also were included as a geochemical variable. As previously described, several of the constituents had concentration data below the MRL; for these cases a value of one-half of the highest MRL was substituted for the censored value. This effectively adjusts lab analyses to a single MRL and sets the censored data to a single value. Selection of the substitution value, be it one-half or one-tenth of the MRL, is arbitrary for the random forest classifier because it classifies on the basis of specific splitting or threshold values, which is explained in detail in the "Algorithm Overview and Application" of the "Random Forest Classifier" section.

The geochemical data were included within the nitrate and arsenic confirmatory classifiers for several reasons. Molybdenum and selenium concentration data were included because, like arsenic, these elements occur as oxyanions in the natural environment and could possibly be affected by similar geochemical controls. Chloride concentration data were included to examine for conservative behavior along flow paths. Dissolved-solids concentrations were included as an indicator of processes that concentrate solutes along flow paths. Orthophosphate concentration data were included because of its similar sorption chemistry to arsenic and because of its use in fertilizers. Dissolved oxygen, iron, manganese, nitrate (only for use in the arsenic classifier) and sulfate concentration data were included as indicators of the redox conditions present. Finally, many biogeochemical processes are influenced by pH, and so pH data also were included in the confirmatory classifiers.

Random Forest Classifier

The random forest classifier was selected as the type of statistical model to use in relating constituent concentrations to explanatory variables representing sources, aquifer susceptibility, and geochemical conditions. The classifier was implemented by using a Fortran program provided by the researchers that developed its algorithm (Breiman and Cutler, 2010). This section provides an overview of the random forest classifier, its strengths and limitations, and details on its application in this study. Additional information on the classifier can be found in Breiman (2001) and Breiman and Cutler (2010), and information on classification trees can be found in Breiman and others (1984) and Hastie and others (2001).

Strengths and Limitations

The primary reason for selecting the random forest classifier to predict concentrations was that it is a rule-based method, which fits the conceptual models. The conceptual models indicate that concentrations of nitrate, arsenic, and other contaminants (1) were dependent on several natural and human-related factors, and (2) had relations with these factors that were generally conditional. That is, concentrations were dependent on a combination of specific source, susceptibility, and geochemical conditions. For example high concentrations are expected if (1) there is a significant source and (2) soil and sediment properties permit transport, and (3) there is sufficient recharge to transport the substance from the land surface to the aquifer, and (4) geochemical conditions are unfavorable to degradation mechanisms. If any one of these four specific sets of conditions is not met, a different outcome will result. Classification trees excel over least squares regression and logistic regression methods for situations where a series of conditions must be met among the explanatory variables to achieve a given response. In these regression models, the effect of each explanatory variable on the response variable is assumed independent of the effects of other explanatory variables, which is inconsistent with the conceptual models that describe the vulnerability of basin-fill aquifers to contamination (Bexfield and others, 2011). While the ability to handle such conditional relations was the primary reason for selecting an algorithm utilizing classification trees, it also should be noted that other significant benefits are that they are well suited for handling analyses with many explanatory variables and are nonparametric and, therefore, do not require assumptions about the underlying distribution of the variables (such as normality). Another benefit is that classification-tree methods are robust where variables are collinear (Piramuthu, 2008). Whereas in logistic regression multicollinearity must be avoided by combining variables or by removing variables from the analysis, classification tree algorithms are not affected by multicollinearity in the training dataset variables. In fact, removal of multicollinearity can result in poor classification performance (Piramuthu, 2008).

In addition to the aforementioned benefits of using a classification tree-based method, some of the beneficial features of random forest classifiers include the following (Breiman and Cutler, 2010):

- The algorithm has the highest prediction accuracy among ensemble classifier algorithms. It generates an internal, unbiased error estimate as the forest building progresses and has methods for balancing error in datasets with unbalanced class populations.
- The algorithm runs efficiently on large datasets and can handle thousands of explanatory variables without variable deletion. Where study objectives include a comparison of the importance of different explanatory variables in the model, the random forest algorithm allows for retaining all significant variables.
- The algorithm provides an effective method for estimating missing data and maintains accuracy when a large proportion of the data are missing.
- The algorithm provides estimates of which variables are important in the classification.
- Generated forests can be saved for future use on other data.

There are several disadvantages to using the random forest classifier and limitations to the classifiers developed in this study. The primary disadvantage of using the random forest classifier is that prediction results come from many classification trees, so there is no single model to directly examine and interpret the decision rules like there is when using a single classification tree or a set of coefficients from a regression model. Such examination of decision rules, coefficients, or model structure is advantageous for meeting the objective of evaluating the conceptual models of factors affecting nitrate and arsenic concentrations (for example, Bexfield and others, 2011). That objective, however, was second in priority to predicting concentrations for the SWPA study area and was still accomplished through less direct evaluation methods that are discussed at the end of this section.

Gashler and others (2008) found that when using datasets with large numbers of irrelevant explanatory variables, as is done in many data mining applications, other algorithms that use ensembles of classification trees can perform better than the random forest algorithm. In this study, irrelevant explanatory variables were avoided by careful selection of variables that are relevant to the conceptual models of factors affecting nitrate and arsenic concentrations (for example, Bexfield and others, 2011). Another disadvantage for this study is that the random forest classifier algorithm does not consider the constituent concentration classes as ordinal; that is, the algorithm does not recognize the order to the classes. Consequently, misclassifications can fall into any one of the incorrect classes without benefit of an algorithm that creates a greater likelihood of falling into a class with concentrations nearest those of the correct class.

It should be recognized also that this is a retrospective observational study. In such investigations, response variable *Y* and potential factors *X* (explanatory variables) are observed or obtained from an existing dataset, and the effects of *X* on *Y* are determined. In some cases there can be strong effects or correlations between *X* and *Y*; however, in such cases the causality of *X* generating a response in *Y* is not by any means assured. In fact, the correlations only corroborate existing hypotheses. In some cases spurious correlations between X_1 and *Y* can arise because the spatial distribution of X_1 follows the spatial distribution of another variable X_2 or the net effect of variables X_2–X_M, which have true causal effects on Y. The larger the number (M) of explanatory variables composing X, the greater likelihood such spurious correlations will occur. There also can be other important factors affecting response variable Y that were not included in the set of explanatory variables X.

Another limitation to the modeling approach is that explanatory variables composing X are not congruently aligned in time and in space with response Y, and this could decrease the accuracy of the classifier and its predictions. Concentrations of arsenic and nitrate represent conditions observed over the course of about three decades (1980–2009), and for any given location, can change with time. Similarly, some explanatory variables, such as nitrogen loading, land use, or recharge variables, represent conditions for a specific year or set of years. Another complication to the modeling approach is the presence of time lags between the change in such time-dependent variables and the corresponding response of nitrate or arsenic concentrations in groundwater to that change. In short, the nitrate and arsenic classifiers were developed by using an implicit assumption of geochemically steady-state conditions when, in fact, transient conditions were likely present, at least in areas affected by water-resources development. Similar to the potential misalignment between X and Y in time, there is also likely misalignment in space. For example, the effects of several variables, such as land use, are represented at the spatial scale of the model grid cell or of the entire alluvial basin. Because concentration is measured for a specific point within the cell, some of the area represented by the explanatory variable occurs downgradient of the well from which the nitrate or arsenic sample was collected and, therefore, should not affect the concentration. Another potential misalignment in space is that source, susceptibility, and geochemical conditions are only represented as explanatory variables for either the cell or the basin the cell is in. In some cases it could be that conditions in an adjacent cell upgradient from the test cell are the most important factors affecting concentrations for the cell of interest, and those are not represented as explanatory variables.

A general limitation to this and other regional-scale water-quality modeling studies is that the models are developed to evaluate variables influencing concentrations across a broad region. Inevitably, there will be some variables that can be important for specific localized areas that were either not considered in the regional model or were included but found to be of lesser importance because these local characteristics were masked by larger regional influences. In some cases, conceptual models can indicate that a certain factor imparts an important regional-scale influence to the constituent of interest; however, data are not always available “wall-to-wall” across the region to adequately represent this factor. For this study, examples of such data limitations include the lack of regional datasets for the spatial distribution of depth to water, occurrence of confining layers, and geochemical data.

Algorithm Overview and Application

Random forest is a classification algorithm that builds many classification trees (Breiman and others, 1984) and assigns the class to an observation that is most often assigned by the ensemble of individual trees (Breiman, 2001; Breiman and Cutler, 2010). For training an individual classification tree, the model space is split into multiple regions by using a set of recursive-binary partition rules. Consider, for example, a training dataset with N observations of the response variable Y and two (M=2) explanatory variables X_1 and X_2. While Y consists of categorical values, values of X can be categorical or numeric. The model space defined by X_1 and X_2 is first split into two regions at X_1=t_1 (fig. 7*A*). As a result of splitting the model space, responses of Y are more homogeneous in each of the two resulting regions than in the previously unpartitioned model space. Next, one of these regions is further split into two regions, and this process is continued until a subsequent split would yield fewer observations in a region than a specified minimum number of observations to remain in each node. In the random forest algorithm, trees are not pruned by removal of any nodes in the tree. The statistical nature of the classification trees is rooted in the algorithm used to grow the tree, where the splitting variables (X_1, X_2, …X_M) chosen at each split and the splitting values (t_1, t_2, … t_i) are recursively determined in a manner so as to minimize heterogeneity and maximize homogeneity of response values in each region of the model space. For i splits, the resulting model space consists of i+1 regions R_1, R_2…R_{i+1} defined by t_1, t_2…t_i (fig. 7*A*). The most common class Y for observations within each region is assigned to represent that region (for example, diamonds for R_1 in figure 7*A*). Typically, there are more regions than classes, so multiple regions can be represented by a given class (R_1 and R_5 in fig. 7*A*). The models are called “classification trees” because of the tree-like appearance when the recursive binary splits are diagramed (fig. 7*B*). Models with three or more explanatory variables (M is greater than or equal to 3) are trained in the same manner as that described here, but are more difficult to geometrically illustrate.

To predict the class of the response variable Y for a new unclassified observation, values for explanatory variables X_1, X_2, …X_M for the observation are moved through the trained classification tree, with decisions made at the splitting nodes of the tree based on each value t_1, t_2…t_i. The decision path ultimately leads to one of the i+1 terminal nodes representing the model-space regions R_1, R_2… R_{i+1} (figs. 7*A* and 7*B*). The most common class for that region determined from model training is assigned as the new observation’s predicted class. For example, a new unclassified observation with values of X_1 and X_2 that fall within region R_3 would be assigned as a circle because 9 of 11 training observations in R_3 are circles.

The random forest classifier grows many classification trees from a single training dataset, and the resulting ensemble of trees constitutes a trained forest. To classify a new object with an input vector of explanatory variables X_M, the object is run through each of the individual classification trees in a trained

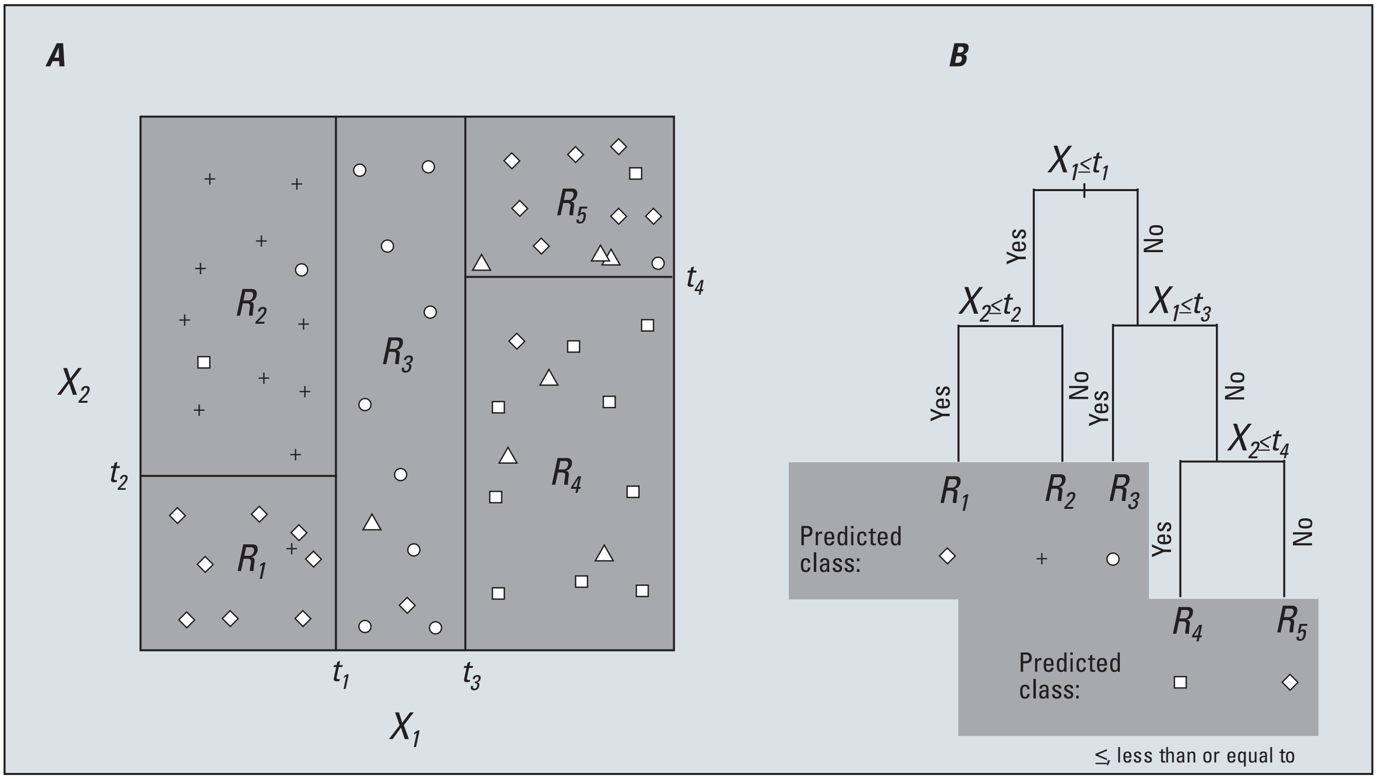


Figure 7. Example of recursive binary partitioning: *A*, shown geometrically; *B*, shown as a classification tree.

forest. A classification is assigned from each tree, and the most commonly assigned class from the ensemble of trees is the class assigned to the object. In some cases, training observations were missing data for water level, well depth, and aquifer-penetration depth, as well as concentrations of various constituents representing geochemical conditions. These missing data were estimated by using the random forest classifier's missing-value replacement function. This method uses "hot-deck imputation" and is described in more detail in Breiman (2001) and Breiman and Cutler (2010).

Individual classification trees are sensitive to the observations and the explanatory variables used to grow each tree. In some cases, small changes in the observations or explanatory variables used in tree growth result in a different series of splits and classification assignments. This instability results primarily from the hierarchical nature of the tree—the effect of an error at the top split is propagated to all of the splits below it. The random forest classifier determines and reduces the variance of prediction estimate by using results from an ensemble of trees rather than a single tree in a method called "bootstrap aggregating" (Breiman, 2001, and Breiman and Cutler, 2010). In bootstrap aggregating, an ensemble of different classification trees is obtained through random selection of observations and of explanatory variables during individual tree growth. This reduces the effects of outliers in the dataset and results in a more robust ensemble of classification trees. Consider a training dataset with N observations. For training an individual tree, N observations are randomly selected, but with replacement, from the original training dataset. This process results in a different set of observations used to grow each tree, and any given dataset contains duplicates of about one-third of the observations as a result of replacement during the selection process. Because some observations are selected more than once, about one-third of the observations are not selected, and these are set aside for evaluating model accuracy, which is discussed later in this section.

Differences in the classification trees composing the forest are generated by using a subset of the explanatory variables for determining each splitting node variable and splitting value also. The number of explanatory variables to subset, $m<<M$, is user specified and held constant throughout the forest growth. The subset variables are randomly selected without replacement for each splitting node of each tree. Breiman (2001) found that the overall error of the trained forest depends on the error rate of individual trees and the correlation between them. A decrease in correlation between individual trees decreases the forest error rate, and decreasing the error of individual trees decreases the forest error rate. While reducing *m* decreases the correlation of individual trees, it increases their error rate. Breiman and Cutler (2010) suggest a trial and error approach for selecting *m* to minimize the forest error rate, stating that theoretically *m* should be about equal to the square root of *M*.

The random forest classifier does not require cross-validation or a separate test set to get an unbiased estimate of the test-set error because it is determined while training the forest (Breiman, 2001). Consequently, this study did not place aside a test set for subsequent evaluation of prediction uncertainties.

For each individual tree constructed, a different sample of N observations is selected with replacement from the original data, which leaves about one-third of the observations out of the sample. The latter are called “out-of-bag” observations. After each tree is constructed, the out-of-bag observations are then run through the tree for classification. In this manner, each observation is left out of about one-third of the classification trees, and these instances form a test set for evaluating the classifier’s accuracy. After all trees have been grown for the forest, the most common predicted class is determined for a given observation while it was held out-of-bag, compared to the true class of the observation, and assessed whether it was classified correctly. The out-of-bag error estimate is equal to the percentage of total observations misclassified. Detailed out-of-bag error estimates are most usefully displayed in a matrix where, for each true class of *Y*, observation counts are tabulated for the predicted classes of *Y*.

Variable Selection and Classifier Goodness-Of-Fit

Source, susceptibility, and geochemical variables were selected for inclusion in the nitrate and arsenic classifiers on the basis of variable importance statistics. The goodness-of-fit for the classifiers was evaluated on the basis of the distribution of misclassification errors with respect to (1) observed concentration class, (2) geographic location, (3) statistical distribution of explanatory variables, and (4) estimated sampling error. Classifier goodness-of fit is assessed primarily by analysis of misclassifications rather than by direct comparison of the distributions of observed and predicted concentration classes for a given region. The reason for this is the distributions of explanatory variables are not identical for the training dataset, which generally represents the part of aquifers actively being used for groundwater supply, and the prediction dataset, which represents all basin-fill aquifers across the SWPA. In particular, explanatory variables tend to reflect greater amounts of human activities, greater recharge rates, and less frequent occurrence of volcanic rocks in the training dataset than in the prediction dataset. Without identical distributions of explanatory variables, concentration distributions for the training observations and predictions are not expected to be identical either, which complicates direct comparison of them.

Variable-importance statistics were used to determine which variables to retain in the classifiers during their development. The variable importance indicates the change in classification accuracy for out-of-bag training observations as a result of perturbing the numerical values for that explanatory variable and, therefore, shows how sensitive the prediction accuracy of *Y* is to that variable. Computation of variable importance follows several steps. After growing a given tree, the values for variable X_i for the out-of-bag observations are randomly permuted to other values. Next, the out-of-bag observations are classified by using the permuted values, and the decrease in the number of assignments for the correct class due to permuting X_i is computed. The more sensitive the classifier prediction accuracy is to the explanatory variable X_i, the more important the variable is, and the larger increase in the misclassification errors of *Y* as a result of permutation. The average, standard deviation, and standard error for the increase in incorrect assignments over all trees in the forest are determined, and then these statistics are used to compute a z-score and significance value for the variable importance (Breiman and Cutler, 2010). The z-scores are reported in the results section as the “standardized importance score.” This procedure is repeated for all M explanatory variables. A higher standardized importance score indicates a greater sensitivity of the classifier’s accuracy to the explanatory variable. Standardized importance scores greater than about 2 indicate that the probability of achieving a similar effect on the classifier predictions by chance with a variable populated with random numbers is less than 5 percent (p is less than 0.05). For standardized importance scores greater than about 3.3, the classifier output reported (censored) p-values as “<0.001.” Most standardized importance scores in the classifier output were greater than 3.3, and so as to allow for differentiation of the sensitivity of the classifier accuracy to each explanatory variable, this study presented the standardized importance scores rather than the p-values.

For preliminary and final classifiers, all variables tested in each classifier were significant to explaining concentration variations (standardized importance scores greater than 2). It was unexpected that all variables would be significant, so the sensitivity of the random forest classifier algorithm to meaningless explanatory variables was tested by replacing the legitimate data for some variables with random numbers. Standardized importance scores for the variables with data replaced by random numbers were all less than 2, thereby confirming the ability of the random forest classifier to detect meaningless explanatory variables.

In addition to testing the ability of the random forest classifier to detect meaningless data, the effects of correlation between the explanatory variables on misclassification rates and standardized importance scores were also evaluated. Some of the explanatory variables are highly correlated because they either provide similar measures of the same factor, for example population and population density, or because they are mathematically related, for example land cover variables and basin water-budget component variables each add up to 100 percent. To assess whether the presence of such correlations had a substantial negative effect on the classifiers, the final nitrate prediction classifier was modified by removing 20 variables from the classifier that were correlated to one or more of the remaining variables. The modified classifier with 38 variables had a comparable misclassification rate to the final classifier (0.3 percent greater), and both classifiers showed a similar order for the ranks of the standardized variable importance scores. Given that the effects were minor for including the 20 additional variables, they were retained in the final classifiers so that they would be available for use in the comparison between the classifiers and the conceptual models.

The goodness-of-fit for the classifiers was assessed on the basis of misclassification errors. To compute misclassification errors, the constituent concentration classes were assigned numbers 1 through 6 for nitrate, and 1 through 7 for arsenic, in accordance with increasing concentrations. The misclassification error for an individual training observation was computed in the same manner as residual errors are computed for regression models:

Observed class = Predicted class + misclassification error,

which can be rearranged:

Misclassification error = Observed class-predicted class

While this approach is standard for statistical models, some find the sign of the errors counter-intuitive. A misclassification error of +1 indicates the prediction is one class less than the observed class (underpredicting), and a misclassification error of -1 indicates the prediction is one class greater than the observed class (overpredicting). Misclassification errors can range within ±5 concentration classes for nitrate, and ±6 concentration classes for arsenic.

The statistical distribution of misclassification errors was evaluated to detect potential for prediction bias. An ideal distribution for misclassification errors would mimic a bell-shaped curve and consist of the largest percentage of errors being equal to zero (correct classification), with a marked decrease in the percentage of errors as the magnitude of the error increases, so that larger errors are less common than smaller errors. In addition, the percentage of positive misclassification errors should be approximately equal to the percentage of negative misclassification errors for an unbiased set of predictions. The random forest classifier algorithm provides a means to weight each class of *Y* so that each class has a similar percentage of misclassifications. While use of such weights generally increases the overall misclassification rate, the benefit is that it provides a mechanism by which each class can be predicted with similar rates of uncertainty. In this study, weights were adjusted manually by iteration to achieve an unbiased distribution of misclassification errors, where the sum of all misclassification errors equaled zero.

The spatial distribution of misclassification errors was examined for geographic patterns that can indicate bias in predictions across large parts of the SWPA study area. General patterns were assessed visually for a qualitative assessment, and for a more quantitative evaluation, the average misclassification error was computed for each basin with more than 15 training observations. If the average misclassification error was greater than 0.50 or less than –0.50, then the basin was identified as having a potential for underpredicted concentrations, or overepredicted concentrations, respectively. The threshold of ±0.50 was selected on the basis of rounding conventions, where numbers greater than this round to ±1.0, indicating bias toward the next larger or smaller class, whereas numbers less than ±0.50 round to 0.0, indicating insufficient bias to change predicted class.

For the prediction classifiers, average misclassification errors were examined across the statistical distribution for each explanatory variable to determine if the classifier was overpredicting or underpredicting concentrations in the SWPA study area relative to high or low values for each variable. In this evaluation, each explanatory variable was examined independently by assigning each observation in the training dataset to one of the six following percentile ranges for that variable: less than 10.0 percent, 10.0 percent to 24.9 percent, 25.0 percent to 49.9 percent, 50.0 percent to 74.9 percent, 75.0 percent to 89.9 percent, and equal to or greater than 90 percent. For each percentile range of each explanatory variable, the average misclassification error and count of the number of training observations was tabulated. If the average error was greater than 0.50 or less than –0.50, then the percentile range for that explanatory variable was declared as having a potential for underpredicted or overpredicted concentrations, respectively. Limited representation by the training dataset of certain environmental conditions in the SWPA study area was identified where less than 100 training observations represented a given explanatory variable percentile range.

The distribution of misclassification errors was compared to the estimated distribution of sampling errors to evaluate the potential limitations imposed on overall classifier accuracy by sampling error. Sources of the sampling error in nitrate and arsenic concentrations include measurement error that arises from field and laboratory procedures, inherit local spatial variation within individual model grid cells, and local temporal variation that could have occurred within individual model grid cells during the study period (1980–2009). Using the training data for model grid cells with two observed concentrations, the distribution of the sampling error for grid cell concentrations was estimated by subtracting one concentration class from the other (table 4). For nitrate, 41 percent of the paired concentration observations were within the same concentration class, and 29 percent had the paired concentration observations differing by one class (±1 class). Arsenic showed similar results where 33 percent of the paired observed concentrations were within the same concentrations class; 36 percent were within one concentration class. Sampling error can be as great as ±5 classes for nitrate and ±6 classes for arsenic. For developing predictive classifiers, the ideal situation would have minimal sampling error—nearly 100 percent of the observations being in the same class and few or no observations in different classes. The estimates above indicate

Table 4. Estimated sampling error for random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.

[**Abbreviations**: ±, plus or minus]

Sampling error	0 (no error)	±1 class	±2 classes	±3 classes	±4 classes	±5 classes	±6 classes
	Percentage of paired training-observation concentrations, by magnitude of sampling error						
Nitrate	41	29	15	8.4	3.2	4.0	No data
Arsenic	33	36	16	7.0	4.2	2.2	1.4

that sampling error for nitrate and arsenic concentrations is high, and that a nitrate classifier without variables explaining within-cell variation would make a correct classification for about 41 percent of the observations at best, or be within ±1 class for about 70 percent of the observations. Aquifer-penetration depth related variables and geochemical variables could explain some variation within a given model grid cell, so inclusion of these variables could potentially increase the correct classification rate to greater than 41 percent. Sampling error is larger for arsenic than for nitrate; consequently, it is possible to achieve a more accurate classifier for nitrate than for arsenic. For a 6-class classifier with uniformly distributed training observations, the likelihood of simply guessing the correct class without a classifier is 1/6 or 17 percent, and, similarly, the likelihood of guessing the correct class for a 7-class classifier is 14 percent. Therefore, for a nitrate classifier without explanatory variables explaining within-cell variation, the correct classification rate can be expected to range between 17 and 41 percent. Similarly, the correct classification rate for an arsenic classifier can be expected to range between 14 and 33 percent.

An estimate of measurement error for nitrate is available from Mueller and Titus (2005), who examined the variation of concentrations in replicate groundwater samples that result from variation in the collection of sample water at a given well and from lab error. The reported standard deviation for the paired replicates was 0.043 mg/L for low concentration (0.02–0.30 mg/L) samples, and 2.9 percent for those with higher concentrations. From these results, it is apparent that measurement error is a small component of the sampling error. The regression analysis previously described for paired samples suggests that temporal trends are not a significant component of the sampling error either, and so it appears that spatial variability is the more significant factor limiting the overall accuracy of the nitrate classifiers. This is likely the case for arsenic classifiers as well.

Sampling error could be reduced by decreasing the number of concentration classes used in the classifier. This approach, however, would also reduce the detail of information provided by the classifier about how concentrations vary across the basin-fill aquifers in the SWPA. Given that spatial variability is a large contributor to the sampling error, another approach is to decrease the cell size of the model grid, for instance from 3-km to 1-km. Such a cell-size reduction, however, would increase computer file sizes, by a factor of 9 for this example, and create a much larger computational burden for developing explanatory variables, managing training and prediction datasets, and training the classifiers.

Evaluation of Conceptual Model

The conceptual models that describe factors affecting nitrate and arsenic concentrations within the 16 case-study basins (Bexfield and others, 2011) were evaluated through examination of (1) standardized importance scores, (2) correlations between the classifier predictions and the explanatory variables, and (3) average predicted classes for selected percentile ranges of each explanatory variable.

The standardized importance scores provide a measure of the sensitivity of the classifier's prediction accuracy to each explanatory variable. Examination of the standardized importance scores allowed for determination of variables in the classifiers, and their corresponding factors in the conceptual models, that were not important for predicting nitrate or arsenic concentrations. Lack of sensitivity by the prediction accuracy to an explanatory variable (standardized importance score <2) contradicts the hypothesis that the variable is an important factor affecting nitrate or arsenic. Sensitivity of the prediction accuracy to an explanatory variable (standardized importance scores >2) corroborates, but does not prove, that the variable is an important factor affecting nitrate or arsenic. Care must be taken not to associate high concentrations of nitrate or arsenic with a high standardized importance score, but rather one should consider a high standardized importance score as indicative of a strong relation between nitrate or arsenic concentration and the explanatory variable.

Investigations of causality between response and explanatory variables usually examine correlations between these variables. Unlike regression models, random forest classifiers do not have model coefficients that can be examined to determine whether concentrations are positively or negatively correlated with each explanatory variable. This is due to the nature of the classifiers, which allows for highly nonlinear relations to occur between the response and explanatory variables. To assess general relations between predicted nitrate or arsenic concentrations and the explanatory variables, the Kendall's tau test was used to evaluate the correlations between predicted classes and explanatory variable data from all 54,854 model grid cells representing basin-fill aquifers in the SWPA study area. For positive correlations, the predicted concentrations are generally higher in locations where the explanatory variable data have high values. Similarly, concentrations are generally lower in locations where explanatory variable data have low values. For negative correlations, the predicted concentrations are generally lower in locations where the explanatory variable data have high values, and concentrations are generally higher in locations where explanatory variable data have low values.

General relations between predicted nitrate or arsenic concentrations and the explanatory variables also were assessed through examination of average predicted classes for selected percentile ranges of each explanatory variable. In this evaluation, each observation in the prediction dataset was assigned one of the following percentile ranges for the distribution of a given explanatory variable: less than 10.0 percent, 10.0 percent to 24.9 percent, 25.0 percent to 49.9 percent, 50.0 percent to 74.9 percent, 75.0 percent to 89.9 percent, and equal to or greater than 90 percent. This was repeated for each explanatory variable. Then for each percentile range of each explanatory variable, an average concentration class was determined and then tabulated. The tabulation shows whether the average predicted concentration class is greater for lesser or for greater values of the explanatory variable. While the tabulation is less robust than Kendall's tau test, it can be used to discover which explanatory variable condition is associated with the highest or lowest average predicted concentration class.

General relations between nitrate and arsenic concentrations and geochemical variables also were evaluated by use of Kendall's tau test for univariate correlations and by examining average nitrate and arsenic concentration classes across selected percentile ranges for the geochemical variables. This was similar to the evaluation for source and susceptibility variables; however, a notable difference in this analysis was that relations were determined using geochemical variable concentrations from the training dataset because of their lack of availability in the prediction dataset.

While both the Kendall's tau test for trend and the examination of average predicted concentration classes across percentile ranges for explanatory variables require considerable computation, these evaluations should be considered more qualitative than quantitative in nature. The reason for this is these two analyses are univariate, that is, they are only examining the relation between a predicted concentration class and a single variable. Concentrations are affected simultaneously by multiple explanatory variables, so variable interactions could convolute the results such that true relations are not reported by these two evaluations.

Nitrate and Arsenic Classifiers and Predicted Concentrations

The prediction and confirmatory classifiers for nitrate and for arsenic performed well for assessing the vulnerability of basin-fill aquifers in the SWPA study area to contamination by these constituents. All four random forest classifiers generally produced unbiased predictions, and misclassification errors for each classifier were about as low as one could expect given the high sampling error. Predictions from the classifiers indicated that for aquifer-penetration depths of 200 ft, only about 2.4 percent of the basin-fill aquifers area exceeded the drinking-water standard for nitrate. About 42.7 percent of the basin-fill aquifers area, however, was predicted to exceed the arsenic drinking-water standard. The nitrate and arsenic classifiers were found to be generally consistent with, and provided additional information and detail for, the conceptual models for natural and human-related factors affecting these constituents as described in Bexfield and others (2011).

All percentile ranges for each explanatory variable were well represented by training observations, and average misclassification errors for the percentile ranges of each variable were generally small, less than ±0.50 in nearly every case. The combination of good representation and low average misclassification errors indicated that predicted nitrate and arsenic concentrations represent observed concentrations in a fair and unbiased manner across the full range of explanatory variable values. The lack of bias in predicted concentrations with respect to location and to different values of the explanatory variables was important not just for creating good model fit, but also because it permitted the use of the predicted concentrations (rather than training observations) in (1) describing the regional distribution of concentrations and (2) examining the relation between concentrations and explanatory variables. Use of the predicted concentrations in these two aforementioned analyses is important because they represent the SWPA study area as a whole, whereas the observed concentrations only represent a subset of the SWPA study area.

For both nitrate and arsenic, the confirmatory classifier accuracy was most sensitive to the explanatory variables that represented geochemical conditions, which is consistent with the conceptual models. In addition, observed nitrate and arsenic concentrations were well correlated to geochemical conditions (table 5). While the confirmatory classifier accuracy was most sensitive to geochemical variables, neither the source variables nor aquifer susceptibility variables were particularly predominant over the other in terms of affecting accuracy. In general, the accuracy of each classifier was more sensitive to variables representing source conditions or aquifer susceptibility conditions at the grid-cell scale than to comparable variables representing basin-scale conditions.

Flow-path variables that represented the model grid cell's position within the basin and aquifer as well as proximity to sources proved useful in the classifiers and led to predictions that revealed spatial patterns in concentrations along likely groundwater flow paths within each basin. For nitrate, concentrations generally decreased along flow paths from the upper basin margin through the middle parts of the basin and ending in the basin lowlands, except where substantial nitrogen loadings occurred in urban and agricultural areas and caused increases in concentration. In contrast to nitrate, concentrations of arsenic generally increased along flow paths, and concentrations were generally higher in basins where recharge is low and in basins surrounded by volcanic or crystalline rocks.

Predicted nitrate and arsenic concentrations are primarily discussed in this report for a single aquifer-penetration depth, 200 ft, because at the regional scale, predicted concentrations generally did not systematically vary by aquifer-penetration depth. This is consistent with the training dataset, which through regression analysis showed only slight decreases in nitrate concentrations and no systematic trends in arsenic concentrations with aquifer-penetration depth. To determine the extent of any systematic variation in predicted concentrations with aquifer-penetration depth, predicted concentration classes were examined for 8 different prescribed aquifer-penetration depths in each cell: 50, 100, 150, 200, 250, 500, 750, and 1,000 ft. While aquifer thickness generally increases laterally away from basin margins, there likely are numerous model grid cells along the basin margins where the prescribed aquifer-penetration depths exceed the actual aquifer thickness. Because of lack of thickness information, no attempt was made to account for this. For each model grid cell, presence of an overall increase, decrease, or lack of change in concentration with aquifer-penetration depth was determined by using

Table 5. Relation between observed nitrate and arsenic concentration classes and geochemical variables representing conditions in basin-fill aquifers of case-study basins in the Southwest Principal Aquifers study area.

[If a difference greater than 0.1 occurred between the sum of the average concentration class for percentiles 0 through 49.9 and the sum of the average concentration class for percentiles 50 through 100, then the predicted nitrate concentration class was deemed greater for lesser values of the explanatory variable. If this difference was less than –0.1, then the predicted nitrate concentration class was deemed greater for greater values of the explanatory variable; otherwise the relation between the nitrate concentration class and the explanatory variables was deemed unclear. **Abbreviations**: ≥, greater than or equal to; <, less than]

Geochemical variable	Average observed concentration class number by percentile range for explanatory variable [1]						Observed concentration class is greater for lesser or for greater values of the geochemical variable	Kendall's tau test on observed class number and explanatory variable		
	<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90		tau	z-score	p-value
Nitrate [2]										
pH	3.1	3.5	3.4	3.0	2.6	2.3	Lesser	–0.13	–14.9	<0.001
Dissolved oxygen	2.1	2.1	3.2	3.5	3.6	3.8	Greater	0.22	16.3	<0.001
Dissolved solids	2.2	2.6	2.9	3.3	3.5	3.6	Greater	0.15	16.4	<0.001
Sulfate	2.2	2.6	3.0	3.2	3.3	3.7	Greater	0.14	16.0	<0.001
Iron	3.2	3.2	3.2	3.2	2.5	1.7	Lesser	–0.13	–20.7	<0.001
Manganese	3.4	3.4	3.4	2.9	2.1	2.0	Lesser	–0.19	–22.8	<0.001
Alkalinity	2.2	2.9	2.7	2.8	3.3	2.7	Greater	0.06	3.6	<0.001
Bicarbonate	2.0	2.5	2.6	2.6	3.2	2.6	Greater	0.07	3.4	<0.001
Orthophosphate	3.2	3.2	3.1	3.2	2.8	2.3	Lesser	–0.09	–8.4	<0.001
Chloride	2.1	2.7	3.0	3.2	3.5	3.4	Greater	0.14	15.4	<0.001
Molybdenum	3.2	3.2	3.2	3.3	3.4	3.6	Greater	0.03	2.7	0.008
Selenium	2.7	2.7	2.7	3.5	4.1	5.0	Greater	0.29	22.8	<0.001
Arsenic [3]										
pH	3.3	3.5	3.5	3.8	4.4	5.2	Greater	0.19	18.5	<0.001
Dissolved oxygen	4.4	4.4	4.3	4.0	3.4	2.8	Lesser	–0.15	–10.5	<0.001
Dissolved solids	3.3	3.7	4.0	4.1	4.2	4.1	Greater	0.07	6.8	<0.001
Nitrate	4.3	4.3	3.9	3.9	3.6	3.5	Lesser	–0.09	–8.2	<0.001
Sulfate	3.8	3.5	3.7	4.3	4.1	4.0	Greater	0.06	5.8	<0.001
Iron	3.8	3.8	3.8	3.8	4.0	4.4	Greater	0.03	4.0	<0.001
Manganese	3.8	3.8	3.8	3.5	4.1	4.4	Greater	0.04	4.5	<0.001
Alkalinity	3.7	3.9	3.8	3.4	3.6	4.3	Unclear	–0.01	–0.3	0.729
Bicarbonate	4.0	4.3	3.8	3.5	3.9	4.7	Unclear	–0.01	–0.4	0.717
Orthophosphate	3.9	3.9	3.1	3.6	3.8	4.7	Greater	0.03	2.8	0.005
Chloride	3.4	3.6	3.9	3.9	4.3	4.4	Greater	0.09	9.0	<0.001
Molybdenum	3.7	3.7	3.7	4.3	4.8	4.8	Greater	0.13	13.0	<0.001
Selenium	3.9	3.9	3.9	3.3	4.2	3.5	Lesser	0.01	1.0	0.299

[1] See appendix 3 for values of the geochemical variable that correspond to each percentile range.

[2] Nitrate concentration ranges for classes 1 through 6 are <0.50, 0.50–0.99, 1.0–1.9, 2.0–4.9, 5.0–9.9, and ≥10 milligrams per liter as nitrogen.

[3] Arsenic concentration ranges for classes 1 through 7 are <1.0, 1.0–1.9, 2.0–2.9, 3.0–4.9, 5.0–9.9, 10–24, and ≥25 micrograms per liter.

linear regression. Only 7.9 percent of the model grid cells were found to have an increasing or decreasing trend in nitrate concentrations with aquifer-penetration depth, and only 11.6 percent of the model grid cells were found to have an increasing or decreasing trend in arsenic concentrations with aquifer-penetration depth. Given that most model grid cells in the basin-fill aquifers showed no distinct increasing or decreasing trend, the discussions on predicted nitrate and arsenic concentrations focus on only one aquifer-penetration depth—200 ft—unless stated otherwise. The value of 200 ft is somewhat deeper than the median aquifer-penetration depth (163 ft) for wells in the training dataset (appendix 3) and, in general, is deep with respect to domestic-supply wells, and shallow with respect to public-supply or irrigation wells.

Nitrate

Two random forest classifiers were developed to predict concentrations and improve the understanding of basin-fill aquifers to nitrate contamination. The prediction classifier primarily was developed to provide information on the spatial distribution of nitrate concentrations in basin-fill aquifers across the SWPA study area and, secondarily, to provide a statistical model that could be compared with and confirm or refute the conceptual model for natural and human-related factors affecting nitrate (Bexfield and others, 2011). The second classifier, the confirmatory classifier, was specifically developed to further evaluate the conceptual model but has limited use for predicting nitrate concentrations on a regional basis.

Nitrate Classifier Descriptions and Goodness-Of-Fit

The nitrate prediction classifier consists of 1,000 individual classification trees and was trained from 5,787 observations of nitrate concentrations and 58 explanatory variables, including 32 that represent source conditions and 26 that represent susceptibility conditions (table 6). Standardized importance scores were greater than 3.3 and indicated that all explanatory variables in the classifier were significant in explaining nitrate concentrations at p-values less than 0.001 (table 6), and consequently, all explanatory variables tested were retained in the classifier. Training specification parameters were adjusted to minimize the misclassification rate and produce minimally biased predictions—these included a minimum node size of 15 observations and 10 randomly selected variables for development of each splitting node in individual trees. Iteratively determined weights for classes 1 through 6 were 2.4, 3.7, 3.0, 3.2, 3.5, and 3.3, respectively.

The nitrate confirmatory classifier consists of 1,000 individual classification trees and was trained from 2,298 observations of nitrate concentrations and 89 explanatory variables, including all those in the prediction classifier plus 19 variables representing susceptibility conditions and 12 variables representing geochemical conditions (table 6). Whereas the training observations for the prediction classifier were from locations distributed throughout the study area, training observations for the confirmatory classifier were limited to the case-study basins (fig. 1) because these have the basin groundwater-budget variable data available. As was the case for the prediction classifier, all explanatory variables tested were significant (p is less than 0.001) and retained in the classifier (table 6). Training specification parameters for the final classifier included a minimum node size of 10 observations and 10 randomly selected variables for development of each splitting node in individual trees. Iteratively determined weights for concentration classes 1–6 were 2.5, 5.0, 3.5, 4.0, 3.5, and 3.0, respectively.

The nitrate classifiers were unbiased and had a good fit of the observed concentration classes as determined from the distribution of misclassification errors with respect to observed concentration class, geographic distribution, statistical distribution of explanatory variables, and estimated sampling error. The prediction classifier placed 42.3 percent of all training observations into the correct class, and 72.5 percent of the training observations into either the correct class, one class above the correct class, or one class below the correct class (table 7). Misclassification errors were generally symmetric about the correct class, having 28.7 percent of the observations misclassified into classes for lower concentrations than the true class and 29.0 percent misclassified into classes for higher concentrations (fig. 8). This symmetry indicates a lack of bias in the classifier toward overpredicting or underpredicting nitrate classes. Further, for a given concentration class, most of the training observations are placed in the correct class, and the number of misclassified observations generally decreases for classes distant from the correct class (table 7, fig. 8). The confirmatory classifier had a lower misclassification rate than the prediction classifier and correctly placed 48.6 percent of all training observations into the correct class and 80.4 percent of the training observations into either the correct class, one class above the correct class, or one class below the correct class (table 7). The reduction in misclassifications is a result of the additional geochemical and susceptibility variables included in the confirmatory classifier (table 6).

The spatial distribution of the misclassification errors from the two nitrate classifiers showed no significant patterns. Errors observed by visual inspection appear random and evenly distributed across the study area (appendix 8). Average misclassifications were computed for 88 basins that had more than 15 training observations for the prediction classifier. Of these, eight (9 percent of those evaluated) have average misclassification errors greater than 0.50, indicating potentially underpredicted concentrations, and six (7 percent of those evaluated) have average misclassification errors less than -0.50, indicating potentially overpredicted concentrations (appendix 4). The Palomas Basin was the only basin in this group with an average misclassification error greater than ±1.00; it was 1.16. For the confirmatory classifier, there were no basins identified as having potential bias on the basis of average misclassification errors (appendix 5).

All percentile ranges for each explanatory variable were well represented by training observations, and in most cases there were more than 100 training observations used for computing the average misclassification error (appendix 6). Thus, there were no variables that lacked representation by training-observation concentrations for low, medium, or high values of the explanatory variables with respect to their distribution across all basin-fill aquifers of the SWPA study area. Average misclassification errors were all within ±0.50 for high, medium, and low values of each explanatory variable, which indicated that predictions were generally unbiased across the range of values occurring in the SWPA study area (appendix 6). The highest average misclassification error was 0.45 for training observations in the 10th to 24.9th percentile range for the percent of urban land in the basin, and the lowest average misclassification error was –0.38 for training observations in the 75th to 89.9th percentile range for percent crystalline rocks surrounding the basin.

The nitrate classifiers have good predictive ability in consideration of the high sampling error detected in the training dataset. As discussed in the approach section, sampling error limits the percent of correctly classified training observations to 41.0 percent. Both nitrate classifiers exceed this percentage (table 7, fig. 8) because they contain additional variables that explain some of the within-cell variation of nitrate concentrations, namely aquifer-penetration depth for the prediction classifier, and the geochemical variables, aquifer-penetration depth, well depth, and depth to water, for the confirmatory classifier. Although breaking up the nitrate concentration classes into 6 categories results in a rather small correct classification rate of 42.3 percent, it increases the number of ways the predictions can be utilized. The ability to correctly classify

Table 6. Standardized importance scores for the prediction and confirmatory random forest classifers of nitrate concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.

[Standardized importance scores greater than 3.3 correspond to p-values less than 0.001 in the standard normal distribution. —, explanatory variable not included in model]

Variable group	Explanatory variable	Prediction classifier		Confirmatory classifier	
		Standardized importance score	Rank, maximum of 58	Standardized importance score	Rank, maximum of 89
Source variables					
Nitrogen loading	Nitrogen, atmospheric	36.06	5	22.53	10
	Nitrogen, farm fertilizer	28.92	20	19.89	18
	Nitrogen, non-farm fertilizer	23.64	40	10.68	53
	Nitrogen, confined manure	26.79	26	16.67	35
	Nitrogen, unconfined manure	29.45	17	14.02	44
	Nitrogen, total	29.91	15	20.08	17
Agricultural, urban, and biotic sources	Biotic community	13.58	58	8.58	64
	Septic/sewer ratio	28.05	24	12.33	48
	Local population	27.08	25	12.04	49
	Local population density	26.48	29	14.17	43
	Basin population	17.73	53	7.27	72
	Basin population density	21.33	46	8.80	61
	Local urban land	25.74	33	13.55	45
	Local agricultural land	25.87	32	15.17	41
	Basin urban land	21.83	44	8.31	66
	Basin agricultural land	19.65	47	7.20	73
	Basin rangeland	18.73	50	11.55	50
	Basin other land cover	23.67	39	7.56	71
Geologic sources	Geology, carbonate rocks	14.89	55	8.99	59
	Geology, crystalline rocks	19.00	49	18.71	25
	Geology, clastic sedimentary rocks	24.00	37	10.34	55
	Geology, mafic volcanic rocks	21.58	45	10.09	56
	Geology, felsic and silicic volcanic rocks	14.06	57	5.52	82
	Geology, intermediate composition volcanic rocks	15.87	54	5.67	80
	Geology, undifferentiated volcanic rocks	14.80	56	5.85	79
	Geology, distance to carbonate rocks	32.35	12	18.39	28
	Geology, distance to crystalline rocks	28.43	22	19.11	22
	Geology, distance to clastic sedimentary rocks	33.48	8	17.05	33
	Geology, distance to mafic volcanic rocks	30.97	13	17.31	32
	Geology, distance to felsic and silicic volcanic rocks	37.49	3	19.33	21
	Geology, distance to intermediate composition volcanic rocks	36.64	4	20.98	14
	Geology, distance to undifferentiated volcanic rocks	33.50	7	18.98	23
Susceptibility variables					
Flow path	Aquifer-penetration depth	35.94	6	19.59	20
	Well depth	—	—	21.29	13
	Water-level depth	—	—	18.77	24
	Land-surface slope	29.73	16	16.55	37
	Land-surface elevation	33.26	9	22.28	11
	Land-surface elevation percentile	46.67	2	25.26	9
	Basin elevation	26.41	31	10.41	54
	Distance to basin margin	23.21	41	16.13	38
Soil properties	Soil, seasonally high water depth	50.10	1	26.53	8
	Soil, hydric	24.32	35	16.80	34
	Soil, hydrologic group A	28.28	23	14.75	42
	Soil, hydrologic group B	26.43	30	18.61	26
	Soil, hydrologic group C	26.71	27	16.56	36
	Soil, hydrologic group D	23.87	38	18.56	27
	Soil, permeability	28.76	21	15.90	39
	Soil, organic material	32.39	11	20.21	16

Table 6. Standardized importance scores for the prediction and confirmatory random forest classifers of nitrate concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Standardized importance scores greater than 3.3 correspond to p-values less than 0.001 in the standard normal distribution. —, explanatory variable not included in model]

Variable group	Explanatory variable	Prediction classifier		Confirmatory classifier	
		Standardized importance score	Rank, maximum of 58	Standardized importance score	Rank, maximum of 89
Susceptibility variables—Continued					
Soil properties—Continued	Soil, clay	29.14	19	17.80	29
	Soil, silt	29.16	18	17.72	30
	Soil, sand	26.61	28	17.71	31
Water use and hydroclimatic	Water-resources development index	18.36	51	6.60	75
	Groundwater use, irrigated agriculture	32.56	10	19.88	19
	Surface-water use, irrigated agriculture	22.62	43	13.41	46
	Groundwater use, public-water supply	17.87	52	12.94	47
	Surface-water use, public-water supply	19.18	48	11.40	51
	Recharge, contributing area	25.45	34	8.84	60
	Recharge, basin	24.05	36	8.76	62
	Potential evapotranspiration	30.82	14	20.96	15
	Mean air temperature	23.17	42	15.53	40
Basin groundwater budget	Recharge, subsurface inflow	—	—	5.19	86
	Recharge, mountain front	—	—	6.17	78
	Recharge, precipitation	—	—	5.37	84
	Recharge, stream infiltration	—	—	5.63	81
	Recharge, irrigation	—	—	6.19	77
	Recharge, artificial	—	—	7.90	68
	Recharge, change	—	—	7.93	67
	Storage, change	—	—	7.86	69
	Recharge, total	—	—	4.52	87
	Discharge, total	—	—	5.40	83
	Discharge, change	—	—	6.37	76
	Discharge, subsurface outflow	—	—	4.41	88
	Discharge, evapotranspiration	—	—	8.75	63
	Discharge, to streams	—	—	5.35	85
	Discharge, to springs and drains	—	—	4.02	89
	Discharge, well withdrawals	—	—	7.83	70
	Residence time	—	—	8.34	65
Geochemical variables					
Geochemical	Groundwater, pH	—	—	28.65	6
	Groundwater, dissolved oxygen	—	—	21.63	12
	Groundwater, dissolved solids	—	—	34.77	1
	Groundwater, nitrate	—	—	—	—
	Groundwater, sulfate	—	—	33.30	3
	Groundwater, iron	—	—	28.03	7
	Groundwater, manganese	—	—	34.24	2
	Groundwater, alkalinity	—	—	10.91	52
	Groundwater, bicarbonate	—	—	9.79	57
	Groundwater, orthophosphate	—	—	9.06	58
	Groundwater, chloride	—	—	30.81	5
	Groundwater, molybdenum	—	—	7.18	74
	Groundwater, selenium	—	—	33.19	4

Table 7. Training-observation classification summary for the prediction and confirmatory random forest classifiers of nitrate concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.

[Dark cell shading indicates correct classification; light shading indicates one class above or below the correct class. Abbreviations: ≥, greater than or equal to; <, less than]

	Observed nitrate class and concentration range, in milligrams per liter as nitrogen						
	1	2	3	4	5	6	All classes
	<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	
Predicted class	Prediction classifier						
	True class, count of training observations						
1	1,096	185	218	112	69	78	1,758
2	188	170	158	72	41	11	640
3	223	176	439	252	72	45	1,207
4	84	50	200	254	132	67	787
5	64	32	94	156	211	148	705
6	81	27	70	83	149	280	690
	Classification summary						
Count of observations in class	1,736	640	1,179	929	674	629	5,787
Percentage of observations in class	30.0	11.1	20.4	16.1	11.6	10.9	100
Percentage of observations classified in correct class	63.1	26.6	37.2	27.3	31.3	44.5	42.3
Percentage of observations classified in correct class, one class above, or one class below	74.0	83.0	67.6	71.3	73.0	68.0	72.5
Predicted class	Confirmatory classifier						
	True class, count of training observations						
1	425	50	50	21	12	22	580
2	84	65	76	30	17	2	274
3	61	63	150	85	26	8	393
4	26	20	74	110	82	31	343
5	14	8	31	61	120	76	310
6	18	9	17	27	81	246	398
	Classification summary						
Count of observations in class	628	215	398	334	338	385	2,298
Percentage of observations in class	27.3	9.4	17.3	14.5	14.7	16.8	100
Percentage of observations classified in correct class	67.7	30.2	37.7	32.9	35.5	63.9	48.6
Percentage of observations classified in correct class, one class above, or one class below	81.1	82.8	75.4	76.6	83.7	83.6	80.4

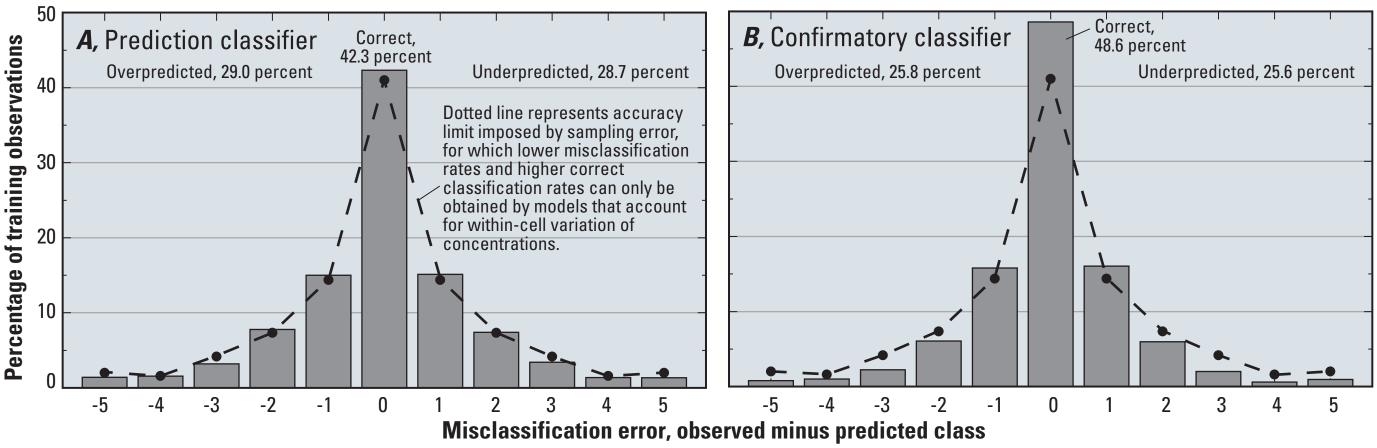
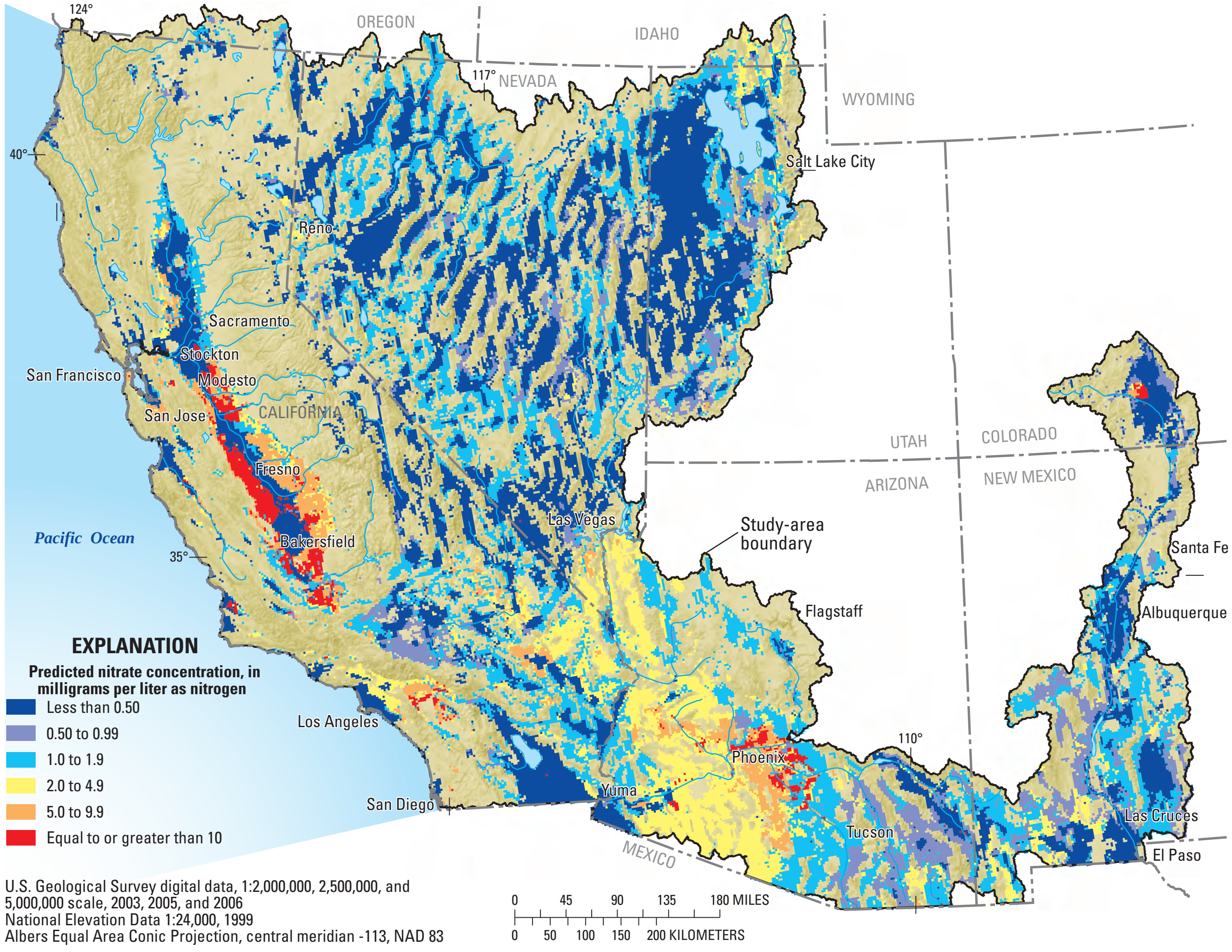


Figure 8. Statistical distribution of misclassification errors for the random forest classifiers of nitrate concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area: A, Prediction classifier; B, Confirmatory classifier.

predicted concentrations increases upon combining concentration classes and this can be useful for certain uses of the predicted concentrations. For instance, if one is distinguishing where nitrate occurs at concentrations less than 5 mg/L from where it is equal to or greater than 5 mg/L, and the error rate associated with that evaluation is needed, data contained in table 7 can be used to determine that error. First, counts of the observed nitrate classes 1 through 4 that were predicted as classes 5 and 6 are summed; this value (607) represents the number of observations misclassified for concentrations less than 5 mg/L. Next, counts of observed nitrate classes 5 and 6 that were predicted as classes 1 through 4 are summed; this value (515) represents the number of observations misclassified for concentrations equal to or greater than 5 mg/L. With 1,122 (607+515) of the 5,787 training observations misclassified, the misclassification rate for this example is 19.4 percent. The correct classification rate for this 2-class example is 80.6 percent, which is nearly double that of the original 6-class scheme (42.3 percent).

Regional Distribution and Depth Variation of Predicted Nitrate Concentrations

The prediction classifier was used in conjunction with an input dataset of explanatory variables to predict nitrate concentration classes across the 190,612 mi^2 (54,854 model grid cells) of basin-fill aquifers in the SWPA study area (fig. 9; appendix 8). While over half (53.3 percent) of the area of basin-fill aquifers are predicted to have nitrate concentrations less than 1.0 mg/L (classes 1 and 2 combined), 2.4 percent (4,530 mi^2) is predicted to equal or exceed the drinking-water standard of 10 mg/L (table 8). Whereas observed nitrate concentrations in the prediction classifier training dataset represent parts of the basin-fill aquifers with groundwater development, the predicted concentrations represent all basin-fill aquifers across the Southwest. The extent of human activities, such as urban land use, agricultural land use, and nitrogen loading, is generally greater for areas represented in the training dataset than for basin-fill aquifers across the SWPA



U.S. Geological Survey digital data, 1:2,000,000, 2,500,000, and 5,000,000 scale, 2003, 2005, and 2006
National Elevation Data 1:24,000, 1999
Albers Equal Area Conic Projection, central meridian -113, NAD 83

Figure 9. Predicted nitrate concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area. For a larger version of this figure, see appendix 8.

Table 8. Statistical distribution of predicted nitrate concentrations, by aquifer-penetration depth, principal aquifer, and state, for basin-fill aquifers in the Southwest Principal Aquifers study area.

[**Abbreviations:** ≥, greater than or equal to; <, less than; %, percent]

Distribution area	Total area for predictions, square miles	Percentage of total area, by nitrate class and concentration range, in milligrams per liter as nitrogen					
		1	2	3	4	5	6
		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10
Distribution by aquifer-penetration depth, entire study area[1]							
50 feet	190,612	43.0	10.4	26.0	13.1	4.5	3.1
100 feet	190,612	42.9	9.8	27.3	12.8	4.4	2.9
150 feet	190,612	42.7	10.2	27.4	12.5	4.5	2.6
200 feet	190,612	42.4	10.9	27.8	12.0	4.6	2.4
250 feet	190,612	41.9	11.5	28.0	11.7	4.6	2.3
500 feet	190,612	42.5	12.3	27.3	11.6	4.5	1.8
750 feet	190,612	42.7	13.6	25.7	11.9	4.2	1.8
1000 feet	190,612	42.4	14.3	25.6	11.8	4.2	1.8
Distribution by principal aquifer[2]							
Basin and Range basin-fill aquifers	138,642	43.0	9.9	29.9	13.3	3.0	0.9
California Coastal Basin aquifers	7,082	52.9	2.0	10.9	16.6	13.1	4.4
Central Valley aquifer system	16,766	33.8	2.0	16.9	8.1	21.9	17.3
Pacific Northwest basin-fill aquifers	3,489	68.2	11.3	18.0	2.5	0.0	0.0
Rio Grande aquifer system	24,634	37.7	24.8	29.3	7.4	0.4	0.5
Distribution by state[2]							
Arizona	36,928	8.6	13.3	30.9	35.8	8.2	3.2
California	52,450	43.9	5.4	23.2	11.0	10.4	6.1
Colorado	3,093	66.1	25.6	1.2	1.2	2.1	3.7
Idaho	890	9.8	11.3	34.0	44.9	0.0	0.0
Nevada	52,030	58.0	8.1	32.0	1.5	0.4	0.0
New Mexico	22,073	33.2	24.4	34.2	8.1	0.1	0.0
Oregon	768	34.4	13.1	52.0	0.5	0.0	0.0
Texas	31	100.0	0.0	0.0	0.0	0.0	0.0
Utah	22,351	65.3	10.5	20.1	3.7	0.3	0.0

[1] About 2.3% of the model cells in the study area have a linear increase in nitrogen concentration class with aquifer-penetration depth ($p<0.05$), and 5.6% have a linear decrease ($p<0.05$).

[2] Predictions are for an aquifer-penetration depth of 200 feet.

study area as a whole (comparison of training and prediction dataset values in appendix 3), and, consequently, the distribution of nitrate concentrations as shown above is higher for the observed nitrate concentrations (table 7) than for the predicted concentrations (table 8).

Concentration classes of predicted nitrate in individual model grid cells generally did not have significant systematic trends across the 8 aquifer-penetration depths, and were found to increase with aquifer-penetration depth in only 2.3 percent of the model grid cells, and decrease in only 5.6 percent of the cells. This is consistent with the observed data (fig. 2*A*), and, consequently, the spatial distribution of predicted nitrate concentrations presented is for just one aquifer-penetration depth—200 ft. For the basin-fill aquifers as a whole, a larger percent by area is predicted to have concentrations equal to or greater than 5.0 mg/L at shallower aquifer-penetration depths than at deeper aquifer-penetration depths. For example, 7.6 percent of the basin-fill aquifer area is predicted to have concentrations equal to or greater than 5.0 mg/L for an aquifer-penetration depth of 50 ft, whereas only 6.0 percent is predicted to have concentrations equal to or greater than 5.0 mg/L for an aquifer-penetration depth of 1,000 ft (classes 5 and 6 combined, table 8). Similarly, 56.7 percent of the basin-fill aquifer area is predicted to have concentrations less than 1.0 mg/L for an aquifer-penetration depth of 1,000 ft, whereas only 53.4 percent is predicted to have concentrations less than 1.0 mg/L for an aquifer-penetration depth of 50 ft.

In most cases, where a different concentration class is predicted for a deeper aquifer-penetration depth than was predicted for shallower aquifer-penetration depth in a given model grid cell, the concentration for the deeper aquifer-penetration depth also is predicted to occur laterally in adjacent grid cells at shallow depths. This relation could represent a natural transition boundary in concentrations that occurs both horizontally and vertically in the aquifer. These transitions between concentrations typically occur over just 1 to 3 grid cells (3 to 9 km), and because there is a horizontal component, they can represent the geochemical evolution of nitrate along the flow path.

In the west-central part of the San Joaquin Valley (appendix 8), a large cluster of model grid cells have a decrease in predicted nitrate concentrations with aquifer-penetration depth, which represents a human-related concentration stratification in the aquifer rather than a natural transition zone as previously described. In this area, the upper part of the aquifer is likely affected by high nitrate loadings from human-related sources, and the lower part generally has not been affected and has low nitrate concentrations. The Corcoran Clay member of the Tulare Formation occurs in this area (Faunt, 2009) and likely provides a confining layer that separates and retards flow between the upper and lower parts of the aquifer. Similar stratification is known to occur in other basins within the SWPA study area, such as in the Salt River Valley west of Phoenix, Arizona; however, the stratification in these areas is not simulated by the random forest classifier (Thiros and others, 2010; Edmonds and Gellenbeck, 2002). A higher density in training observations at different aquifer-penetration depths or addition of other variables not included in this classifier that represent the presence of confining layers would be needed for such stratification to be captured and predicted by the classifier.

The predicted nitrate concentrations exhibit substantial regional-scale variation throughout the basin-fill aquifers in the study area, and this variation is quite evident through examination of the distribution of concentrations by principal aquifer and by state (table 8). Of the principal aquifers, predicted nitrate concentrations are generally highest in the Central Valley aquifer system. This principal aquifer has the largest percentage, 39.2 percent, of its area predicted to be equal to or greater than 5.0 mg/L, and the smallest percentage, 35.8 percent, of its area predicted to be less than 1.0 mg/L (table 8). Predicted concentrations are generally lowest in the Pacific Northwest basin-fill aquifers, which have no areas predicted to have concentrations equal to or greater than 5.0 mg/L and 68.2 percent of its area predicted to be less than 0.50 mg/L.

Of the states within the SWPA study area, California has the largest percentage of its area of basin-fill aquifers, 16.5 percent, predicted to be equal to or greater than 5.0 mg/L, (table 8). Basins in California that had areas predicted to equal or exceed the 10 mg/L drinking-water standard include the Twentynine Palms Area; the Monterey, San Diego Coastal, San Francisco Bay Peninsula, Santa Ana Inland, and San Jacinto Basins; and the Bicycle, Cuyama, Livermore and Sunol, Imperial, Middle Salinas River, Sacramento, San Joaquin, Santa Clara River, Santa Maria River, Suisun- Fairfield, and Temecula Valleys (appendices 7 and 8).

Arizona had the second largest portion of its area of basin-fill aquifers equal or exceed 5.0 mg/L, at 11.4 percent (classes 5 and 6 combined, table 8). It is also quite notable that nearly 36 percent of the area of basin-fill aquifers in Arizona have concentrations between 2.0 and 4.9 mg/L—likely a result of nitrogen loading by desert legumes (discussed in the next section). Basins in Arizona that contained areas predicted to equal or exceeded the 10 mg/L drinking-water standard include Avra Valley, Eloy Area, Gila Bend Basin, Harquahala Basin, King and San Cristobal Valleys, Lower Verde River Basin, McMullen Valley, Palomas and Sentinal Plains, Paradise Valley, Renegras Plain, Salt River Valley (both Chandler and Phoenix areas), San Simon Valley, Stanfield Area, and Waterman Wash (appendices 7 and 8).

Compared to California and Arizona, predicted nitrate concentrations overall were low for the basin-fill aquifers in Colorado, Nevada, and Utah, where over half of the area has concentrations predicted to be less than 0.50 mg/L (table 8). Basins in these states, and in New Mexico, that contained areas predicted to equal or exceed the 10 mg/L drinking-water standard include the San Luis Valley within Colorado; the Socorro Basin in New Mexico; Pahrump, Silver State and Quinn River, and Spanish Springs Valleys in Nevada; and Utah Valley in Utah (appendices 7 and 8).

Comparison of the Nitrate Classifiers and Conceptual Model

In the conceptual model developed by Bexfield and others (2011), occurrence of nitrate concentrations greater than 5.0 mg/L is generally controlled by the presence of human-related sources of nitrogen on the land surface, transport to the aquifer by natural and human-related recharge mechanisms, and persistence in the aquifer as a result of favorable geochemical conditions. Analysis of (1) the standardized importance scores and (2) the univariate correlations between predicted nitrate class and the explanatory variables indicates that the prediction and confirmatory classifiers are consistent with, and provide additional detail to, the conceptual model.

The standardized importance scores indicated that, for both the prediction and confirmatory classifiers, concentrations of nitrate are affected more by local conditions than by basin-scale conditions. For the confirmatory classifier, prediction accuracy was most sensitive to several of the geochemical variables, as indicated by the top rankings of the standardized importance scores (table 6). For example, dissolved solids, manganese, sulfate, selenium, chloride, pH, iron, and dissolved oxygen have larger standardized importance scores than nearly all of the source or susceptibility variables. Concentrations of these constituents likely vary within each grid cell, and, therefore, they represent localized conditions.

While the confirmatory classifier prediction accuracy was found to be most sensitive to geochemical conditions, source and susceptibility variables were also important. Variables representing source and susceptibility conditions within the 3-km model grid cell were more important to prediction accuracy than comparable measures for the entire basin. For example, in the confirmatory classifier, population and population density for the cell (ranks 49 and 43, respectively) were ranked as more important than basin population and population density (ranks 72 and 61, respectively; table 6). Similarly local urban land and agricultural land (ranks 45 and 41, respectively) were ranked as more important than basin urban land and agricultural land (ranks 66 and 73, respectively). This pattern in the ranking of standardized importance scores for the agricultural and urban source variables also is observed for the prediction classifier (table 6). Also, for both classifiers, the nitrogen loading variables that represent actual flux rates of nitrogen to the land surface generally are ranked as more important than the agricultural and urban source variables, such as population and percentage of a specific land use within a cell or basin, which represent surrogates for the flux rates (table 6). Another example of the importance in spatial scale represented by the explanatory variables is that the four variables representing groundwater and surface-water use for agricultural and urban purposes within the model grid cells have larger (more significant) standardized importance scores (ranks between 19 and 51) than the basin-scale water-budget terms, even discharge by well withdrawals (ranked 70; confirmatory classifier data in table 6).

Relations between predicted nitrate class and the explanatory variables were generally consistent with what would be expected on the basis of the conceptual model. Where the univariate tau correlations were positive, the conceptual model generally indicated that nitrate concentrations increased as a result of increasing the magnitude of the natural or human-related factor being considered. Conversely, where correlations were negative, the conceptual model indicated a decrease in nitrate concentrations.

The conceptual model includes natural and human-related sources of nitrate (Bexfield and others, 2011). Natural sources include natural recharge of water, which carries nitrate that is contained in precipitation, accumulated in the soil-zone, or fixed by vegetation. Human-related sources include agricultural irrigation water infiltrating through fertilized fields, agricultural wastewater (including that from irrigation and from confined-feeding operations), public-supply water infiltrating through fertilized turf areas, urban runoff infiltrating as diffuse recharge, wastewater infiltrating from septic systems or sewer lines, and treated urban wastewater infiltrating through streams or irrigated fields. The nitrogen loading variables and their univariate tau correlations with predicted nitrate class are, with one exception, positive (table 9) and, therefore, consistent with nitrate sources in the conceptual model. While the p-values indicate tau for these variables is highly significant, they rank somewhat low. This likely occurs because nitrogen inputs are low for most model grid cells. The significance of the nitrogen inputs in terms of their effect is seen more clearly for average predicted classes across the range of values for the inputs. For example, the average nitrate class generally increases with higher rates of nitrogen loading and is greater than 3.0 where loading rates exceed the 90th percentile value for farm fertilizer, confined manure, and total nitrogen inputs (table 9). These are some of the highest average nitrate classes for any percentile range and variable shown in table 9.

The correlation between atmospheric nitrogen is negative and, therefore, contradicts the conceptual model. Similar to the correlation, average predicted nitrate class for percentile ranges generally decreases with higher atmospheric input rates (table 9). Closer inspection, however, shows that concentrations decrease at lower deposition rates but then increase for higher deposition rates. The negative correlation is probably spurious and a result of univariate analysis not considering other source and susceptibility variables that have a larger effect on nitrate in areas where atmospheric deposition rates are low.

Biotic community (fig. 5) was a significant source variable and serves as a surrogate for the different amounts of nitrate transported to the aquifer by each community. In the classifiers, biotic community is represented as categorical-type data, which precludes a correlation test to predicted nitrate. The standardized importance scores of both classifiers (8.58 and 13.58; table 6) indicate prediction accuracy is dependent on this source variable.

As part of the nitrogen cycle, nitrate is fixed primarily by bacteria in biological soil crusts (Eskew and Ting, 1978; Belnap and others, 2005) or by bacteria contained in the root nodules of desert legumes, such as mesquite, ironwood, smoke, and palo verde trees (Virginia, 1986). Virginia (1986) found that annual nitrogen fixation rates by root nodules for a mesquite woodland were 4,000 to 5,000 kilograms per square kilometer (kg/km^2). Although much of this nitrate is consumed by the tree (Virginia, 1986), this rate is considerably larger than the 95th percentile for farm fertilizer application in model grid cells—9,493 kg/yr per cell (appendix 3), or about 1,050 kg/km^2 per year. Schlesinger and others (1999) estimated desert shrublands in southern New Mexico lose 43 kg/km^2 per year of nitrogen carried in runoff, and 25 kg/km^2 per year is lost from desert grasslands. Some of this nitrogen enters the aquifer as runoff becomes groundwater recharge. While desert legumes flourish in the Sonoran Desert, their abundance in the Mojave Desert is low; therefore, nitrate contributions from them are expected to be low (Belnap and others, 2008). Microbial activity in desert soils is closely related to the timing, intensity, and amount of precipitation (Belnap and others, 2005), and flourishes during episodic wet pulses. Water from these precipitation events, if not utilized by the plant and soil community or evaporated, can transport nitrate as it infiltrates to deeper soils and recharges the aquifer or becomes runoff in streams, where it can infiltrate and recharge the aquifer (Belnap and others, 2005; Virginia, 1986). Walvoord and others (2003) found substantial accumulation of nitrate in the unsaturated zone beneath soils in the Chihuanhuan, Mojave, and Sonoran Deserts and concluded that this posed a concern for groundwater contamination, especially if subsequent development of those lands created additional recharge that could transport the nitrate into the aquifer.

An examination of the distribution of predicted concentrations for areas with minimal agricultural or urban development indicates that predicted nitrate concentrations in groundwater are generally less than 2.0 mg/L in most (22 of 26) biotic communities above basin-fill aquifers in the SWPA study area (table 10). Four communities, however, have predicted nitrate concentrations that equal or exceed 2.0 mg/L in more than 10 percent of their minimally developed land (table 10): Semi-desert Grassland, 12 percent; Mojave Desertscrub, 22 percent; Sonoran Desertscrub-Arizona Uplands, 48 percent; and Sonoran Desertscrub-Lower Colorado River Valley, 54 percent. Most of area within the Mojave Desertscrub community with predicted nitrate concentrations equal to or greater than 2.0 mg/L occurs in the eastern part near the Colorado River (appendix 8, fig. 5), which is adjacent to Sonoran Desertscrub

Table 9. Relation between predicted nitrate concentrations and explanatory variables representing conditions for basin-fill aquifers of the Southwest Principal Aquifers study area.

[If a difference greater than 0.1 occurred between the sum of the average concentration class for percentiles 0 through 49.9 and the sum of the average concentration class for percentiles 50 through 100, then the predicted nitrate concentration class was deemed greater for lesser values of the explanatory variable. If this difference was less than –0.1, then the predicted nitrate concentration class was deemed greater for greater values of the explanatory variable; otherwise the relation between the nitrate concentration class and the explanatory variables was deemed unclear. **Abbreviations**: ≥, greater than or equal to; <, less than]

Variable group	Explanatory variable	Average predicted nitrate concentration class number by percentile range for explanatory variable [1,2]						Observed concentration class is greater for lesser or for greater values of the geochemical variable	Kendall's tau test on predicted nitrogen class number and explanatory variable			
		<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90		tau	Rank, maximum of 56	z-score	p-value
Source variables												
Nitrogen loading	Nitrogen, atmospheric	2.6	2.5	2.3	2.1	2.4	2.5	Lesser	–0.06	30	–23.6	<0.001
	Nitrogen, farm fertilizer	2.2	2.2	2.2	2.0	2.3	3.3	Greater	0.06	31	27.0	<0.001
	Nitrogen, non-farm fertilizer	2.2	2.2	2.2	2.2	2.8	2.8	Greater	0.05	34	29.4	<0.001
	Nitrogen, confined manure	2.2	2.2	2.2	2.0	2.3	3.3	Greater	0.06	32	26.2	<0.001
	Nitrogen, unconfined manure	2.3	3.0	2.1	2.1	2.6	2.8	Unclear	0.03	40	13.0	<0.001
	Nitrogen, total	2.8	2.4	2.0	2.0	2.4	3.3	Greater	0.01	51	4.6	<0.001
Agricultural, urban, and biotic sources	Septic/sewer ratio	2.3	2.4	2.4	2.3	2.4	2.2	Lesser	–0.02	48	–7.2	<0.001
	Local population	2.1	2.1	2.1	2.3	2.7	2.9	Greater	0.09	23	37.1	<0.001
	Local population density	2.1	2.1	2.1	2.3	2.7	2.9	Greater	0.09	22	37.1	<0.001
	Basin population	1.9	2.0	2.5	2.2	2.1	3.4	Greater	0.08	25	30.4	<0.001
	Basin population density	1.9	1.9	2.4	2.3	2.0	3.0	Greater	0.10	21	35.4	<0.001
	Local urban land	2.1	2.1	2.1	2.3	2.8	2.8	Greater	0.08	24	33.8	<0.001
	Local agricultural land	2.2	2.2	2.2	2.1	2.3	3.2	Greater	0.04	37	20.0	<0.001
	Basin urban land	2.3	2.0	2.2	2.3	2.9	2.5	Greater	0.08	27	28.1	<0.001
	Basin agricultural land	2.4	2.3	2.1	2.1	2.4	3.2	Greater	0.02	44	8.6	<0.001
	Basin rangeland	3.1	2.3	1.9	2.3	2.2	2.6	Lesser	0.00	53	–0.6	0.551
	Basin other land cover	2.9	2.3	2.3	2.1	2.5	1.8	Lesser	–0.11	18	–40.5	<0.001
Geologic sources	Geology, carbonate rocks	2.5	3.9	2.8	2.2	1.7	1.8	Lesser	–0.15	10	–57.1	<0.001
	Geology, crystalline rocks	1.8	1.9	2.1	2.4	3.1	2.8	Greater	0.18	8	66.5	<0.001
	Geology, clastic sedimentary rocks	2.8	2.7	2.1	2.4	2.0	2.0	Lesser	–0.10	19	–37.3	<0.001
	Geology, mafic volcanic rocks	1.9	2.5	2.4	2.4	2.2	2.8	Greater	0.08	26	29.6	<0.001
	Geology, felsic and silicic volcanic rocks	2.7	2.7	2.4	2.1	2.1	1.9	Lesser	–0.11	15	–42.4	<0.001
	Geology, intermediate composition volcanic rocks	2.5	2.5	1.7	2.3	2.5	2.7	Greater	0.05	35	18.1	<0.001
	Geology, undifferentiated volcanic rocks	2.3	2.3	2.3	1.9	2.3	2.9	Unclear	0.00	55	0.2	0.855
	Geology, distance to carbonate rocks	2.0	1.9	2.1	2.7	2.7	2.3	Greater	0.11	17	40.7	<0.001
	Geology, distance to crystalline rocks	2.7	2.6	2.4	2.1	2.1	2.3	Lesser	–0.11	16	–41.2	<0.001
	Geology, distance to clastic sedimentary rocks	2.4	2.3	2.2	2.2	2.3	2.9	Greater	0.03	43	9.7	<0.001
	Geology, distance to mafic volcanic rocks	2.5	2.6	2.4	2.2	2.2	2.0	Lesser	–0.10	20	–35.6	<0.001
	Geology, distance to felsic and silicic volcanic rocks	2.2	2.1	2.1	2.3	2.6	2.9	Greater	0.07	29	24.9	<0.001
	Geology, distance to intermediate composition volcanic rocks	2.5	2.4	2.3	2.0	2.4	2.9	Greater	–0.02	46	–7.7	<0.001
	Geology, distance to undifferentiated volcanic rocks	2.3	2.3	2.4	2.3	2.4	2.3	Unclear	0.01	52	2.7	0.006

Table 9. Relation between predicted nitrate concentrations and explanatory variables representing conditions for basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[If a difference greater than 0.1 occurred between the sum of the average concentration class for percentiles 0 through 49.9 and the sum of the average concentration class for percentiles 50 through 100, then the predicted nitrate concentration class was deemed greater for lesser values of the explanatory variable. If this difference was less than –0.1, then the predicted nitrate concentration class was deemed greater for greater values of the explanatory variable; otherwise the relation between the nitrate concentration class and the explanatory variables was deemed unclear. **Abbreviations**: ≥, greater than or equal to; <, less than]

Variable group	Explanatory variable	Average predicted nitrate concentration class number by percentile range for explanatory variable [1,2]						Observed concentration class is greater for lesser or for greater values of the geochemical variable	Kendall's tau test on predicted nitrogen class number and explanatory variable			
		<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90		tau	Rank, maximum of 56	z-score	p-value
Susceptibility variables												
Flow path	Land-surface slope	2.0	2.3	2.4	2.4	2.4	2.2	Greater	0.05	33	18.5	<0.001
	Land-surface elevation	2.3	3.6	2.4	1.9	2.0	1.8	Lesser	–0.17	9	–64.5	<0.001
	Land-surface elevation percentile	1.7	2.0	2.3	2.6	2.6	2.5	Greater	0.15	11	55.5	<0.001
	Basin elevation	2.8	3.4	2.4	2.0	1.9	1.7	Lesser	–0.20	5	–75.0	<0.001
	Distance to basin margin	2.3	2.3	2.3	2.3	2.3	2.5	Greater	–0.02	45	–7.8	<0.001
Soil properties	Soil, seasonally high water depth	1.2	1.6	2.3	2.7	2.7	2.7	Greater	0.25	1	97.0	<0.001
	Soil, hydric	2.7	2.7	2.7	1.9	1.7	1.2	Lesser	–0.22	4	–96.8	<0.001
	Soil, hydrologic group A	2.8	2.8	2.2	2.0	2.0	2.3	Lesser	–0.13	13	–48.7	<0.001
	Soil, hydrologic group B	1.7	2.1	2.1	2.3	2.8	3.3	Greater	0.20	6	74.3	<0.001
	Soil, hydrologic group C	2.7	2.0	2.1	2.4	2.3	2.3	Greater	–0.01	50	–4.9	<0.001
	Soil, hydrologic group D	3.4	2.5	2.4	2.1	2.0	1.8	Lesser	–0.20	7	–73.3	<0.001
	Soil, permeability	2.4	2.6	2.5	2.0	2.2	2.3	Lesser	–0.07	28	–27.3	<0.001
	Soil, organic material	2.2	2.3	2.5	2.3	2.3	2.2	Lesser	0.00	54	–0.4	0.695
	Soil, clay	2.3	2.3	2.3	2.5	2.4	1.9	Lesser	–0.02	47	–7.2	<0.001
	Soil, silt	2.3	2.7	2.7	2.2	1.7	2.0	Lesser	–0.13	12	–48.9	<0.001
	Soil, sand	1.6	2.1	2.3	2.6	2.6	2.4	Greater	0.12	14	45.2	<0.001
Water use and hydroclimatic	Water-resources development index	2.4	2.1	2.3	2.2	2.5	2.9	Greater	0.04	38	14.8	<0.001
	Groundwater use, irrigated agriculture	2.2	2.2	2.2	2.0	2.2	3.4	Greater	0.05	36	21.0	<0.001
	Surface-water use, irrigated agriculture	2.2	2.2	2.2	2.1	2.4	3.0	Greater	0.04	39	15.8	<0.001
	Groundwater use, public water supply	2.3	2.3	2.3	2.3	2.3	2.8	Greater	0.03	41	21.4	<0.001
	Surface-water use, public water supply	2.3	2.3	2.3	2.3	2.3	2.9	Greater	0.03	42	22.2	<0.001
	Recharge, contributing area	2.7	2.3	2.2	2.0	2.3	3.0	Unclear	0.00	56	–0.1	0.938
	Recharge, basin	2.7	2.3	2.1	2.0	2.8	2.3	Unclear	0.02	49	6.4	<0.001
	Potential evapotranspiration	1.7	1.7	1.9	2.6	3.5	2.4	Greater	0.23	2	86.9	<0.001
	Mean air temperature	1.8	1.9	1.8	2.5	3.2	3.1	Greater	0.23	3	87.1	<0.001

[1] See appendix 3 for values of the explanatory variable that correspond to each percentile range.

[2] Concentration ranges for classes 1 through 6 are <0.50, 0.50-0.99, 1.0–1.9, 2.0–4.9, 5.0–9.9, and ≥10 milligrams per liter as nitrogen.

communities. These results indicate that natural nitrate loading to the groundwater is much higher in these four biotic communities than in the remaining 22 communities, especially in the two in the Sonoran Desert, which together cover about 31,000 mi^2, or about 16 percent of the basin-fill aquifers' area in the SWPA study area.

The distributions of nitrate concentrations for natural areas with minimal land development (table 10) allow for establishment of relative background concentrations. Following the precedent of Nolan and Hitt (2003), relative background concentration in this report represents a concentration predominantly resulting from natural processes plus an extraneous component from low-level influence of human activities. For most biotic communities, relative background concentrations of nitrate are less than 2.0 mg/L. The Semidesert Grassland, Mojave Desertscrub, Sonoran Desertscrub-Arizona Uplands, and Sonoran Desertscrub-Lower Colorado River Valley biotic communities, however, have relative background concentrations less than 5.0 mg/L.

Correlations were mostly positive between predicted nitrate and the agricultural and urban source variables (table 9) and, therefore, consistent with the Bexfield and others (2011) conceptual model. The sum of the four basin land uses must total 100 percent, so when urban or agricultural land uses are large percentages, rangeland and other land uses are small percentages of the total. Consequently, with the positive correlation between nitrate class and agricultural and urban land uses, the correlation is negative between nitrate class and

Table 10. Statistical distribution of predicted nitrate concentrations for relative background conditions, by selected biotic community, in basin-fill aquifers of the Southwest Principal Aquifers study area.

[Biotic communities from Brown and others (2007). **Abbreviations**: ID, identifier; mg/L, milligrams per liter; ≥, greater than or equal to <, less than; —, not applicable]

Biotic community		Total area in Southwest Principal Aquifer study area, square miles	Subset of model grid cells representative of background conditions where agricultural and urban land is less than 5 percent of cell area								
ID	Name		Area, square miles	Percentage in area by predicted nitrate concentration class						Relative background nitrate concentration threshold that 90 percent of the area does not exceed, in mg/L as nitrogen
				1	2	3	4	5	6	
				<0.50 mg/L as nitrogen	0.50–0.99 mg/L as nitrogen	1.0–1.9 mg/L as nitrogen	2.0–4.9 mg/L as nitrogen	5.0–9.9 mg/L as nitrogen	≥10 mg/L as nitrogen	
5	Great Basin Conifer Woodland	14,129	12,510	42	19	37	3	0	0	2.0
6	Great Basin Shrub-Grassland	6,536	3,635	59	3	35	3	0	0	2.0
8	Great Basin Desertscrub	54,549	45,584	64	8	28	0	0	0	2.0
17	Mojave Desertscrub	25,026	21,058	44	5	29	18	4	0	5.0
18	Plains Grassland, Shortgrass communities	5,167	3,864	33	25	41	1	0	0	2.0
19	Semidesert Grassland	14,539	12,471	17	29	41	12	0	0	5.0
22	Sonoran Desertscrub– Arizona uplands	9,341	7,614	7	7	37	40	8	0	5.0
23	Sonoran Desertscrub–Lower Colorado River Valley	21,520	14,150	19	1	26	46	8	0	5.0
24	Chihuahuan Desertscrub Cochise-Tranpecos	8,236	6,752	43	26	25	5	0	0	2.0
—	Remaining 17 communities	31,250	6,512	50	18	27	5	1	0	2.0

basin rangeland and basin other land cover (table 9). Average predicted nitrate class is generally greater for greater values of local and basin population and population density, local and basin agricultural land, and local and basin urban land. The average nitrate class for these variables is nearly 3.0 or greater for where loading rates exceed the 90th percentile value (table 9) and is greater than most other average nitrate classes shown in table 9.

The relative background nitrate concentrations established for the different biotic communities provide a benchmark for comparison of concentrations in areas with agricultural or urban land use. The percentage of model grid cells that exceed the relative background concentration was tabulated for different amounts of agricultural or urban land use within a model grid cell (table 11). In general, the larger the percentages of agricultural or urban land use within a grid cell, the greater the percentage of model grid cells that exceed relative background concentrations (table 11). Areas dominated by agricultural lands and areas dominated by urban lands both have similar percentages of cells exceeding relative background concentrations (table 11). Of all model grid cells developed entirely for agricultural or urban land uses, which are those on the bottom-left to upper-right diagonal in table 11, about 48 percent exceeded relative background concentrations. For the basin-fill aquifers in the SWPA study area, about 19,000 mi^2 are predicted to have nitrate concentrations that exceed the relative background concentrations listed in table 10. This represents about 10 percent of the basin-fill aquifers' area, and about 34 percent of the area with more than 5-percent agricultural or urban land use within a given model grid cell.

Geologic source variables all had significant standardized importance scores (table 6), indicating the accuracy of both classifiers was sensitive to these variables. While geologic sources in bedrock generally are not considered by the conceptual model as significant sources of nitrate, it is likely that these variables are surrogates for geochemical conditions for the model grid cells, such as pH, alkalinity, or dissolved-solids content, which are associated with certain rock types. Note that for each geologic unit, the correlation for the distance to the rocks is opposite of the correlation for the percent of the unit in the surrounding bedrock. This is expected because if a given rock type is associated with higher concentrations of nitrate, then the percentage measure should be positively correlated with nitrate, and the distance measure should be negatively correlated. Geologic variables represented by distance to the units have some of the highest standardized importance scores in the prediction classifier (ranks 3–22, table 6). These variables could be providing flow-path information, in addition to geochemical information, because the distance from the cell to rocks is equal to or greater than the distance to the basin margin, and several of the flow-path variables also are ranked highly in the prediction classifier.

Table 11. Percentage of basin-fill aquifer model grid cells in the Southwest Principal Aquifers study area that are predicted to exceed the relative background nitrate concentration, by percentage of agricultural and urban land use in the model grid cell.

[Relative background concentration threshold varies by biotic community and is listed in table 10. Example: Of all model grid cells in the study area that have 50–55 percent agricultural land, and 45–50 percent urban land, 64 percent of them exceed the relative background concentration observed in undeveloped conditions. **Abbreviations**: X, no model grid cells with the indicated agricultural and urban land-use conditions available for computing the percentage; —, no data because the sum of agricultural and urban land use exceeds 100 percent]

		Percentage of agricultural land in model grid cell																			
		0–5	**5–10**	**10–15**	**15–20**	**20–25**	**25–30**	**30–35**	**35–40**	**40–45**	**45–50**	**50–55**	**55–60**	**60–65**	**65–70**	**70–75**	**75–80**	**80–85**	**85–90**	**90–95**	**95–100**
Percentage of urban land in model grid cell	**0–5**	3	5	8	9	10	12	12	17	20	23	19	21	22	20	22	20	28	36	53	48
	5–10	5	5	14	10	15	24	23	25	35	30	32	31	32	28	34	40	40	50	67	—
	10–15	5	16	18	19	17	12	19	35	26	53	38	25	37	37	41	46	59	60	—	—
	15–20	3	9	13	26	6	27	X	47	56	48	40	43	21	40	55	47	50	—	—	—
	20–25	9	5	X	20	33	27	8	38	67	35	50	35	20	39	43	46	—	—	—	—
	25–30	12	6	X	10	X	29	X	50	36	54	18	43	41	45	59	—	—	—	—	—
	30–35	12	18	38	30	X	29	17	57	67	29	80	55	55	38	—	—	—	—	—	—
	35–40	12	27	11	22	X	38	31	18	50	71	60	43	50	—	—	—	—	—	—	—
	40–45	3	18	43	80	40	50	25	33	X	25	58	38	—	—	—	—	—	—	—	—
	45–50	19	27	33	17	33	40	17	45	55	11	64	—	—	—	—	—	—	—	—	—
	50–55	10	73	17	X	20	X	75	50	54	67	—	—	—	—	—	—	—	—	—	—
	55–60	15	X	25	33	25	38	50	60	50	—	—	—	—	—	—	—	—	—	—	—
	60–65	21	36	X	38	50	X	46	57	—	—	—	—	—	—	—	—	—	—	—	—
	65–70	25	25	17	50	50	63	43	—	—	—	—	—	—	—	—	—	—	—	—	—
	70–75	25	29	25	25	56	22	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	75–80	35	50	25	36	29	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	80–85	42	46	70	40	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	85–90	39	31	45	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	90–95	48	50	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	95–100	49	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

The conceptual model (Bexfield and others, 2011) for the vulnerability of basin-fill aquifers to nitrate contamination considers several natural susceptibility conditions, such as high evapotranspiration of recharge water, presence of confining layers or upward hydraulic gradients to impede downward migration from nitrate sources, and high evapotranspiration of shallow discharging groundwater. Human-related factors that are conceptualized to affect aquifer susceptibility to nitrate contamination include high evapotranspiration of artificial recharge, high rates of artificial recharge and associated young groundwater ages, application of artificial recharge to areas with nitrate build-up in the unsaturated zone from previous agriculture, accumulation of human-related compounds in areas with large amounts of artificial recharge and shallow water tables, and high degree of hydrologic system modification—especially increased amounts of recharge to and discharge from the aquifer.

The accuracy of the classifiers was highly sensitive to the flow-path variables, including land-surface elevation percentile, which had the second highest standardized importance score for the prediction classifier (table 6). Land-surface elevation percentile, land-surface slope, and distance to the basin margin provide an indication of the location within a basin (fig. 6) and, therefore, likely serve to account for various processes that occur along groundwater flow paths from the basin margin to the basin center. Such processes include, but are not limited to, natural recharge and discharge because these tend to occur along the margin and in the center of the basin, respectively, and both directly affect the transport of nitrate to and from the aquifer. Tau is positive and average nitrate concentration classes increase with elevation percentile (table 9), indicating nitrogen concentrations are lower in the basin lowlands than in the upper basin margins and, therefore, likely decrease with groundwater flow.

Predicted nitrate concentrations were highly correlated to seasonally high soil-water depth and hydric (water-saturated) soils, having the highest and 4th highest ranked correlation in table 9. The correlations and average nitrate concentration classes indicate predicted nitrate concentrations tend to be lower for areas with shallower seasonally high water depths or where soils are hydric, and concentrations tend to be higher for areas with deeper seasonally high water depths or where hydric soils are absent. In the prediction classifier, both variables are likely serving to distinguish low nitrate concentrations in wetland areas and floodplains with shallow aquifers from the remaining areas without such shallow groundwater, where concentrations are often higher. The average nitrate class for the 10th percentile or lesser values of seasonally high soil-water depth and the average nitrate class for the 90th percentile or greater values of hydric soils are among the lowest average nitrate predictions for any percentile range or variable in table 9. Bexfield and others (2011) do not discuss in detail nitrate conditions for soils with hydric conditions or shallow seasonally high water depths, however other studies also found low nitrate concentrations for these conditions. For example, in a study of nutrients in groundwater across the United States, Mueller and others (1995) found that in agricultural areas, nitrate concentrations were lower where seasonally high water depth was less than 5 ft as compared to areas where seasonally high water depth was greater than 5 ft. Reasons for the lower concentrations at the shallow depths are several: (1) the areas are likely discharging groundwater, indicating gradients are upward rather than downward and not conducive to downward transport of nitrogen from surficial sources (Bexfield and others, 2011; Burow and others, 2008); (2) wetland areas and those with shallow groundwater typically have substantial natural vegetation or cultivated crops that can consume nitrate as a nutrient for their growth; (3) wetland areas, especially those with organic matter in the soils, have an increased potential for loss of nitrate through denitrification (Nolan and Hitt, 2006; Hanson and others, 1994); and (4) the predominant form of nitrogen could be ammonia rather than nitrate, as was found by Hamilton and others (1993).

For other variables representing soil conditions, correlations with predicted nitrate generally indicate that nitrate concentrations are lower in areas with poor infiltration and higher in areas with good soil drainage and, therefore, potential for transmittance of water to deeper depths, which is consistent with results from other studies (Mueller and others, 1995; Nolan and Hitt, 2006; Rupert, 2003). For example, the correlations are negative (table 9) for soil hydrologic groups C and D, which have low infiltration rates (table 3), but the correlation is positive for soil hydrologic group B, which has moderate infiltration rates. Similarly, the correlations are negative for clay and silt (table 9), which generally are considered to have lower infiltration rates than sand, for which the correlation is positive. The correlations are negative for permeability and soil hydrologic group A (table 9), however, which indicates lower nitrate concentrations where infiltration rates are high, and it is not clear why this inconsistency occurs.

The hydroclimatic variables of mean air temperature and potential evapotranspiration were positively correlated with predicted nitrate concentrations, indicating that concentrations are generally higher where the climate is warmer and has more potential for evaporation (table 9). Average predicted nitrate classes show this same trend. The positive correlation for both these variables, which are correlated with each other, could result from increased microbial activity in warmer climates because nitrogen fixation rates are known to increase with temperature (Stark, 1996). Another plausible explanation for the positive correlation is evaporative-concentration effects, where groundwater is evaporated from the aquifer, thereby removing water from the system and, consequently, increasing nitrate concentrations. Nitrate concentrations are negatively correlated with recharge in the contributing area, which indicates that concentrations are higher in basins with low natural recharge. Basins with low recharge also tend to be warmer and have a greater potential for evapotranspiration; therefore, the negative correlation with recharge in the contributing area is consistent with the positive correlations with temperature and evapotranspiration.

Water-use variables were all positively correlated with predicted nitrate concentrations and are, therefore, consistent with the conceptual model (Bexfield and others, 2011). Average predicted nitrate classes are greater for greater values of all the water-use variables (table 9). In fact, the average predicted nitrate class for the 90th and greater percentiles for groundwater use, 3.4, is among the highest averages in table 9. These variables all represent water sources that can transport nitrate from land-surface sources to the aquifer. The standardized importance scores indicate that the accuracy of the confirmatory classifier was more sensitive to the four variables for groundwater and surface-water use for irrigated agriculture or public-water supply than to the basin-scale water-resources index, or any of the basin groundwater-budget variables (table 6). In fact, as a group of variables, the accuracy of the confirmatory classifier was least sensitive to the basin groundwater-budget variables, likely because they are basin-scale variables. This indicates that the improvement in the correct classification rate of 42.3 percent for the prediction classifier to 48.6 percent for the confirmatory classifier (table 7) is largely due to the inclusion of the geochemical variables in the confirmatory classifier rather than the basin groundwater budget variables.

While geochemical variables could not be used in the prediction classifier, several have the highest standardized importance scores for the confirmatory classifier, indicating the high sensitivity of the accuracy of the confirmatory classifier to that group of variables (table 6). Average observed nitrate concentration class for the geochemical-variable percentile ranges shows the effects of redox conditions, and is high where dissolved oxygen concentrations are high (oxic), and low where dissolved iron and manganese concentrations are high (reducing). For example, the average observed nitrogen class is 1.7, corresponding to 0.50 mg/L, where iron concentrations are greater than the 90th percentile value for iron (160 µg/L; table 5 and appendix 3). Dissolved solids, sulfate, selenium, and chloride are among the highest five standardized importance scores (table 6), and all are positively correlated with nitrate (table 5). These variables could be distinguishing conditions in agricultural areas where application of irrigation water has transported these constituents and nitrate from the unsaturated zone and concentrated them within the upper part of the aquifer.

Effects of Selected Natural and Human-Related Factors on Predicted Nitrate Concentration

The previous section discussed univariate correlations (positive or negative; strong or weak) between the explanatory variables and predicted nitrate concentration class. To fully understand the effect the variables have on predicted nitrate concentrations, however, the distribution of nitrate concentrations resulting from specific conditions must be described. Examination of appendix 8 and tables 6 and 9 indicates that some of the most important natural and human-related source and susceptibility conditions affecting the spatial distribution of predicted nitrate concentrations in the basin-fill aquifers across the SWPA study area include the general position in the basin and location along a groundwater flow path, land development, biotic community, and presence of wetlands in lowland areas. To illustrate the size of the effect that these factors have on predicted nitrate concentrations, model grid cells representing the basin-fill aquifers were categorized on the basis of four explanatory variables:

- Land-surface elevation percentile (greater than 75 percent to indicate basin margin, 10 to 75 percent to indicate the middle parts of the basin, or less than 10 percent to indicate the basin lowlands).
- Agricultural and urban land within the model grid cell (less than 5 percent to indicate minimal development, 5 to 50 percent to indicate moderate development, and greater than 50 percent to indicate highly developed areas).
- Biotic community (Sonoran Desertscrub, or all communities excluding Sonoran Desertscrub).
- Hydric soils (greater than 10 percent in model grid cell to indicate presence of wetlands, only assessed for cells classified as basin lowlands).

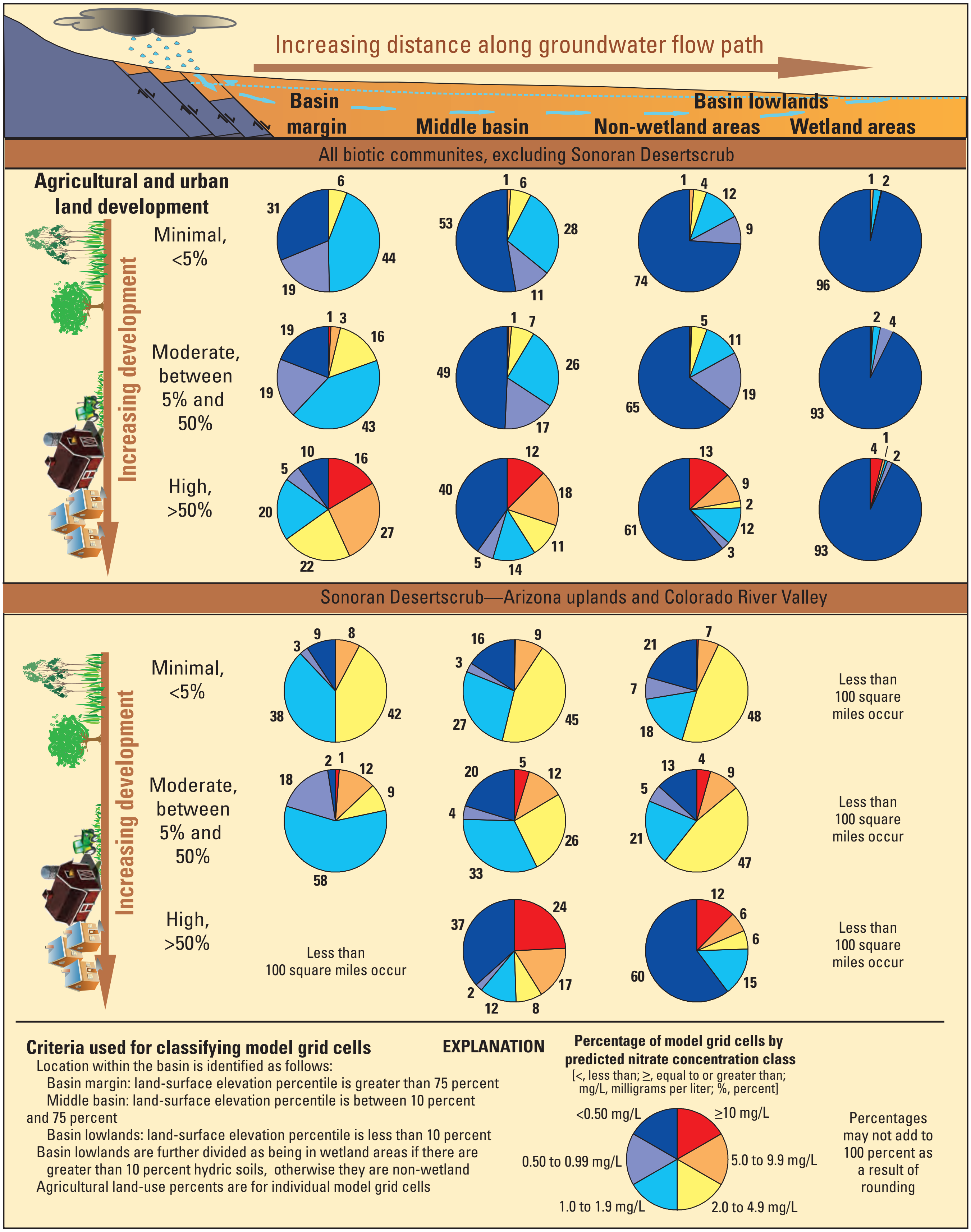
Classification of the model grid cells on the basis of these criteria resulted in 24 possible categories of cells, 4 of which had less than 100 mi^2 of basin-fill aquifer per category and, therefore, were excluded from this analysis. For each category, the statistical distribution of predicted nitrate concentrations was determined and illustrated as a pie chart (fig. 10).

Comparison of the statistical distributions for the 20 categories of cells (fig. 10) shows that nitrate concentrations generally decrease along groundwater flow paths from the basin margin to the basin lowlands for non-Sonoran Desertscrub communities, increase with land development, are higher in the Sonoran Desertscrub biotic communities than in other biotic communities, and are smallest in basin lowlands that have wetlands. Groundwater flow paths generally start at the upper basin margins where groundwater is recharged along the mountain front by precipitation. Predicted nitrate concentrations within the upper basin margins are generally similar to concentrations of inorganic nitrogen (nitrate plus ammonium) dissolved in precipitation, which typically are between about 0.50 and 2.0 mg/L (fig. 11; National Atmospheric Deposition Program, 2010). Evaporative-concentration and nitrogen-cycling processes within the biotic communities can increase or decrease nitrate concentrations from precipitation and runoff water prior to recharge.

The statistical distribution of predicted nitrate concentrations shifts toward lower concentrations from the upper basin margins to the basin lowlands, which likely results from denitrification along the groundwater flow path (fig. 10). Alternatively, groundwater age generally increases along the flow path from the upper basin margin to the basin lowlands, and the same spatial pattern in groundwater nitrate concentrations could result if nitrate concentrations in recharge water have increased over the ages. In minimally developed, non-Sonoran Desertscrub biotic communities, the percent area with concentrations less than 0.50 mg/L increases from 31 percent in the upper basin margins to 74 percent in non-wetland basin lowland areas, and increases to 96 percent in the basin lowland areas that have wetlands. Similar but less dramatic shifts toward lower concentrations are observed from the upper basin margins to basin lowlands within the individual categories for moderate and high amounts of development in non-Sonoran Desertscrub communities (fig. 10). Decreases in predicted nitrate concentrations from the upper basin margins to the basin lowlands are less apparent for Sonoran Desertscrub communities (fig. 10). In the minimally developed areas in Sonoran Desertscrub communities, the percent area with concentrations less than 0.50 mg/L only increases from 9 percent in the upper basin margins to 21 percent in non-wetland, basin lowland areas. Within the category of moderately developed areas in the Sonoran Desertscrub communities, the overall distribution of predicted nitrate concentrations shifts toward higher concentrations in the basin lowlands, which is contrary to the pattern observed for other biotic communities and amounts of land development.

As shown previously in table 10, the distribution of nitrate concentrations for minimally developed areas varies by biotic community and is generally shifted toward higher concentrations for Sonoran Desertscrub communities compared to the other communities (fig. 10). For minimally developed areas, the percent of the area near the upper basin margins predicted to have concentrations equal to or greater than 2.0 mg/L is only 6 percent for non-Sonoran Desertscrub communities, but 50 percent for Sonoran Desertscrub communities. Similarly, the percent of the area in the middle of the basin predicted to have concentrations equal to or greater than 2.0 mg/L is only 7 percent for non-Sonoran Desertscrub communities, but 54 percent for Sonoran Desertscrub communities. The percent of the area in the non-wetland basin lowlands predicted to have concentrations equal to or greater than 2.0 mg/L is only 5 percent for non-Sonoran Desertscrub communities, but 55 percent for Sonoran Desertscrub communities.

The statistical distribution of predicted nitrate concentrations shifts toward higher concentrations with increased land development, which likely results from the additional nitrogen inputs from human-related sources and processes that can facilitate transfer from the land surface to the aquifer (fig. 10). For example, in areas near the upper basin margin and having



Criteria used for classifying model grid cells

Location within the basin is identified as follows:

Basin margin: land-surface elevation percentile is greater than 75 percent

Middle basin: land-surface elevation percentile is between 10 percent and 75 percent

Basin lowlands: land-surface elevation percentile is less than 10 percent

Basin lowlands are further divided as being in wetland areas if there are greater than 10 percent hydric soils, otherwise they are non-wetland

Agricultural land-use percents are for individual model grid cells

Figure 10. Distribution of predicted nitrate concentrations as a function of distance along generalized groundwater flow path, land development, presence of wetlands, and biotic community for basin-fill aquifers of the Southwest Principal Aquifers study area.

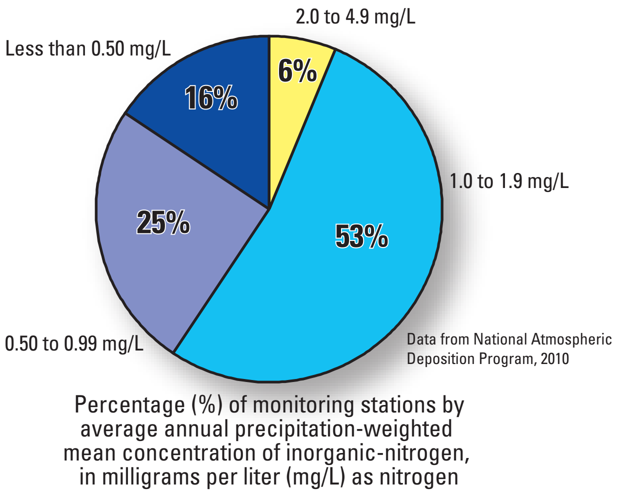


Figure 11. Statistical distribution of the mean inorganic-nitrogen concentration for atmospheric-deposition monitoring sites in the Southwest Principal Aquifers study area.

non-Sonoran Desertscrub biotic communities, only 6 percent of the cells equal or exceed 2.0 mg/L where land is minimally developed, but 65 percent of the cells equal or exceed 2.0 mg/L where land is highly developed. Similar but less dramatic shifts toward higher concentrations are observed for increased land development in the middle and lowland parts of basins, except in the lowlands where there are wetlands. Nitrate concentrations equal to or greater than 2.0 mg/L in non-Sonoran Desertscrub areas, and concentrations equal to or greater than 5.0 mg/L in Sonoran Desertscrub communities are largely only found in areas with more than 5 percent agricultural or urban land development. About 15 percent of the area with more than 50 percent of the land developed for agricultural and urban uses is predicted to have nitrate concentrations equal to or greater than 10 mg/L, although this percentage varies from 12 to 24 percent depending on biotic community and location within the basin (fig. 10).

Nearly all wetland areas in the basin lowlands have predicted concentrations less than 0.50 mg/L, regardless of the amount of land development (fig. 10). As discussed previously, the lower concentrations in areas with wetlands could be present because (1) groundwater likely discharges in those areas, indicating gradients are upward rather than downward so are not conducive to downward transport of nitrogen from surficial sources; (2) natural vegetation or cultivated crops could have consumed nitrate as a nutrient for their growth; (3) nitrate was lost by denitrification; or (4) the predominant form of nitrogen could be ammonia rather than nitrate.

Nitrate Summary and Vulnerability Assessment

The random forest classifiers provide information on the spatial distribution of nitrate within the upper 200 ft of basin-fill aquifers (190,612 mi^2) across the SWPA study area and allow for a general assessment of the vulnerability of basin-fill aquifers to nitrate contamination. The classifiers were effectively trained to the relations between observed nitrate concentrations and the natural and human-related factors important to nitrate occurrence. This enabled extrapolation of nitrate concentrations from areas where concentration conditions were measured and known into areas where data were unavailable and unknown. The ability of the model to predict nitrate concentrations across the study area within plus or minus one concentration class was 72.5 percent; the relatively low prediction accuracy for actual concentration class results largely from natural spatial variability and the use of six concentration classes. The use of six concentration classes, however, provided a detailed characterization of the distribution of nitrate concentrations throughout the SWPA within reasonable accuracy for such a large area. Analysis of the misclassifications indicated the model was unbiased spatially and unbiased across the distribution of values for the explanatory variables.

While the training observations indicate nitrate concentrations were equal to or exceeded 10 mg/L in 10.9 percent of the groundwater samples, use of the prediction classifier to extrapolate concentrations across the SWPA study area revealed that only about 2.4 percent of the study area underlain by basin-fill aquifers exceeds this concentration, and 93.0 percent of the area has less than 5.0 mg/L of nitrogen in the groundwater samples:

Nitrate concentration class, mg/L of nitrogen	>0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10
Percent training observations in concentration class, generally representing part of aquifers with groundwater development, from table 7 (n = 5,787)	30.0	11.1	20.4	16.1	11.6	10.9
Percent of basin-fill aquifer area in Southwest Principal Aquifer study area predicted for concentration class, from table 8 (190,612 miles^2)	42.4	10.9	27.8	12.0	4.6	2.4

Such differences in the distribution of observed and predicted nitrate concentrations are expected and result from the fact that the prediction dataset represents the full extent of basin-fill aquifers in the SWPA study area, whereas the training dataset represents a subset of those aquifers where observations were available, and each dataset has somewhat different but overlapping distributions of source and aquifer-susceptibility variables that affect nitrate in groundwater.

Relative background concentrations of nitrate in groundwater in undeveloped land-use settings were determined to be less than 2.0 mg/L for most biotic communities overlaying basin-fill aquifers, except for the Semidesert Grassland, Mojave Desertscrub, Sonoran Desertscrub-Arizona Uplands, and Sonoran Desertscrub-Lower Colorado River Valley communities, where relative background concentrations were estimated to be less than 5.0 mg/L. Nitrate concentrations greater than these relative background concentrations are largely found in areas with agricultural or urban land development.

Concentrations of nitrate in the basin-fill aquifers were predicted to exceed relative background concentrations in about 34 percent of areas having more than 5 percent agricultural or urban development. Exceedance of relative background concentrations increased with the amount of agricultural or urban development. For areas entirely developed for agricultural or urban land uses, nitrate concentrations in the basin-fill aquifers were predicted to exceed relative background concentrations in nearly half (48 percent) of those lands. About 15 percent of the basin-fill aquifers' area with more than half of the land developed for agricultural and urban uses was predicted to have nitrate concentrations equal to or exceed 10 mg/L. Predicted concentrations generally decreased along groundwater flow paths from the basin margin to the basin lowlands. Nearly all wetland areas in the basin lowlands have concentrations less than 0.50 mg/L, regardless of the amount of land development.

A further understanding of conditions that render the basin-fill aquifers in the SWPA study area vulnerable to nitrate contamination was gained from an analysis of the correlations between the predicted concentrations and the explanatory variables (table 9), as well as correlations between observed nitrate and other constituent concentrations (table 5) in the training dataset. These univariate correlations indicate that areas are more likely to have higher concentrations and, therefore, are generally more vulnerable to nitrate contamination, where one or more of the following conditions occur:

- Land is used for agricultural or urban purposes, especially where fertilizers are used or where there are livestock.
- Nitrogen is fixed by natural vegetation, such as legumes in the Sonoran Desert.
- Soils are present that have textures which favor water infiltration, lack hydric conditions, or lack organic material.
- Water-use rates are high from groundwater or surface-water supplies for agricultural purposes or for public-water supply.
- Natural recharge is low in the contributing parts of the hydrogeologic areas.
- Mean air temperatures and potential evapotranspiration are high.
- The contributing part of the hydrogeologic area has an abundance of crystalline, mafic volcanic, intermediate composition volcanic, and undifferentiated volcanic rocks, which likely produce geochemically favorable conditions.
- The groundwater is oxic.

These source, aquifer susceptibility, and geochemical conditions associated with the vulnerability of basin-fill aquifers to nitrate contamination, as determined by the random forest classifier results, are consistent with the conceptual model of natural and human-related factors that affect nitrate concentrations as described by Bexfield and others (2011).

Arsenic

Similar to the nitrate classifiers previously described, two random forest classifiers were developed to assess natural and human-related factors influencing the distribution of arsenic within the SWPA study area. The prediction classifier was developed for use in predicting arsenic concentrations and to assess aquifer vulnerability to arsenic enrichment in alluvial basin areas within the SWPA study area where arsenic data were unavailable. The confirmatory classifier evaluates the current understanding of the occurrence and environmental fate of arsenic within 16 case-study basins within the SWPA study area, as represented by the conceptual model of Bexfield and others (2011). Generally, the two random forest classifiers indicate that arsenic enrichment is influenced by geologic sources, recharge conditions, and position along a groundwater flow path within a basin. The confirmatory classifier supported the principal findings of the conceptual model described by Bexfield and others (2011), and identified geologic sources, groundwater residence time (position along a flow path), and geochemical characteristics as important influences on the occurrence, transport, and fate of arsenic.

Arsenic Classifier Descriptions and Goodness-Of-Fit

The arsenic prediction classifier was trained from 4,162 observations of arsenic concentrations and 53 explanatory variables that represent source and aquifer susceptibility conditions (table 12, appendix 1). Arsenic concentrations were partitioned into seven concentration classes for the classifier (table 13). The percent of observations within each concentration class in the training dataset used for the prediction classifier was fairly uniform and ranged from about 11 to 17 percent, with the greatest percentage occurring for arsenic concentrations between 5.0 and 9.9 µg/L (class 5, table 13). Differences in the number of observations representing each concentration class can contribute to uneven misclassification rates for each class, so each class was weighted. The weights used for the prediction classifier concentration classes 1 through 7 were 1.5, 1.9, 1.9, 1.5, 1.4, 1.6, and 1.8, respectively.

All 53 possible explanatory variables tested were found to have standardized importance scores greater than 2 ($p<0.05$) during the training of the prediction classifier and, therefore, were retained in the final classifier (table 12). Geochemical and selected basin-scale variables were not available throughout the entire SWPA study area and, therefore, were not included in the prediction classifier. The number of explanatory variables randomly chosen for each tree generated in the prediction classifier was 18 (about 35 percent of the total available). A minimum of 10 observations was required for each node in each tree generated.

The confirmatory classifier was developed for 16 case-study basins within the SWPA study area (fig. 1) from 1,851 observations partitioned into the same 7 concentration classes used for the prediction classifier (table 13). Generally, the

Table 12. Standardized importance scores for the prediction and confirmatory random forest classifiers of arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.

[Standardized importance scores greater than 3.3 correspond to p-values less than 0.001 in the standard normal distribution. —, explanatory variable not included in model]

Variable group	Explanatory variable	Prediction classifier		Confirmatory classifier	
		Standardized importance score	Rank, maximum of 53	Standardized importance score	Rank, maximum of 85
Source variables					
Agricultral, urban, and biotic sources	Biotic community	11.93	53	6.13	68
	Septic/sewer ratio	28.52	19	17.05	22
	Local population	26.25	30	12.68	42
	Local population density	27.52	25	13.19	39
	Basin population	15.53	50	7.56	56
	Basin population density	17.65	43	5.61	70
	Local urban land	26.93	28	11.73	45
	Local agricultural land	26.59	29	17.41	21
	Basin urban land	20.73	38	6.89	62
	Basin agricultural land	19.24	41	7.52	57
	Basin rangeland	15.77	48	7.39	59
	Basin other land cover	15.44	51	5.16	75
Geologic sources	Geology, carbonate rocks	20.85	37	5.12	76
	Geology, crystalline rocks	21.61	36	7.37	60
	Geology, clastic sedimentary rocks	20.43	39	6.43	64
	Geology, mafic volcanic rocks	31.47	13	19.66	14
	Geology, felsic and silicic volcanic rocks	16.66	45	3.63	82
	Geology, intermediate composition volcanic rocks	24.74	32	3.61	83
	Geology, undifferentiated volcanic rocks	15.05	52	4.55	78
	Geology, distance to carbonate rocks	34.13	8	22.66	12
	Geology, distance to crystalline rocks	36.88	5	28.00	3
	Geology, distance to clastic sedimentary rocks	32.27	11	16.32	27
	Geology, distance to mafic volcanic rocks	35.58	7	22.02	13
	Geology, distance to felsic and silicic volcanic rocks	30.27	17	25.62	9
	Geology, distance to intermediate composition volcanic rocks	36.49	6	32.06	2
	Geology, distance to undifferentiated volcanic rocks	38.65	4	26.15	7
	Soil and rock equivalent uranium-238 concentration	22.84	35	14.11	36
Susceptibility variables					
Flow path	Aquifer-penetration depth	27.92	23	13.04	41
	Well depth	—	—	16.18	28
	Water-level depth	—	—	16.43	25
	Land-surface slope	32.48	10	25.18	11
	Land-surface elevation	38.70	3	25.70	8
	Land-surface elevation percentile	41.65	2	26.16	6
	Basin elevation	18.76	42	7.49	58
	Distance to basin margin	31.39	14	19.54	15
Soil properties	Soil, seasonally high water depth	28.17	21	16.04	29
	Soil, hydric	16.71	44	7.22	61
	Soil, hydrologic group A	23.00	34	13.09	40
	Soil, hydrologic group B	27.07	27	14.49	34
	Soil, hydrologic group C	27.76	24	17.77	18
	Soil, hydrologic group D	27.11	26	16.89	23
	Soil, permeability	28.21	20	15.87	30
	Soil, organic material	33.18	9	18.78	16
	Soil, clay	31.05	15	15.15	33
	Soil, silt	30.78	16	17.63	19
	Soil, sand	29.23	18	16.68	24

Table 12. Standardized importance scores for the prediction and confirmatory random forest classifiers of arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Standardized importance scores greater than 3.3 correspond to p-values less than 0.001 in the standard normal distribution. —, explanatory variable not included in model]

Variable group	Explanatory variable	Prediction classifier		Confirmatory classifier	
		Standardized importance score	Rank, maximum of 53	Standardized importance score	Rank, maximum of 85
Susceptibility variables—Continued					
Water use and hydroclimatic	Water-resources development index	15.54	49	6.51	63
	Groundwater use, irrigated agriculture	24.77	31	13.53	37
	Surface-water use, irrigated agriculture	24.55	33	16.38	26
	Groundwater use, public water supply	19.41	40	10.59	47
	Surface-water use, public water supply	15.95	47	9.30	49
	Recharge, contributing area	32.04	12	6.01	69
	Recharge, basin	28.09	22	7.61	55
	Potential evapotranspiration	42.47	1	25.48	10
	Mean air temperature	16.45	46	10.95	46
Basin groundwater budget	Recharge, subsurface inflow	—	—	6.42	65
	Recharge, mountain front	—	—	10.21	48
	Recharge, precipitation	—	—	4.48	79
	Recharge, stream infiltration	—	—	4.26	80
	Recharge, irrigation	—	—	5.45	72
	Recharge, artificial	—	—	6.19	67
	Recharge, change	—	—	8.97	51
	Storage, change	—	—	9.20	50
	Recharge, total	—	—	5.41	73
	Discharge, total	—	—	6.40	66
	Discharge, change	—	—	11.90	44
	Discharge, subsurface outflow	—	—	3.94	81
	Discharge, evapotranspiration	—	—	5.08	77
	Discharge, to streams	—	—	3.60	84
	Discharge, to springs and drains	—	—	3.19	85
	Discharge, well withdrawals	—	—	5.55	71
	Residence time	—	—	5.31	74
Geochemical variables					
Geochemical	Groundwater, pH	—	—	39.34	1
	Groundwater, dissolved oxygen	—	—	12.21	43
	Groundwater, dissolved solids	—	—	15.64	31
	Groundwater, nitrate	—	—	26.26	5
	Groundwater, sulfate	—	—	15.36	32
	Groundwater, iron	—	—	14.15	35
	Groundwater, manganese	—	—	17.59	20
	Groundwater, alkalinity	—	—	7.89	52
	Groundwater, bicarbonate	—	—	7.80	54
	Groundwater, orthophosphate	—	—	27.80	4
	Groundwater, chloride	—	—	18.39	17
	Groundwater, molybdenum	—	—	13.27	38
	Groundwater, selenium	—	—	7.87	53

Table 13. Training observation classification summary for the prediction and confirmatory random forest classifiers of arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.

[Dark shading indicates correct classification; light shading indicates one class above or below the correct class. **Abbreviations**: ≥, greater than or equal to, <, less than]

	Observed arsenic class and concentration range, in micrograms per liter							
	1	2	3	4	5	6	7	All classes
	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10–24	≥25	
Predicted class	Prediction classifier							
	True class, count of training observations							
1	300	110	77	42	45	29	15	618
2	142	167	132	94	58	30	22	645
3	90	138	161	142	75	58	25	689
4	32	58	108	181	130	47	25	581
5	19	24	28	114	214	107	46	552
6	22	21	36	65	143	209	113	609
7	27	10	23	40	57	111	200	468
	Classification summary							
Count of observations in class	632	528	565	678	722	591	446	4,162
Percentage of observations in class	15.2	12.7	13.6	16.3	17.3	14.2	10.7	100
Percentage of observations classified in correct class	47.5	31.6	28.5	26.7	29.6	35.4	44.8	34.4
Percentage of observations classified in correct class, one class above, or one class below	69.9	78.6	71.0	64.5	67.5	72.3	70.2	70.2
Predicted class	Confirmatory classifier							
	True class, count of training observations							
1	151	54	37	21	13	6	6	288
2	67	103	48	25	13	6	4	266
3	43	71	98	76	35	12	5	340
4	7	34	72	96	67	17	7	300
5	5	7	11	62	89	47	12	233
6	7	11	18	32	65	82	41	256
7	9	7	8	16	29	39	60	168
	Classification summary							
Count of observations in class	289	287	292	328	311	209	135	1,851
Percentage of observations in class	15.6	15.5	15.8	17.7	16.8	11.3	7.3	100
Percentage of observations classified in correct class	52.2	35.9	33.6	29.3	28.6	39.2	44.4	36.7
Percentage of observations classified in correct class, one class above, or one class below	75.4	79.4	74.7	71.3	71.1	80.4	74.8	75.0

distribution of observations among the lower 6 concentration classes was similar (11 to 18 percent); however, the highest concentration class (class 7, equal to or greater than 25 µg/L) was represented by about 7 percent of the 1,851 observations. During classifier optimization, the following weights were used to compensate for these differences in the number of observations within each of the concentration classes 1 through 7: 2.0, 2.2, 2.3, 2.2, 2.0, 2.6, and 2.8, respectively.

The confirmatory classifier was trained from observations of arsenic concentrations and 85 explanatory variables, which included geochemical data as well as basin-scale variables that were determined during the development of the conceptual model (Bexfield and others, 2011; table 12). All variables in the confirmatory classifier were significant, with standardized importance scores greater than 2, and, therefore, retained in the final classifier. The number of variables randomly chosen for each tree generated in the confirmatory classifier was 30 (about 35 percent of the total available). Similar to the prediction classifier, a minimum of 10 observations was required for each node in each tree generated.

The primary differences between the prediction and confirmatory random forest classifiers were the additional 2,311 observations available in the dataset used to train the prediction classifier and the availability of geochemical and selected basin-scale variables for the confirmatory classifier (table 12). Generally, the standardized importance scores for the 53 variables that were available for both the prediction and confirmatory classifiers were ranked in similar order for both classifiers (table 12), which indicates stability between the classifiers with respect to the variables used.

The classifiers were generally unbiased and demonstrated a good fit for the observed concentration classes as determined from the distribution of misclassification errors with respect to observed concentration class, geographic location, statistical distribution of explanatory variables, and estimated sampling error. Overall, about 34.4 percent of the 4,162 observations used to train the prediction classifier were placed in the correct category, and 70.2 percent were correctly placed within plus or minus one category (table 13). About 31.4 percent of the observations were overpredicted and about 34.2 percent underpredicted, indicating a lack of bias in the classifier toward overpredicting or underpredicting arsenic concentration classes. Correct classification rates for a given class (1 through 7) ranged from 26.7 to 47.5 percent (table 13). Further, for a given concentration class, most of the training observations are placed in the correct class, and the number of misclassified observations generally decreases for classes distant from the correct class (table 13, fig. 12). For the confirmatory classifier, 36.7 percent of the training observations overall were properly placed into each of the seven arsenic concentration classes (fig. 12, table 13). The proper placement of observations within the correct concentration class, 1 through 7, ranged from 28.6 to 52.2 percent. When considering the placement of observations within plus or minus one concentration class, the overall correct classification improved across individual concentration classes (71.1 to 80.4 percent), as well as overall (75.0 percent; table 13). Similar to the prediction classifier, the confirmatory classifier also was unbiased, having 33.5 and 29.8 percent of the observations were placed in concentration classes lower than or greater than the actual concentration class, respectively (fig. 12). The increase in correct classification rates (plus or minus one class) from the prediction classifier (70.2 percent) to the confirmatory classifier (75.0 percent) is due to the inclusion of geochemical and select susceptibility data in the confirmatory classifier (table 13).

The spatial distribution of the misclassification errors from the two classifiers generally showed no significant regional patterns. Misclassification errors for the prediction classifier observed by visual inspection appear to be random and evenly distributed across the study area (appendix 9). Average misclassification errors were examined for 64 basins where there were at least 15 observations of arsenic; 16 of the basins characterized could have underpredicted arsenic concentrations, and 3 could have overpredicted arsenic concentrations (appendix 4). Of the basins with potential for underprediction, several had average misclassification errors that were only slightly greater than 0.50, and only Avra Valley and Cache Valley had average misclassification errors greater than 1.00, which indicates that, if present, bias is generally one concentration class or less. For the confirmatory classifier, average misclassification errors indicated arsenic concentrations could be overestimated in Eagle Valley (average equal to -1.00) and underpredicted for the San Jacinto Basin (average equal to 0.72; appendix 5).

All percentile ranges for each explanatory variable were represented by training observations, and, in most cases, there were more than 100 training observations used for computing the average misclassification error (appendix 6). Thus, there were no variables that lacked representation by training observation concentrations for low, medium, or high values of the explanatory variables with respect to their distribution across all basin-fill aquifers of the SWPA study area. Average misclassification errors were within ±0.50 for high, medium, and low values of nearly all explanatory variables, which indicated that predictions were generally unbiased across the range of values occurring in the SWPA study area (appendix 6). Average misclassification errors were greater than 0.50 for training observations in the less than 10th percentile range for basin

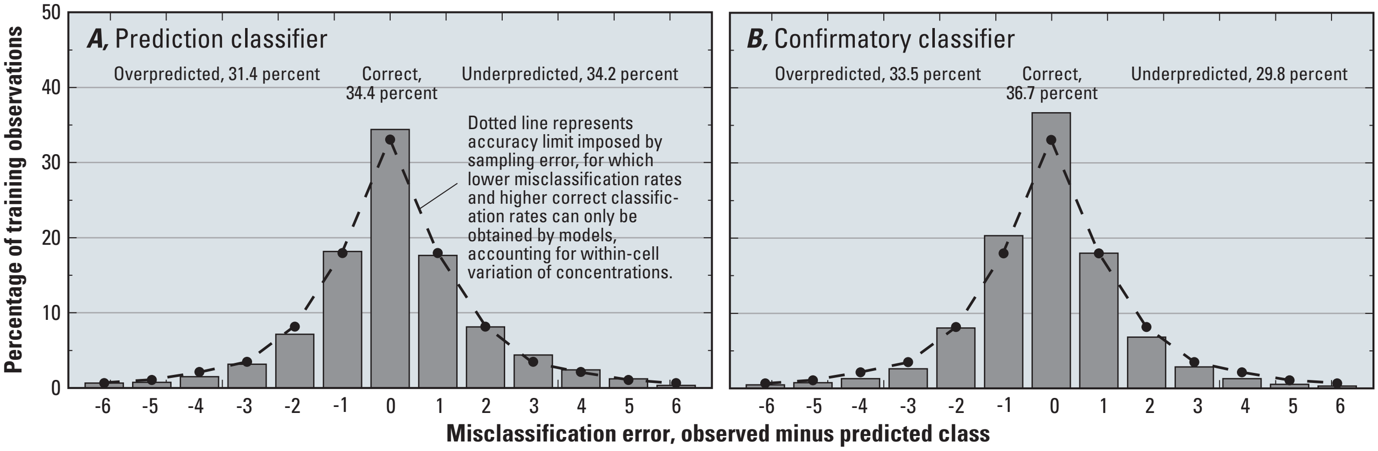


Figure 12. Statistical distribution of misclassification errors for random forest classifiers of arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area: A, Prediction classifier; B, Confirmatory classifier.

population (0.57), basin population density (0.69), and mean air temperature (0.70). These results indicate that in areas where there are relatively low rates of urbanization and air temperatures, arsenic concentrations can be underpredicted; however, these three average errors exceeding 0.50 could be due, in part, to the low number (less than 100) of training observations representing these percentile ranges (appendix 6). The lowest average misclassification error was −0.40 for training observations in the less than 10th percentile range for the water-resources development index, indicating overprediction of arsenic concentrations could occur in areas with low groundwater development. The low number of observations in this percentile group for the water-resources development index, however, could be influencing the misclassification error determined.

The arsenic classifiers have good predictive ability in consideration of the high sampling error detected in the training dataset. As discussed in the "Variable Selection and Goodness-of-Fit" part of the approach and methods section, sampling error limits the percent of correctly classified training observations to 33.0 percent. Both nitrate classifiers exceed this percentage (table 13, fig. 12) because they contain additional variables that explain some of the within-cell variation in nitrate concentrations, namely, aquifer-penetration depth for the prediction classifier and the geochemical variables, aquifer-penetration depth, well depth, and depth to water, for the confirmatory classifier. Although breaking up the arsenic concentration classes into seven categories results in a rather small correct classification rate of 34.4 percent, it increases the number of ways the predictions can be utilized. The ability to correctly classify predicted concentrations increases upon combining concentration classes and this can be useful for certain uses of the predicted concentrations. For instance, if one is distinguishing where arsenic occurs at concentrations less than 5 µg/L from where it is equal to or greater than 5 µg/L, and the error rate associated with that evaluation is needed, data contained in table 13 can be used to determine that error. First, data under the observed arsenic classes 1 through 4 that were predicted as classes 5 through 7 are summed: this value (429) represents the number of observations misclassified for concentrations less than 5 µg/L. Data under the observed arsenic classes 5 through 7 that were predicted as classes 1 through 4 are summed: this value (559) represents the number of observations misclassified for concentrations equal to or greater than 5 µg/L. With 988 (429 + 559) of the 4,162 training observations misclassified, the misclassification rate for this example is 23.9 percent. The correct classification rate for this example is 76.1 percent, which is more than double that of the original seven-class scheme (34.4 percent).

Regional Distribution and Depth Variation of Predicted Arsenic Concentrations

The random forest classifier, developed to predict arsenic concentrations throughout the 190,612 mi^2 of basin-fill aquifers (54,854 model grid cells) in the SWPA study area at an aquifer-penetration depth of 200 ft, indicated that 42.7 percent of the area is predicted to have groundwater with arsenic concentrations equal to or greater than the 10 µg/L drinking-water standard (table 14; fig. 13; appendix 9). Of the 4,162 observed (measured) concentrations of arsenic, 24.9 percent are equal to or greater than 10 µg/L (table 13; appendix 9). Differences in the statistical distributions of arsenic concentrations between the training dataset (table 13) and the prediction dataset (table 14) occur because they are two different statistical populations with somewhat different distributions of values for explanatory variables (appendix 3). Consequently, it would not be expected that both populations would provide the same statistical distribution of arsenic concentration.

As was found for the nitrate predictions, only a small percentage of the model grid cells had distinct trends in predicted arsenic concentration with aquifer-penetration depth; this is consistent with the training dataset (fig. 4). Of the 54,854 grid cells representing basin-fill aquifers, predicted arsenic concentrations systematically increased with aquifer-penetration depth in 6.9 percent of the cells and systematically decreased in 4.7 percent of the cells. Given the minimal influence of aquifer-penetration depth on predicted arsenic occurrence, the spatial distribution of arsenic concentrations was based on an aquifer-penetration depth of 200 feet because this represents an aquifer-penetration depth between those for domestic and water supply wells (table 14; fig. 13; appendix 9). Although concentrations were determined from the prediction classifier, these results also support the general findings summarized by Bexfield and others (2011) for the SWPA case-study basins that all ranges in arsenic concentrations are found at all depths within the basin-fill aquifers.

Broad areas showing predicted arsenic concentrations equal to or greater than 10 µg/L are, for the most part, sparsely populated and located in southeastern California, western Nevada, southwestern Arizona, and northwestern Utah (fig. 13; appendix 9). Localized areas in the San Joaquin and Sacramento Basins of the Central Valley, California, and the Middle Rio Grande Basin, New Mexico, also are predicted to have arsenic concentrations equal to or greater than 10 µg/L (appendix 9). For states representing greater than 10 percent of the SWPA study area, about 22 to 56 percent of the area of basin-fill aquifers located in each state had predicted arsenic concentrations equal to or greater than 10 µg/L (table 14). Nevada had the highest predicted area exceeding the drinking-water standard (56 percent); New Mexico had the lowest (22 percent). The highest percentages of the basin-fill aquifers' area with predicted arsenic concentrations equal to or greater than 25 µg/L were located in California (24 percent; 12,745 mi^2), Utah (23 percent; 5,208 mi^2), and Arizona (20 percent; 7,386 mi^2). Of the 32 (out of 422) basins where at least 75 percent of the basin was predicted to have arsenic concentrations equal to or greater than 25 µg/L, 23 are located in California (5,563 mi^2 total area), 6 in Arizona (4,903 mi^2 total area), and 3 in Nevada (2,113 mi^2 total area; appendices 7 and 9).

When evaluated by area, about 39 percent of the basin-fill aquifers in the SWPA are predicted to yield groundwater with

Table 14. Statistical distribution of predicted arsenic concentrations, by aquifer-penetration depth, principal aquifer, and state, for basin-fill aquifers in the Southwest Principal Aquifers study area.

[**Abbreviations:** ≥, greater than or equal to; <, less than; %, percent]

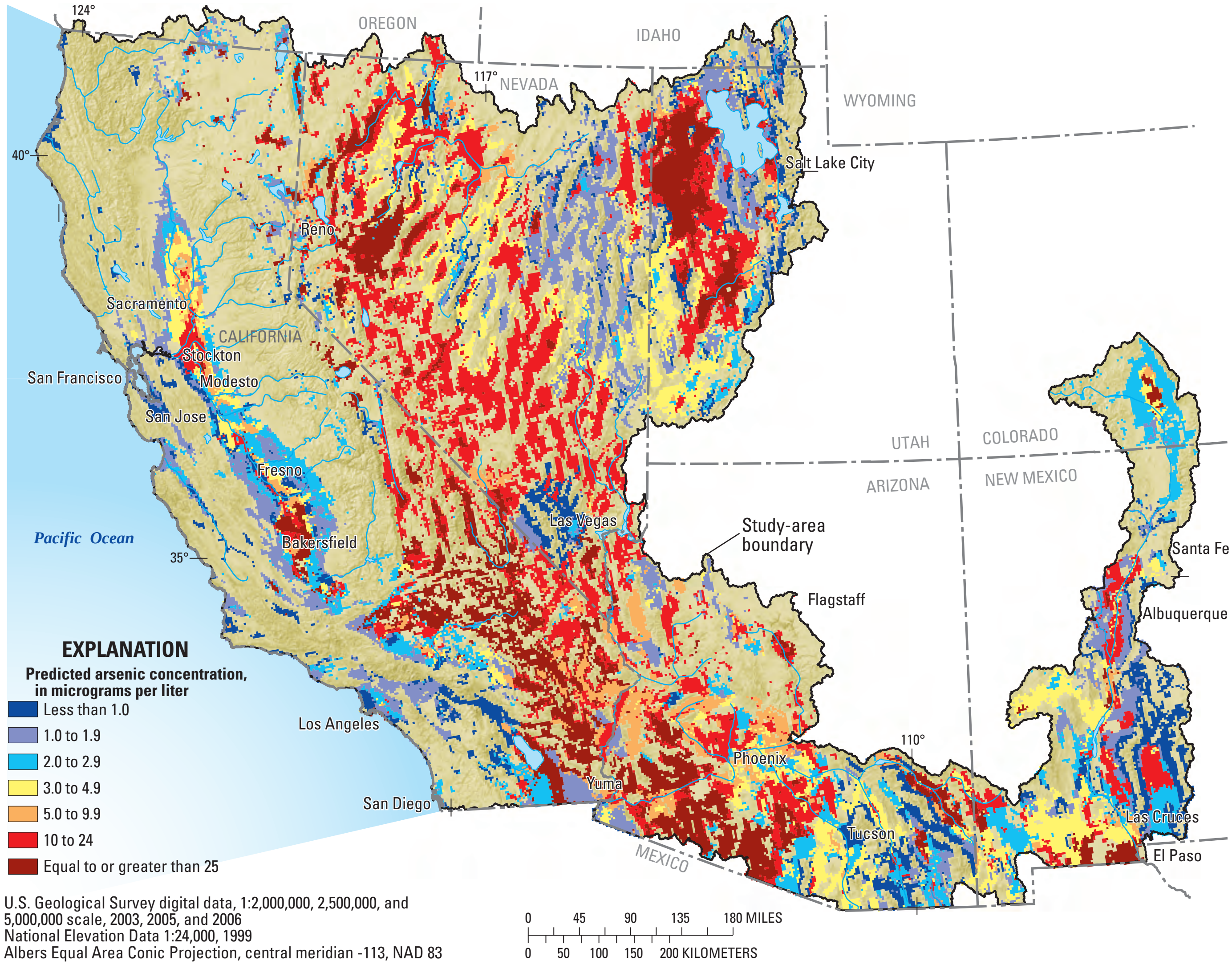
Distribution area	Total area for predictions, square miles	Percentage of total area, by arsenic class and concentration range, in micrograms per liter						
		1	2	3	4	5	6	7
		<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10–24	≥25
Distribution by aquifer-penetration depth, entire study area[1]								
50 feet	190,612	10.4	15.2	11.8	14.3	5.8	24.0	18.6
100 feet	190,612	10.6	15.5	11.9	13.4	6.3	23.6	18.7
150 feet	190,612	10.7	15.9	11.3	13.1	6.7	24.0	18.3
200 feet	190,612	11.1	15.6	10.8	12.7	7.0	25.8	16.9
250 feet	190,612	11.1	15.6	10.9	12.4	7.1	26.2	16.7
500 feet	190,612	10.4	16.0	10.3	11.4	6.8	28.9	16.3
750 feet	190,612	10.3	15.7	10.0	10.5	6.9	30.3	16.4
1,000 feet	190,612	10.1	15.7	9.5	9.9	6.6	31.8	16.6
Distribution by principal aquifer[2]								
Basin and Range basin-fill aquifers	138,642	8.9	12.3	6.2	11.7	7.9	32.1	20.9
California Coastal basin aquifers	7,082	42.3	43.6	11.2	1.4	0.4	0.2	0.9
Central Valley aquifer system	16,766	7.2	27.7	26.8	15.1	9.9	8.0	5.3
Pacific Northwest basin-fill aquifers	3,489	5.2	40.2	22.7	2.2	2.5	3.8	23.4
Rio Grande aquifer system	24,634	17.8	15.0	23.8	21.5	2.8	13.2	6.0
Distribution by state[2]								
Arizona	36,928	9.3	8.3	10.3	13.7	15.2	23.2	20.0
California	52,450	11.0	20.5	15.6	6.2	7.9	14.6	24.3
Colorado	3,093	0.0	4.7	74.8	10.0	3.3	0.0	7.2
Idaho	890	33.2	50.8	7.8	2.3	5.5	0.4	0.0
Nevada	52,030	8.2	15.5	2.6	13.7	4.0	45.7	10.3
New Mexico	22,073	20.0	16.0	16.2	22.9	2.9	16.2	5.8
Oregon	768	0.5	44.3	21.7	0.5	0.0	29.9	3.2
Texas	31	0.0	0.0	0.0	0.0	11.1	0.0	88.9
Utah	22,351	13.4	15.7	5.0	15.1	3.5	24.0	23.3

[1] About 6.9% of the model cells in the study area have a linear increase in arsenic concentration class with aquifer-penetration depth ($p<0.05$), and 4.7% have a linear decrease ($p<0.05$).

[2] Predictions are for an aquifer-penetration depth of 200 feet.

arsenic concentrations between 1.0 and 4.9 µg/L (table 14). Of the five principal aquifers, four have more than 50 percent of their area predicted to have arsenic concentrations in this range. The exception is the Basin and Range basin-fill aquifers, which are predicted to have about 30 percent (41,870 mi^2) of the area between 1.0 and 4.9 µg/L, and about 61 percent (84,430 mi^2) of the area with arsenic concentrations equal to or greater than 5.0 µg/L. Regionally, the predicted arsenic concentrations for all basin-fill aquifers in the SWPA study area appeared to follow a similar distribution pattern in percent occurrence to that found for the Basin and Range basin-fill aquifers, but this is likely because this area covered about 73 percent of the entire SWPA study area.

The Great Basin covers about 140,000 mi^2 (70 percent) of the Basin and Range Physiographic Province (Schaefer and others, 2006). Geologically, the western portion of the Great Basin is characterized by marine sedimentary and volcanic bedrock, where basin-fill aquifers are predicted to have higher arsenic concentrations. Lower arsenic concentrations are predicted for basin-fill aquifers in the eastern Great Basin, where bedrock is composed primarily of clastic sedimentary and carbonate rocks (Harrill and Prudic, 1998; appendix 9). The relatively low concentrations of arsenic observed in the vicinity of Las Vegas and in eastern Nevada can be attributed to the predominance of carbonate bedrock in those areas. Most of the basins in the southernmost 82,000 mi^2 of the Basin and Range Physiographic Province, residing largely in Arizona, are composed of clastic sediments, evaporites, volcanic rocks, and alluvium (Robertson, 1989). Many shallow groundwater samples in this area have arsenic concentrations that exceed 10 µg/L (Robertson, 1989).



U.S. Geological Survey digital data, 1:2,000,000, 2,500,000, and 5,000,000 scale, 2003, 2005, and 2006
National Elevation Data 1:24,000, 1999
Albers Equal Area Conic Projection, central meridian -113, NAD 83

Figure 13. Predicted arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area. For a larger version of this figure, see appendix 9.

In the Central Valley, California, two distinct areas have predicted arsenic concentrations that exceed 10 µg/L—the deltaic region at the confluence of the Sacramento and San Joaquin Rivers and the Tulare Lake area at the southern end of the San Joaquin Basin (appendix 9). Predicted arsenic concentrations from this study corroborate conclusions from Izbicki and others (2008), who found that concentrations of arsenic were lower near groundwater recharge areas along the foothills of the Sierra Nevada, and concentrations were higher in deeper wells at the downgradient end of long flow paths near the margin of the San Joaquin Delta. The Tulare Lake area covers about 310 mi^2 in the southern end of the Central Valley and is mostly used for the cultivation of cotton (Schroeder and others, 1988). Welch and others (1999) attribute the high arsenic concentrations in the area to natural sources and evapotranspiration. Fujii and Swain (1995) indicated that trace element enrichment in the area, including arsenic, primarily was driven by two processes: evaporative concentration and prevailing redox reactions.

In the Middle Rio Grande Basin of central New Mexico, Bexfield and Plummer (2003) used a detailed hydrogeologic framework to determine the general source and distribution of arsenic within the Santa Fe Group aquifer system of this basin. The primary sources of arsenic within the Middle Rio Grande Basin are silicic volcanic rocks and mineralized water originating deep within the basin and migrating upward along faults and other major structural features. Arsenic concentrations within the Middle Rio Grande Basin ranged from less than 1 to 600 µg/L, with the highest concentrations typically occurring in the northwestern and central portions of the basin (Bexfield and Plummer, 2003). The predictions of arsenic concentrations determined by using the random-forest classifier produced results similar to those found by Bexfield and Plummer (2003).

Comparison of the Arsenic Classifiers and Conceptual Model

In brief, the conceptual model developed by Bexfield and others (2011) identified geologic sources, residence time, redox conditions, and pH as important parameters to consider when examining the occurrence and transport of arsenic in 16 case-study basins within the SWPA study area (table 2). Although the classifiers were not as sensitive to basin-scale variables as local-scale characteristics, the results of the modeling efforts herein are consistent with this conceptual model.

The standardized importance scores indicate that both arsenic classifiers were generally more sensitive to geologic variables than other source variables, which shows geology has significant influence on arsenic concentrations in basin-fill aquifers of the SWPA study area (table 12). In most cases, the classifiers' accuracy was more sensitive to variables representing the distance to a geologic unit than to variables representing the percent area of a geologic unit in the bedrock surrounding a given basin (table 12). The strong negative correlations between arsenic predictions and the distance to mafic volcanic rocks, intermediate composition volcanic rocks, felsic/silicic volcanic rocks and crystalline rocks (table 15) indicate that these rocks are important sources of arsenic, and concentrations decrease with increasing distance from them. The significance of these geologic units as sources is seen for average predicted classes across the range of values for the distances to the geologic unit also. For example, average arsenic classes generally are greater for shorter distances than longer distances to mafic, intermediate, felsic and silicic composition volcanic rocks (table 15). For example, average arsenic class is 5.2 (table 15) where distance to mafic volcanic rocks is less than the 10th percentile value (6 km, from appendix 3). Likewise, the average arsenic class is 5.6 where percentage of mafic volcanic rocks in the surrounding bedrock exceeds the 90th percentile value (26.3 percent, from appendix 3).

Positive correlations between arsenic predictions and distance to carbonate rocks and clastic sedimentary rocks (table 15) indicate that arsenic concentrations increase with greater distance from these rocks, which indicates these types of rocks are not a significant source of arsenic. Likewise, predicted arsenic concentration is negatively correlated to percent area of carbonate and clastic sedimentary rocks (table 15). Predicted arsenic concentration is negatively correlated to percent area of carbonate and clastic sedimentary rocks (table 15), in part, because where there are less of these rocks, there are more crystalline, mafic volcanic, intermediate composition volcanic, or felsic/silicic volcanic rocks, which are associated with high arsenic concentrations. Areas with greater than 20 percent abundance of rocks classified as undifferentiated volcanic rocks are located in northern California, with lesser amounts in the south-central Sierra Nevada Mountains, Utah, and New Mexico (appendix 12). As was for the case of carbonate and clastic sedimentary rocks, the correlation between arsenic concentrations and percent areal coverage was negative and with distance from undifferentiated volcanic rocks was positive. These correlations could be influenced by the general locations of this geologic unit within the study area where other processes are of greater importance than source.

The correlations between arsenic predictions and rock type corroborate findings from previous studies. Welch and others (1988) found that geologic materials are the major sources of arsenic in the western United States. Robertson (1989) concluded that the highest arsenic concentrations in Arizona basins occur where basins are bounded by volcanic rocks, especially those of mafic and intermediate composition. Spatially complete datasets of arsenic content in rocks and sediments across the SWPA study area potentially could provide better information than the geologic variables used in the arsenic classifiers; however, such datasets were not available. Woolson and others (1977, p.17) determined the typical reported range of arsenic content for igneous rocks. Arsenic concentrations for mafic volcanic rocks generally have been reported to range from 0.06 to 113 µg/g and can contain a higher arsenic content than the felsic/silicic volcanic rocks, which have been generally reported to range from 0.2 to 13.8 µg/g. For comparison, carbonate rocks have been found to contain from 0.1 to 20.1 µg/g arsenic (Welch and others, 1988). As can be seen, arsenic concentrations even within a given rock type can be highly variable.

Unlike nitrate, arsenic-deposition data were not available from the National Atmospheric Deposition Program, and, therefore, atmospheric deposition could not be evaluated as a potential arsenic source. Other studies have shown, however, that atmospheric deposition of arsenic can be correlated to the deposition of sulfate bearing particles associated with power plant emissions (Heit and others, 1981; Nriagu, 1983; Smith and others, 1987). One study found arsenic concentrations in rainfall ranged from 0.1 µg/L in rural areas to 5 µg/L in urban areas (Galloway and others, 1982).

Correlations between arsenic predictions and agricultural and urban source variables such as population, agricultural land, and urban land were relatively strong and, in most cases, negative (table 15). Correlations were generally stronger for model grid cell-scale variables of land use than for basin-scale variables, indicating the greater sensitivity to localized conditions than basin-wide conditions. The negative correlations indicate that agricultural and urban lands are not general sources of arsenic, but rather, there could be processes, such as incidental recharge from precipitation in the contributing bedrock areas of the basins, that serve to reduce arsenic concentrations by flushing arsenic out of the system to streams or to adjacent basins (see text box "Have human activities in agricultural and urban areas affected arsenic concentrations in basin-fill aquifers of the Southwest?"). Average arsenic classes are low, typically near 3.0, where agricultural and urban source variables are high and exceed the 90th percentile for these variables. This indicates that, on average, groundwater from urban and agricultural areas is likely to have arsenic concentrations near 2.0 to 2.9 µg/L.

Generally, coarser grained sediments exist along upper basin margins and finer-grained sediments exist near the

Table 15. Relation between predicted arsenic concentrations and explanatory variables representing conditions for basin-fill aquifers of the Southwest Principal Aquifers study area.

[If a difference greater than 0.1 occurred between the sum of the average concentration class for percentiles 0 through 49.9 and the sum of the average concentration class for percentiles 50 through 100, then the predicted arsenic concentration class was deemed greater for lesser values of the explanatory variable. If this difference was less than –0.1, then the predicted arsenic concentration class was deemed greater for greater values of the explanatory variable; otherwise the relation between the arsenic concentration class and the explanatory variables was deemed unclear. **Abbreviations:** ≥, greater than or equal to; <, less than]

Variable group	Explanatory variable	Average predited arsenic concentration class number by percentile range for explanatory variable[1,2]						Observed concentration class is greater for lesser or for greater values of the geochemical variable	Kendall's tau test on predicted arsenic class number and explanatory variable			
		<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90		tau	Rank, maximum of 51	z-score	p-value
						Source variables						
Agricultural, urban, and biotic sources	Septic/sewer ratio	4.8	4.4	4.4	4.4	3.7	4.1	Lesser	–0.07	37	–25.7	<0.001
	Local population	4.8	4.8	4.8	4.0	3.8	3.1	Lesser	–0.16	12	–63.8	<0.001
	Local population density	4.8	4.8	4.8	4.0	3.8	3.1	Lesser	–0.16	13	–63.8	<0.001
	Basin population	4.9	4.8	4.8	4.4	3.3	3.3	Lesser	–0.18	10	–64.0	<0.001
	Basin population density	4.8	4.8	5.2	4.0	3.9	3.1	Lesser	–0.20	7	–70.7	<0.001
	Local urban land	4.7	4.7	4.7	4.2	3.9	3.1	Lesser	–0.15	18	–57.6	<0.001
	Local agricultural land	4.6	4.6	4.6	4.0	3.8	3.6	Lesser	–0.11	27	–47.4	<0.001
	Basin urban land	4.7	5.4	4.6	4.3	3.6	2.9	Lesser	–0.21	5	–76.6	<0.001
	Basin agricultural land	5.7	4.9	4.5	4.1	3.8	3.2	Lesser	–0.20	8	–69.9	<0.001
	Basin rangeland	3.1	3.3	4.1	4.4	5.7	5.6	Greater	0.29	3	102.0	<0.001
	Basin other land cover	5.9	5.0	4.7	4.3	3.5	2.8	Lesser	–0.26	4	–94.2	<0.001
Geologic sources	Geology, carbonate rocks	4.5	6.6	4.5	4.5	3.9	3.7	Lesser	–0.07	36	–26.1	<0.001
	Geology, crystalline rocks	3.9	4.4	4.2	4.6	4.3	4.5	Greater	0.04	45	14.4	<0.001
	Geology, clastic sedimentary rocks	5.3	4.8	4.3	4.2	4.4	3.2	Lesser	–0.16	14	–56.6	<0.001
	Geology, mafic volcanic rocks	3.6	3.5	4.3	4.4	4.8	5.6	Greater	0.16	11	58.2	<0.001
	Geology, felsic and silicic volcanic rocks	4.2	4.2	4.2	4.4	4.5	4.7	Greater	0.04	44	15.2	<0.001
	Geology, intermediate composition volcanic rocks	3.9	3.9	4.5	4.2	4.6	5.3	Greater	0.11	29	38.6	<0.001
	Geology, undifferentiated volcanic rocks	4.6	4.6	4.6	4.1	4.3	3.8	Lesser	–0.07	35	–28.9	<0.001
	Geology, distance to carbonate rocks	3.3	3.9	4.6	4.9	4.8	3.3	Greater	0.07	38	25.3	<0.001
	Geology, distance to crystalline rocks	4.2	4.4	4.6	4.4	4.1	4.2	Lesser	–0.02	50	–8.4	<0.001
	Geology, distance to clastic sedimentary rocks	3.5	3.9	4.2	4.4	4.9	5.0	Greater	0.14	20	50.2	<0.001
	Geology, distance to mafic volcanic rocks	5.2	5.1	4.6	3.9	3.8	3.5	Lesser	–0.19	9	–66.2	<0.001
	Geology, distance to felsic and silicic volcanic rocks	4.5	4.6	4.4	4.5	4.8	2.8	Lesser	–0.06	42	–20.5	<0.001
	Geology, distance to intermediate composition volcanic rocks	4.8	4.7	4.5	4.6	4.1	2.8	Lesser	–0.12	25	–41.2	<0.001
	Geology, distance to undifferentiated volcanic rocks	4.3	4.2	4.1	4.7	4.4	4.3	Greater	0.03	48	10.8	<0.001
	Soil and rock equivalent uranium-238 concentration	3.6	3.8	4.2	4.6	4.9	4.7	Greater	0.14	21	48.8	<0.001

Table 15. Relation between predicted arsenic concentrations and explanatory variables representing conditions for basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[If a difference greater than 0.1 occurred between the sum of the average concentration class for percentiles 0 through 49.9 and the sum of the average concentration class for percentiles 50 through 100, then the predicted arsenic concentration class was deemed greater for lesser values of the explanatory variable. If this difference was less than –0.1, then the predicted arsenic concentration class was deemed greater for greater values of the explanatory variable; otherwise the relation between the arsenic concentration class and the explanatory variables was deemed unclear. **Abbreviations:** ≥, greater than or equal to; <, less than]

Variable group	Explanatory variable	Average predited arsenic concentration class number by percentile range for explanatory variable[1,2]						Observed concentration class is greater for lesser or for greater values of the geochemical variable	Kendall's tau test on predicted arsenic class number and explanatory variable			
		<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90		tau	Rank, maximum of 51	z-score	p-value
							Susceptibility variables					
Flow path	Land-surface slope	5.0	4.7	4.4	4.3	4.1	3.6	Lesser	–0.12	24	–42.9	<0.001
	Land-surface elevation	3.6	4.7	5.0	4.6	3.8	3.2	Lesser	–0.10	31	–37.2	<0.001
	Land-surface elevation percentile	5.2	4.8	4.5	4.1	3.8	3.6	Lesser	–0.16	15	–56.6	<0.001
	Basin elevation	3.3	5.3	4.9	4.4	4.1	3.4	Lesser	–0.07	39	–25.0	<0.001
	Distance to basin margin	4.0	4.0	4.3	4.5	4.6	4.6	Greater	0.08	34	28.5	<0.001
Soil properties	Soil, seasonally high water depth	5.0	4.3	4.0	4.4	4.4	4.4	Lesser	–0.03	49	–10.8	<0.001
	Soil, hydric	4.3	4.3	4.3	3.9	3.9	5.4	Greater	0.04	47	16.5	<0.001
	Soil, hydrologic group A	4.0	4.0	3.9	4.3	5.2	5.2	Greater	0.14	19	52.4	<0.001
	Soil, hydrologic group B	5.0	4.2	4.3	4.2	4.4	4.3	Lesser	–0.04	46	–13.6	<0.001
	Soil, hydrologic group C	4.8	4.6	4.5	4.3	3.8	3.7	Lesser	–0.11	26	–41.1	<0.001
	Soil, hydrologic group D	4.5	4.5	4.3	4.2	4.0	5.0	Unclear	0.00	51	0.2	0.806
	Soil, permeability	3.6	3.9	4.1	4.5	4.9	5.3	Greater	0.16	16	55.6	<0.001
	Soil, organic material	4.4	4.6	5.0	4.5	3.5	3.1	Lesser	–0.15	17	–52.6	<0.001
	Soil, clay	5.0	4.7	4.4	3.8	3.9	4.9	Lesser	–0.07	40	–24.7	<0.001
	Soil, silt	4.9	4.7	4.5	3.8	4.5	3.9	Lesser	–0.09	33	–31.5	<0.001
	Soil, sand	4.8	3.8	3.8	4.4	4.7	5.1	Greater	0.09	32	33.0	<0.001
Water use and hydroclimatic	Water-resources development index	6.0	5.0	4.5	3.7	4.0	3.5	Lesser	–0.20	6	–71.7	<0.001
	Groundwater use, irrigated agriculture	4.6	4.6	4.6	3.9	3.9	3.6	Lesser	–0.11	28	–45.2	<0.001
	Surface-water use, irrigated agriculture	4.6	4.6	4.6	4.0	3.8	3.7	Lesser	–0.11	30	–45.0	<0.001
	Groundwater use, public water supply	4.5	4.5	4.5	4.5	4.5	3.0	Lesser	–0.06	41	–44.0	<0.001
	Surface-water use, public water supply	4.4	4.4	4.4	4.4	4.4	3.0	Lesser	–0.05	43	–43.0	<0.001
	Recharge, contributing area	5.9	6.0	5.0	3.3	2.9	3.2	Lesser	–0.37	1	–132.6	<0.001
	Recharge, basin	6.0	5.8	4.8	3.7	3.1	3.0	Lesser	–0.37	2	–130.7	<0.001
	Potential evapotranspiration	3.2	4.4	4.4	3.8	5.0	5.5	Greater	0.13	22	47.9	<0.001
	Mean air temperature	3.3	3.8	4.7	3.7	4.9	5.4	Greater	0.13	23	45.2	<0.001

[1] See appendix 3 for values of the explanatory variable that correspond to each percentile range.

[2] Concentration ranges for classes 1 through 7 are <1.0, 1.0–1.9, 2.0–2.9, 3.0–4.9, 5.0–9.9, 10–24, ≥25 micrograms per liter.

basin center or low-lying areas (Thiros and others, 2010). As mentioned previously in the nitrate section, variables representing important soil characteristics, including permeability and percent clay, silt, and sand, can be used to provide a measure for the ease with which water can move through the soil surface and, potentially, into the aquifer system. The correlations between predicted arsenic and soil properties are inconsistent with respect to recharge characteristics (table 15). Arsenic predictions generally are positively correlated with variables indicative of more permeable environments (soil permeability, hydrologic group A, and percent sand), which indicates higher arsenic concentrations are likely to be found in areas with more permeable soils. If areas of relatively high permeability are associated with those with enhanced recharge, natural or otherwise, the recharge component potentially could contribute to higher arsenic concentrations and, therefore, a positive correlation also should exist between predicted arsenic and at least some of the variables indicative

Have human activities in agricultural and urban areas affected arsenic concentrations across the Southwest?

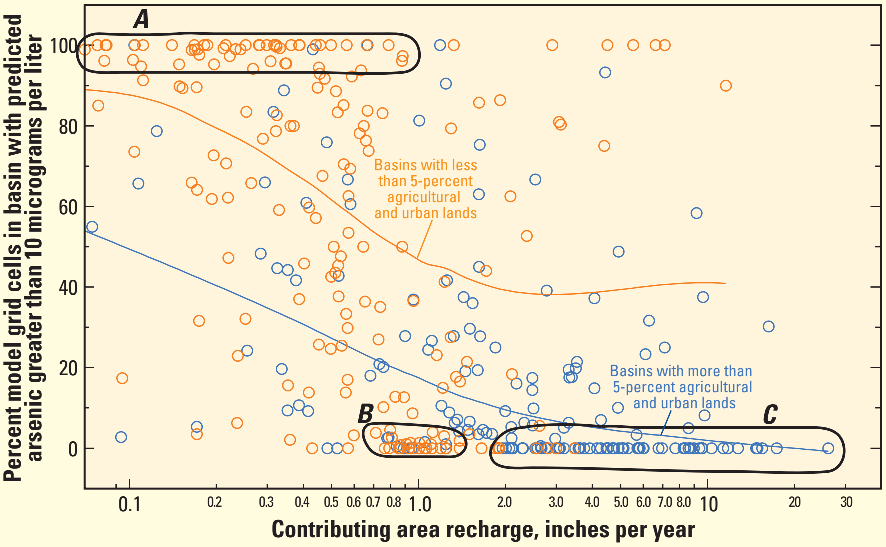
An initial inspection of the graph below shows that the percentage, by area, of a given basin that is predicted by the random forest classifier to exceed the 10 µg/L drinking water standard for arsenic decreases with an increase in the natural recharge within the contributing area of that basin. Further inspection shows that basins with developed lands (shown in blue, and having 5 percent or more land developed for agricultural or urban uses) generally have small percentages of their area predicted to exceed 10 µg/L. In contrast, basins with minimal agricultural and urban land development (shown in orange, and having less than 5 percent land developed for agricultural or urban uses) tend to have substantial percentages of their area predicted to exceed 10 µg/L. The solid lines help illustrate this relation and are locally-weighted scatter-plot smooths for the points representing the two different sets of basins. A hastily arrived at conclusion, based on the graph below and the negative correlations between predicted arsenic and land-use variables (table 10), is that land development lowers arsenic concentrations. Further examination of the data points in the graph, however, indicates that this relation is spurious, at least at the regional scale for the Southwest.

Basins in **group *A*** are predominately located in the western part of the Basin and Range Physiographic Province and have little natural recharge from precipitation and little or no groundwater outflow. Most of the basins in this group are undeveloped, in part because of their dry character. The lack of land development in these basins was not the cause of high arsenic concentrations in these basins. Rather, arsenic concentrations are mostly above 10 µg/L because of the presence of volcanic or crystalline rocks in the surrounding bedrock, the lack of groundwater flushing, which is characterized by low natural recharge from precipitation in the contributing area and also by little or no groundwater outflow, or both.

Basins in **group *B*** are predominately from the carbonate province of the Great Basin, near the Nevada-Utah border. Natural recharge from precipitation in the contributing area is moderate, ranging from about 0.7 to 1.7 inches per year. Arsenic concentrations in these basins are generally less than 10 µg/L because of the predominance of carbonate or clastic sedimentary rocks, variable presence of volcanic or crystalline rocks surrounding these basins, and the flushing of groundwater from these basins to neighboring basins.

Most of the basins in **group *C*** are located in California, primarily in northern California or near the coast. These basins generally have more natural recharge from precipitation compared to other basins in the Southwest Principal Aquifer study area. Most of these basins are urbanized or are agricultural, in part, because water is more readily available, which attracted people to settle there. Arsenic concentrations are relatively low in these basins (mostly less than 10 µg/L) because of the relatively high groundwater recharge conditions and groundwater drainage to other basins or to the ocean.

Thus, in summary for the Southwest Principal Aquifer study area as a region, the apparent causal relation between land use and predicted arsenic concentrations is spurious and is likely to simply be a relict of human tendency to settle and develop lands in basins that have larger amounts of contributing area recharge because of the greater availability of water. In localized areas, however, recharge from human activities can change the geochemistry of the aquifer and result in increased or decreased concentrations.



of recharge characteristics (e.g. recharge of contributing area, groundwater and(or) surface water used for irrigation). Basin-scale recharge-related variables derived by Bexfield and others (2011) were not available throughout the SWPA study area and, therefore, were unavailable for use in the prediction classifier. The confirmatory classifier accuracy was not strongly sensitive to basin-scale recharge variables (ranks ranging from 48 to 80 out of 85, table 12), however, and predicted arsenic concentrations from the prediction classifier were negatively correlated to basin and contributing area recharge, indicating that lower concentrations of arsenic are likely to be found in areas of greater recharge. A possible reason for the inconsistency between the recharge and soil characteristic correlations to predicted arsenic is that the soil characteristics are for surficial soil conditions and do not necessarily represent conditions deeper in the basin-fill sediments. On a regional scale, it appears that the leaching of arsenic from surficial soils does not provide a strong signature; however, some researchers have found arsenic leaching to be a contributory source of arsenic in localized irrigated areas (Busbee and others, 2009).

The two variables with the strongest correlation to predicted arsenic concentration were recharge for the contributing area and recharge for the basin (table 15). Where recharge is low, less than the 25th percentiles for either variable, the average arsenic concentration class is high, about 6.0 (table 15). The prediction classifier accuracy was also relatively sensitive to recharge for the contributing area and recharge for the basin (rank 12 and 22, respectively, out of 53); however, the confirmatory classifier accuracy was less sensitive to these variables (rank 69 and 55, respectively, out of 85; table 12). It is unclear why the importance of recharge is enhanced in the prediction classifier and not within the confirmatory classifier. One possible explanation is that these recharge variables, if correlated to recharge from irrigated agriculture, could be related to geochemical controls for which information was available for the confirmatory classifier. For example, Jurgens and others (2009) studied the influence of recharge and groundwater pumping on uranium occurrence in the eastern San Joaquin Valley, California. These researchers concluded that the application of irrigation water has changed shallow groundwater chemistry and the rate of shallow groundwater movement to deeper parts of the aquifer in that area. Additionally, aquifer systems with adequate recharge could possess the ability to flush solutes out of the basin as interbasin flow or discharge to streams flowing into adjacent basins or, in some cases, to the ocean (see text box “Have human activities in agricultural and urban areas affected arsenic concentrations across the Southwest?”). It is important to note, however, that these possible explanations for the importance of aquifer recharge to arsenic concentrations are not necessarily exclusive of one another.

Predicted arsenic concentrations were positively correlated with potential evapotranspiration (ranked 22, table 15). Average arsenic class was near 5.0, corresponding to 5.0–9.9 µg/L, where potential evapotranspiration exceeded the 75th percentile value, 1,443 mm/yr. Although evapotranspiration from the basin groundwater budget was not a sensitive characteristic within the confirmatory classifier (ranked 77), potential evapotranspiration (ranked 1 in the prediction classifier and 10 in the confirmatory classifier) could be acting as a surrogate, indicating areas of discharge in the SWPA study area and associated geochemical conditions, or as a surrogate for low recharge (table 12). Largely on the basis of field reconnaissance, Harrill and Prudic (1998) concluded that most of the groundwater discharge within the Great Basin is through evapotranspiration at topographically low areas of valleys where the water table is relatively close to the land surface. In low-lying areas within the basins of the SWPA study area, where the water table is relatively close to the land surface, evapotranspiration has been found to increase solutes in the underlying aquifer (Anning and others, 2007). A study investigating water budget and quality in an unpopulated terminally closed basin in Nevada showed a gradient in arsenic concentration from lower concentrations in deeper wells located along the upper basin margins to higher concentrations in shallow wells on the playa (see text box “Arsenic accumulation underneath playas”). The surrounding geology is predominately felsic/silicic rhyolitic tuff (Mankhemthong and others, 2008). Groundwater flows into the basin from surrounding valleys and discharges largely by evapotranspiration within the playa area (Harrill and Hines, 1995). Welch and others (2000) concluded that volcanic rocks, pH, and evaporative concentration contribute to the high arsenic concentrations in groundwater in the western United States.

The conceptual model developed by Bexfield and others (2011) also identified residence time as an important factor with respect to arsenic occurrence. Groundwater recharge in many of the basins occurs along upper basin margins, where the subsurface is largely composed of coarse-grained and poorly sorted material conducive to downward hydraulic gradients. Within the random forest classifier, relatively high values for land-surface elevation percentile and land-surface slopes were used as surrogates representing basin-fill locations near mountain front recharge areas. Additional sources of aquifer recharge can be in the form of inflow from adjacent basins and through the infiltration of precipitation on valley floors and stream channels. The amount of time a volume of water takes to move through an aquifer system from the point of recharge to a point of discharge is termed “residence time.” Generally, the longer a volume of groundwater is in contact with sediments, the greater the likelihood the groundwater will become enriched in some constituents that are present in or sorbed to those substrates. Groundwater recharged near upper basin margins tends to be relatively dilute in naturally occurring constituents. As the water moves through the system from upper basin margins toward basin lowlands, over time, these constituents can become increasingly concentrated.

Small values of land-surface elevation percentiles and land-surface slopes were used to indicate basin lowland areas, away from upper basin margins and, therefore, where groundwater residence times tend to be longer. The prediction and confirmatory classifier accuracies were sensitive to land-surface elevation percentile (ranked 2 and 6, respectively).

Arsenic accumulation underneath playas

The largest area where arsenic concentrations are predicted to equal or exceed 10 µg/L occurs within the Basin and Range Physiographic Province, which contains many terminal lakes and playas that can concentrate arsenic concentrations through evapotranspiration. At the end of the last ice age, Pleistocene-age lakes in the western United States, such as Lahontan, Bonneville, and Mojave I/II, covered vast areas where only remnant perennial lakes now remain and have no outflow, including, for example, Pyramid Lake, Mono Lake, and Great Salt Lake (Reheis, 1999; Morrison, 1991). When the climate changed and evaporative losses from these ancient lakes exceeded inflow, lake elevations declined and shorelines receded. Faulting and geologic features of low permeability throughout the area resulted in the formation of topographically closed basins, most of which contain playas. Playas are topographically low areas that retain ephemeral water originating from periodic events and are characterized by fine-grained sediments and a relatively shallow water table compared to other parts of the basin (Planert and Williams, 1995). The depth to groundwater maintained by playas depends largely on the climate and playa sediments, and whether there is subsurface flow to an adjacent basin (Planert and Williams, 1995; Tyler and others, 2005). In areas where there is vegetation, water depth also can be influenced by the type of vegetation present (Laio and others, 2009). In closed basins where there is no surface or groundwater outflow, arsenic and other solutes accumulate.

As groundwater moves through the aquifer from an area of recharge near the basin margin to low-lying areas, solutes are leached from aquifer sediments into the groundwater, increasing in concentration along the way. In a closed basin, such as the Dixie Valley pictured below, evapotranspiration processes also concentrate solutes in the shallow groundwater. On the open playa, a crust of salt often forms at the surface as shallow groundwater is transported to the surface and evaporates. In Dixie Valley, arsenic concentrations in groundwater samples collected near the basin margins averaged 9.1 µg/L; shallow groundwater collected from underneath the Dixie Valley playa had arsenic concentrations averaging 12 mg/L, about 1300 times higher than the average concentration in groundwater near the margins (Jena Huntington, U.S. Geological Survey, written commun., 2010).

Accumulation of salts on the playa surface in Dixie Valley, Nevada. [Pictures taken by Jena Huntington and Michael Rosen, U.S. Geological Survey, 2008]

The classifier accuracies were also sensitive to land-surface slope and distance to basin margin (table 12). There is a negative relation between predicted arsenic concentrations and both land-surface slope and land-surface elevation percentile (table 15). These results indicate that, at the upper basin margins where slopes and relative elevations are greatest, arsenic concentrations likely will be lower than in the basin lowlands. Average arsenic class increases from 3.6 at the highest elevation percentiles to 5.2 for the lowest elevation percentiles (table 15). A positive correlation exists between distance to basin margin and arsenic concentration class, indicating that the further out from the basin margin, the higher the arsenic concentrations are likely to be. These results are consistent with a flow-path related influence to the distribution of arsenic within the SWPA study area. Groundwater residence time, as a basin-scale variable, was not a very sensitive indicator within the confirmatory random forest classifier for which this variable was available (ranked 74, table 12). As mentioned previously for the nitrate classifiers, arsenic classifier accuracies were not as sensitive to basin-scale variables as variables scaled to the model grid cell. The lack of sensitivity to basin-scale variables, such as groundwater residence time, is likely to be the result, in part, of greater variability inherent to the variable within each individual basin that is not captured in the relatively large-scale estimates that often are totals for a basin. These results indicate that as groundwater moves along a flow path away from upper basin margins into lowland areas with relatively low slope, arsenic concentrations are likely to increase. These findings are consistent with those of Robertson (1989), who found that arsenic concentrations were generally higher near basin centers in Arizona. Hinkle and others (2009) found that arsenic concentrations will most likely increase with increasing residence time (greater than 200 years, in some cases) in aquifers with arsenic-containing sediment.

The confirmatory classifier accuracy was most sensitive to geochemical variables, particularly to pH (ranked 1; table 12). Other important geochemical variables included concentrations of orthophosphate (ranked 4), nitrate (ranked 5), chloride (ranked 17), and manganese (ranked 20). Univariate correlations were positive for pH, dissolved solids, sulfate, iron, manganese, orthophosphate, chloride, and molybdenum, and negative for dissolved oxygen and nitrate (table 5). Detailed examination of these results indicate that pH, which affects arsenic sorption to aquifer materials, is likely the predominant geochemical factor affecting arsenic concentrations in basin-fill aquifers of the SWPA study area. Other factors that appear to be less dominant at the regional scale, but could be important at the basin to local scale, include reductive dissolution of iron and manganese oxides with subsequent release of arsenic to the groundwater and competitive sorption between arsenic and phosphorus.

The sorption of arsenic to aquifer substrates is largely influenced by pH primarily because of the charge imparted to substrate surfaces and the dominant arsenic species present in the system (see text box “Arsenic and iron oxide interactions under different pH and redox conditions”), which is likely to be arsenate. The importance of pH to classifier prediction accuracy (table 12) is likely related to this. Although arsenic-speciation data were unavailable, given that nitrate predominates over nitrite in most of the groundwater samples included in this assessment, the redox conditions are likely oxidizing with respect to the arsenite/arsenate redox couple (Cherry and others, 1979). Robertson (1989, fig. 4, table 1) assessed arsenic speciation and its associated geochemical conditions in shallow basin-fill aquifers in Arizona, and found that conditions were favorable to support arsenate and that concentrations were largely controlled by sorption mechanisms. Similar findings were reported by Bexfield and Plummer (2003) for the Middle Rio Grande Basin, New Mexico. Theoretically, the charge of the arsenate species ($H_2AsO_4^-$ and $HAsO_4^{2-}$) is negative within this range of pH, whereas arsenite ($H_3AsO_3^o$) remains largely uncharged (Stollenwerk, 2003). At basic pH (greater than 8.0), the negative charge theoretically imparted on metal oxides, such as those of aluminum, iron, and manganese (Anderson and others, 1976; Davis and Leckie, 1978; Driehaus and others, 1995), present within an aquifer can repel the negatively charged arsenate oxyanion. This limits the arsenate adsorption capacity of aquifer substrates, thereby keeping arsenic in solution (see text box “Arsenic and iron oxide interactions under different pH and redox conditions”).

Observed relations between arsenic concentrations and pH are consistent with the results of the geochemical studies mentioned in the previous paragraph. Arsenic concentration class and pH are strongly and positively correlated (table 5), which is consistent with the limitation on adsorption capacity with increasing pH. More specifically, the average observed arsenic concentration class was between 3.3 and 3.5 where pH was in the less than the 10th percentile (less than 7.0 pH units), in the 10th to 24.9th percentile (7.0 to 7.3), or in the 25th to 49.9th percentile (7.0 to 7.6). In contrast, where pH was greater than the 90th percentile value (8.2), the average arsenic concentration class was 5.2.

Analysis of the classifier results indicates that reductive dissolution of iron or manganese oxides is not, at the regional scale, as prevalent a mechanism for releasing arsenic to groundwater as sorption processes influenced by pH under oxidizing conditions. Locally, however, reducing conditions can dissolve these oxides and, subsequently, release arsenic, along with the iron or manganese, into the groundwater. McMahon and Chapelle (2007) found that reducing environments were associated with higher arsenic concentrations in aquifer systems studied throughout the United States. They present a redox classification scheme that considers the influence of biological processes in reducing environments, where nitrate is a preferred electron acceptor over manganese and iron. Using their redox classification scheme, the redox condition was oxidizing with respect to nitrate, manganese, and iron redox couples for most of the basin-fill aquifer model grid cells for which the necessary data were available. This is consistent with McMahon and Chapelle’s (2007) findings that manganese and iron reducing conditions were less common in aquifers in the west than in other principal aquifers in the

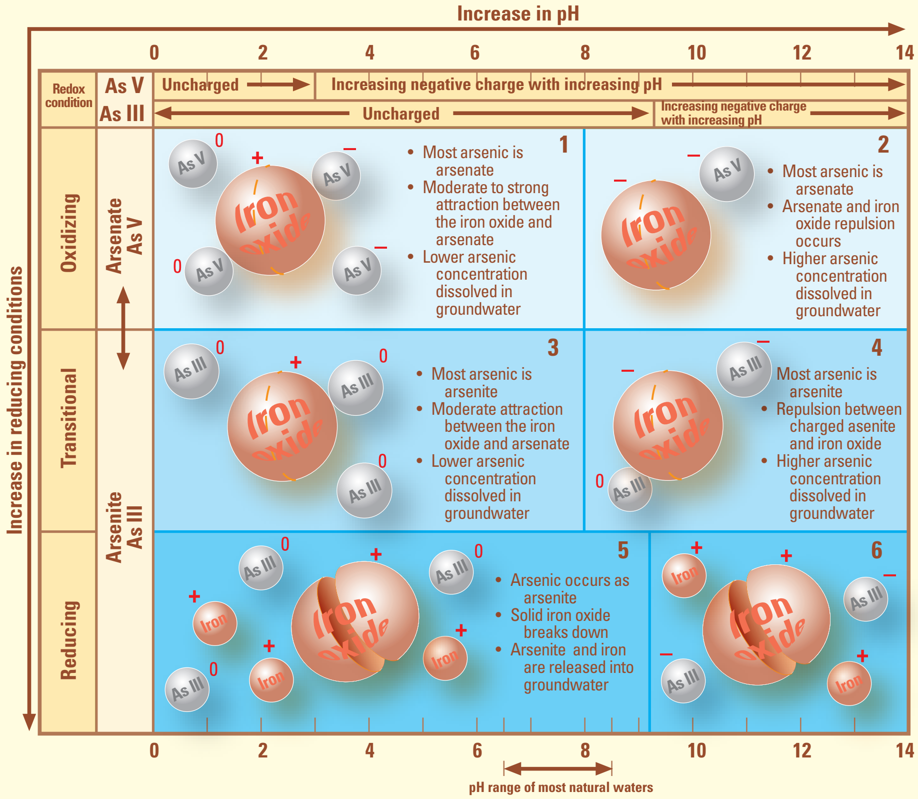
Arsenic and iron oxide interactions under different pH and redox conditions

The geochemical processes most important to arsenic removal and release in groundwater are sorption, co-precipitation, and reductive dissolution. The interaction of arsenic with iron oxide (rust) is one of the most important processes by which groundwater can become either enriched or depleted in arsenic. The interaction is controlled largely by the availability of dissolved oxygen and the pH of the groundwater, and the underlying principle is charge repulsion and attraction between the iron-oxide surface and the arsenic molecule. In some instances, the organic arsenic species (carbon containing), monomethylarsonic acid (MMA) and dimethylarsenic acid (DMA), can be found; MMA and DMA generally occur at very low concentrations relative to the inorganic forms of arsenic. The following discussion is a synthesis of information from research regarding the behavior of arsenic in groundwater systems. A literature review of these processes and those in more complex systems is covered in greater detail in Smedley and Kinniburgh (2002).

In a simplified system, as shown by conditions 1 though 6 in the illustration below, with sufficient dissolved oxygen present (oxidizing conditions), the oxidized form of arsenic (arsenate) predominates. When the pH of water is less than 8, the surface of iron oxide theoretically will carry a positive charge (denoted by “+”), and arsenate, As(V), will be either uncharged (denoted by “0”) or carry a negative charge (denoted by “–“). The two constituents are attracted, and the arsenic essentially attaches to the iron-oxide surface (adsorbs), which removes it from solution (Condition 1). Alternatively, when the pH of water is above 8 under oxidizing conditions, the iron-oxide surface becomes negative. Arsenate is negative also; therefore, repulsion occurs between arsenate and the iron-oxide surface, and arsenic remains dissolved in water (Condition 2).

If conditions are sufficiently reducing, the iron oxide breaks apart by a process termed reductive dissolution. During dissolution, any constituents attached to the oxide are released into the water along with the iron (Conditions 5 and 6). At pH values greater than 9.3, arsenite, As(III), the reduced form of arsenic, becomes negatively charged (Condition 6). Generally, under reducing conditions, both arsenite and iron are in solution, and the attraction or repulsion between arsenic and iron is of less importance.

In a simplified system, as presented here, when redox conditions transition from reducing to oxidizing, the rate at which the reduced forms of iron and arsenic become oxidized is increasingly important. Reduced, or ferrous, iron, oxidizes to ferric iron more rapidly than arsenite oxidizes to arsenate. Under these conditions, where the redox state of the groundwater could be promoting the formation of iron oxide, arsenite can still predominate, or contribute substantially, to the overall arsenic species present in solution (or adsorbed to the iron-oxide surface). Similar to oxidizing conditions, when the pH of water is less than 8, the surface of iron oxide is positive; however, aqueous arsenite is uncharged. A moderate attraction occurs between positively charged iron oxide and uncharged arsenite, and some arsenic is removed from the water as it attaches to the oxide surface (Condition 3). Under these transitional redox conditions, where the pH of water is above 8, iron oxide becomes negatively charged, and any uncharged arsenite in solution remains somewhat attracted to the oxide and will attach to the oxide surface; however, at pH values greater than 9.3, there is no longer an attraction between the negatively charged iron-oxide surface and the now negatively charged arsenite molecule, so arsenite remains in solution (Condition 4).


Increase in pH
0
2
4
6
8
10
12
14
Redox condition
As V
Uncharged
Increasing negative charge with increasing pH
As III
Uncharged
Increasing negative charge with increasing pH
Increase in reducing conditions
Oxidizing
Transitional
Reducing
Arsenate As V
Arsenite As III
1
• Most arsenic is arsenate
• Moderate to strong attraction between the iron oxide and arsenate
• Lower arsenic concentration dissolved in groundwater
2
• Most arsenic is arsenate
• Arsenate and iron oxide repulsion occurs
• Higher arsenic concentration dissolved in groundwater
3
• Most arsenic is arsenite
• Moderate attraction between the iron oxide and arsenate
• Lower arsenic concentration dissolved in groundwater
4
Most arsenic is arsenite
• Repulsion between charged asenite and iron oxide
• Higher arsenic concentration dissolved in groundwater
5
• Arsenic occurs as arsenite
• Solid iron oxide breaks down
• Arsenite and iron are released into groundwater
6
Iron oxide
As V
As III
Iron
0
+
–
pH range of most natural waters

United States. Prevalence of oxic conditions likely favor sorption of arsenic to iron and manganese oxides, making it likely to be a predominant factor affecting arsenic occurrence in SWPA basin-fill aquifers at a regional scale.

Average observed arsenic concentration class for geochemical variable percentile ranges shows the effects of redox conditions on arsenic concentration. Average observed arsenic concentration class is relatively low where dissolved-oxygen concentrations or nitrate concentrations are high and indicative of oxic conditions. For example, the average observed arsenic concentration class is 2.8 (table 5) where dissolved-oxygen concentration is greater than the 90th percentile value (7.3 mg/L, appendix 3). Where dissolved-oxygen concentration is less than the 25th percentile, which corresponds to less than 1.0 mg/L (appendix 3), the average observed arsenic concentration class is 4.4 (table 5). The average observed arsenic concentration class is also 4.4 (table 5) when iron or manganese concentrations are high, greater than their 90th percentile values (160 µg/L and 270 µg/L, respectively; appendix 3), and likely indicative of reducing conditions.

The difference in the average observed arsenic concentration class between the 10th percentile values and 90th percentile values is greater for pH (1.9 classes) than the difference for dissolved oxygen (1.6), iron (0.6), and manganese (0.6; table 5). In addition, the correlation as measured by Kendall's tau is much stronger for pH (0.19) than for dissolved oxygen (–0.15), iron (0.03), and manganese (0.04; table 5). This greater difference in average observed arsenic concentration class and stronger correlations, as well as a greater standardized importance value for pH (table 12), indicates that reductive dissolution of iron or manganese oxides with subsequent release of arsenic to groundwater is not, at the regional scale, as prevalent a mechanism for releasing arsenic to groundwater as sorption processes influenced by pH. This, however, does not necessarily preclude that locally, in certain areas, reducing conditions could enhance the release of arsenic into groundwater.

In addition to pH, the presence of orthophosphate (ranked 4 in importance out of 85; table 12) has been shown to influence the quantity of arsenate adsorbed to sediments because orthophosphate and arsenate are both oxyanions that compete for the same adsorption sites. In some instances, depending on soil type, orthophosphate and arsenic concentrations, and equilibration time, orthophosphate can replace arsenic or inhibit its adsorption to soil substrates (Peryea, 1991; Welch and others, 2000; Hongshao and Stanforth, 2001; Zeng and others, 2008), thereby releasing or keeping arsenic in solution. Where orthophosphate concentrations are greater than the 90th percentile (0.14 mg/L), the average observed arsenic concentration class was also high, 4.7. Unlike the other variables describing general surficial soil characteristics, the prediction and confirmatory classifiers were relatively sensitive to soil organic matter (table 12). Although a causal relation between fulvic and humic acids and arsenic sorption mechanisms is not yet fully understood, in some cases organic matter has been proposed to enhance the desorption of arsenic from aquifer matrices (Smedley and Kinniburgh, 2002).

Some researchers have suggested that arsenic can be leached from aquifer substrates by bicarbonate/carbonate (Kim and others, 2000), and, therefore, any changes in alkalinity could affect arsenic concentrations. The confirmatory classifier accuracy was not sensitive to alkalinity, and the correlation between alkalinity and arsenic concentration class was not significant (table 5). Therefore, on a regional basis, it appears unlikely that a change in the carbonate-system is influencing arsenic occurrence.

The accuracy of the confirmatory classifier was relatively sensitive to chloride (ranked 17 out of 85), which could be an indicator of evaporative effects or of geothermal waters; however, there was little correlation between chloride concentration and arsenic concentration class (+0.09, table 5). Welch and others (2000) suggest that both evaporative concentration of shallow groundwater and limited adsorption of arsenic on aquifer substrates, contribute to high arsenic concentrations in groundwater in relatively arid areas.

Effects of Selected Natural and Human-Related Factors on Predicted Arsenic Concentration

The conceptual model developed by Bexfield and others (2011) qualitatively examined the occurrence, fate, and transport of arsenic in the 16 case-study basins in the SWPA study area. As previously discussed, geologic sources, flow-path characteristics, and geochemical conditions are all important to consider when evaluating arsenic in basin-fill aquifers (appendix 9; table 12). In addition, Smedley and Kinniburgh (2002) proposed two governing factors involved in generating high-arsenic groundwater on a regional scale: (1) favorable geochemical conditions necessary to release arsenic from aquifer substrates and (2) lack of adequate flushing of groundwater from the aquifer system. Groundwater flushing, as it is considered for the SWPA study area, occurs where groundwater discharge mechanisms transport arsenic out of a basin through basin-fill deposits or consolidated rocks to adjacent basins, streams flowing into adjacent basins, or to the ocean. Factors that influence these two mechanisms include climate (aridity), geologic structure, and aquifer properties. Aquifer properties that enhance the movement of groundwater into and through the system include a relatively shallow unsaturated zone (with permeable soil characteristics), high hydraulic conductivity, high hydraulic gradients (vertical head pressures and horizontal flow), and the distribution of these characteristics along groundwater flow paths (Bexfield and others, 2011). To illustrate the effect of some of these factors on arsenic concentrations in the SWPA basin-fill aquifers, predicted arsenic concentrations were evaluated by categorizing each model grid cell by the following three explanatory variables:

- Land-surface elevation percentile (greater than 75 percent to indicate basin margin, 10 to 75 percent to indicate the middle parts of the basin, or less than 10 percent to indicate the basin lowlands).

- Predominant geologic characteristics (assignment made such that greater than 50 percent of the bedrock in the hydrogeologic area surrounding the basin was characterized as a particular type of rock: volcanic, crystalline, carbonate, or clastic sedimentary).
- Contributing area recharge rates (low recharge conditions were considered less than 1.7 in/yr and high recharge conditions were considered equal to or greater than 1.7 in/yr because 1.7 in/yr represents the 75th percentile for this explanatory variable; prediction classifier dataset, appendix 3).

Classification of the model grid cells on the basis of these criteria resulted in 24 possible categories representative of four different geologic settings, two recharge conditions, and three possible locations along a generalized flow path (margin, middle, and lowlands). The category representing clastic sedimentary bedrock, low recharge conditions, and lowland basin-fill environments only covered 83 mi^2 and, therefore, was excluded from the analysis; the remaining 23 categories represent 250 to 18,722 mi^2 of basin-fill aquifers throughout the SWPA. For each of the 23 categories, the distribution of predicted arsenic concentrations was determined and illustrated in figure 14. Land use was not considered in these generalized models because the relations between land use and predicted arsenic concentrations were considered spurious, with inherent characteristics of these basins likely the underlying reason for the correlations (see text box "Have human activities in agricultural and urban areas affected arsenic concentrations across the Southwest?").

As discussed previously, different rock types can contain varying amounts of arsenic. This information was used, in the most general sense, by comparing the predicted concentrations of arsenic in areas surrounded by different geologic units. Each model grid cell was assessed and categorized for geologic characteristics: volcanic, crystalline, carbonate, or clastic sedimentary rocks, where volcanic rocks were represented by the combined data for intermediate composition, mafic, felsic/silicic, and undifferentiated volcanic rocks. Most of the 54,854 model grid cells (190,612 mi^2) representing basin-fill aquifers were classified as being predominantly influenced by volcanic (30 percent) or by crystalline (28 percent) rocks. Carbonate rocks influenced 11 percent and clastic sedimentary rocks influenced 5 percent. About 26 percent of the basin-fill aquifer area remained unclassified because of a lack of a predominant rock type.

The results of the prediction classifier for arsenic within the SWPA study area indicated that the distribution of predicted arsenic concentrations is largely influenced by the contributing recharge rate in a given area (table 12). The majority of the classified cells in the SWPA study area (75 percent) were representative of low recharge conditions (less than 1.7 in/yr). Most recharge in the study area occurs near mountain fronts as precipitation infiltrating through rock and soil.

Land-surface elevation percentile was chosen to represent the relative position of a cell along a generalized groundwater flow path within the basin-fill aquifer environment. Cells with higher relative elevation are located in areas with generally greater recharge rates (nearer upper basin margins) than cells at lower relative elevation, located in the basin lowlands where discharge is likely to occur through evapotranspiration, seepage to stream channels, or outflow to adjacent downgradient basins. It is recognized that other points of discharge can occur within a basin in areas other than topographically low points, such as through springs; however, for the purposes of evaluating predicted arsenic concentrations along the generalized flow path presented here, such discharge points were not considered.

Generally, predicted arsenic concentrations are higher in areas surrounded by volcanic and crystalline rocks than areas surrounded by carbonate and clastic sedimentary rocks (fig. 14). Although the distribution of predicted arsenic concentrations varies among the various geologic settings and recharge conditions, arsenic concentrations generally increase from the upper basin margins to the basin lowlands (fig. 14). For description of this, see text box "Arsenic accumulation underneath playas." In areas of relatively low recharge where volcanic or crystalline rocks predominate, the average percent of predicted arsenic concentrations that are equal to or greater than 10 µg/L increases from about 48 percent near the upper basin margins to 56 percent in the middle part of flow path and, eventually, to 65 percent in the basin lowlands. Under similar recharge conditions in areas predominately composed of carbonate and clastic sedimentary rocks, the average percentage of predicted arsenic concentrations equal to or greater than 10 µg/L is generally lower than those described above. There is still an increase, however, with increasing distance from upper basin margin areas from about 30 percent at the upper basin margins to about 39 percent in the middle part of the flow path and, eventually, to about 80 percent in the lowland areas primarily influenced by carbonate geology (fig. 14). About 72 percent of the data included in the equal to or greater than 25 µg/L concentration class for the carbonate, low recharge, lowlands category comes from Utah, and approximately three quarters of these data are representative of the Great Salt Lake Desert, the terminus of the groundwater flow system for that area. Although the data representing this category and concentration class are predominately from a terminal area, this information emphasizes the importance of flow path and closed conditions with respect to arsenic enrichment.

Under relatively high recharge conditions (equal to or greater than 1.7 in/yr), the percent of predicted arsenic concentrations that equal or exceed 10 µg/L, although increasing along the generalized flow path, is less than that determined under low recharge conditions (fig. 14). In areas surrounded predominately by volcanic rocks, the fraction of arsenic concentrations equal to or greater than 10 µg/L that move downgradient from upper basin margins to the middle part of flow path, and eventually move to basin lowlands, is 13, 22, and 44 percent, respectively. A similar relation in predicted arsenic concentrations equal to or greater than 10 µg/L was

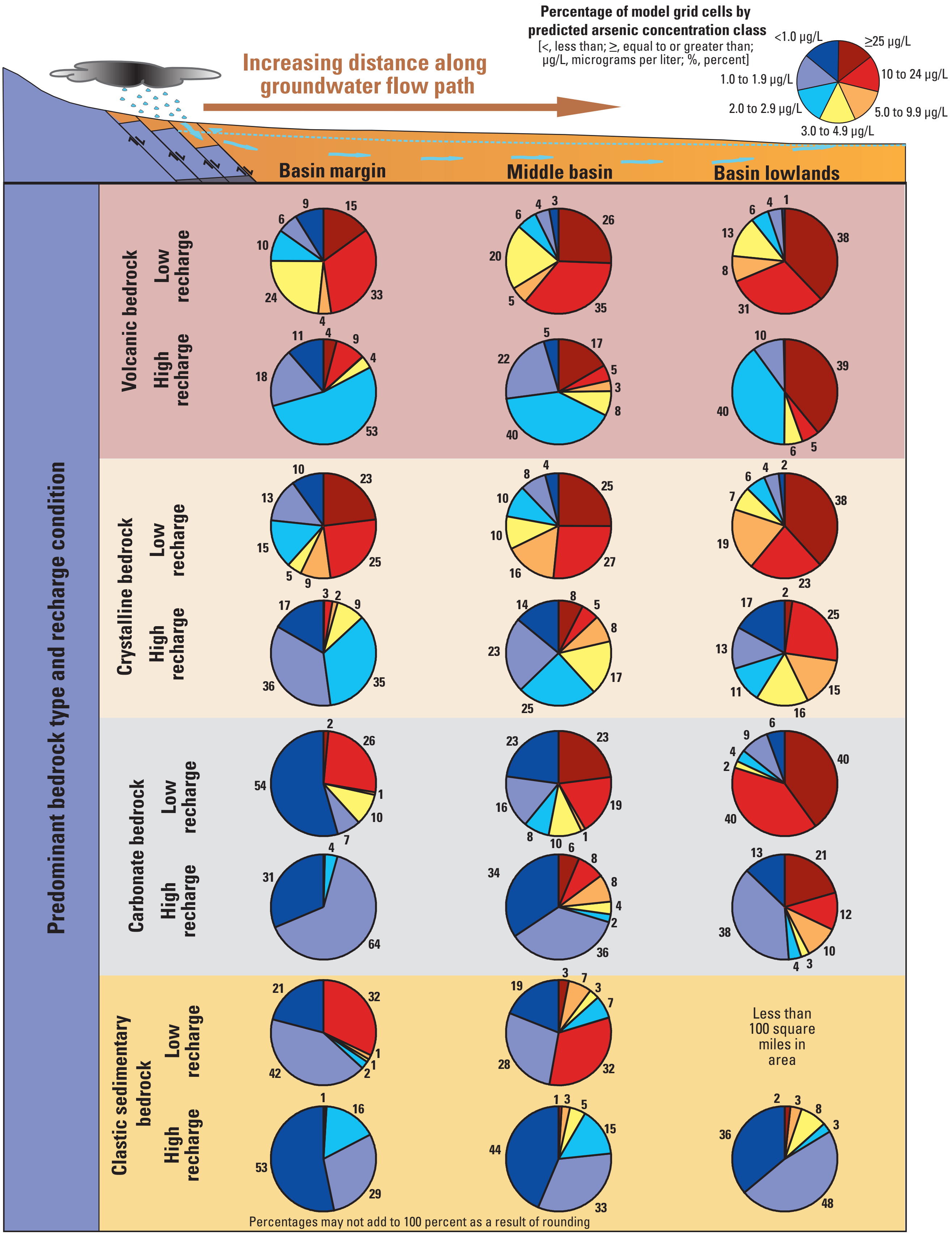


Figure 14. Distribution of predicted arsenic concentrations as a function of distance along generalized groundwater flow path, geologic setting, and recharge characteristics for basin-fill aquifers of the Southwest Principal Aquifers study area.

observed with position in flow path for areas with crystalline rocks, but with overall lower fractions of concentrations above the standard (3, 13, and 27 percent, respectively). Under the high recharge condition, areas with carbonate rocks appear to show a similar gradient to that shown for areas with crystalline rocks, where arsenic concentrations equal to or greater than 10 µg/L increase from upper basin margins (less than 1 percent) to basin lowlands (33 percent). Basin-fill aquifers influenced predominately by clastic sedimentary bedrock did not show as much of an increase in arsenic concentration along the general flow path; however, data availability was limited for this category.

The simplified evaluations of distribution patterns in arsenic concentration class along flow paths (fig. 14) presented for each combination of geologic and recharge scenarios covering at least 100 mi^2 of SWPA basin-fill area do not account for heterogeneity in geologic composition or geochemical conditions that occur in nature. It is important to consider that arsenic generally does not behave conservatively, and there will be areas of enrichment and attenuation along any particular flow path. Although the relations described previously are not distinctly apparent along flow paths in all basins, in a general sense, these relations can be important for consideration by water-resource managers and can be used in conjunction with other research concerning the distribution of arsenic in basin-fill aquifers of the SWPA study area.

Arsenic Summary and Vulnerability Assessment

The random forest classifiers provided information on the spatial distribution of arsenic within the upper 200 ft of basin-fill aquifers (190,612 mi^2) and allowed for a general assessment of the vulnerability of aquifers throughout the SWPA study area to arsenic enrichment. The classifiers were effectively trained to the relations between observed arsenic concentrations and factors important to the occurrence of arsenic, and this enabled the extrapolation of predicted arsenic concentrations from areas where concentrations were measured and known into areas where data were unavailable and unknown. The ability of the model to predict arsenic concentrations across the study area within plus or minus one concentration class was 70.2 percent; the relatively low prediction accuracy for actual concentration class results largely from natural spatial variability and the use of seven concentration classes. The use of seven concentration classes, however, provided a somewhat detailed characterization of the distribution of arsenic concentrations throughout the SWPA within reasonable accuracy for such a large area. Analysis of the misclassifications indicated the model was unbiased spatially and unbiased across the distribution of values for the explanatory variables.

While the training observations indicate arsenic concentrations equal or exceed 10 µg/L in 24.9 percent of the groundwater samples, use of the prediction classifier to extrapolate concentrations across the SWPA study area revealed that 42.7 percent of the area underlain by basin-fill aquifers exceeds this concentration, and 50.2 percent of the area has concentrations less than 5.0 µg/L:

Arsenic concentration class, µg/L	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10–24	≥25
Percent training observations in concentration class, generally representing part of aquifers with groundwater development, from table 13 (n=4,162)	15.2	12.7	13.6	16.3	17.3	14.2	10.7
Percent of basin-fill aquifer area in Southwest Principal Aquifer study area predicted for concentration class, from table 14 (190,612 miles^2)	11.1	15.6	10.8	12.7	7.0	25.8	16.9

Such differences in the distributions of observed and predicted arsenic concentrations are expected and result from the fact that the prediction dataset represents the full extent of basin-fill aquifers in the SWPA study area, whereas the training dataset represents a subset of those aquifers where observations were available, and each dataset has somewhat different but overlapping distributions of source and aquifer-susceptibility variables that affect arsenic in groundwater.

The largest area where arsenic concentrations in groundwater were predicted to be equal to or greater than the drinking-water standard of 10 µg/L was in the Basin and Range basin-fill aquifer. Spatially, the Basin and Range basin-fill aquifers compose about 73 percent of the regional study area, and much of the area is undeveloped and is largely unused or used as open rangeland. Distribution patterns with aquifer-penetration depth obtained from the random forest classifiers support the conceptual model findings indicating that arsenic concentrations exceeding 10 µg/L can occur at various aquifer-penetration depths throughout the SWPA (Bexfield and others, 2011).

Within a given basin, predicted concentrations generally increased along groundwater flow paths from the upper basin margins to the basin lowlands, with greater concentrations associated with basin-fill sediments derived from surrounding mountains predominately composed of volcanic or crystalline bedrock. Basins surrounded by carbonate rocks characteristically showed lower concentrations of arsenic in the basin-fill aquifers. Although areas developed for agricultural or urban use had lower arsenic concentrations, this appears to be largely an artifact of the hydrogeologic nature of the areas developed. Generally, the more developed areas have higher rates of recharge and probably greater flushing of solutes out of the basin either to rivers or to the ocean. In contrast, basins with lower rates of recharge, and likely correspondingly lower flushing of solutes, tend to be less developed and generally are located in areas with relatively high potential evapotranspiration rates.

A further understanding of conditions that render the basin-fill aquifers in the SWPA study area vulnerable to arsenic enrichment was gained from an analysis of the correlations between the predicted concentrations and the explanatory variables (table 12), as well as correlations between observed arsenic and other constituent concentrations (table 5) in the training dataset. These univariate correlations indicate that higher arsenic concentrations are more likely to be found in areas where the following conditions exist:

- Basins are surrounded by mafic volcanic bedrock, felsic/silicic volcanic bedrock, or crystalline bedrock.
- Groundwater flow paths are long.
- There is a general lack of groundwater flushing as indicated by low rates of natural recharge, high potential evapotranspiration rates, and minimal or altogether absent groundwater flow out of the basin.
- Geochemical conditions favor the release of arsenic from aquifer substrates to surrounding groundwater, especially where pH is basic (above 8.0) and, in some localized areas, where reducing conditions prevail or where competitive adsorption with orthophosphate could be occurring.

The sources, aquifer susceptibility, and geochemical conditions associated with the vulnerability of basin-fill aquifers to arsenic enrichment, as determined by the random forest classifier results, are consistent with the conceptual model of natural and human-related factors that affect arsenic concentrations described by Bexfield and others (2011).

Summary and Conclusions

Human-health concerns and economic considerations associated with meeting drinking-water standards motivated a study of the vulnerability of groundwater to nitrate and arsenic contamination in basin-fill aquifers in the SWPA study area. Statistical models that used the random forest classifier algorithm were developed to predict concentrations of these two contaminants across basin-fill aquifers in the study area. Analysis of the classifiers indicated (1) good agreement with conceptual models for the natural and human-related factors affecting nitrate and arsenic and (2) that the classifier predictions were unbiased and reasonably precise, especially in consideration of the inherent spatial variability exhibited by the contaminants. Classifier predictions indicate that only a small percentage of the area of basin-fill aquifers has concentrations of nitrate (2.4 percent, or 4,530 mi^2) equal to or greater than the 10 mg/L U.S. Environmental Protection Agency drinking-water standard. For arsenic, however, a considerable percentage (42.7 percent, or 81,430 mi^2) of the area of the basin-fill aquifers equals or exceeds the drinking-water standard of 10 µg/L.

Areas predicted to exceed the nitrate drinking-water standard are generally developed, especially for irrigated agriculture, but are also in more urbanized locations such as Modesto, Phoenix, and suburbs east of Los Angeles. While population densities are much smaller in agricultural than in urban areas, high nitrate concentrations underlying agricultural landscapes could be problematic with respect to public supply for large populations if those lands are eventually converted to urban uses. For the areas affected by high nitrate concentrations in agricultural land use settings, fertilizer and livestock manure are significant sources and are typically mitigated with best management practices. Large tracks of land in the Sonoran Desert with nitrate concentrations between 2.0 and 5.0 mg/L, however, appear to be affected by natural nitrogen fixation by legumes and present a more challenging condition for nitrogen management.

Arsenic in groundwater is derived primarily from natural sources, namely the basin-fill sediments and the parent bedrock from which the sediments were derived. While most of the area that is predicted to have arsenic concentrations equal to or greater than the current drinking-water standard of 10 µg/L is sparsely populated, major population centers are not necessarily unaffected. Areas within or adjacent to the metropolitan areas of Albuquerque, Bakersfield, Phoenix, Reno, Sacramento, Salt Lake City, and Stockton have arsenic concentrations above the drinking-water standard, which could affect future groundwater development as these cities grow.

References Cited

Anderson, M.A., Ferguson, J.F., and Gavis, J., 1976, Arsenate adsorption on amorphous aluminum hydroxide: Journal of Colloid and Interface Science, v. 54, no. 3, p. 391–399.

Anning, D.W., Bauch, N.J., Gerner, S.J., Flynn, M.E., Hamlin, S.N., Moore, S.J., Schaefer, D.H., Anderholm, S.K., and Spangler, L.E., 2007, Dissolved solids in basin-fill aquifers and streams in the Southwestern United States: U.S. Geological Survey Scientific Investigations Report 2006–5315, 336 p., available at URL *http://pubs.usgs.gov/sir/2006/5315/.*

Anning, D.W., and Konieczki, A.D., 2005, Classification of hydrogeologic areas and hydrogeologic flow systems in the Basin and Range Physiographic Province, Southwestern United States: U.S. Geological Survey Professional Paper 1702, 37 p., available at URL *http://pubs.usgs.gov/pp/2005/pp1702/.*

Belnap, Jayne, Webb, R.H., Miller, D.M., Miller, M.E., DeFalco, L.A., Medica, P.A., Brooks, M.L., Esque, T.C., and Bedford, D.R., 2008, Monitoring ecosystem quality and function in arid settings of the Mojave Desert: U.S. Geological Survey Scientific Investigations Report 2008–5064, 119 p., available at URL *http://pubs.usgs.gov/sir/2008/5064/.*

Belnap, Jayne, Welter, J.R., Grimm, N.B., Barger, Nichole, and Ludwig, J.A., 2005, Linkages between microbial and hydrologic processes in arid and semiarid watersheds: Ecology, v. 86, no. 2, p. 298–307.

Bexfield, L. M., and Plummer, L. N., 2003, Occurrence of arsenic in ground water of the Middle Rio Grande Basin, central New Mexico, chap. 11 in Welch, A. H., and Stollenwerk, K. G., eds., Arsenic in Ground Water: Geochemistry and Occurrence, Kluwer Academic Publishers, p. 295–327.

Bexfield, L.M., Thiros, S.A., Anning, D.W., Huntington, J.M., and McKinney, T.S., 2011, Effects of natural and human factors on groundwater quality of basin-fill aquifers in the Southwestern United States—Conceptual models for selected contaminants: U.S. Geological Survey Scientific Investigations Report 2011–5020, 90 p., available at URL *http://pubs.usgs.gov/sir/2011/5020/pdf/sir20115020.pdf*.

Breiman, Leo, 2001, Random Forests: Machine Learning v. 45, no. 1, p. 5–32, available at URL *http://www.springerlink.com/content/u0p06167n6173512/*.

Breiman, Leo, and Cutler, Adele, 2010, Random Forests, accessed March 17, 2010 at *http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm*.

Breiman, Leo, Friedman, Jerome, Stone, Charles, and Olshen, R.A., 1984, Classification and Regression Trees: New York, Chapman and Hill, 368 p.

Brown, D.E., Unmack, P.J., and Brennan, T.C., 2007, Digitized map of biotic communities for plotting and comparing distributions of North American animals: The Southwestern Naturalist, 52: p. 610–616, digital data available at URL *http://www.peter.unmack.net/biotic/*.

Burow, K.R., Shelton, J.L., and Dubrovsky, N.M., 2008, Regional nitrate and pesticide trends in the eastern San Joaquin Valley, California: Journal of Environmental Quality, v. 37, special submissions, p. 249–268.

Busbee, M.W., Kocar, B.D., and Benner, S.G., 2009, Irrigation produces elevated arsenic in the underlying groundwater of a semi-arid basin in southwestern Idaho: Applied Geochemistry, v. 24, p. 843–859.

Cherry, J.A., Shaikh, A.U., Tallman, D.E., and Nicholson, R.V., 1979, Arsenic species as an indicator of redox conditions in groundwater: Journal of Hydrology, v. 43, p. 373–392.

Davis, J.A., and Leckie, J.O., 1978, Surface ionization and complexation at the oxide/water interact: II. Surface properties of amorphous iron oxyhydroxide and adsorption of metal ions: Journal of Colloid and Interface Science, v. 67, no. 1, p. 90–107.

Driehaus, W., Seith, R., and Jekel, M., 1995, Oxidation of arsenate(III) with manganese oxides in water treatment: Water Research, v. 29, no. 1, p. 297–305.

Dubrovsky, N.M., Burow, K.R., Clark, G.M., Gronberg, J.M., Hamilton P.A., Hitt, K.J., Mueller, D.K., Munn, M.D., Nolan, B.T., Puckett, L.J., Rupert, M.G., Short, T.M., Spahr, N.E., Sprague, L.A., and Wilber, W.G., 2010, The quality of our Nation's waters—Nutrients in the Nation's streams and groundwater, 1992–2004: U.S. Geological Survey Circular 1350, 174 p., additional information available at URL *http://water.usgs.gov/nawqa/nutrients/pubs/circ1350*.

Edmonds, R.J., and Gellenbeck, D.J., 2002, Ground-water quality in the West Salt River Valley, Arizona, 1996–98—Relations to hydrogeology, water use, and land use: U.S. Geological Survey Water Resources Investigations Report 01–4126, 60 p., available at URL *http://az.water.usgs.gov/pubs/WRIR01-4126intro.html*.

Eskew, D.L., and Ting, I.P., 1978, Nitrogen fixation by legumes and blue-green algal-lichen crusts in a Colorado Desert environment: American Journal of Botany, v. 65, no. 8, p. 850–856.

Faunt, C.C., ed., 2009, Groundwater availability of the Central Valley Aquifer, California: U.S. Geological Survey Professional Paper 1766, 225 p., available at URL *http://pubs.usgs.gov/pp/1766/*.

Flint, A.L., and Flint, L.E., 2007, Application of the basin characterization model to estimate in-place recharge and runoff potential in the Basin and Range Carbonate-Rock Aquifer System, White Pine County, Nevada, and adjacent areas in Nevada and Utah: U.S. Geological Survey Scientific Investigations Report 2007–5099, 21 p., available at URL *http://pubs.usgs.gov/sir/2007/5099/*.

Focazio, M.J., Reilly, T.E., Rupert, M.G., and Helsel, D.R., 2002, Assessing ground-water vulnerability to contamination—Providing scientifically defensible information for decision makers: U.S. Geological Survey Circular 1224, 33 p., available at URL *http://pubs.usgs.gov/circ/2002/circ1224/*.

Frey, M.M., and Edwards, M.A., 1997, Surveying arsenic occurrence: Journal of the American Water Works Association, v. 89, p. 105–117.

Fujii, Roger, and Swain,W.C., 1995, Areal distribution of selected trace elements, salinity, and major ions in shallow ground water, Tulare Basin, Southern San Joaquin Valley, California: U.S. Geological Survey Water-Resources Investigations Report 95–4048, 67 p.

Galloway, J.N., Thornton, J.D., Norton, S.A., Valchok, H.L., and McLean, R.A.N., 1982, Trace metals in atmospheric deposition: A review and assessment: Atmospheric Environment, v. 16, p. 1677–1700.

Gashler, Mike, Giraud-Carrier, Christophe, and Martinez, Tony, 2008, Decision tree ensemble: small heterogeneous is better than large homogeneous [abs.] in 7th International Conference on Machine Learning and Applications, San Diego, California, 2008, Institute of Electrical and Electronics Engineers Computer Society, p. 900–905.

Green, G.N., 1992, The digital geologic map of Colorado: U.S. Geological Survey Open-File Report 92–0507, available in ArcINFO format at *http://pubs.usgs.gov/of/1992/ofr-92-0507/*.

Green, G.N., and Jones, G.E., 1997, The digital geologic map of New Mexico: U.S. Geological Survey Open-File Report 97–0052, 1:500,000 scale, available in ArcINFO Format at *http://pubs.usgs.gov/of/1997/ofr-97-0052/new_mex.htm*.

Hamilton, P.A., Denver, J.M., Phillips, P.J., and Shedlock, R.J., 1993, Water-quality assessment of the Delmarva Peninsula, Delaware, Maryland, and Virginia—Effects of agricultural activities on, and distribution of, nitrate and other inorganic constituents in the surficial aquifer: U.S. Geological Survey Open-File Report 93–40, 95 p., available at URL *http://pubs.er.usgs.gov/publication/ofr9340*.

Hanson, G.C., Groffman, P.M., and Gold, A.J., 1994, Denitrification in riparian wetlands receiving high and low groundwater nitrate inputs: Journal of Environmental Quality, v. 23, p. 917–922.

Harrill, J.R., and Hines, L.B., 1995, Estimated natural groundwater recharge, discharge, and budget for the Dixie Valley Area, west-central Nevada: U.S. Geological Survey Water-Resources Investigations Report 95–4052, 12 p.

Harrill, J.R., and Prudic, D.E., 1998, Aquifer systems in the Great Basin region of Nevada, Utah, and adjacent States: Summary Report U.S. Geological Survey Professional Paper 1409–A, 66 p., available at URL *http://pubs.er.usgs.gov/publication/pp1409A*.

Hastie, Trevor, Tibshirani, Robert, and Friedman, Jerome, 2001, The elements of statistical learning; data mining, inference, and prediction: New York, Springer-Verlag, 533 p.

Heit, M., Tan, Y., Klusek, C., and Burke, J.C., 1981, Anthropogenic trace elements and polycyclic aromatic hydrocarbons levels in sediment cores from two lakes in the Adirondack Acid Lake Region: Water, Air, and Soil Pollution, v. 15, p. 441–464.

Hinkle, S.R., Kauffman, L.J., Thomas, M.A., Brown, C.J., McCarthy, K.A., Eberts, S.M., Rosen, M.R., and Katz, B,G., 2009, Combining particle-tracking and geochemical data to assess public supply well vulnerability to arsenic and uranium: Journal of Hydrology, v. 376, p. 132–142.

Hirschberg, D.M., and Pitts, S.G., 2000, Digital geologic map of Arizona: a digital database from the 1983 printing of Wilson, Moore and Cooper: U.S. Geological Survey Open-File Report 2000–409, 1:500,000, available at URL *http://geopubs.wr.usgs.gov/open-file/of00-409/*.

Hitt, K.J., 1997, Unpublished digital data for sewage disposal method for census block groups provided electronically, data provided is an extraction from U.S. Bureau of the Census, 1992, Census of population and housing, 1990—Summary tape file 3*A* on CD-ROM (machine-readable data file): Washington, D.C., The Bureau of the Census.

Homer, Collin, Huang, Chengquan, Yang, Limin, Wylie, Bruce, and Coan, Michael, 2004, Development of a 2001 National Landcover Database for the United States: Photogrammetric Engineering and Remote Sensing, v. 70, no. 7, July 2004, p. 829–840, available at URL *http://www.mrlc.gov/pdf/July_PERS.pdf*.

Hongshao, Zhao, and Stanforth, Robert, 2001, Competitive adsorption of phosphate and arsenate on goethite: Environmental Science and Technology, v. 35, p. 4753–4757.

Izbicki, J.A., Stamos, C.L., Metzger, L.F., Halford, K.J., Kulp, T.R., and Bennett, G.L., 2008, Source, distribution, and management of arsenic in water from wells, eastern San Joaquin ground-water subbasin, California: U.S. Geological Survey Open-File Report 2008–1272, 8 p.

Johnson, B.R., and Raines, G.L., 1996, Digital representation of the Idaho state geologic map: a contribution to the Interior Columbia River Basin Ecosystem Management Project: U.S. Geological Survey Open-File Report 95–690, scale 1:500,000, available at URL *http://pubs.usgs.gov/of/1995/of95-690/*.

Jurgens, B.C., Fram, M.S., Belitz, Kenneth, Burow, K.R., and Landon, M.K., 2009, Effects of Ground-Water Development on Uranium: Central Valley, California, USA: Ground Water, Published online: September 28, 2009, DOI: 10.1111/j.1745-6584.2009.00635.x, available at URL *http://onlinelibrary.wiley.com/doi/10.1111/j.1745-6584.2009.00635.x/full*.

Kim, Myoung-Jin, Nriagu, Jerome, Haack, Sheridan, 2000, Carbonate ions and arsenic dissolution by groundwater: Environmental Science and Technology, v. 34, p. 3094–3100.

Kucks, R.P., 2005, Terrestrial radioactivity and gamma-ray exposure in the United States and Canada: gridded geographic images: U.S. Geological Survey Mineral Resources On-Line Spatial Data, available at URL *http://tin.er.usgs.gov/metadata/narad.faq.html*.

Laio, Francesco, Tamea, Stefania, Ridolfi, Luca, D'Odorico, Paolo, and Rodriguez-Iturbe, Ignacio, 2009, Ecohydrology of groundwater-dependent ecosystems: 1. Stochastic water table dynamics: Water Resources Research, v. 45, W05419, 13 p.

Landon, M.K., Jurgens, B.C., Katz, B.G., Eberts, S.M., Burrow, K.R., and Crandall, C.A., 2010, Depth-dependent sampling to identify short-circuiting pathways to public-supply wells in multiple aquifer settings in the United States: Hydrogeology Journal, v. 18, no. 3, p. 577–593, DOI: 10.1007/s10040-009-0531-2, 17 p.

Lapham, W.W., Hamilton, P.A., and Myers, D.N., 2005, National Water-Quality Assessment Program—Cycle II regional assessments of aquifers: U.S. Geological Survey Fact Sheet 2005–3013, available at URL *http://pubs.usgs.gov/fs/2005/3013/*.

Lopes, T.J., 2006, Quality of Nevada's aquifers and their susceptibility to contamination, 1990–2004: U.S. Geological Survey Scientific Investigations Report 2006–5127, 52 p., available at URL *http://pubs.usgs.gov/sir/2006/5127/*.

Mankhemthong, N., Oppliger, G.L., and Aslett, Z., 2008, Structural localization of two low temperature geothermal systems within the gravity defined linkage between Dixie Valley and Fairview Valley, Nevada, USA: Geothermal Resources Council Transactions, v. 32, p. 291–295.

Maupin, M.A., and Arnold, T.L., 2010, Estimates for self-supplied domestic withdrawals and population served for selected principal aquifers, calendar year 2005: U.S. Geological Survey Open-File Report 2010–1223, 10 p., available at URL *http://pubs.usgs.gov/of/2010/1223/*.

Maupin, M.A., and Barber, N.L., 2005, Estimated withdrawals from principal aquifers in the United States, 2000: U.S. Geological Survey Circular 1279, 46 p., available at URL *http://pubs.usgs.gov/circ/2005/1279/*.

Maxey, G.B., and Eakin, T.E., 1949, Groundwater in the White River Valley, White Pine, Nye, and Lincoln counties, Nevada: State of Nevada, Office of the State Engineer, Water Resources Bulletin No. 8, 59 p.

McCoy, J., Johnston, K., Kopp, S., Borup, B., and Willison, J., 2001, Using ArcGIS TM Spatial Analyst: Redlands, California, Environmental Systems Research Institute, 232 p.

McKinney, T.S., and Anning, D.W., 2009, Geospatial data to support analysis of water-quality conditions in basin-fill aquifers in the southwestern United States: U.S. Geological Survey Scientific Investigations Report 2008–5239, 16 p., available at URL *http://pubs.usgs.gov/sir/2008/5239/*.

McMahon, P.B., and Chappelle, F.H., 2007, Redox processes and water quality of selected principal aquifer systems: Ground Water, v. 46, issue 2, p. 259–271.

Miller, J.A., 1999, Ground water atlas of the United States: U.S. Geological Survey Hydrologic Atlas 730, available at URL *http://pubs.usgs.gov/ha/ha730/index.html*.

Morrison, R.B., 1991, Quaternary stratigraphic, hydrologic, and climatic history of the Great Basin, with emphasis on Lakes Lahontan, Bonneville, and Tecopa, *in* Morrison, R.B., ed., Quarternary Nonglacial Geology: Conterminous U.S.: Boulder, Colorado, Geological Society of America, p. 283–320.

Mueller, D.K., Hamilton, P.A., Helsel, D.R., Hitt, K.J., and Ruddy, B.C., 1995, Nutrients in groundwater and surface water of the United States—an analysis of data through 1992: U.S. Geological Survey Water-Resources Investigations Report 95–4031, 74 p.

Mueller, D.K., and Titus, C.J., 2005, Quality of nutrient data from streams and ground water sampled during water years 1992–2001: U.S. Geological Survey Scientific Investigations Report 2005–5106, 27 p., available at URL *http://pubs.usgs.gov/sir/2005/5106/pdf/sir2005-5106.pdf*.

National Atmospheric Deposition Program, 2010, National Trends Network Data, available at URL *http://nadp.sws.uiuc.edu/data/*.

Natural Resources Conservation Service, 2010, U.S. General Soil Map (STATSGO2) for the United States: United States Department of Agriculture, accessed September 2010 at *http://soildatamart.nrcs.usda.gov*.

National Research Council, 2001, Arsenic in drinking water—2001 update: National Academy Press, Washington D.C., 244 p.

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.

Nolan, B.T., and Hitt, K.J., 2006, Vulnerability of shallow groundwater and drinking-water wells to nitrate in the United States: Environmental Science and Technology, v. 40, no. 24, p. 7834–7840.

Nriagu, J.O., 1983, Arsenic enrichment in lakes near the smelters at Sudbury, Ontario: Geochimica et Cosmochimica Acta, v. 47, p. 1523–1526.

Oak Ridge National Laboratory, 2005, LandScan^{TM} Global Population Database, accessed September 2006, at *http://www.ornl.gov/landscan/*.

Paul, A.P., Seiler, R.L., Rowe, T.G., and Rosen, M.R., 2007, Effects of agriculture and urbanization on quality of shallow ground water in the arid to semiarid western United States, 1993–2004: U.S. Geological Survey Scientific Investigations Report 2007–5179, 56 p., available at URL *http://pubs.usgs.gov/sir/2007/5179/*.

Peryea, F.J., 1991, Phosphate-induced release of arsenic from soils contaminated with lead arsenate: Soil Science Society of America Journal, v. 55, p. 1301–1306.

Piramuthu, Selwyn, 2008, Input data for decision trees: Expert Systems with Applications, v. 34, no. 2, p. 1220–1226.

Planert, Michael, and Williams, J.S., 1995, Groundwater atlas of the United States, Segment 1—California and Nevada: U.S. Geological Survey Hydrologic Investigations Atlas 730–B, available online at *http://pubs.usgs.gov/ha/ha730/gwa.html.*

Priestley, C.H.B., and Taylor, R.J., 1972, On the assessment of surface heat flux and evaporation using large-scale parameters: Monthly Weather Review, v. 100, p. 81–92.

PRISM Group, 2004a, Precipitation-elevation regression on independent slopes model (PRISM): Oregon State University, precipitation data accessed on April 12, 2005, at *http://www.ocs.oregonstate.edu/prism/.*

PRISM Group, 2004b, Precipitation-elevation regression on independent slopes model (PRISM): Oregon State University, temperature data accessed on April 10, 2010, at *http://www.ocs.oregonstate.edu/prism/.*

Rahman, Tauhidur, and Uhlman, Kristine, 2009, Predicting groundwater vulnerability to nitrate in Arizona: University of Arizona College of Agriculture and Life Sciences, Water Resources Research Center WSP-TRIF Project P06-04 Final Report, accessed March 17, 2010 at *http://www.srnr.arizona.edu/nemo/review/Nitrate/.*

Ramsey, R.D., 1996, Digital compilation of geologic map of Utah *by* Hintze, L.F., Willis, G.C., Laes, D.Y.M., Sprinkle, D.A., and Brown, K.D., U.S. Geological Survey Digital Data Series, DDS–41.

Reheis, Marith, 1999, Extent of Pleistocene lakes in the western Great Basin: U.S. Geological Survey Miscellaneous Field Studies Map, MF–2323.

Robertson, F.N., 1989, Arsenic in groundwater under oxidizing conditions, Southwest United States: Environmental Geochemistry and Health, v. 11, no.3 / 4, p. 171–185.

Robertson, F.N., 1991, Geochemistry of ground water in alluvial basins of Arizona and adjacent parts of Nevada, New Mexico, and California: U.S. Geological Survey Professional Paper 1406–C, 90 p., available at URL *http://pubs.er.usgs.gov/publication/pp1406C.*

Rosen, M.R., 2003, Trends in nitrate and dissolved-solids concentrations in ground water, Carson Valley, Douglas County, Nevada, 1985–2001: U.S. Geological Survey Water-Resources Investigations Report 03–4152, 6 p.

Ruddy, B.C., Lorenz, D.L., and Mueller, D.K., 2006, County-level estimates of nutrient inputs to the land surface of the conterminous United States, 1982–2001: U.S. Geological Survey Scientific Investigations Report 2006–5012, 17 p., available at URL *http://pubs.usgs.gov/sir/2006/5012/.*

Rupert, M.G., 2003, Probability of detecting atrazine/desethyl-atrazine and elevated concentrations of nitrate in groundwater in Colorado: U.S. Geological Survey Water-Resources Investigations Report 02–4269, 35 p., available at URL *http://pubs.usgs.gov/wri/wri02-4269/.*

Saucedo, G. L., Bedford, B. R., Raines, G. L., Miller, R. J., and Wentworth, C. M., 2000, Modified from California Division of Mines and Geology, CD-ROM 2000–007, GIS data for the geologic map of California.

Schaefer, D.H., Thiros, S.A., and Rosen, M.R., 2006, Groundwater quality in the carbonate-rock aquifer of the Great Basin, Nevada and Utah, 2003: U.S. Geological Survey Scientific Investigations Report 2005–5232, 32 p., available at URL *http://pubs.usgs.gov/sir/2005/5232/.*

Schlesinger, W. H., Abrahams, A. D., Parsons, A. J., and J. Wainwright, 1999, Nutrient losses in runoff from grassland and shrubland habitats in Southern New Mexico: I. Rainfall simulation experiments: Biogeochemistry, v. 45, p. 21–34.

Schroeder, R.A., Palawski, D.U., and Skorupa, J.P., 1988, Reconnaissance investigation of water quality, bottom sediment, and biota associated with irrigation drainage in the Tulare Lake Bed area, southern San Joaquin Valley, California, 1986–87: U.S. Geological Survey Water-Resources Investigations Report 88–4001, 86 p.

Smedley, P.L., and Kinniburgh, D.G., 2002, A review of the source, behavior and distribution of arsenic in natural waters: Applied Geochemistry, v. 17, no. 5, p. 517–568.

Smith, R.A., Alexander, R.B., and Wolman, M.G., 1987, Water-quality trends in the nation’s rivers: Science, v. 235, p. 1607–1615.

Stark, J.M., 1996, Modeling the temperature response of nitrification: Biogeochemistry, v. 35, no. 3, p. 433–445.

Stollenwerk, K.G., 2003, Geochemical processes controlling transport of arsenic in groundwater: a review of adsorption, chap. 3 in Welch, A.H. and Stollenwerk, K.G., eds., Arsenic in ground water: Kluwer Academic Publishers, Boston, p.67–100.

Thiros, S.A., Bexfield, L.M., Anning, D.W., and Huntington, J.M., 2010, Conceptual understanding and groundwater quality of selected basin-fill aquifers in the Southwestern United States: U.S. Geological Professional Paper 1781, 288 p., available at URL *http://pubs.usgs.gov/pp/1781/.*

Thomas, J.M., Welch, A.H., and Dettinger, M.D., 1996, Geochemistry and isotope hydrology of representative aquifers in the Great Basin Region of Nevada, Utah, and adjacent States: U.S. Geological Survey Professional Paper 1409–C, 100 p., available at URL *http://pubs.er.usgs.gov/publication/pp1409C*.

Turner, R.M., and Bawic, W.J., 1996, Digital Map of Nevada at scale of 1:500,000: U.S. Geological Survey Digital Data Series DDS–41.

Tyler, S.W., Munoz, J.F., and Wood, W.W., 2005, The response of playa and sabkha hydraulics and mineralogy to climate forcing: Ground Water, v. 44, no. 3, p. 329–338.

U.S. Bureau of the Census, 1992, Census of population and housing, 1990—Summary tape file 3*A* on CD-ROM (machine-readable data file): Washington, D.C., U.S. Bureau of the Census.

U.S. Environmental Protection Agency, 2009, National primary drinking water regulations, accessed November 2009 at *http://www.epa.gov/safewater/contaminants/index.html*.

U.S. Geological Survey, 1999, National Elevation Dataset, accessed January 2006, at *http://ned.usgs.gov*.

U.S. Geological Survey, 2003a, Principal aquifers, *in* National Atlas of the United States of America, 1 sheet, accessed May 6, 2008 at *http://nationalatlas.gov/mld/aquifrp.html*.

U.S. Geological Survey, 2003b, National Land Cover Database (NLCD 2001), accessed February 2007 at *http://www.mrlc.gov/nlcd_multizone_map.php*.

U.S. Geological Survey, 2004, Estimated use of water in the United States: county-level data for 2000, available at URL *http://water.usgs.gov/watuse/data/2000/index.html*.

U.S. Geological Survey, 2005, Elevation derivatives for national applications (EDNA), accessed March 2006 at *http://edna.usgs.gov/*.

U.S. Geological Survey, 2008, National Land Cover Database 2001 (NLCD01) Tile 3, Southwestern United States: NLCD01_3, edition 1, U.S. Geological Survey Data Series 383C, raster digital data, available at URL *http://water.usgs.gov/GIS/metadata/usgswrd/XML/nlcd01_3.xml*.

U.S. Geological Survey, 2010a, National Water-Quality Assessment Program Nutrients National Synthesis Project, accessed March 17, 2010 at *http://water.usgs.gov/nawqa/nutrients/*.

U.S. Geological Survey, 2010b, National Water Information System (NWIS), accessed January 2010 at *http://waterdata.usgs.gov/nwis*.

Virginia, R.A., 1986, Soil development under legume tree canopies: Forest Ecology and Management, v. 16, p. 69–79.

Walker, G.W., MacLeod, N.S., Miller, R.J., Raines, G.L., and Connors, K.A., 2003, Spatial digital database for the geologic map of Oregon: U.S. Geological Survey Open-File Report 2003–67, scale 1:500,000, available at URL *http://pubs.usgs.gov/of/2003/of03-067/*.

Walvoord, M.A., Phillips, F.M., Stonestrom, D.A., Evans, R.D., Hartsough, P.C., Newman, B.D., and Striegl, R.G., 2003, A reservoir of nitrate beneath desert soils: Science, v. 302, no. 5647, p. 1021–1024.

Ward, M. H., DeKok, T.M., Levallois, Patrick, Brender, Jean, Gulis, Gabriel, Nolan, B.T., and VanDerslice, James, 2005, Workgroup Report: Drinking-water nitrate and health—recent findings and research needs: Environmental Health Perspectives, v. 113, no. 11, p. 1607–1614.

Welch, A.H., Helsel, D.R., Focazio, M.J., and Watkins, S.A., 1999, Arsenic in ground water supplies of the United States: Conference proceedings of the Third International Conference on Arsenic Exposure and Health Effects, Calderon, R.L., Abernathy, C.O., and Chapelle, W.R., eds., San Diego, CA, July 12–15, 1998, Proceedings, p. 9–17.

Welch, A.H., Lico, M.S., and Hughes, J.L., 1988, Arsenic in ground water of the western United States: Ground Water, v. 26, no. 3, p. 333–347.

Welch, A.H., Westjohn, D.B., Helsel, D.R., and Wanty, R.B., 2000, Arsenic in groundwater of the United States: occurrence and geochemistry: Ground Water, v. 38, no. 4, p. 589–604.

Wolock, D.M., 1997, STATSGO soil characteristics for the conterminous United States: U.S. Geological Survey Open-File Report 97–656.

Woolson, E.A., Moore, L., Fleischer, M., Kearney, P.C., Buck, W.B., Peoples, S.A., and Calvert, C.C., 1977, Distribution of arsenic in the environment, in Medical and biological effects of environmental pollutants—Arsenic: National Academy of Sciences, United States, Washington D.C., p. 16–79.

Zeng, Hui, Fisher, Brian, and Giammar, D.E., 2008, Individual and competitive adsorption of arsenate and phosphate to a high-surface-area iron oxide-based sorbent: Environmental Science and Technology, v. 42, p. 147–152.

Appendixes 1 and 2.

Appendix 1. Explanatory variable data and observed, predicted, and misclassification error data for observations in the training dataset for random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers in the Southwest Principal Aquifers study area.

Appendix 2. Explanatory variable data and predicted concentration class data for observations in the prediction dataset for random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.

Appendixes 1 and 2 are available as a single Microsoft Excel file. This Excel file can be downloaded at *http://pubs.usgs.gov/sir/2012/5065/*.

Appendix 3.

Appendix 3. Summary statistics for explanatory variables in the prediction and training datasets used in the random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.

[Prediction dataset contains 54,854 grid cells; training dataset contains 6,244 observations and in several cases 2 per grid cell. **Abbreviations:** acre-ft/yr, acre foot per year; F, Fahrenheit; ft, foot; in/hr, inch per hour; kg/yr, kilogram per year; km, kilometer; km^2, square kilometer; mm, millimeter; mg/L, milligram per liter; ppm, part per million; U, uranium; <, less than; µg/L, microgram per liter]

Variable	Units	Dataset	Observa-tions (percent)	Minimum	Percentile							Maximum
					5	10	25	50	75	90	95	
Nitrogen, atmospheric	kg/yr	Prediction	100	0	55	64	88	118	147	185	208	309
Nitrogen, atmospheric	kg/yr	Training	99.9	47	69	75	96	136	181	205	232	309
Nitrogen, farm fertilizer	kg/yr	Prediction	100	0	0	0	0	0	32	4,862	9,493	72,967
Nitrogen, farm fertilizer	kg/yr	Training	99.7	0	0	0	0	656	5,905	10,235	11,884	68,577
Nitrogen, non-farm fertilizer	kg/yr	Prediction	100	0	0	0	0	0	0	6	84	4,144
Nitrogen, non-farm fertilizer	kg/yr	Training	100	0	0	0	0	2	90	556	1,217	3,660
Nitrogen, confined manure	kg/yr	Prediction	100	0	0	0	0	0	11	792	2,617	42,421
Nitrogen, confined manure	kg/yr	Training	100	0	0	0	0	193	1,452	3,529	5,312	42,421
Nitrogen, unconfined manure	kg/yr	Prediction	100	0	0	0	1	38	253	669	986	4,207
Nitrogen, unconfined manure	kg/yr	Training	100	0	0	6	51	313	718	1,268	1,721	2,831
Nitrogen, total	kg/yr	Prediction	100	0	66	84	113	185	659	7,460	12,265	75,441
Nitrogen, total	kg/yr	Training	99.7	54	112	157	436	2,490	9,198	14,367	17,813	70,322
Septic/sewer ratio	dimension-less	Prediction	100	0	0.19	11.00	55.87	89.00	97.34	100	100	100
Septic/sewer ratio	dimension-less	Training	100	0	0.91	4.00	32.00	77.37	95.74	99.88	100	100
Local population	persons	Prediction	100	0	0	0	0	0	15	248	1,889	129,812
Local population	persons	Training	100	0	0	0	14	149	2,165	13,579	22,576	126,291
Local population density	persons/km^2	Prediction	100	0	0	0	0	0	2	28	210	14,424
Local population density	persons/km^2	Training	100	0	0	0	2	17	241	1,509	2,508	14,032
Basin population	persons	Prediction	100	0	9	44	529	5,181	67,437	1,900,542	3,770,045	5,105,938
Basin population	persons	Training	100	0	940	2,663	13,911	178,377	1,900,542	3,770,045	3,770,045	5,105,938
Basin population density	persons/km^2	Prediction	100	0	0	0	0	3	24	102	268	4,006
Basin population density	persons/km^2	Training	100	0	1	1	7	95	235	809	1,101	4,006
Local urban land	percent	Prediction	100	0	0	0	0	0	3.7	9.8	26.3	100
Local urban land	percent	Training	100	0	0	0.1	3.6	7.9	27.9	82.2	97.2	100
Local agricultural land	percent	Prediction	100	0	0	0	0	0	1.1	57.2	85.8	100
Local agricultural land	percent	Training	100	0	0	0	0	11.6	58.1	87.7	92.8	98.8
Basin urban land	percent	Prediction	100	0	0.1	0.2	0.5	1.3	4.8	11.8	22.6	97.3
Basin urban land	percent	Training	100	0	0.8	1.1	2.9	10.2	15.8	47.2	63.3	97.3
Basin agricultural land	percent	Prediction	100	0	0	0	0.2	1.6	9.3	37.8	62.5	71.9
Basin agricultural land	percent	Training	100	0	0	0.2	1.7	9.3	30.5	62.5	62.5	62.5

Appendix 3. Summary statistics for explanatory variables in the prediction and training datasets used in the random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Prediction dataset contains 54,854 grid cells; training dataset contains 6,244 observations and in several cases 2 per grid cell. **Abbreviations:** acre-ft/yr, acre foot per year; F, Fahrenheit; ft, foot; in/hr, inch per hour; kg/yr, kilogram per year; km, kilometer; km^2, square kilometer; mm, millimeter; mg/L, milligram per liter; ppm, part per million; U, uranium; <, less than; µg/L, microgram per liter]

Variable	Units	Dataset	Observa-tions (percent)	Minimum	Percentile							Maximum
					5	10	25	50	75	90	95	
Basin rangeland	percent	Prediction	100	0	22.5	29.8	75.0	92.1	97.9	99.2	99.5	100
Basin rangeland	percent	Training	100	1.3	16.4	20.9	26.6	63.4	89.2	95.4	97.6	100
Basin other land cover	percent	Prediction	100	0	0	0.1	0.4	2.0	4.8	10.5	15.2	81.8
Basin other land cover	percent	Training	100	0	0.1	0.3	1.0	3.3	7.2	15.2	20.2	81.8
Geology, carbonate rocks	percent	Prediction	100	0	0	0	0	3.1	23.0	52.6	65.2	97.8
Geology, carbonate rocks	percent	Training	100	0	0	0	0	0.2	8.6	29.1	52.6	94.9
Geology, crystalline rocks	percent	Prediction	100	0	0.9	2.3	8.4	20.8	55.7	70.2	84.5	99.9
Geology, crystalline rocks	percent	Training	100	0	1.3	2.7	12.8	36.1	68.0	85.8	86.3	99.9
Geology, clastic sedimentary rocks	percent	Prediction	100	0	0	0.7	4.6	12.6	22.0	35.2	50.1	98.8
Geology, clastic sedimentary rocks	percent	Training	100	0	0.3	1.0	6.1	18.3	26.0	53.4	85.4	98.8
Geology, mafic volcanic rocks	percent	Prediction	100	0	0	0	0.2	2.6	13.6	26.3	44.0	98.9
Geology, mafic volcanic rocks	percent	Training	100	0	0	0	0.3	1.9	11.8	21.5	33.0	83.9
Geology, felsic and silicic volcanic rocks	percent	Prediction	100	0	0	0	0	5.1	22.5	41.5	56.4	93.7
Geology, felsic and silicic volcanic rocks	percent	Training	100	0	0	0	0	0.2	10.2	21.0	28.5	93.7
Geology, intermediate composition volcanic rocks	percent	Prediction	100	0	0	0	0	2.3	13.2	25.9	34.1	98.9
Geology, intermediate composition volcanic rocks	percent	Training	100	0	0	0	0	0	8.6	25.3	38.8	79.9
Geology, undifferentiated volcanic rocks	percent	Prediction	100	0	0	0	0	0	3.5	9.6	17.3	44.0
Geology, undifferentiated volcanic rocks	percent	Training	100	0	0	0	0	0.4	9.1	10.0	32.4	35.6
Geology, distance to carbonate rocks	km	Prediction	100	0	3	4	9	25	69	106	131	251
Geology, distance to carbonate rocks	km	Training	100	0	6	8	17	60	104	136	154	244
Geology, distance to crystalline rocks	km	Prediction	100	0	3	3	6	12	21	33	40	94
Geology, distance to crystalline rocks	km	Training	100	0	3	4	7	13	27	40	46	78
Geology, distance to clastic sedimentary rocks	km	Prediction	100	0	3	3	7	12	21	33	40	81
Geology, distance to clastic sedimentary rocks	km	Training	100	0	3	3	7	12	21	34	40	81
Geology, distance to mafic volcanic rocks	km	Prediction	100	0	3	6	12	25	48	75	90	182

Appendix 3. Summary statistics for explanatory variables in the prediction and training datasets used in the random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Prediction dataset contains 54,854 grid cells; training dataset contains 6,244 observations and in several cases 2 per grid cell. **Abbreviations:** acre-ft/yr, acre foot per year; F, Fahrenheit; ft, foot; in/hr, inch per hour; kg/yr, kilogram per year; km, kilometer; km^2, square kilometer; mm, millimeter; mg/L, milligram per liter; ppm, part per million; U, uranium; <, less than; µg/L, microgram per liter]

Variable	Units	Dataset	Observations (percent)	Minimum	Percentile							Maximum
					5	10	25	50	75	90	95	
Geology, distance to mafic volcanic rocks	km	Training	100	0	6	8	15	28	48	69	82	174
Geology, distance to felsic and silicic volcanic rocks	km	Prediction	100	0	3	4	9	24	68	194	247	408
Geology, distance to felsic and silicic volcanic rocks	km	Training	100	0	7	9	19	49	204	264	312	404
Geology, distance to intermediate composition volcanic rocks	km	Prediction	100	0	4	6	11	27	90	195	246	413
Geology, distance to intermediate composition volcanic rocks	km	Training	100	0	6	9	18	119	203	264	315	411
Geology, distance to undifferentiated volcanic rocks	km	Prediction	100	0	8	13	28	68	166	246	276	456
Geology, distance to undifferentiated volcanic rocks	km	Training	100	3	9	13	25	51	143	239	262	389
Soil and rock equivalent uranium-238	ppm equivalent U	Prediction	97.3	–0.08	1.11	1.42	1.94	2.51	3.12	3.76	4.20	27.56
Soil and rock equivalent uranium-238	ppm equivalent U	Training	96.0	0.12	1.20	1.45	1.90	2.44	2.93	3.45	3.81	10.08
Aquifer-penetration depth	ft	Training	68.2	0	11	16	60	163	380	679	884	3,499
Well depth	ft	Training	82.9	0	20	33	120	290	570	946	1,150	7,230
Water-level depth	ft	Training	72.8	–91	2	6	18	70	179	333	435	1,045
Land-surface slope	degrees	Prediction	100	0	0.1	0.2	0.5	1.4	3.7	7.1	9.5	29.1
Land-surface slope	degrees	Training	100	0	0.1	0.1	0.3	0.7	2.0	4.6	6.7	21.9
Land-surface elevation percentile	dimension-less	Prediction	100	0	5.4	10.4	25.4	50.3	75.4	90.4	95.4	100
Land-surface elevation percentile	dimension-less	Training	100	0.2	4.4	8.0	18.3	37.1	58.9	77.8	87.5	100
Land-surface elevation	ft	Prediction	100	–82	31	86	492	1,267	1,550	1,845	2,013	3,605
Land-surface elevation	ft	Training	100	–80	13	27	98	631	1,311	1,534	1,750	2,516
Basin elevation	ft	Prediction	100	0	96	187	494	1,295	1,531	1,793	1,944	2,418
Basin elevation	ft	Training	100	19	64	96	187	713	1,396	1,595	1,765	2,418
Distance to basin margin	km	Prediction	100	0	0	3	3	6	11	17	24	50
Distance to basin margin	km	Training	100	0	3	3	6	9	15	23	30	50
Soil, seasonally high water depth	ft	Prediction	100	0	2.60	4.28	5.65	6.00	6.00	6.00	6.00	6.00
Soil, seasonally high water depth	ft	Training	100	0.03	3.53	4.17	5.20	5.91	6.00	6.00	6.00	6.00
Soil, hydric	percent	Prediction	100	0	0	0	0	0	1.0	13.4	39.1	100
Soil, hydric	percent	Training	100	0	0	0	0	0	2.0	10.0	21.6	89.6
Soil, hydrologic group A	percent	Prediction	100	0	0	0	0	2.1	14.6	42.9	63.7	100
Soil, hydrologic group A	percent	Training	100	0	0	0	0	1.4	12.0	32.6	46.0	100
Soil, hydrologic group B	percent	Prediction	100	0	0.1	4.5	19.0	39.7	65.1	83.4	95.2	100
Soil, hydrologic group B	percent	Training	100	0	2.0	6.7	22.5	45.8	71.0	89.0	100	100
Soil, hydrologic group C	percent	Prediction	100	0	0	0	1.0	8.0	23.2	41.4	55.1	100

Appendix 3. Summary statistics for explanatory variables in the prediction and training datasets used in the random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Prediction dataset contains 54,854 grid cells; training dataset contains 6,244 observations and in several cases 2 per grid cell. **Abbreviations:** acre-ft/yr, acre foot per year; F, Fahrenheit; ft, foot; in/hr, inch per hour; kg/yr, kilogram per year; km, kilometer; km^2, square kilometer; mm, millimeter; mg/L, milligram per liter; ppm, part per million; U, uranium; <, less than; µg/L, microgram per liter]

Variable	Units	Dataset	Observa-tions (percent)	Minimum	Percentile							Maximum
					5	10	25	50	75	90	95	
Soil, hydrologic group C	percent	Training	100	0	0	0	2.0	11.1	28.6	50.0	65.0	100
Soil, hydrologic group D	percent	Prediction	100	0	0	0	6.0	21.4	46.6	70.2	82.0	100
Soil, hydrologic group D	percent	Training	100	0	0	0	2.2	15.4	35.3	62.9	79.0	100
Soil, permeability	in/hr	Prediction	100	0	0.5	0.8	1.6	3.4	5.9	10.1	12.5	19.0
Soil, permeability	in/hr	Training	100	0	0.3	0.6	1.3	2.8	5.3	9.6	10.8	15.3
Soil, organic material	percent	Prediction	100	0	0.1	0.1	0.2	0.3	0.4	0.7	0.9	13.2
Soil, organic material	percent	Training	100	0	0.1	0.1	0.2	0.3	0.5	0.8	1.0	10.4
Soil, clay	percent	Prediction	100	0	6.5	9.0	12.6	17.3	24.5	31.2	34.5	55.1
Soil, clay	percent	Training	100	0.1	7.9	10.1	13.4	19.3	25.9	32.4	36.6	51.3
Soil, silt	percent	Prediction	100	0	19.5	23.0	29.3	35.7	42.1	46.5	50.0	76.2
Soil, silt	percent	Training	100	0.2	20.9	24.2	29.3	34.6	40.4	45.6	48.9	67.6
Soil, sand	percent	Prediction	100	0	18.3	23.7	33.8	44.7	54.7	65.5	72.4	87.7
Soil, sand	percent	Training	100	0.3	20.0	23.4	32.4	44.0	54.0	64.4	68.7	83.8
Water-resources development index	dimension-less	Prediction	100	0	0	2.09	3.52	4.61	5.37	6.23	7.26	7.26
Water-resources development index	dimension-less	Training	100	0	3.61	4.01	4.78	5.40	6.00	7.26	7.26	7.26
Groundwater use, irrigated agriculture	acre-ft/yr	Prediction	100	0	0	0	0	0	42.95	1,239.19	2,442.96	7,299.79
Groundwater use, irrigated agriculture	acre-ft/yr	Training	100	0	0	0	0	185.81	1,572.31	2,862.48	3,054.29	7,299.79
Surface-water use, irrigated agriculture	acre-ft/yr	Prediction	100	0	0	0	0	0	59.36	2,156.64	3,236.53	12,518.17
Surface-water use, irrigated agriculture	acre-ft/yr	Training	100	0	0	0	0	351.68	2,134.81	3,476.91	4,519.10	12,518.17
Groundwater use, public water supply	acre-ft/yr	Prediction	100	0	0	0	0	0	0	0	144.68	9,751.78
Groundwater use, public water supply	acre-ft/yr	Training	100	0	0	0	0	0	174.84	1,311.39	2,236.20	7,092.67
Surface-water use, public water supply	acre-ft/yr	Prediction	100	0	0	0	0	0	0	0	42.29	21,940.15
Surface-water use, public water supply	acre-ft/yr	Training	100	0	0	0	0	0	60.17	1,649.22	3,165.89	21,940.15
Recharge, contributing area	in/yr	Prediction	100	0	0	0.01	0.22	0.75	1.72	4.06	5.12	26.16
Recharge, contributing area	in/yr	Training	100	0	0.02	0.09	0.53	2.10	4.06	5.77	9.48	26.16
Recharge, basin	in/yr	Prediction	100	0	0	0	0.07	0.40	0.82	1.68	3.10	17.48
Recharge, basin	in/yr	Training	100	0	0.01	0.02	0.25	0.82	1.68	3.50	5.06	17.48
Potential evapotranspiration	mm	Prediction	100	0	933	965	1,048	1,292	1,443	1,508	1,533	1,647
Potential evapotranspiration	mm	Training	100	0	979	1,003	1,195	1,367	1,460	1,492	1,513	1,642
Mean air temperature	degrees F	Prediction	100	31.10	44.60	45.50	49.10	58.10	63.50	69.80	71.60	75.20

Appendix 3. Summary statistics for explanatory variables in the prediction and training datasets used in the random forest classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Prediction dataset contains 54,854 grid cells; training dataset contains 6,244 observations and in several cases 2 per grid cell. **Abbreviations:** acre-ft/yr, acre foot per year; F, Fahrenheit; ft, foot; in/hr, inch per hour; kg/yr, kilogram per year; km, kilometer; km^2, square kilometer; mm, millimeter; mg/L, milligram per liter; ppm, part per million; U, uranium; <, less than; µg/L, microgram per liter]

Variable	Units	Dataset	Observa-tions (percent)	Minimum	Percentile							Maximum
					5	10	25	50	75	90	95	
Mean air temperature	degrees F	Training	100	40.10	47.30	49.10	54.50	60.80	63.50	69.80	70.70	74.30
Recharge, subsurface inflow	percent	Training	39.7	0	0	0	0	0	3.0	7.1	13.6	13.6
Recharge, mountain front	percent	Training	39.7	0	0	0	0	1.3	30.8	49.8	55.0	73.7
Recharge, precipitation	percent	Training	39.7	0	0	0	1.5	11.3	11.3	11.3	21.1	22.9
Recharge, stream infiltration	percent	Training	39.7	0	0.3	9.7	10.9	19.5	19.5	29.9	66.5	66.5
Recharge, irrigation	percent	Training	39.7	1.6	10.0	17.3	31.4	67.4	69.2	69.2	69.2	69.2
Recharge, artificial	percent	Training	39.7	0	0	0	0	0	3.7	10.0	18.8	49.1
Recharge, change	percent	Training	39.7	10.8	37.8	37.9	85.6	565.0	565.0	565.0	572.1	572.1
Storage, change	percent	Training	39.7	–1,300,000	–1,300,000	–1,300,000	–1,300,000	–112,000	–19,600	0	700	7,000
Recharge, total	acre-ft/yr	Training	39.7	3,500	40,000	80,000	335,000	1,274,985	13,300,000	13,300,000	13,300,000	13,300,000
Discharge, total	acre-ft/yr	Training	39.7	7,200	54,500	73,000	335,000	1,341,000	14,600,000	14,600,000	14,600,000	14,600,000
Discharge, change	percent	Training	39.7	22.2	37.8	48.5	142.2	630.0	630.0	630.0	670.6	670.6
Discharge, subsurface outflow	percent	Training	39.7	0	0	0	0	0	1.3	3.6	3.6	26.4
Discharge, evapotranspiration	percent	Training	39.7	0	0	2.9	11.4	20.5	20.5	34.1	38.6	38.6
Discharge, to streams	percent	Training	39.7	0	0	0	1.0	15.8	15.8	17.1	46.1	46.1
Discharge, to springs and drains	percent	Training	39.7	0	0	0	0	0	0	9.1	59.3	59.3
Discharge, well withdrawals	percent	Training	39.7	26.1	26.1	33.1	63.0	63.7	74.9	94.8	97.6	100
Residence time	years	Training	39.7	450	450	800	1,000	1,000	1,800	2,900	8,000	21,000
Groundwater, pH	standard units	Training	94.3	3.8	6.8	7.0	7.3	7.6	7.9	8.2	8.6	10.9
Groundwater, dissolved oxygen	mg/L	Training	39.9	<1.0	<1.0	<1.0	<1.0	3.6	5.8	7.3	8.3	50
Groundwater, dissolved solids	mg/L	Training	88.1	32	165	204	287	490	993	2,460	4,520	162,000
Groundwater, nitrate	mg/L as nitrogen	Training	92.8	<0.10	<0.10	<0.10	0.32	1.5	4.5	11	17	1,600
Groundwater, sulfate	mg/L as sulfate	Training	100	<1.0	<1.0	2.5	16	56	190	630	1,500	65,000
Groundwater, iron	mg/L	Training	84.4	<50	<50	<50	<50	<50	<50	160	480	210,000
Groundwater, manganese	mg/L	Training	80.0	<10	<10	<10	<10	<10	35	270	705	67,000
Groundwater, alkalinity	mg/L	Training	28.8	11	67	81	117	165	233	328	426	18,400
Groundwater, bicarbonate	mg/L	Training	17.3	14	77	102	142	198	278	395	550	2,800
Groundwater, orthophosphate	mg/L as phosphate	Training	66.0	<0.01	<0.01	<0.01	<0.01	0.02	0.051	0.14	0.31	28
Groundwater, chloride	mg/L	Training	92.7	<5.0	<5.0	6.86	15.5	50	190	590	1,200	83,000
Groundwater, molybdenum	mg/L	Training	44.9	<10	<10	<10	<10	<10	11	39	110	28,000
Groundwater, selenium	mg/L	Training	39.9	<1.0	<1.0	<1.0	<1.0	<1.0	2	7	19	4,400

Appendix 4.

Appendix 4. Count of training observations and average misclassification errors by basin for the prediction classifers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.

[Basins with 15 or more observations were assessed for a potential for bias in nitrate or arsenic predictions. A potential for bias in predictions for a basin was noted where the average error was greater than 0.50 (predictions potentially underestimated for that basin) or less than –0.50 (predictions potentially overestimated for that basin). Under, underestimate; over, overestimate; –, no data]

Alluvial basin name	Alluvial basin number	Nitrate			Arsenic		
		Count of training observations	Average misclassification error	Potential bias	Count of training observations	Average misclassification error	Potential bias
Agua Fria River Basin	257	9	–0.22	—	6	0.33	—
Aguirre Valley	316	2	–1.00	—	2	0.00	—
Albuquerque-Belen Basin	218	112	0.54	Under	93	0.05	—
Alkali Spring Valley	148	1	0.00	—	0	—	—
Altar Valley	325	11	–0.09	—	4	1.00	—
Amargosa Desert	186	10	0.10	—	5	–0.20	—
Animas Basin	318	12	0.25	—	1	3.00	—
Antelope Valley	72	1	–2.00	—	0	—	—
Antelope Valley	116	2	0.00	—	1	–1.00	—
Antelope Valley	230	80	0.09	—	58	–0.16	—
Aravaipa Valley	293	14	0.00	—	14	0.29	—
Avra Valley	308	24	0.38	—	16	1.06	Under
Baboquivari and Tecolote Valleys	322	2	0.00	—	2	0.00	—
Beaver Valley	136	3	2.00	—	3	0.33	—
Bicycle Valley	221	8	0.00	—	6	–0.67	—
Big Chino Basin	213	29	0.07	—	28	0.04	—
Big Sandy River Basin	233	20	0.05	—	9	–0.11	—
Big Smoky Valley–Northern part	98	3	–0.33	—	0	—	—
Big Smoky Valley–Tonapah Flat	120	1	3.00	—	0	—	—
Borrego Valley	284	18	0.44	—	0	—	—
Bristol Valley	249	1	1.00	—	0	—	—
Burro Creek Basin	247	0	—	—	1	1.00	—
Butler Valley	272	11	–0.64	—	10	0.10	—
Cache Valley	1	12	0.25	—	17	1.12	Under
Calleguas–Oxnard Basin	357	66	0.39	—	10	0.10	—
Carson Desert	66	60	0.52	Under	101	–0.81	Over
Carson Valley	99	64	0.16	—	44	0.07	—
Cave Valley	121	1	2.00	—	3	–0.33	—
Cedar City Valley	151	33	0.15	—	6	0.00	—
Cedar Valley	64	6	–0.17	—	5	0.00	—
Central California Coastal Basin	424	2	0.00	—	0	—	—
Churchill Valley	96	11	–0.18	—	11	–0.36	—
Cienega Creek Basin	324	5	0.20	—	5	–0.20	—
Coachella Valley	270	31	0.19	—	25	0.64	Under
Coal Valley	143	2	0.50	—	2	1.00	—
Coastal Plain of Los Angeles	352	69	–0.04	—	73	0.66	Under
Concord-Pittsburg Area	380	2	1.00	—	0	—	—
Coyote Spring Valley	176	4	0.50	—	1	0.00	—
Crater Flat	182	2	0.00	—	2	0.00	—
Cronise Valley	234	14	0.07	—	10	0.00	—
Curlew Valley	328	11	0.18	—	8	0.50	—
Cuyama Valley	363	16	0.00	—	4	0.00	—
Dale Valley	264	1	–3.00	—	0	—	—
Dayton Valley	420	18	0.39	—	18	–0.22	—
Death Valley	170	5	0.00	—	0	—	—
Deep Creek Valley	78	0	—	—	2	0.00	—
Detrital Valley	205	15	0.33	—	19	–0.84	Over

Appendix 4. Count of training observations and average misclassification errors by basin for the prediction classifers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Basins with 15 or more observations were assessed for a potential for bias in nitrate or arsenic predictions. A potential for bias in predictions for a basin was noted where the average error was greater than 0.50 (predictions potentially underestimated for that basin) or less than –0.50 (predictions potentially overestimated for that basin). Under, underestimate; over, overestimate; –, no data]

Alluvial basin name	Alluvial basin number	Nitrate			Arsenic		
		Count of training observations	Average misclassification error	Potential bias	Count of training observations	Average misclassification error	Potential bias
Diamond Valley	73	2	0.50	—	0	—	—
Dixie Valley	71	18	0.50	—	0	—	—
Donnelly Wash	301	2	1.00	—	0	—	—
Douglas Basin	326	29	0.00	—	25	0.64	Under
Dripping Springs Wash	294	6	0.33	—	6	0.00	—
Duncan Basin	300	13	0.00	—	12	–0.17	—
Eagle Valley	421	16	0.44	—	15	0.47	—
East Bay Plain	381	22	0.32	—	23	0.39	—
East Shore Area	333	78	0.24	—	21	0.76	Under
Edwards Creek Valley	92	1	3.00	—	0	—	—
Eel River Basin	406	24	0.63	Under	12	1.08	—
Eldorado Valley	206	1	1.00	—	2	–3.00	—
Eloy Area	303	101	–0.33	—	71	–0.24	—
Engle Basin	285	14	0.86	—	5	0.20	—
Escalante Desert	142	17	0.24	—	10	0.90	—
Espanola Basin	211	33	0.15	—	21	0.10	—
Fairview Valley	101	1	1.00	—	0	—	—
Fenner Valley	239	7	–0.29	—	4	–1.00	—
Fernley Area	79	4	1.00	—	5	0.00	—
Fish Lake Valley	149	1	0.00	—	0	—	—
Fortymile Wash	173	5	0.00	—	5	0.20	—
Freemont Valley	220	21	0.62	Under	7	0.14	—
Garnet and Hidden Valleys	194	4	0.50	—	2	–1.00	—
Gila Bend Basin	298	39	0.00	—	39	–0.08	—
Grass Valley	31	1	–2.00	—	0	—	—
Great Salt Lake	331	6	0.00	—	5	0.60	—
Grouse Creek Valley	12	2	0.50	—	1	–1.00	—
Growler Valley	313	2	–2.00	—	1	–3.00	—
Harper Valley	232	16	0.19	—	12	–1.17	—
Harquahala Basin	281	77	–0.32	—	48	–0.10	—
Honey Lake Valley	398	27	0.33	—	18	0.78	Under
Hualapai Basin	215	20	0.00	—	17	0.47	—
Hualapi Flat	28	1	0.00	—	0	—	—
Humboldt River Basin–Boulder Flat Segment	37	8	–0.25	—	7	0.29	—
Humboldt River Basin–Lovelock Segment	56	2	0.00	—	1	1.00	—
Humboldt River Basin–Red House Segment	26	1	0.00	—	1	2.00	—
Imperial Valley	287	9	1.56	—	11	–0.18	—
Independence Valley	32	2	0.00	—	0	—	—
Indian Springs Valley	181	2	0.00	—	1	–5.00	—
Indian Wells Valley	197	58	0.07	—	8	0.50	—
Jersey Valley	65	1	0.00	—	0	—	—
Johnson Valley	256	9	0.11	—	6	–0.67	—
Jornada del Muerto Basin–Northern Part	263	7	0.86	—	5	–0.20	—
King and San Cristobal Valleys	290	21	–0.05	—	13	–0.38	—
Kings River and Desert Valleys	5	11	0.91	—	11	–0.36	—
Kirkland Creek Basin	251	3	–0.33	—	3	–1.67	—
La Jencia Basin	261	5	–0.20	—	2	1.00	—
La Posa Plain	269	15	–0.27	—	15	–0.40	—
Lake Mead Basin	195	3	0.33	—	3	0.33	—
Lake Pleasant	340	3	0.67	—	4	0.00	—

Appendix 4. Count of training observations and average misclassification errors by basin for the prediction classifers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Basins with 15 or more observations were assessed for a potential for bias in nitrate or arsenic predictions. A potential for bias in predictions for a basin was noted where the average error was greater than 0.50 (predictions potentially underestimated for that basin) or less than –0.50 (predictions potentially overestimated for that basin). Under, underestimate; over, overestimate; –, no data]

Alluvial basin name	Alluvial basin number	Nitrate			Arsenic		
		Count of training observations	Average misclassification error	Potential bias	Count of training observations	Average misclassification error	Potential bias
Lake Valley	122	2	1.50	—	3	1.33	—
Lanfair Valley	235	2	0.00	—	2	0.00	—
Las Vegas Valley	192	84	0.25	—	53	0.40	—
Leamington Canyon Area	94	14	–0.50	—	2	–0.50	—
Lechuquilla Desert	309	15	0.47	—	12	0.00	—
Lemmon Valley	83	4	0.00	—	0	—	—
Little Chino Basin	250	16	–0.06	—	9	–0.11	—
Livermore and Sunol Valleys	376	42	–0.71	Over	0	—	—
Long Valley	77	1	–2.00	—	0	—	—
Long Valley	155	4	0.00	—	5	0.20	—
Lordsburg Basin	317	8	0.38	—	1	–1.00	—
Lower Bear River Basin	327	14	0.43	—	17	0.82	Under
Lower Bill Williams River Basin	255	3	0.00	—	3	–2.00	—
Lower Mohave River Valley	241	62	0.29	—	47	–0.23	—
Lower San Pedro River Basin	292	42	–0.07	—	39	0.18	—
Lower Verde River Basin	339	13	0.08	—	10	0.20	—
Lucerne Valley	252	34	0.47	—	33	0.58	Under
Mad–Redwood Basin	412	9	0.22	—	8	0.13	—
McMullen Valley	271	36	–0.33	—	26	–0.19	—
Mercury Valley	189	1	0.00	—	1	1.00	—
Mesilla Basin	315	36	0.36	—	13	–0.85	—
Middle Hassayampa River Basin	276	5	0.00	—	0	—	—
Middle Reese River Valley	69	1	1.00	—	0	—	—
Milford Area	125	17	0.29	—	9	–0.11	—
Mimbres River Basin	302	20	0.25	—	15	0.00	—
Mohave River Valley	244	13	0.69	—	10	0.10	—
Mohave Valley	240	23	–0.09	—	19	–0.32	—
Mohawk Valley	305	8	–0.25	—	7	–0.43	—
Montecello–Cuchillo Basin	278	2	2.50	—	2	–1.00	—
Monterey Basin	373	42	0.52	Under	26	0.12	—
Muddy River Springs Area	185	1	0.00	—	0	—	—
North Ivanpah Valley	207	1	3.00	—	0	—	—
North Piute Valley	223	1	0.00	—	1	0.00	—
North Railroad Valley	110	3	–0.33	—	0	—	—
North Spring Valley	86	4	0.75	—	5	1.00	—
Northern Coastal Basins	400	4	1.00	—	3	0.00	—
Northern Juab Valley	91	18	0.33	—	3	0.33	—
Oasis Valley	172	1	0.00	—	0	—	—
Orocopia Valley	280	1	–1.00	—	0	—	—
Pahranagat Valley	161	1	0.00	—	0	—	—
Pahrump Valley	196	5	2.00	—	1	–5.00	—
Palo Verde Valley	279	1	–2.00	—	1	–3.00	—
Palomas and Sentinal Plains	288	24	0.13	—	22	–0.36	—
Palomas Basin	295	43	1.16	Under	36	–0.17	—
Panaca Valley	156	1	–2.00	—	0	—	—
Paradise Valley	342	21	–0.05	—	8	0.25	—
Parker and Vidal Valleys	267	1	1.00	—	2	0.00	—
Parowan Valley	147	14	–0.50	—	10	0.80	—
Pavant Valley	112	37	0.14	—	34	0.15	—
Penoyer Valley	152	1	2.00	—	0	—	—

Appendix 4. Count of training observations and average misclassification errors by basin for the prediction classifers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Basins with 15 or more observations were assessed for a potential for bias in nitrate or arsenic predictions. A potential for bias in predictions for a basin was noted where the average error was greater than 0.50 (predictions potentially underestimated for that basin) or less than –0.50 (predictions potentially overestimated for that basin). Under, underestimate; over, overestimate; –, no data]

Alluvial basin name	Alluvial basin number	Nitrate			Arsenic		
		Count of training observations	Average misclassification error	Potential bias	Count of training observations	Average misclassification error	Potential bias
Pilot Valley	27	1	–2.00	—	0	—	—
Pine Forest Valley	7	2	0.00	—	1	–1.00	—
Pine Valley	52	0	—	—	1	2.00	—
Pinto Basin	273	1	–1.00	—	0	—	—
Playas Basin	323	2	1.50	—	0	—	—
Pleasant Valley	58	1	0.00	—	0	—	—
Pocatello and Blue Creek Valleys	329	3	0.33	—	1	2.00	—
Quijotoa Valley	320	4	0.25	—	4	–0.50	—
Red Pass Valley	222	1	4.00	—	1	1.00	—
Renegras Plain	277	51	–0.02	—	51	–0.39	—
Rock Creek Valley	33	1	0.00	—	1	–2.00	—
Rock Valley	187	2	0.00	—	2	0.00	—
Ruby Valley	43	7	0.71	—	1	3.00	—
Rush Valley	63	35	0.49	—	22	0.86	Under
Sacramento Valley	231	25	–0.44	—	24	–0.50	—
Sacramento Valley	405	310	0.11	—	244	0.07	—
Safford Valley	289	54	–0.19	—	50	–0.60	Over
Salt Lake Valley	46	98	0.02	—	83	0.07	—
Salt River Valley–Chandler Area	343	109	–0.50	Over	61	0.00	—
Salt River Valley–Phoenix Area	341	177	–0.45	—	168	–0.04	—
San Agustin Basin	266	6	0.83	—	5	0.00	—
San Antonio Creek Valley	362	5	0.00	—	1	0.00	—
San Diego Coastal Basins	345	19	–0.16	—	16	0.75	Under
San Emidio Desert	45	1	2.00	—	0	—	—
San Fernando Valley	356	1	3.00	—	0	—	—
San Francisco Bay Peninsula Basins	377	24	–0.04	—	0	—	—
San Gabriel Valley	354	1	–3.00	—	0	—	—
San Jacinto Basin	350	77	–0.51	Over	61	0.64	Under
San Joaquin Valley	370	863	–0.57	Over	820	0.15	—
San Luis Rey–Escondido Coastal Basin	346	7	–0.14	—	0	—	—
San Luis Valley	423	139	0.13	—	13	0.46	—
San Marcial Basin	274	10	0.10	—	4	0.25	—
San Mateo Coastal Basins	375	1	3.00	—	0	—	—
San Ramon Valley	379	1	–2.00	—	0	—	—
San Simon Valley	310	52	0.52	Under	50	0.08	—
Santa Ana Coastal Basin	351	66	0.27	—	41	0.66	Under
Santa Ana Inland Basin	353	130	–0.62	Over	63	0.06	—
Santa Barbara Coastal Basins	360	14	0.36	—	4	0.00	—
Santa Clara River Valley	359	23	0.04	—	0	—	—
Santa Clara Valley	374	26	–0.42	—	19	0.16	—
Santa Margarita Valley	347	2	0.00	—	2	0.00	—
Santa Maria River Valley	364	19	0.11	—	0	—	—
Santa Rosa Valley	311	3	0.67	—	4	0.75	—
Santa Rosa Valley	391	1	2.00	—	0	—	—
Santa Rosa Vallley	392	7	0.14	—	2	0.00	—
Santa Ynez River Valley	361	51	0.16	—	4	1.00	—
Sarcobatus Flat	167	1	–1.00	—	0	—	—
Sevier Desert	89	83	0.41	—	68	–0.40	—
Shasta Lake Area	407	4	0.25	—	0	—	—
Shasta Valley	413	1	1.00	—	0	—	—

Appendix 4. Count of training observations and average misclassification errors by basin for the prediction classifers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.—Continued

[Basins with 15 or more observations were assessed for a potential for bias in nitrate or arsenic predictions. A potential for bias in predictions for a basin was noted where the average error was greater than 0.50 (predictions potentially underestimated for that basin) or less than –0.50 (predictions potentially overestimated for that basin). Under, underestimate; over, overestimate; –, no data]

Alluvial basin name	Alluvial basin number	Nitrate			Arsenic		
		Count of training observations	Average misclassification error	Potential bias	Count of training observations	Average misclassification error	Potential bias
Silver State and Quinn River Valleys	3	4	1.75	—	1	1.00	—
Skull Valley	335	5	1.00	—	5	1.00	—
Smith River Basin	419	5	1.00	—	5	0.80	—
Smith Valley	111	7	1.43	—	0	—	—
Smoke Creek Desert	34	11	0.18	—	9	–0.78	—
Snake Valley	84	15	0.07	—	21	–0.33	—
Socorro Basin	268	19	0.47	—	20	–0.45	—
Sonoma Valley	386	6	0.33	—	5	2.00	—
South Butte Valley	74	1	–2.00	—	0	—	—
South Ivanpah Valley	219	0	—	—	3	–0.67	—
South Owens Valley	177	6	0.33	—	1	4.00	—
South Piute Valley	242	1	0.00	—	1	1.00	—
South Tikapoo Valley	174	1	3.00	—	0	—	—
Spanish Springs Valley	422	15	0.13	—	3	–0.67	—
Stanfield Area	299	45	–0.69	Over	29	–0.21	—
Steptoe Valley	68	9	–0.22	—	2	0.00	—
Stone Cabin Valley	128	1	–2.00	—	0	—	—
Suisun-Fairfield Valley	384	2	2.50	—	1	0.00	—
Summit Lake Valley	16	2	0.00	—	2	0.00	—
Tahoe Valley	389	11	0.00	—	4	1.50	—
Temecula Valley	348	20	–0.40	—	10	0.90	—
Tooele Valley	51	55	0.04	—	18	0.33	—
Truckee River Basin–Reno/Sparks Segment	90	33	–0.18	—	31	–0.13	—
Truckee River Basin–Tracy Segment	93	2	–0.50	—	2	0.50	—
Truxton Wash	224	7	–0.14	—	4	0.00	—
Tularosa Basin	265	48	0.38	—	32	0.16	—
Tule Valley	97	1	0.00	—	0	—	—
Twentynine Palms Area	253	65	0.02	—	49	0.47	—
Ukiah Valley	395	11	0.45	—	7	0.57	—
Upper Cache Creek Basin	394	6	0.50	—	4	0.50	—
Upper Hassayampa River Basin	262	12	0.00	—	9	0.00	—
Upper Humboldt River Basin	15	1	1.00	—	0	—	—
Upper Mohave River Valley	243	102	0.22	—	96	0.22	—
Upper Reese River Valley	81	1	–1.00	—	0	—	—
Upper San Pedro River Basin	321	66	0.15	—	74	0.57	Under
Upper Santa Cruz River Basin	314	48	0.10	—	45	0.53	Under
Utah Valley	61	80	0.06	—	38	0.76	Under
Vallecito, Carrizo, and Coyote Wells Valleys	304	14	–0.07	—	0	—	—
Valley of the Ajo	319	2	0.00	—	2	0.00	—
Vekol Valley	307	10	–0.90	—	8	0.00	—
Verde Valley	236	51	0.16	—	38	–0.45	—
Walker Lake Valley	106	2	1.00	—	0	—	—
Waterman Wash	296	14	0.21	—	9	0.11	—
White River Valley	108	2	0.00	—	0	—	—
Willcox Basin	312	49	0.45	—	48	0.44	—
Willow Creek Valley	25	1	–2.00	—	1	2.00	—
Yuma Basin	306	37	0.68	Under	32	–0.47	—
Yuma Wash	297	8	–0.50	—	0	—	—

Appendix 5

Appendix 5. Count of training observations and average misclassification errors by basin for the confirmatory classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.

[Basins with 15 or more observations were assessed for a potential for bias in nitrate or arsenic predictions. A potential for bias in predictions for a basin was noted where the average error was greater than 0.50 (predictions potentially underestimated for that basin) or less than –0.50 (predictions potentially overestimated for that basin). Under, underestimate; over, overestimate; –, no data]

Alluvial basin name	Nitrate			Arsenic		
	Count	Average misclassification error	Potential bias	Count	Average misclassification error	Potential bias
Albuquerque-Belen Basin	112	0.44	—	93	–0.44	—
Carson Valley	64	0.16	—	44	–0.45	—
Eagle Valley	16	0.25	—	15	–1.00	Over
Las Vegas Valley	84	0.07	—	53	0.40	—
Sacramento Valley	310	0.13	—	244	–0.16	—
Salt Lake Valley	98	–0.03	—	83	–0.19	—
Salt River Valley–Phoenix Area	177	–0.09	—	168	–0.24	—
San Jacinto Basin	77	–0.25	—	61	0.72	Under
San Joaquin Valley	863	–0.17	—	820	–0.09	—
San Luis Valley	139	0.25	—	13	–0.23	—
Santa Ana Coastal Basin	66	0.21	—	41	0.41	—
Santa Ana Inland Basin	130	–0.26	—	63	0.25	—
Spanish Springs Valley	15	0.33	—	3	–0.33	—
Truckee River Basin–Reno/Sparks Segment	33	0.03	—	31	–0.10	—
Upper San Pedro River Basin	66	0.21	—	74	0.22	—
Upper Santa Cruz River Basin	48	0.17	—	45	0.49	—

Appendix 6

Appendix 6. Average misclassification error and count of training observations by percentile range for explanatory variables used in the prediction classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifer study area.

[Percentile ranges are for the Southwest Principal Aquifer study area and values bounding the percentile range are listed in appendix 3 under the prediction dataset. Average misclassifications greater than 0.5 or less than –0.5 are listed in **bold** type and represent potential cases where model predictions could be biased. Positive average misclassifications indicate underprediction, and negative average misclassifications indicate overprediction. **Abbreviations:** <, less than; ≥, greater than or equal to; *, cases where the explanatory variable has the same value for many grid cells such that the multiple percentile ranges share the same value for the explanatory variable. Often, the value is zero. Because the same value occurs for multiple percentile ranges, the observation counts are reported for each percentile ranges sharing the same bounding value; —, variable not used in arsenic classifier]

Variable group	Explanatory variable	Nitrate prediction classifier											
		Average misclassification error, by percentile range for explanatory variable						Count of training observations, by percentile range for explanatory variable					
		<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90	<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90
Source variables													
Nitrogen loading	Nitrogen, atmospheric	0.11	–0.07	0.11	0.18	–0.15	–0.07	167	1,022	862	1,105	1,333	1,298
	Nitrogen, farm fertilizer	0.14	0.14	0.14	0.18	0.12	–0.35	1,637*	1,637*	1,637*	258	2,191	1,701
	Nitrogen, non-farm fertilizer	0.06	0.06	0.06	0.06	–0.11	–0.05	2,536*	2,536*	2,536*	2,536*	751	2,500
	Nitrogen, confined manure	0.13	0.13	0.13	0.09	0.12	–0.26	1,667*	1,667*	1,667*	241	1,844	2,035
	Nitrogen, unconfined manure	0.06	0.07	0.07	0.13	–0.03	–0.17	258	55	971	1,358	1,548	1,597
	Nitrogen, total	–0.11	0.06	0.16	0.23	0.10	–0.33	119	199	444	967	2,283	1,775
Agricultural, urban, and biotic sources	Septic/sewer ratio	0.01	–0.01	–0.14	0.03	0.06	0.15	850	1,159	1,508	1,140	574	556
	Local population	0.21	0.21	0.21	0.08	–0.08	–0.04	589*	589*	589*	883	1,741	2,574
	Local population density	0.21	0.21	0.21	0.08	–0.08	–0.04	589*	589*	589*	883	1,741	2,574
	Basin population	0.27	0.12	0.18	0.18	0.06	–0.37	55	152	617	1,534	1,829	1,600
	Basin population density	0.36	0.18	0.16	0.17	0.19	–0.20	56	125	583	1,254	911	2,858
	Local urban land	0.22	0.22	0.22	0.12	–0.13	–0.02	566*	566*	566*	915	1,781	2,525
	Local agricultural land	0.10	0.10	0.10	0.09	0.13	–0.36	1,753*	1,753*	1,753*	391	2,122	1,521
	Basin urban land	0.15	0.45	0.20	0.19	–0.23	–0.03	92	53	540	1,318	1,743	2,041
	Basin agricultural land	0.17	–0.01	0.20	0.15	–0.02	–0.36	240	304	814	1,440	1,770	1,219
	Basin rangeland	–0.23	–0.06	0.20	0.22	0.05	0.10	1,867	1,523	1,205	952	159	81
	Basin other land cover	0.04	0.07	0.10	–0.01	–0.29	0.20	289	633	1,192	1,213	1,463	997
Geologic sources	Geology, carbonate rocks	0.01	0.16	–0.22	0.09	0.23	0.33	1,941	25	1,841	1,127	554	299
	Geology, crystalline rocks	0.25	0.16	0.26	0.07	–0.38	–0.20	530	518	1,070	1,429	1,175	1,065
	Geology, clastic sedimentary rocks	–0.11	–0.17	0.07	–0.18	0.16	0.20	425	873	1,108	1,549	771	1,061
	Geology, mafic volcanic rocks	0.15	–0.13	–0.24	0.12	0.11	0.07	815	566	1,713	1,432	805	456
	Geology, felsic and silicic volcanic rocks	–0.12	–0.12	0.01	0.15	0.08	0.41	2,869*	2,869*	1,155	1,164	462	137
	Geology, intermediate composition volcanic rocks	–0.10	–0.10	0.35	–0.07	0.09	0.14	3,086*	3,086*	505	1,064	584	548
	Geology, undifferentiated volcanic rocks	0.01	0.01	0.01	0.15	0.11	–0.22	2,513*	2,513*	2,513*	1,088	685	1,501
	Geology, distance to carbonate rocks	0.19	0.24	0.17	–0.09	–0.11	–0.09	244	475	1,031	1,368	1,278	1,391
	Geology, distance to crystalline rocks	0.11	–0.14	–0.02	0.09	–0.01	–0.15	490	368	1,639	1,332	932	1,026
	Geology, distance to clastic sedimentary rocks	0.10	–0.02	0.02	0.04	–0.02	–0.27	690	728	1,396	1,446	908	619
	Geology, distance to mafic volcanic rocks	0.16	0.09	0.06	–0.05	–0.13	–0.16	397	693	1,579	1,650	1,047	421
	Geology, distance to felsic and silicic volcanic rocks	0.41	0.15	0.13	0.01	0.01	–0.22	145	349	1,255	1,347	1,038	1,653
	Geology, distance to intermediate composition volcanic rocks	0.09	0.07	0.15	0.04	–0.01	–0.19	323	374	1,206	875	1,385	1,624
	Geology, distance to undifferentiated volcanic rocks	0.35	0.01	–0.05	–0.06	–0.10	–0.05	533	1,177	1,785	1,018	785	489
	Soil and rock equivalent uranium 283	—	—	—	—	—	—	—	—	—	—	—	—

Appendix 6. Average misclassification error and count of training observations by percentile range for explanatory variables used in the prediction classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifer study area.—Continued

[Percentile ranges are for the Southwest Principal Aquifer study area and values bounding the percentile range are listed in appendix 3 under the prediction dataset. Average misclassifications greater than 0.5 or less than –0.5 are listed in **bold** type and represent potential cases where model predictions could be biased. Positive average misclassifications indicate underprediction, and negative average misclassifications indicate overprediction. **Abbreviations:** <, less than; ≥, greater than or equal to; *, cases where the explanatory variable has the same value for many grid cells such that the multiple percentile ranges share the same value for the explanatory variable. Often, the value is zero. Because the same value occurs for multiple percentile ranges, the observation counts are reported for each percentile ranges sharing the same bounding value; —, variable not used in arsenic classifier]

Variable group	Explanatory variable	Nitrate prediction classifier											
		Average misclassification error, by percentile range for explanatory variable						Count of training observations, by percentile range for explanatory variable					
		<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90	<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90
		Susceptibility variables											
Flow path	Aquifer-penetration depth	0.02	0.24	0.06	–0.05	–0.21	–0.29	2,242	587	968	988	602	400
	Land-surface slope	–0.22	–0.09	0.01	0.10	0.23	0.04	841	1,459	1,640	1,056	539	252
	Land-surface elevation	–0.12	–0.36	0.17	0.24	0.22	0.14	1,409	1,342	1,459	1,068	285	224
	Land-surface elevation percentile	0.16	0.07	–0.07	–0.11	–0.07	0.07	820	1,182	1,827	1,287	474	197
	Basin elevation	–0.20	–0.22	0.03	0.26	0.28	0.16	1,649	835	1,660	944	449	250
	Distance to basin margin	0.14	0.14	0.09	0.02	–0.01	–0.22	971*	971*	442	1,967	1,255	1,152
Soil properties	Soil, seasonally high water depth	0.29	0.11	–0.12	0.17	–0.12	–0.12	674	1,324	1,464	188	2,137*	2,137*
	Soil, hydric	–0.15	–0.15	–0.15	0.26	0.19	0.38	3,700*	3,700*	3,700*	376	1,266	445
	Soil, hydrologic group A	–0.08	–0.08	–0.18	0.15	0.04	0.11	2,335*	2,335*	839	1,328	885	400
	Soil, hydrologic group B	0.21	0.00	0.12	0.05	–0.11	–0.33	433	830	1,198	1,515	1,016	795
	Soil, hydrologic group C	–0.13	0.14	–0.03	0.07	–0.04	0.00	981	264	1,206	1,544	1,014	778
	Soil, hydrologic group D	–0.28	–0.10	0.11	0.08	0.04	0.05	1,155	657	1,674	1,211	654	436
	Soil, permeability	–0.09	–0.07	–0.02	0.05	0.00	0.13	900	1,006	1,488	1,111	754	528
	Soil, organic material	0.08	0.14	–0.10	–0.10	–0.06	0.12	809	682	1,102	1,335	1,090	769
	Soil, clay	–0.05	0.10	–0.02	–0.04	0.05	–0.12	443	850	1,029	1,769	963	733
	Soil, silt	0.16	0.00	–0.11	–0.02	0.09	0.02	505	921	1,744	1,456	682	479
	Soil, sand	–0.02	–0.01	0.05	–0.08	0.02	0.00	597	1,025	1,367	1,394	859	545
Water use and hydroclimatic	Water-resources development index	0.13	0.08	0.19	0.14	–0.04	–0.32	80	157	868	1,657	1,704	1,321
	Groundwater use, irrigated agriculture	0.08	0.08	0.08	0.17	0.15	–0.33	1,802*	1,802*	1,802*	437	1,838	1,710
	Surface-water use, irrigated agriculture	0.08	0.08	0.08	0.15	0.06	–0.27	1,802*	1,802*	1,802*	491	2,029	1,465
	Groundwater use, public water supply	0.00	0.00	0.00	0.00	0.00	–0.03	3,638*	3,638*	3,638*	3,638*	3,638*	2,149
	Surface-water use, public water supply	0.01	0.01	0.01	0.01	0.01	–0.05	3,826*	3,826*	3,826*	3,826*	3,826*	1,961
	Recharge, contributing area	0.21	0.04	–0.05	0.31	0.09	–0.21	208	564	1,004	841	1,110	2,060
	Recharge, basin	0.30	0.11	–0.08	0.34	–0.23	0.05	188	625	1,068	702	1,753	1,451
	Potential evapotranspiration	0.18	0.16	0.24	–0.08	–0.24	0.20	147	716	1,035	2,021	1,500	368
	Mean air temperature	0.16	0.17	0.27	–0.05	–0.17	–0.22	210	224	1,237	2,308	1,045	763

Appendix 6. Average misclassification error and count of training observations by percentile range for explanatory variables used in the prediction classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifer study area.—Continued

[Percentile ranges are for the Southwest Principal Aquifer study area and values bounding the percentile range are listed in appendix 3 under the prediction dataset. Average misclassifications greater than 0.5 or less than –0.5 are listed in **bold** type and represent potential cases where model predictions could be biased. Positive average misclassifications indicate underprediction, and negative average misclassifications indicate overprediction. **Abbreviations:** <, less than; ≥, greater than or equal to; *, cases where the explanatory variable has the same value for many grid cells such that the multiple percentile ranges share the same value for the explanatory variable. Often, the value is zero. Because the same value occurs for multiple percentile ranges, the observation counts are reported for each percentile ranges sharing the same bounding value; —, variable not used in arsenic classifier]

Variable group	Explanatory variable	Arsenic prediction classifier											
		Average misclassification error, by percentile range for explanatory variable						Count of training observations, by percentile range for explanatory variable					
		<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90	<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90
		Source variables											
Nitrogen loading	Nitrogen, atmospheric	—	—	—	—	—	—	—	—	—	—	—	—
	Nitrogen, farm fertilizer	—	—	—	—	—	—	—	—	—	—	—	—
	Nitrogen, non-farm fertilizer	—	—	—	—	—	—	—	—	—	—	—	—
	Nitrogen, confined manure	—	—	—	—	—	—	—	—	—	—	—	—
	Nitrogen, unconfined manure	—	—	—	—	—	—	—	—	—	—	—	—
	Nitrogen, total	—	—	—	—	—	—	—	—	—	—	—	—
Agricultural, urban, and biotic sources	Septic/sewer ratio	0.25	0.13	0.04	0.14	0.04	–0.10	585	759	1,173	812	446	387
	Local population	–0.03	–0.03	–0.03	–0.01	0.05	0.19	418*	418*	418*	641	1,290	1,813
	Local population density	–0.03	–0.03	–0.03	–0.01	0.05	0.19	418*	418*	418*	641	1,290	1,813
	Basin population	**0.57**	–0.35	–0.01	–0.04	0.22	0.15	30	95	439	1,118	1,078	1,402
	Basin population density	**0.69**	–0.38	–0.16	0.02	–0.02	0.21	29	86	415	931	600	2,101
	Local urban land	–0.08	–0.08	–0.08	–0.02	0.06	0.19	393*	393*	393*	662	1,357	1,750
	Local agricultural land	0.19	0.19	0.19	–0.07	0.04	0.09	1,209*	1,209*	1,209*	293	1,445	1,215
	Basin urban land	–0.26	0.50	–0.01	–0.04	0.06	0.26	61	22	402	937	1,426	1,314
	Basin agricultural land	0.06	0.21	0.13	–0.04	0.12	0.16	163	228	564	1,113	995	1,099
	Basin rangeland	0.21	0.18	–0.09	0.06	–0.28	–0.15	1,472	861	986	696	106	41
	Basin other land cover	–0.18	–0.07	0.16	–0.07	0.13	0.43	247	363	934	887	1,245	486
Geologic sources	Geology, carbonate rocks	0.14	–0.36	0.04	0.02	0.15	0.41	1,259	22	1,590	664	439	188
	Geology, crystalline rocks	0.16	0.36	–0.02	0.04	0.13	0.11	238	300	776	1,044	1,031	773
	Geology, clastic sedimentary rocks	–0.22	–0.03	0.18	0.10	0.02	0.36	325	596	750	1,289	656	546
	Geology, mafic volcanic rocks	0.40	0.39	0.12	0.04	–0.25	–0.16	518	311	1,460	1,078	484	311
	Geology, felsic and silicic volcanic rocks	0.17	0.17	0.03	–0.07	0.13	0.42	2,097*	2,097*	861	896	224	84
	Geology, intermediate composition volcanic rocks	0.23	0.23	–0.17	0.13	–0.08	–0.29	2,154*	2,154*	362	812	374	460
	Geology, undifferentiated volcanic rocks	0.05	0.05	0.05	0.05	0.22	0.13	1,891*	1,891*	1,891*	549	422	1,300
	Geology, distance to carbonate rocks	0.40	0.17	0.06	–0.02	–0.06	0.33	168	312	742	914	1,092	934
	Geology, distance to crystalline rocks	0.15	0.30	0.08	0.06	0.15	0.02	331	240	1,165	931	666	829
	Geology, distance to clastic sedimentary rocks	0.34	0.16	0.23	0.09	–0.13	–0.13	399	471	1,006	1,082	698	506
	Geology, distance to mafic volcanic rocks	–0.13	–0.06	–0.04	0.19	0.24	0.21	301	504	1,034	1,166	789	368
	Geology, distance to felsic and silicic volcanic rocks	0.26	0.30	–0.01	–0.07	0.04	0.29	91	224	925	969	810	1,143
	Geology, distance to intermediate composition volcanic rocks	–0.02	0.04	–0.09	–0.06	0.17	0.30	226	284	917	631	981	1,123
	Geology, distance to undifferentiated volcanic rocks	0.31	0.07	0.13	–0.09	0.13	0.15	355	769	1,292	872	552	322
	Soil and rock equivalent uranium 283	0.29	0.11	0.06	0.09	0.03	–0.07	494	735	1,067	1,099	535	232

Appendix 6. Average misclassification error and count of training observations by percentile range for explanatory variables used in the prediction classifiers of nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifer study area.—Continued

[Percentile ranges are for the Southwest Principal Aquifer study area and values bounding the percentile range are listed in appendix 3 under the prediction dataset. Average misclassifications greater than 0.5 or less than –0.5 are listed in **bold** type and represent potential cases where model predictions could be biased. Positive average misclassifications indicate underprediction, and negative average misclassifications indicate overprediction. **Abbreviations:** <, less than; ≥, greater than or equal to; *, cases where the explanatory variable has the same value for many grid cells such that the multiple percentile ranges share the same value for the explanatory variable. Often, the value is zero. Because the same value occurs for multiple percentile ranges, the observation counts are reported for each percentile ranges sharing the same bounding value; —, variable not used in arsenic classifier]

Variable group	Explanatory variable	Arsenic prediction classifier											
		Average misclassification error, by percentile range for explanatory variable						Count of training observations, by percentile range for explanatory variable					
		<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90	<10	10–24.9	25–49.9	50–74.9	75–89.9	≥90
		Susceptibility variables											
Flow path	Aquifer-penetration depth	0.06	0.07	0.10	0.01	0.18	0.36	1,607	381	726	692	451	305
	Land-surface slope	0.01	0.03	0.09	0.18	0.23	0.25	681	1,123	1,124	744	340	150
	Land-surface elevation	0.16	0.03	–0.02	0.12	0.42	0.35	1,141	894	1,131	754	173	69
	Land-surface elevation percentile	–0.05	0.04	0.09	0.13	0.31	0.30	611	855	1,316	932	315	133
	Basin elevation	0.18	–0.06	–0.03	0.21	0.13	0.45	1,359	539	1,216	674	292	82
	Distance to basin margin	0.28	0.28	0.17	0.14	–0.04	0.03	577*	577*	302	1,314	964	1,005
Soil properties	Soil, seasonally high water depth	0.07	0.08	0.16	–0.09	0.07	0.07	496	819	1,045	151	1,651*	1,651*
	Soil, hydric	0.07	0.07	0.07	0.11	0.24	–0.07	2,844*	2,844*	2,844*	261	746	311
	Soil, hydrologic group A	0.05	0.05	0.19	0.19	0.04	–0.05	1,851*	1,851*	626	880	566	239
	Soil, hydrologic group B	–0.12	0.07	0.11	0.16	0.10	0.08	328	687	805	1,034	681	627
	Soil, hydrologic group C	0.03	0.23	0.05	0.14	0.08	0.13	770	183	908	951	694	656
	Soil, hydrologic group D	0.09	0.15	0.06	0.11	0.16	–0.01	884	438	1,089	906	489	356
	Soil, permeability	0.27	0.06	0.05	0.08	0.06	–0.01	730	709	1,120	778	495	330
	Soil, organic material	–0.01	0.00	0.01	0.15	0.17	0.20	544	497	836	987	777	521
	Soil, clay	0.18	–0.02	0.13	0.12	0.12	0.04	237	625	720	1,246	736	598
	Soil, silt	–0.02	0.00	0.14	0.08	0.09	0.26	299	666	1,365	1,051	482	299
	Soil, sand	0.06	0.11	0.11	0.14	–0.02	0.09	422	763	982	1,044	620	331
Water use and hydroclimatic	Water-resources development index	–0.40	–0.21	0.10	0.05	0.15	0.13	55	108	555	1,159	1,197	1,088
	Groundwater use, irrigated agriculture	0.20	0.20	0.20	–0.03	0.03	0.08	1,250*	1250*	1250*	320	1,292	1,300
	Surface-water use, irrigated agriculture	0.20	0.20	0.20	0.05	0.00	0.09	1,250*	1,250*	1,250*	398	1,321	1,193
	Groundwater use, public water supply	0.04	0.04	0.04	0.04	0.04	0.18	2,644*	2,644*	2,644*	2,644*	2,644*	1,518
	Surface-water use, public water supply	0.04	0.04	0.04	0.04	0.04	0.21	2,810*	2,810*	2,810*	2,810*	2,810*	1,352
	Recharge, contributing area	–0.38	–0.26	–0.12	0.19	0.39	0.17	162	434	729	648	681	1,508
	Recharge, basin	–0.38	–0.22	–0.04	0.02	0.27	0.23	143	476	776	538	1,375	854
	Potential evapotranspiration	0.50	0.28	–0.09	0.15	0.05	–0.03	105	367	731	1,591	1,115	253
	Mean air temperature	**0.70**	0.50	–0.01	0.18	0.07	–0.16	44	165	821	1,809	753	570

Appendix 7

Appendix 7. Statistical distribution of predicted nitrate and arsenic concentration classes, by basin, for basin-fill aquifers in the Southwest Principal Aquifers study area.

[Predictions are for aquifer-penetration depth of 200 feet. States: AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; NV, Nevada; NM, New Mexico; OR, Oregon; TX, Texas; UT, Utah. Nitrate concentration class ranges shown have units of milligrams per liter as nitrogen; arsenic concentration class ranges have units of micrograms per liter. **Abbreviations:** ≥, greater than or equal to; <, less than; X, state the majority of the basin is within; —, state basin is not within]

Aluvial basin		States where basin is									Predicted area (square miles)	Percentage of basin by predicted nitrate class						Percentage of basin by predicted arsenic class						
												1	2	3	4	5	6	1	2	3	4	5	6	7
Name	Number	AZ	CA	CO	ID	NV	NM	OR	TX	UT		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10– 24	≥25
Adobe Lake Valley	144	—	X	—	—	—	—	—	—	—	49	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	14.3	0.0	0.0	64.3	21.4
Agua Fria River Basin	257	X	—	—	—	—	—	—	—	—	49	0.0	7.1	64.3	28.6	0.0	0.0	0.0	7.1	28.6	14.3	28.6	21.4	0.0
Aguirre Valley	316	X	—	—	—	—	—	—	—	—	674	0.5	0.0	71.1	26.8	1.5	0.0	0.0	0.0	47.9	6.7	0.0	22.2	23.2
Albuquerque–Belen Basin	218	—	—	—	—	—	X	—	—	—	3,082	64.6	5.2	28.7	1.2	0.2	0.0	11.4	44.5	0.9	4.2	2.1	34.7	2.1
Aliso–San Onofre Coastal Basins	349	—	X	—	—	—	—	—	—	—	59	82.4	0.0	0.0	17.6	0.0	0.0	52.9	47.1	0.0	0.0	0.0	0.0	0.0
Alkali Spring Valley	148	—	—	—	—	X	—	—	—	—	191	7.3	0.0	92.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.2	1.8
Altar Valley	325	X	—	—	—	—	—	—	—	—	511	0.0	0.7	93.2	6.1	0.0	0.0	12.9	0.0	0.0	76.2	0.7	10.2	0.0
Amargosa Desert	186	—	X	—	—	X	—	—	—	—	827	96.2	0.4	3.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	26.5	53.8	19.7
Animas Basin	318	—	—	—	—	—	X	—	—	—	705	39.4	5.4	30.5	24.6	0.0	0.0	2.5	7.4	3.4	54.7	4.4	7.9	19.7
Antelope Valley	62	—	—	—	—	X	—	—	—	—	233	22.4	13.4	64.2	0.0	0.0	0.0	1.5	3.0	0.0	35.8	0.0	59.7	0.0
Antelope Valley	72	—	—	—	—	X	—	—	—	—	281	45.7	34.6	19.8	0.0	0.0	0.0	13.6	0.0	0.0	46.9	2.5	37.0	0.0
Antelope Valley	102	—	—	—	—	X	—	—	—	—	247	100.0	0.0	0.0	0.0	0.0	0.0	0.0	23.9	0.0	5.6	0.0	70.4	0.0
Antelope Valley	116	—	X	—	—	X	—	—	—	—	76	31.8	18.2	50.0	0.0	0.0	0.0	9.1	81.8	9.1	0.0	0.0	0.0	0.0
Antelope Valley	230	—	X	—	—	—	—	—	—	—	1,755	25.1	41.4	28.1	2.2	3.2	0.0	5.1	23.2	21.8	0.8	6.3	13.5	29.3
Aravaipa Valley	293	X	—	—	—	—	—	—	—	—	247	0.0	57.7	42.3	0.0	0.0	0.0	18.3	0.0	67.6	14.1	0.0	0.0	0.0
Arroyo Seco Basin	286	—	X	—	—	—	—	—	—	—	313	0.0	0.0	73.3	26.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.4	95.6
Avra Valley	308	X	—	—	—	—	—	—	—	—	966	0.0	1.1	89.2	6.5	2.2	1.1	1.1	0.0	25.9	42.8	10.1	19.1	1.1
Baboquivari and Tecolote Valleys	322	X	—	—	—	—	—	—	—	—	994	1.4	0.3	55.9	42.3	0.0	0.0	11.2	0.0	37.4	10.1	0.0	21.7	19.6
Beaver Valley	136	—	—	—	—	—	—	—	—	X	205	16.9	42.4	33.9	6.8	0.0	0.0	0.0	0.0	59.3	35.6	0.0	5.1	0.0
Bicycle Valley	221	—	X	—	—	—	—	—	—	—	136	2.6	0.0	46.2	20.5	28.2	2.6	0.0	0.0	0.0	0.0	0.0	15.4	84.6
Big Chino Basin	213	X	—	—	—	—	—	—	—	—	643	0.0	0.0	97.8	2.2	0.0	0.0	0.5	57.8	0.5	13.0	9.7	17.8	0.5
Big Sandy River Basin	233	X	—	—	—	—	—	—	—	—	612	0.0	5.7	78.4	15.9	0.0	0.0	0.0	6.3	0.0	0.0	7.4	86.4	0.0
Big Smoky Valley–Northern part	98	—	—	—	—	X	—	—	—	—	695	90.5	8.5	1.0	0.0	0.0	0.0	26.0	3.5	2.0	52.0	0.0	15.5	1.0
Big Smoky Valley–Tonapah Flat	120	—	—	—	—	X	—	—	—	—	910	42.7	5.7	50.4	1.1	0.0	0.0	0.0	0.0	0.0	0.8	0.0	84.7	14.5
Black Rock Desert	14	—	—	—	—	X	—	—	—	—	1,393	64.6	0.0	35.4	0.0	0.0	0.0	0.5	0.0	4.2	18.0	0.5	42.1	34.7
Bodega Basin	387	—	X	—	—	—	—	—	—	—	10	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Borrego Valley	284	—	X	—	—	—	—	—	—	—	521	54.7	20.0	25.3	0.0	0.0	0.0	53.3	0.0	31.3	0.0	12.7	1.3	1.3
Bridgeport Valley	378	—	X	—	—	—	—	—	—	—	118	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.9	97.1
Bristol Valley	249	—	X	—	—	—	—	—	—	—	799	18.7	0.0	18.3	16.5	46.5	0.0	0.0	0.0	0.0	0.0	8.7	5.2	86.1
Broadwell Valley	245	—	X	—	—	—	—	—	—	—	136	23.1	0.0	59.0	17.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Buena Vista Valley	49	—	—	—	—	X	—	—	—	—	455	79.4	7.6	13.0	0.0	0.0	0.0	17.6	0.0	0.0	36.6	0.0	35.9	9.9
Buffalo Valley	44	—	—	—	—	X	—	—	—	—	344	43.4	0.0	56.6	0.0	0.0	0.0	1.0	0.0	14.1	37.4	11.1	28.3	8.1
Bullard Wash	260	X	—	—	—	—	—	—	—	—	455	0.0	0.0	13.7	86.3	0.0	0.0	0.0	0.0	13.7	0.0	9.9	76.3	0.0
Burro Creek Basin	247	X	—	—	—	—	—	—	—	—	59	0.0	0.0	23.5	76.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	58.8	41.2

Appendix 7. Statistical distribution of predicted nitrate and arsenic concentration classes, by basin, for basin-fill aquifers in the Southwest Principal Aquifers study area.—Continued

[Predictions are for aquifer-penetration depth of 200 feet. States: AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; NV, Nevada; NM, New Mexico; OR, Oregon; TX, Texas; UT, Utah. Nitrate concentration class ranges shown have units of milligrams per liter as nitrogen; arsenic concentration class ranges have units of micrograms per liter. **Abbreviations:** ≥, greater than or equal to; <, less than; X, state the majority of the basin is within; —, state basin is not within]

Aluvial basin		States where basin is									Predicted area (square miles)	Percentage of basin by predicted nitrate class						Percentage of basin by predicted arsenic class						
												1	2	3	4	5	6	1	2	3	4	5	6	7
Name	Number	AZ	CA	CO	ID	NV	NM	OR	TX	UT		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10– 24	≥25
Butler Valley	272	X	—	—	—	—	—	—	—	—	195	23.2	0.0	21.4	55.4	0.0	0.0	42.9	0.0	21.4	3.6	0.0	32.1	0.0
Butte Creek Basin	415	—	X	—	—	—	—	—	—	—	149	81.4	0.0	16.3	2.3	0.0	0.0	0.0	27.9	34.9	0.0	0.0	2.3	34.9
Cache Valley	1	—	—	—	X	—	—	—	—	X	726	28.2	5.7	22.0	44.0	0.0	0.0	56.0	37.3	1.4	0.0	1.9	3.3	0.0
Cactus Flat	154	—	—	—	—	X	—	—	—	—	306	22.7	0.0	77.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.9	1.1
Cadiz Valley	259	—	X	—	—	—	—	—	—	—	414	39.5	0.0	50.4	10.1	0.0	0.0	0.0	0.0	0.0	0.0	28.6	4.2	67.2
California Valley	203	—	X	—	—	—	—	—	—	—	76	27.3	0.0	72.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.5	50.0	45.5
California Wash	190	—	—	—	—	X	—	—	—	—	188	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Calleguas-Oxnard Basin	357	—	X	—	—	—	—	—	—	—	236	79.4	0.0	2.9	14.7	2.9	0.0	8.8	39.7	51.5	0.0	0.0	0.0	0.0
Carico Lake Valley	67	—	—	—	—	X	—	—	—	—	202	27.6	0.0	72.4	0.0	0.0	0.0	0.0	0.0	0.0	86.2	0.0	13.8	0.0
Carrizo Plain	365	—	X	—	—	—	—	—	—	—	195	39.3	0.0	51.8	8.9	0.0	0.0	1.8	64.3	12.5	0.0	1.8	0.0	19.6
Carson Desert	66	—	—	—	—	X	—	—	—	—	1,803	95.4	0.2	4.4	0.0	0.0	0.0	0.0	0.0	1.3	0.0	0.0	11.2	87.5
Carson Valley	99	—	X	—	—	X	—	—	—	—	208	40.0	30.0	30.0	0.0	0.0	0.0	50.0	10.0	5.0	10.0	1.7	5.0	18.3
Cave Valley	121	—	—	—	—	X	—	—	—	—	233	29.9	9.0	61.2	0.0	0.0	0.0	11.9	88.1	0.0	0.0	0.0	0.0	0.0
Cedar City Valley	151	—	—	—	—	—	—	—	—	X	247	15.5	50.7	18.3	5.6	9.9	0.0	46.5	0.0	40.8	8.5	4.2	0.0	0.0
Cedar Valley	64	—	—	—	—	—	—	—	—	X	219	22.2	71.4	6.3	0.0	0.0	0.0	39.7	17.5	0.0	6.3	19.0	17.5	0.0
Central California Coastal Basin	424	—	X	—	—	—	—	—	—	—	97	100.0	0.0	0.0	0.0	0.0	0.0	53.6	39.3	7.1	0.0	0.0	0.0	0.0
Chemehuevi Valley	254	—	X	—	—	—	—	—	—	—	327	0.0	0.0	1.1	92.6	6.4	0.0	0.0	0.0	0.0	0.0	1.1	36.2	62.8
Chicago Valley	200	—	X	—	—	—	—	—	—	—	80	39.1	0.0	60.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	52.2	47.8
Cholame Valley	366	—	X	—	—	—	—	—	—	—	35	60.0	0.0	40.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Chuckwalla Valley	275	—	X	—	—	—	—	—	—	—	1,011	27.1	0.0	70.8	2.1	0.0	0.0	0.0	0.0	0.0	0.0	9.6	10.0	80.4
Churchill Valley	96	—	—	—	—	X	—	—	—	—	188	68.5	0.0	27.8	3.7	0.0	0.0	0.0	0.0	0.0	0.0	24.1	50.0	25.9
Cienega Creek Basin	324	X	—	—	—	—	—	—	—	—	438	4.8	6.3	78.6	10.3	0.0	0.0	23.0	19.0	3.2	54.0	0.8	0.0	0.0
Clayton Valley	150	—	—	—	—	X	—	—	—	—	271	23.1	0.0	76.9	0.0	0.0	0.0	0.0	0.0	1.3	0.0	2.6	94.9	1.3
Clover Valley	40	—	—	—	—	X	—	—	—	—	341	49.0	7.1	43.9	0.0	0.0	0.0	25.5	53.1	2.0	0.0	0.0	19.4	0.0
Clover Valley	164	—	—	—	—	X	—	—	—	—	76	0.0	0.0	100.0	0.0	0.0	0.0	22.7	72.7	4.5	0.0	0.0	0.0	0.0
Coachella Valley	270	—	X	—	—	—	—	—	—	—	712	59.0	6.8	33.7	0.0	0.5	0.0	86.8	2.0	2.4	0.0	7.8	1.0	0.0
Coal Valley	143	—	—	—	—	X	—	—	—	—	292	46.4	32.1	0.0	21.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Coastal Plain of Los Angeles	352	—	X	—	—	—	—	—	—	—	539	72.3	0.0	0.0	27.7	0.0	0.0	37.4	31.0	22.6	3.2	1.3	2.6	1.9
Columbus Salt Marsh Valley	134	—	—	—	—	X	—	—	—	—	177	43.1	0.0	56.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	72.5	27.5
Concord–Pittsburg Area	380	—	X	—	—	—	—	—	—	—	87	40.0	0.0	24.0	0.0	36.0	0.0	36.0	64.0	0.0	0.0	0.0	0.0	0.0
Continental Lake Valley	2	—	—	—	—	X	—	X	—	—	83	58.3	0.0	41.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Coyote Lake Valley	237	—	X	—	—	—	—	—	—	—	136	33.3	0.0	59.0	0.0	7.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Coyote Spring Valley	176	—	—	—	—	X	—	—	—	—	327	98.9	0.0	1.1	0.0	0.0	0.0	0.0	1.1	0.0	0.0	0.0	98.9	0.0
Crater Flat	182	—	—	—	—	X	—	—	—	—	97	89.3	0.0	10.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	39.3	60.7	0.0
Crescent Valley	55	—	—	—	—	X	—	—	—	—	410	44.9	0.0	55.1	0.0	0.0	0.0	7.6	0.0	0.0	64.4	2.5	24.6	0.8
Cronise Valley	234	—	X	—	—	—	—	—	—	—	177	13.7	2.0	45.1	39.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.9	94.1
Cuddeback Valley	226	—	X	—	—	—	—	—	—	—	111	12.5	6.3	68.8	0.0	12.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0

Appendix 7. Statistical distribution of predicted nitrate and arsenic concentration classes, by basin, for basin-fill aquifers in the Southwest Principal Aquifers study area.—Continued

[Predictions are for aquifer-penetration depth of 200 feet. States: AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; NV, Nevada; NM, New Mexico; OR, Oregon; TX, Texas; UT, Utah. Nitrate concentration class ranges shown have units of milligrams per liter as nitrogen; arsenic concentration class ranges have units of micrograms per liter. **Abbreviations:** ≥, greater than or equal to; <, less than; X, state the majority of the basin is within; —, state basin is not within]

Aluvial basin		States where basin is									Predicted area (square miles)	Percentage of basin by predicted nitrate class						Percentage of basin by predicted arsenic class						
												1	2	3	4	5	6	1	2	3	4	5	6	7
Name	Number	AZ	CA	CO	ID	NV	NM	OR	TX	UT		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10– 24	≥25
Curlew Valley	328	—	—	—	X	—	—	—	—	X	897	32.9	9.3	54.7	3.1	0.0	0.0	10.5	67.4	6.2	3.9	5.4	0.4	6.2
Cuyama Valley	363	—	X	—	—	—	—	—	—	—	149	32.6	18.6	30.2	7.0	7.0	4.7	97.7	0.0	0.0	0.0	0.0	0.0	2.3
Dale Valley	264	—	X	—	—	—	—	—	—	—	313	33.3	0.0	66.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	25.6	74.4
Darwin Plateau Basins	191	—	X	—	—	—	—	—	—	—	115	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Dayton Valley	420	—	—	—	—	X	—	—	—	—	87	24.0	36.0	32.0	8.0	0.0	0.0	8.0	0.0	8.0	0.0	48.0	28.0	8.0
Death Valley	170	—	X	—	—	X	—	—	—	—	1,407	72.6	0.0	26.9	0.5	0.0	0.0	0.0	0.0	0.0	0.0	4.4	39.5	56.0
Deep Creek Valley	78	—	—	—	—	X	—	—	—	X	191	50.9	30.9	18.2	0.0	0.0	0.0	16.4	83.6	0.0	0.0	0.0	0.0	0.0
Deep Springs Valley	165	—	X	—	—	—	—	—	—	—	45	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Delamar Valley	163	—	—	—	—	X	—	—	—	—	222	7.8	0.0	92.2	0.0	0.0	0.0	0.0	0.0	0.0	6.3	0.0	93.8	0.0
Detrital Valley	205	X	—	—	—	—	—	—	—	—	420	0.0	0.0	0.0	95.9	4.1	0.0	0.0	0.8	9.9	0.8	5.0	83.5	0.0
Diamond Valley	73	—	—	—	—	X	—	—	—	—	525	84.1	6.6	8.6	0.0	0.7	0.0	15.2	53.6	0.0	3.3	0.0	12.6	15.2
Dixie Creek and Tenmile Creek Basin	42	—	—	—	—	X	—	—	—	—	146	28.6	0.0	71.4	0.0	0.0	0.0	73.8	26.2	0.0	0.0	0.0	0.0	0.0
Dixie Valley	71	—	—	—	—	X	—	—	—	—	678	63.6	16.4	20.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	53.8	45.1
Donnelly Wash	301	X	—	—	—	—	—	—	—	—	69	0.0	0.0	5.0	75.0	20.0	0.0	60.0	0.0	40.0	0.0	0.0	0.0	0.0
Douglas Basin	326	X	—	—	—	—	—	—	—	—	876	6.3	16.3	55.2	21.4	0.8	0.0	8.7	32.9	46.8	3.6	4.4	3.2	0.4
Dripping Springs Wash	294	X	—	—	—	—	—	—	—	—	76	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Dry Lake Valley	133	—	—	—	—	X	—	—	—	—	521	55.3	8.0	36.7	0.0	0.0	0.0	4.0	8.0	2.7	35.3	0.0	48.0	2.0
Duck Lake Valley	23	—	—	—	—	X	—	—	—	—	83	95.8	0.0	4.2	0.0	0.0	0.0	0.0	0.0	8.3	0.0	29.2	41.7	20.8
Dugway Valley	70	—	—	—	—	—	—	—	—	X	827	58.0	8.0	34.0	0.0	0.0	0.0	11.8	5.9	0.0	4.2	0.0	66.0	12.2
Duncan Basin	300	X	—	—	—	—	X	—	—	—	740	4.2	23.0	69.0	2.8	0.9	0.0	1.9	0.0	2.8	15.5	0.5	68.1	11.3
Eagle Valley	421	—	—	—	—	X	—	—	—	—	28	62.5	0.0	12.5	25.0	0.0	0.0	0.0	25.0	37.5	0.0	12.5	12.5	12.5
Eagle-Rose–Dry Valley	153	—	—	—	—	X	—	—	—	—	87	0.0	16.0	84.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
East Bay Plain	381	—	X	—	—	—	—	—	—	—	215	80.6	1.6	9.7	4.8	3.2	0.0	45.2	51.6	3.2	0.0	0.0	0.0	0.0
East Park Valley	11	—	—	—	—	—	—	—	—	X	261	36.0	9.3	54.7	0.0	0.0	0.0	24.0	29.3	2.7	17.3	0.0	26.7	0.0
East Pilot Knob and Brown Mountain valleys	212	—	X	—	—	—	—	—	—	—	146	16.7	0.0	47.6	16.7	19.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
East Shore area	333	—	—	—	—	—	—	—	—	X	483	67.6	15.8	15.8	0.7	0.0	0.0	45.3	15.8	2.9	1.4	2.9	27.3	4.3
East Soda Spring Valley	124	—	—	—	—	X	—	—	—	—	76	63.6	0.0	36.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
East Walker River Basin	117	—	X	—	—	X	—	—	—	—	181	25.0	1.9	73.1	0.0	0.0	0.0	21.2	0.0	9.6	30.8	1.9	34.6	1.9
Edwards Creek Valley	92	—	—	—	—	X	—	—	—	—	219	60.3	12.7	25.4	1.6	0.0	0.0	0.0	3.2	3.2	60.3	0.0	25.4	7.9
Eel River Basin	406	—	X	—	—	—	—	—	—	—	167	83.3	8.3	2.1	4.2	2.1	0.0	93.8	2.1	4.2	0.0	0.0	0.0	0.0
Eldorado Valley	206	—	—	—	—	X	—	—	—	—	306	0.0	0.0	21.6	29.5	48.9	0.0	1.1	0.0	0.0	0.0	0.0	29.5	69.3
Eloy area	303	X	—	—	—	—	—	—	—	—	1,175	0.0	0.9	65.1	8.9	7.4	17.8	0.0	0.6	11.8	21.0	5.6	54.7	6.2
Emigrant Valley	168	—	—	—	—	X	—	—	—	—	410	59.3	0.0	40.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.2	0.8
Engle Basin	285	—	—	—	—	—	X	—	—	—	327	6.4	70.2	12.8	9.6	1.1	0.0	1.1	14.9	53.2	17.0	0.0	13.8	0.0
Escalante Desert	142	—	—	—	—	X	—	—	—	X	1,275	26.4	42.8	26.2	4.6	0.0	0.0	1.1	0.3	12.5	77.4	1.6	7.1	0.0
Espanola Basin	211	—	—	—	—	—	X	—	—	—	705	28.1	44.3	24.6	3.0	0.0	0.0	30.5	26.6	21.7	21.2	0.0	0.0	0.0

Appendix 7. Statistical distribution of predicted nitrate and arsenic concentration classes, by basin, for basin-fill aquifers in the Southwest Principal Aquifers study area.—Continued

[Predictions are for aquifer-penetration depth of 200 feet. States: AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; NV, Nevada; NM, New Mexico; OR, Oregon; TX, Texas; UT, Utah. Nitrate concentration class ranges shown have units of milligrams per liter as nitrogen; arsenic concentration class ranges have units of micrograms per liter. **Abbreviations:** ≥, greater than or equal to; <, less than; X, state the majority of the basin is within; —, state basin is not within]

Aluvial basin		States where basin is									Predicted area (square miles)	Percentage of basin by predicted nitrate class						Percentage of basin by predicted arsenic class						
												1	2	3	4	5	6	1	2	3	4	5	6	7
Name	Number	AZ	CA	CO	ID	NV	NM	OR	TX	UT		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10– 24	≥25
Eureka Valley	169	—	X	—	—	—	—	—	—	—	195	87.5	0.0	12.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Fairview Valley	101	—	—	—	—	X	—	—	—	—	142	46.3	7.3	46.3	0.0	0.0	0.0	0.0	0.0	0.0	34.1	0.0	39.0	26.8
Fenner Valley	239	—	X	—	—	—	—	—	—	—	747	0.0	4.7	9.8	84.7	0.9	0.0	0.0	2.8	7.0	0.0	9.8	73.0	7.4
Fernley area	79	—	—	—	—	X	—	—	—	—	177	76.5	0.0	23.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.8	90.2
Fish Lake Valley	149	—	X	—	—	X	—	—	—	—	400	82.6	12.2	5.2	0.0	0.0	0.0	0.0	0.0	2.6	0.0	13.9	83.5	0.0
Fish Springs Flat	80	—	—	—	—	—	—	—	—	X	455	97.7	0.0	2.3	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	78.6	20.6
Fortymile Wash	173	—	—	—	—	X	—	—	—	—	264	2.6	0.0	97.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	38.2	44.7	17.1
Freemont Valley	220	—	X	—	—	—	—	—	—	—	375	41.7	11.1	46.3	0.9	0.0	0.0	0.0	1.9	38.9	8.3	3.7	39.8	7.4
Frenchman Flat	179	—	—	—	—	X	—	—	—	—	205	40.7	0.0	59.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Gabbs Valley	109	—	—	—	—	X	—	—	—	—	580	26.3	0.0	73.7	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.0	95.2	3.6
Garden Valley	139	—	—	—	—	X	—	—	—	—	299	36.0	64.0	0.0	0.0	0.0	0.0	0.0	52.3	0.0	47.7	0.0	0.0	0.0
Garfield Flat	130	—	—	—	—	X	—	—	—	—	28	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Garnet and Hidden valleys	194	—	—	—	—	X	—	—	—	—	139	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.0	27.5	62.5	0.0
Gila Bend Basin	298	X	—	—	—	—	—	—	—	—	817	0.0	0.0	1.3	82.6	14.9	1.3	0.0	0.0	6.0	41.3	27.7	23.0	2.1
Gold Flat	158	—	—	—	—	X	—	—	—	—	337	21.6	0.0	78.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	99.0	0.0
Goldstone Valley	229	—	X	—	—	—	—	—	—	—	28	0.0	0.0	50.0	37.5	12.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Goose Lake Valley	416	—	X	—	—	—	—	X	—	—	285	45.1	20.7	32.9	1.2	0.0	0.0	1.2	50.0	24.4	4.9	0.0	1.2	18.3
Goshute Valley	36	—	—	—	—	X	—	—	—	—	681	47.4	0.0	52.6	0.0	0.0	0.0	4.1	15.3	0.0	11.2	0.0	69.4	0.0
Granite Springs Valley	48	—	—	—	—	X	—	—	—	—	528	9.9	0.0	90.1	0.0	0.0	0.0	9.2	0.0	0.0	25.0	0.0	61.2	4.6
Grass Valley	31	—	—	—	—	X	—	—	—	—	271	52.6	0.0	47.4	0.0	0.0	0.0	9.0	17.9	3.8	32.1	19.2	17.9	0.0
Grass Valley	76	—	—	—	—	X	—	—	—	—	347	56.0	2.0	42.0	0.0	0.0	0.0	11.0	0.0	3.0	54.0	5.0	21.0	6.0
Great Salt Lake	331	—	—	—	—	—	—	—	—	X	466	65.7	14.2	18.7	1.5	0.0	0.0	3.0	94.0	0.0	0.0	1.5	0.0	1.5
Great Salt Lake Desert	332	—	—	—	—	X	—	—	—	X	4,302	91.0	0.7	8.3	0.0	0.0	0.0	4.6	0.2	0.0	0.0	0.0	24.0	71.2
Gridley Lake Valley	8	—	—	—	—	X	—	—	—	—	38	9.1	0.0	90.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Grouse Creek Valley	12	—	—	—	—	X	—	—	—	X	219	47.6	11.1	41.3	0.0	0.0	0.0	3.2	30.2	0.0	7.9	46.0	9.5	3.2
Growler Valley	313	X	—	—	—	—	—	—	—	—	612	0.0	0.0	1.1	98.9	0.0	0.0	0.0	0.0	14.2	0.0	0.0	9.1	76.7
Hansel and Northern Rozel Flat	330	—	—	—	—	—	—	—	—	X	188	29.6	0.0	66.7	3.7	0.0	0.0	7.4	42.6	0.0	18.5	3.7	7.4	20.4
Hardscrabble area	10	—	—	—	—	X	—	—	—	—	3	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Harper Valley	232	—	X	—	—	—	—	—	—	—	532	71.9	2.0	21.6	3.9	0.7	0.0	0.0	0.0	3.3	1.3	0.7	1.3	93.5
Harquahala Basin	281	X	—	—	—	—	—	—	—	—	841	0.0	0.0	0.0	59.5	34.7	5.8	0.0	0.0	0.0	10.7	34.3	53.7	1.2
Hidden Valley	210	—	—	—	—	X	—	—	—	—	10	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
High Rock Lake Valley	338	—	—	—	—	X	—	—	—	—	14	75.0	0.0	25.0	0.0	0.0	0.0	0.0	0.0	50.0	0.0	0.0	50.0	0.0
Honey Lake Valley	398	—	X	—	—	X	—	—	—	—	619	62.9	15.7	8.4	12.9	0.0	0.0	0.0	46.1	19.1	1.1	14.6	1.7	17.4
Hot Creek Valley	118	—	—	—	—	X	—	—	—	—	518	47.0	32.2	20.8	0.0	0.0	0.0	0.0	0.0	0.0	2.7	0.0	97.3	0.0
Hualapai Basin	215	X	—	—	—	—	—	—	—	—	650	0.0	0.0	87.7	12.3	0.0	0.0	0.0	84.0	1.6	10.7	3.7	0.0	0.0
Hualapi Flat	28	—	—	—	—	X	—	—	—	—	90	92.3	0.0	7.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	38.5	61.5

Appendix 7. Statistical distribution of predicted nitrate and arsenic concentration classes, by basin, for basin-fill aquifers in the Southwest Principal Aquifers study area.—Continued

[Predictions are for aquifer-penetration depth of 200 feet. States: AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; NV, Nevada; NM, New Mexico; OR, Oregon; TX, Texas; UT, Utah. Nitrate concentration class ranges shown have units of milligrams per liter as nitrogen; arsenic concentration class ranges have units of micrograms per liter. **Abbreviations:** ≥, greater than or equal to; <, less than; X, state the majority of the basin is within; —, state basin is not within]

Aluvial basin		States where basin is									Predicted area (square miles)	Percentage of basin by predicted nitrate class						Percentage of basin by predicted arsenic class						
												1	2	3	4	5	6	1	2	3	4	5	6	7
Name	Number	AZ	CA	CO	ID	NV	NM	OR	TX	UT		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10– 24	≥25
Humboldt River Basin–Boulder Flat segment	37	—	—	—	—	X	—	—	—	—	424	54.1	10.7	35.2	0.0	0.0	0.0	0.0	0.0	0.0	0.8	88.5	9.8	0.8
Humboldt River Basin–Elko and Mary’s Creek segment	39	—	—	—	—	X	—	—	—	—	56	75.0	0.0	25.0	0.0	0.0	0.0	6.3	25.0	0.0	0.0	68.8	0.0	0.0
Humboldt River Basin–Imlay segment	41	—	—	—	—	X	—	—	—	—	455	48.9	0.8	50.4	0.0	0.0	0.0	19.8	0.8	0.0	16.0	0.8	62.6	0.0
Humboldt River Basin–Lovelock segment	56	—	—	—	—	X	—	—	—	—	480	84.8	5.8	9.4	0.0	0.0	0.0	4.3	0.0	0.0	25.4	4.3	10.1	55.8
Humboldt River Basin-Red House segment	26	—	—	—	—	X	—	—	—	—	910	56.5	0.0	43.5	0.0	0.0	0.0	0.8	0.0	1.1	11.8	1.1	84.7	0.4
Humboldt River Basin–Winnemucca segment	30	—	—	—	—	X	—	—	—	—	229	59.1	0.0	40.9	0.0	0.0	0.0	0.0	0.0	0.0	4.5	34.8	59.1	1.5
Huntington Valley	54	—	—	—	—	X	—	—	—	—	504	32.4	2.1	65.5	0.0	0.0	0.0	21.4	66.2	0.0	3.4	5.5	3.4	0.0
Huntoon Valley	137	—	—	—	—	X	—	—	—	—	17	40.0	0.0	60.0	0.0	0.0	0.0	0.0	0.0	20.0	0.0	0.0	80.0	0.0
Imperial Valley	287	—	X	—	—	—	—	—	—	—	2,196	87.7	0.0	7.6	4.6	0.0	0.2	0.5	32.9	21.7	0.0	6.0	0.3	38.6
Independence Valley	32	—	—	—	—	X	—	—	—	—	354	59.8	0.0	40.2	0.0	0.0	0.0	1.0	95.1	0.0	0.0	0.0	3.9	0.0
Indian Springs Valley	181	—	—	—	—	X	—	—	—	—	379	100.0	0.0	0.0	0.0	0.0	0.0	84.4	0.0	0.0	0.0	0.0	15.6	0.0
Indian Valley	397	—	X	—	—	—	—	—	—	—	59	82.4	0.0	17.6	0.0	0.0	0.0	23.5	70.6	5.9	0.0	0.0	0.0	0.0
Indian Wells Valley	197	—	X	—	—	—	—	—	—	—	539	45.8	12.9	41.3	0.0	0.0	0.0	0.0	1.9	14.2	10.3	0.0	28.4	45.2
Ione Valley	113	—	—	—	—	X	—	—	—	—	261	70.7	8.0	21.3	0.0	0.0	0.0	1.3	0.0	0.0	18.7	0.0	80.0	0.0
Jakes Valley	103	—	—	—	—	X	—	—	—	—	195	42.9	5.4	51.8	0.0	0.0	0.0	0.0	92.9	0.0	7.1	0.0	0.0	0.0
Jean Lake Valley	208	—	—	—	—	X	—	—	—	—	69	80.0	0.0	20.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0	0.0	95.0	0.0
Jersey Valley	65	—	—	—	—	X	—	—	—	—	49	57.1	0.0	42.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Johnson Valley	256	—	X	—	—	—	—	—	—	—	202	27.6	34.5	37.9	0.0	0.0	0.0	10.3	0.0	72.4	13.8	0.0	1.7	1.7
Jornada del Muerto Basin–Northern part	263	—	—	—	—	—	X	—	—	—	1,525	22.6	4.1	72.2	0.9	0.2	0.0	40.5	46.5	0.0	0.0	0.2	12.8	0.0
Jornada del Muerto Basin–Southern part	291	—	—	—	—	—	X	—	—	—	1,088	12.8	18.5	59.7	8.9	0.0	0.0	20.8	16.6	11.2	3.2	0.6	47.6	0.0
Kane Springs Valley	171	—	—	—	—	X	—	—	—	—	69	10.0	0.0	90.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0	80.0	0.0
Kawich Valley	160	—	—	—	—	X	—	—	—	—	191	25.5	1.8	72.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Kelso Valley	238	—	X	—	—	—	—	—	—	—	396	32.5	11.4	50.9	5.3	0.0	0.0	0.0	0.0	0.0	0.0	7.0	45.6	47.4
King and San Cristobal valleys	290	X	—	—	—	—	—	—	—	—	1,161	8.4	0.0	24.6	62.9	0.0	4.2	0.0	0.0	2.4	0.6	15.6	5.1	76.3
Kings River and Desert valleys	5	—	—	—	—	X	—	—	—	—	1,008	69.7	0.3	30.0	0.0	0.0	0.0	1.4	0.0	0.3	44.8	0.0	50.3	3.1
Kirkland Creek Basin	251	X	—	—	—	—	—	—	—	—	73	0.0	0.0	61.9	38.1	0.0	0.0	4.8	0.0	14.3	0.0	0.0	81.0	0.0
Klamath Basin	417	—	X	—	—	—	—	X	—	—	14	75.0	25.0	0.0	0.0	0.0	0.0	25.0	25.0	50.0	0.0	0.0	0.0	0.0
Kobeh Valley	88	—	—	—	—	X	—	—	—	—	598	76.2	1.7	22.1	0.0	0.0	0.0	0.6	2.9	0.0	22.7	0.0	73.8	0.0
Kumiva Valley	47	—	—	—	—	X	—	—	—	—	212	13.1	0.0	85.2	1.6	0.0	0.0	13.1	0.0	4.9	3.3	0.0	75.4	3.3
La Jencia Basin	261	—	—	—	—	—	X	—	—	—	320	1.1	6.5	85.9	6.5	0.0	0.0	5.4	65.2	3.3	23.9	2.2	0.0	0.0
La Posa Plain	269	X	—	—	—	—	—	—	—	—	851	0.0	0.0	17.1	73.1	9.8	0.0	0.0	0.0	1.2	1.6	82.9	13.1	1.2
Lake Almanor Valley	399	—	X	—	—	—	—	—	—	—	35	90.0	0.0	10.0	0.0	0.0	0.0	0.0	10.0	0.0	0.0	0.0	30.0	60.0

Appendix 7. Statistical distribution of predicted nitrate and arsenic concentration classes, by basin, for basin-fill aquifers in the Southwest Principal Aquifers study area.—Continued

[Predictions are for aquifer-penetration depth of 200 feet. States: AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; NV, Nevada; NM, New Mexico; OR, Oregon; TX, Texas; UT, Utah. Nitrate concentration class ranges shown have units of milligrams per liter as nitrogen; arsenic concentration class ranges have units of micrograms per liter. **Abbreviations:** ≥, greater than or equal to; <, less than; X, state the majority of the basin is within; —, state basin is not within]

Aluvial basin		States where basin is									Predicted area (square miles)	Percentage of basin by predicted nitrate class						Percentage of basin by predicted arsenic class						
												1	2	3	4	5	6	1	2	3	4	5	6	7
Name	Number	AZ	CA	CO	ID	NV	NM	OR	TX	UT		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10– 24	≥25
Lake Havasu Basin	258	X	X	—	—	—	—	—	—	—	163	2.1	0.0	80.9	17.0	0.0	0.0	0.0	4.3	8.5	0.0	46.8	40.4	0.0
Lake Mead Basin	195	X	—	—	—	X	—	—	—	—	400	16.5	5.2	25.2	53.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Lake Mohave Basin	204	X	—	—	—	X	—	—	—	—	337	9.3	0.0	0.0	88.7	2.1	0.0	0.0	0.0	0.0	0.0	0.0	76.3	23.7
Lake Pleasant	340	X	—	—	—	—	—	—	—	—	184	0.0	0.0	64.2	35.8	0.0	0.0	0.0	0.0	0.0	1.9	92.5	3.8	1.9
Lake Valley	122	—	—	—	—	X	—	—	—	—	365	38.1	8.6	51.4	1.9	0.0	0.0	5.7	31.4	27.6	30.5	3.8	1.0	0.0
Lanfair Valley	235	—	X	—	—	—	—	—	—	—	160	2.2	0.0	78.3	19.6	0.0	0.0	0.0	0.0	0.0	8.7	73.9	17.4	0.0
Las Vegas Valley	192	—	—	—	—	X	—	—	—	—	942	57.2	31.7	9.6	0.0	1.5	0.0	70.5	9.2	10.3	0.0	0.7	1.5	7.7
Lavic Valley	246	—	X	—	—	—	—	—	—	—	149	37.2	0.0	53.5	7.0	2.3	0.0	0.0	0.0	0.0	0.0	0.0	7.0	93.0
Leach Valley	217	—	X	—	—	—	—	—	—	—	80	0.0	0.0	43.5	21.7	34.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Leamington Canyon area	94	—	—	—	—	—	—	—	—	X	358	19.4	11.7	57.3	11.7	0.0	0.0	68.9	31.1	0.0	0.0	0.0	0.0	0.0
Lechuquilla Desert	309	X	—	—	—	—	—	—	—	—	438	34.9	1.6	15.1	38.1	10.3	0.0	0.0	0.0	0.0	0.0	0.0	65.9	34.1
Lemmon Valley	83	—	—	—	—	X	—	—	—	—	83	8.3	0.0	8.3	83.3	0.0	0.0	8.3	62.5	25.0	0.0	0.0	4.2	0.0
Lida Valley	159	—	—	—	—	X	—	—	—	—	302	14.9	0.0	85.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.3	87.4	2.3
Little Chino Basin	250	X	—	—	—	—	—	—	—	—	139	0.0	0.0	92.5	7.5	0.0	0.0	0.0	0.0	65.0	2.5	30.0	2.5	0.0
Little Fish Lake Valley	114	—	—	—	—	X	—	—	—	—	191	74.5	25.5	0.0	0.0	0.0	0.0	7.3	74.5	0.0	18.2	0.0	0.0	0.0
Little Humboldt Valley	9	—	—	—	—	X	—	—	—	—	174	10.0	0.0	90.0	0.0	0.0	0.0	14.0	2.0	10.0	10.0	20.0	40.0	4.0
Livermore and Sunol valleys	376	—	X	—	—	—	—	—	—	—	73	4.8	4.8	9.5	4.8	52.4	23.8	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Lockwood Valley	368	—	X	—	—	—	—	—	—	—	38	45.5	0.0	9.1	45.5	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Long Valley	77	—	—	—	—	X	—	—	—	—	341	73.5	15.3	11.2	0.0	0.0	0.0	1.0	98.0	0.0	0.0	0.0	1.0	0.0
Long Valley	155	—	X	—	—	—	—	—	—	—	122	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.9	97.1
Long Valley	336	—	—	—	—	X	—	—	—	—	268	88.3	0.0	3.9	7.8	0.0	0.0	0.0	0.0	2.6	0.0	1.3	74.0	22.1
Long Valley	393	—	X	—	—	X	—	—	—	—	66	21.1	0.0	68.4	10.5	0.0	0.0	21.1	21.1	5.3	0.0	0.0	52.6	0.0
Lordsburg Basin	317	—	—	—	—	—	X	—	—	—	914	14.4	17.1	68.1	0.0	0.4	0.0	0.0	0.8	89.7	4.9	0.0	4.6	0.0
Lost Lake and Owl Lake valleys	209	—	X	—	—	—	—	—	—	—	35	0.0	0.0	40.0	50.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Lost River Basin	418	—	X	—	—	—	—	X	—	—	497	57.3	11.9	30.8	0.0	0.0	0.0	0.0	68.5	25.2	0.0	0.0	0.0	6.3
Lower Amargosa Valley	199	—	X	—	—	—	—	—	—	—	281	44.4	0.0	55.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.3	87.7
Lower Bear River Basin	327	—	—	—	X	—	—	—	—	X	928	48.7	5.2	17.2	28.8	0.0	0.0	26.2	37.1	4.1	0.0	12.7	5.6	14.2
Lower Bill Williams River Basin	255	X	—	—	—	—	—	—	—	—	417	2.5	0.0	33.3	64.2	0.0	0.0	3.3	29.2	0.8	0.0	24.2	42.5	0.0
Lower Meadow Valley Wash	166	—	—	—	—	X	—	—	—	—	361	89.4	0.0	10.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Lower Moapa Valley	184	—	—	—	—	X	—	—	—	—	146	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	95.2	4.8
Lower Mohave River Valley	241	—	X	—	—	—	—	—	—	—	417	47.5	15.0	33.3	4.2	0.0	0.0	0.8	5.8	9.2	3.3	1.7	5.8	73.3
Lower Pit River Basin	408	—	X	—	—	—	—	—	—	—	111	75.0	3.1	21.9	0.0	0.0	0.0	0.0	12.5	50.0	0.0	0.0	3.1	34.4
Lower Reese River Valley	50	—	—	—	—	X	—	—	—	—	243	67.1	0.0	32.9	0.0	0.0	0.0	0.0	0.0	4.3	0.0	7.1	88.6	0.0
Lower San Pedro River Basin	292	X	—	—	—	—	—	—	—	—	938	1.9	60.0	28.5	9.6	0.0	0.0	18.1	6.3	4.1	64.1	1.9	5.6	0.0
Lower Verde River Basin	339	X	—	—	—	—	—	—	—	—	389	6.3	11.6	69.6	11.6	0.0	0.9	0.0	0.0	0.0	0.0	25.0	75.0	0.0
Lucerne Valley	252	—	X	—	—	—	—	—	—	—	278	45.0	20.0	33.8	1.3	0.0	0.0	2.5	25.0	52.5	8.8	5.0	1.3	5.0
Mad-Redwood Basin	412	—	X	—	—	—	—	—	—	—	174	78.0	18.0	0.0	2.0	2.0	0.0	98.0	2.0	0.0	0.0	0.0	0.0	0.0

Appendix 7. Statistical distribution of predicted nitrate and arsenic concentration classes, by basin, for basin-fill aquifers in the Southwest Principal Aquifers study area.—Continued

[Predictions are for aquifer-penetration depth of 200 feet. States: AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; NV, Nevada; NM, New Mexico; OR, Oregon; TX, Texas; UT, Utah. Nitrate concentration class ranges shown have units of milligrams per liter as nitrogen; arsenic concentration class ranges have units of micrograms per liter. **Abbreviations:** ≥, greater than or equal to; <, less than; X, state the majority of the basin is within; —, state basin is not within]

Aluvial basin		States where basin is									Predicted area (square miles)	Percentage of basin by predicted nitrate class						Percentage of basin by predicted arsenic class						
												1	2	3	4	5	6	1	2	3	4	5	6	7
Name	Number	AZ	CA	CO	ID	NV	NM	OR	TX	UT		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10– 24	≥25
Madeline Plains	403	—	X	—	—	—	—	—	—	—	295	84.7	9.4	5.9	0.0	0.0	0.0	0.0	11.8	9.4	3.5	0.0	27.1	48.2
Maggie Creek Basin	29	—	—	—	—	X	—	—	—	—	3	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
Malibu Coastal Basins	355	—	X	—	—	—	—	—	—	—	3	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Mason Valley	107	—	—	—	—	X	—	—	—	—	309	57.3	6.7	36.0	0.0	0.0	0.0	1.1	0.0	0.0	4.5	15.7	11.2	67.4
Massacre Lake Valley	337	—	—	—	—	X	—	—	—	—	42	58.3	0.0	33.3	8.3	0.0	0.0	0.0	0.0	8.3	0.0	0.0	91.7	0.0
McMullen Valley	271	X	—	—	—	—	—	—	—	—	556	0.0	0.0	0.0	98.1	0.6	1.3	0.0	0.0	6.3	1.9	82.5	9.4	0.0
Mercury Valley	189	—	—	—	—	X	—	—	—	—	69	100.0	0.0	0.0	0.0	0.0	0.0	10.0	0.0	0.0	0.0	5.0	85.0	0.0
Mesilla Basin	315	—	—	—	—	—	X	—	X	—	1,230	84.5	14.7	0.6	0.3	0.0	0.0	3.4	0.0	5.6	28.8	17.5	2.3	42.4
Mesquite Valley	201	—	X	—	—	X	—	—	—	—	271	88.5	0.0	11.5	0.0	0.0	0.0	32.1	1.3	0.0	0.0	2.6	62.8	1.3
Middle Hassayampa River Basin	276	X	—	—	—	—	—	—	—	—	528	0.0	2.0	1.3	50.0	46.7	0.0	0.0	0.0	1.3	1.3	8.6	87.5	1.3
Middle Reese River Valley	69	—	—	—	—	X	—	—	—	—	125	16.7	8.3	50.0	25.0	0.0	0.0	2.8	0.0	0.0	44.4	11.1	41.7	0.0
Middle Salinas River Valley	369	—	X	—	—	—	—	—	—	—	257	36.5	0.0	5.4	17.6	32.4	8.1	48.6	43.2	8.1	0.0	0.0	0.0	0.0
Milford area	125	—	—	—	—	—	—	—	—	X	764	18.6	42.3	36.8	2.3	0.0	0.0	7.3	25.9	21.4	44.5	0.5	0.5	0.0
Mimbres River Basin	302	—	—	—	—	—	X	—	—	—	3,253	28.8	31.7	14.9	24.3	0.3	0.0	7.4	6.2	4.9	56.9	1.5	11.1	12.0
Mohave River Valley	244	—	X	—	—	—	—	—	—	—	181	1.9	44.2	17.3	30.8	5.8	0.0	0.0	7.7	69.2	3.8	17.3	1.9	0.0
Mohave Valley	240	X	X	—	—	X	—	—	—	—	615	22.6	0.0	36.7	40.1	0.6	0.0	3.4	10.7	5.1	0.6	35.6	43.5	1.1
Mohawk Valley	305	X	—	—	—	—	—	—	—	—	761	10.5	0.0	7.3	81.7	0.5	0.0	0.0	0.0	0.0	5.0	29.7	13.7	51.6
Monitor Valley	104	—	—	—	—	X	—	—	—	—	490	70.9	29.1	0.0	0.0	0.0	0.0	4.3	39.7	2.8	53.2	0.0	0.0	0.0
Mono Valley	132	—	X	—	—	X	—	—	—	—	247	63.4	16.9	19.7	0.0	0.0	0.0	1.4	0.0	11.3	7.0	0.0	25.4	54.9
Monte Cristo Valley	129	—	—	—	—	X	—	—	—	—	146	26.2	0.0	73.8	0.0	0.0	0.0	2.4	0.0	0.0	0.0	0.0	95.2	2.4
Montecello-Cuchillo Basin	278	—	—	—	—	—	X	—	—	—	292	0.0	91.7	7.1	1.2	0.0	0.0	0.0	64.3	33.3	2.4	0.0	0.0	0.0
Monterey Basin	373	—	X	—	—	—	—	—	—	—	500	92.4	0.7	0.7	2.1	2.1	2.1	14.6	55.6	27.1	2.8	0.0	0.0	0.0
Mud Meadow Creek Basin	18	—	—	—	—	X	—	—	—	—	188	90.7	0.0	9.3	0.0	0.0	0.0	0.0	0.0	5.6	0.0	0.0	70.4	24.1
Muddy River Springs area	185	—	—	—	—	X	—	—	—	—	14	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Napa Valley	388	—	X	—	—	—	—	—	—	—	66	73.7	0.0	26.3	0.0	0.0	0.0	0.0	84.2	15.8	0.0	0.0	0.0	0.0
Newark Valley	87	—	—	—	—	X	—	—	—	—	528	78.9	7.2	13.8	0.0	0.0	0.0	1.3	95.4	0.7	0.0	0.0	2.0	0.7
North Butte Valley	60	—	—	—	—	X	—	—	—	—	191	83.6	0.0	16.4	0.0	0.0	0.0	5.5	94.5	0.0	0.0	0.0	0.0	0.0
North Fork Humboldt River Basin and Lamoille Valley	21	—	—	—	—	X	—	—	—	—	132	68.4	7.9	23.7	0.0	0.0	0.0	47.4	52.6	0.0	0.0	0.0	0.0	0.0
North Ivanpah Valley	207	—	—	—	—	X	—	—	—	—	132	94.7	0.0	2.6	2.6	0.0	0.0	68.4	0.0	0.0	0.0	0.0	31.6	0.0
North Little Smoky Valley	105	—	—	—	—	X	—	—	—	—	382	90.0	6.4	3.6	0.0	0.0	0.0	0.0	0.0	0.0	4.5	0.0	95.5	0.0
North Owens Valley	146	—	X	—	—	X	—	—	—	—	410	87.3	12.7	0.0	0.0	0.0	0.0	35.6	13.6	19.5	1.7	0.0	28.8	0.8
North Piute Valley	223	—	X	—	—	X	—	—	—	—	368	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
North Railroad Valley	110	—	—	—	—	X	—	—	—	—	1,241	63.0	31.9	5.0	0.0	0.0	0.0	1.4	0.0	0.0	31.1	0.0	63.3	4.2
North Spring Valley	86	—	—	—	—	X	—	—	—	—	1,004	76.5	21.1	2.4	0.0	0.0	0.0	9.3	59.2	6.2	17.0	8.3	0.0	0.0
North Three Lakes Valley	180	—	—	—	—	X	—	—	—	—	163	100.0	0.0	0.0	0.0	0.0	0.0	10.6	0.0	0.0	0.0	0.0	89.4	0.0
North Tikapoo Valley	162	—	—	—	—	X	—	—	—	—	347	18.0	26.0	55.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0

Appendix 7. Statistical distribution of predicted nitrate and arsenic concentration classes, by basin, for basin-fill aquifers in the Southwest Principal Aquifers study area.—Continued

[Predictions are for aquifer-penetration depth of 200 feet. States: AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; NV, Nevada; NM, New Mexico; OR, Oregon; TX, Texas; UT, Utah. Nitrate concentration class ranges shown have units of milligrams per liter as nitrogen; arsenic concentration class ranges have units of micrograms per liter. **Abbreviations:** ≥, greater than or equal to; <, less than; X, state the majority of the basin is within; —, state basin is not within]

Aluvial basin		States where basin is									Predicted area (square miles)	Percentage of basin by predicted nitrate class						Percentage of basin by predicted arsenic class						
												1	2	3	4	5	6	1	2	3	4	5	6	7
Name	Number	AZ	CA	CO	ID	NV	NM	OR	TX	UT		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10– 24	≥25
Northern Coastal Basins	400	—	X	—	—	—	—	—	—	—	90	73.1	0.0	0.0	23.1	3.8	0.0	96.2	3.8	0.0	0.0	0.0	0.0	0.0
Northern Juab Valley	91	—	—	—	—	—	—	—	—	X	122	17.1	5.7	48.6	28.6	0.0	0.0	74.3	25.7	0.0	0.0	0.0	0.0	0.0
Novato Vallley	382	—	X	—	—	—	—	—	—	—	42	91.7	0.0	0.0	8.3	0.0	0.0	58.3	33.3	8.3	0.0	0.0	0.0	0.0
Oasis Valley	172	—	—	—	—	X	—	—	—	—	38	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	27.3	72.7	0.0
Ogilby Valley	282	—	X	—	—	—	—	—	—	—	83	8.3	0.0	37.5	54.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Orocopia Valley	280	—	X	—	—	—	—	—	—	—	73	4.8	0.0	95.2	0.0	0.0	0.0	4.8	0.0	0.0	0.0	4.8	4.8	85.7
Pahranagat Valley	161	—	—	—	—	X	—	—	—	—	327	89.4	4.3	6.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Pahroc Valley	140	—	—	—	—	X	—	—	—	—	247	64.8	25.4	9.9	0.0	0.0	0.0	2.8	0.0	0.0	4.2	0.0	93.0	0.0
Pahrump Valley	196	—	X	—	—	X	—	—	—	—	653	99.5	0.0	0.0	0.0	0.0	0.5	37.2	58.0	0.0	0.0	0.0	2.1	2.7
Palo Verde Valley	279	X	X	—	—	—	—	—	—	—	858	43.3	0.0	56.7	0.0	0.0	0.0	0.0	38.1	0.0	0.0	11.7	35.2	15.0
Palomas and Sentinal plains	288	X	—	—	—	—	—	—	—	—	1,442	0.0	0.0	0.0	98.1	0.0	1.9	0.0	0.0	0.0	0.0	0.5	18.1	81.4
Palomas Basin	295	—	—	—	—	—	X	—	—	—	1,018	18.8	53.9	7.2	20.1	0.0	0.0	24.9	6.5	36.9	20.5	0.7	10.6	0.0
Panaca Valley	156	—	—	—	—	X	—	—	—	—	139	12.5	2.5	85.0	0.0	0.0	0.0	37.5	47.5	12.5	0.0	0.0	2.5	0.0
Panamint Valley	193	—	X	—	—	—	—	—	—	—	379	89.9	1.8	8.3	0.0	0.0	0.0	0.0	0.0	0.9	0.0	2.8	41.3	55.0
Paradise Valley	17	—	—	—	—	X	—	—	—	—	389	47.3	6.3	46.4	0.0	0.0	0.0	12.5	9.8	18.8	0.0	21.4	33.0	4.5
Paradise Valley	342	X	—	—	—	—	—	—	—	—	275	0.0	1.3	91.1	5.1	0.0	2.5	0.0	0.0	0.0	0.0	91.1	7.6	1.3
Parker and Vidal valleys	267	X	X	—	—	—	—	—	—	—	751	23.6	0.5	43.1	31.5	1.4	0.0	0.0	0.0	0.0	0.9	52.8	31.5	14.8
Parowan Valley	147	—	—	—	—	—	—	—	—	X	177	23.5	5.9	70.6	0.0	0.0	0.0	0.0	0.0	90.2	0.0	0.0	9.8	0.0
Patterson Valley	141	—	—	—	—	X	—	—	—	—	222	21.9	6.3	71.9	0.0	0.0	0.0	1.6	98.4	0.0	0.0	0.0	0.0	0.0
Pavant Valley	112	—	—	—	—	—	—	—	—	X	372	14.0	0.9	70.1	8.4	6.5	0.0	32.7	6.5	50.5	10.3	0.0	0.0	0.0
Penoyer Valley	152	—	—	—	—	X	—	—	—	—	441	55.9	0.8	43.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	96.9	3.1
Petaluma Valley	385	—	X	—	—	—	—	—	—	—	45	100.0	0.0	0.0	0.0	0.0	0.0	53.8	46.2	0.0	0.0	0.0	0.0	0.0
Pilot Creek Valley	35	—	—	—	—	X	—	—	—	—	240	65.2	0.0	34.8	0.0	0.0	0.0	34.8	40.6	0.0	0.0	0.0	24.6	0.0
Pilot Valley	27	—	—	—	—	X	—	—	—	X	400	87.8	0.0	12.2	0.0	0.0	0.0	8.7	0.0	0.0	1.7	0.0	42.6	47.0
Pine Forest Valley	7	—	—	—	—	X	—	—	—	—	261	49.3	0.0	50.7	0.0	0.0	0.0	0.0	0.0	8.0	10.7	0.0	81.3	0.0
Pine Valley	52	—	—	—	—	X	—	—	—	—	330	29.5	1.1	69.5	0.0	0.0	0.0	4.2	3.2	0.0	9.5	0.0	83.2	0.0
Pine Valley	126	—	—	—	—	—	—	—	—	X	459	98.5	0.8	0.8	0.0	0.0	0.0	5.3	50.8	0.0	41.7	1.5	0.8	0.0
Pinto Basin	273	—	X	—	—	—	—	—	—	—	281	0.0	0.0	69.1	13.6	17.3	0.0	0.0	0.0	4.9	0.0	4.9	33.3	56.8
Playas Basin	323	—	—	—	—	—	X	—	—	—	587	15.4	1.2	40.2	43.2	0.0	0.0	8.3	16.6	26.0	44.4	1.8	3.0	0.0
Pleasant Valley	58	—	—	—	—	X	—	—	—	—	108	74.2	0.0	25.8	0.0	0.0	0.0	0.0	0.0	6.5	90.3	0.0	3.2	0.0
Pocatello and Blue Creek valleys	329	—	—	—	X	—	—	—	—	X	205	3.4	0.0	16.9	79.7	0.0	0.0	39.0	16.9	6.8	33.9	3.4	0.0	0.0
Pyramid Lake Valley	59	—	—	—	—	X	—	—	—	—	382	94.5	0.0	5.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.5	64.5	30.9
Quijotoa Valley	320	X	—	—	—	—	—	—	—	—	865	2.0	0.0	0.0	98.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	1.2	98.4
Race Track and Hidden Valleys	183	—	X	—	—	—	—	—	—	—	45	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Ralston Valley	127	—	—	—	—	X	—	—	—	—	542	52.6	33.3	14.1	0.0	0.0	0.0	0.6	0.0	0.0	37.2	0.0	59.6	2.6
Red Pass Valley	222	—	X	—	—	—	—	—	—	—	125	2.8	0.0	55.6	5.6	36.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Renegras Plain	277	X	—	—	—	—	—	—	—	—	709	0.0	0.0	0.0	63.7	34.3	2.0	0.0	0.0	0.0	0.0	5.9	31.4	62.7

Appendix 7. Statistical distribution of predicted nitrate and arsenic concentration classes, by basin, for basin-fill aquifers in the Southwest Principal Aquifers study area.—Continued

[Predictions are for aquifer-penetration depth of 200 feet. States: AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; NV, Nevada; NM, New Mexico; OR, Oregon; TX, Texas; UT, Utah. Nitrate concentration class ranges shown have units of milligrams per liter as nitrogen; arsenic concentration class ranges have units of micrograms per liter. **Abbreviations:** ≥, greater than or equal to; <, less than; X, state the majority of the basin is within; —, state basin is not within]

Aluvial basin		States where basin is									Predicted area (square miles)	Percentage of basin by predicted nitrate class						Percentage of basin by predicted arsenic class						
												1	2	3	4	5	6	1	2	3	4	5	6	7
Name	Number	AZ	CA	CO	ID	NV	NM	OR	TX	UT		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10– 24	≥25
Rhodes Salt Marsh Valley	131	—	—	—	—	X	—	—	—	—	73	14.3	0.0	85.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.8	95.2	0.0
Riggs Valley	225	—	X	—	—	—	—	—	—	—	146	35.7	0.0	35.7	16.7	11.9	0.0	0.0	0.0	0.0	0.0	0.0	31.0	69.0
Rock Creek Valley	33	—	—	—	—	X	—	—	—	—	122	25.7	0.0	74.3	0.0	0.0	0.0	0.0	0.0	0.0	74.3	0.0	20.0	5.7
Rock Valley	187	—	—	—	—	X	—	—	—	—	97	85.7	0.0	14.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	85.7	14.3
Rose Valley	198	—	X	—	—	—	—	—	—	—	59	47.1	41.2	11.8	0.0	0.0	0.0	0.0	0.0	5.9	0.0	0.0	88.2	5.9
Ruby Valley	43	—	—	—	—	X	—	—	—	—	639	75.5	20.1	4.3	0.0	0.0	0.0	19.0	58.7	11.4	0.0	1.6	9.2	0.0
Rush Valley	63	—	—	—	—	—	—	—	—	X	507	22.6	48.6	27.4	1.4	0.0	0.0	5.5	75.3	4.1	0.0	0.7	6.2	8.2
Sacramento Valley	231	X	—	—	—	—	—	—	—	—	674	0.0	0.0	1.0	99.0	0.0	0.0	0.0	20.1	0.5	0.5	76.8	2.1	0.0
Sacramento Valley	405	—	X	—	—	—	—	—	—	—	4,351	50.9	3.8	28.8	7.9	8.1	0.4	0.5	30.4	14.5	29.3	15.7	9.1	0.5
Safford Valley	289	X	—	—	—	—	—	—	—	—	1,088	72.8	13.7	1.6	6.1	5.8	0.0	32.3	0.0	1.6	3.5	17.6	9.3	35.8
Saline Valley	178	—	X	—	—	—	—	—	—	—	240	98.6	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.7	27.5	63.8
Salt Lake Valley	46	—	—	—	—	—	—	—	—	X	414	34.5	0.8	44.5	17.6	2.5	0.0	23.5	3.4	1.7	1.7	21.0	34.5	14.3
Salt River Valley–Chandler area	343	X	—	—	—	—	—	—	—	—	1,015	0.0	4.5	42.1	15.4	5.8	32.2	0.0	4.8	18.5	33.9	21.9	20.2	0.7
Salt River Valley–Phoenix area	341	X	—	—	—	—	—	—	—	—	1,178	0.0	14.5	20.6	9.1	31.0	24.8	0.0	1.2	21.2	18.6	34.8	22.7	1.5
Salton Sea	283	—	X	—	—	—	—	—	—	—	295	94.1	0.0	5.9	0.0	0.0	0.0	0.0	0.0	4.7	0.0	10.6	22.4	62.4
San Agustin Basin	266	—	—	—	—	—	X	—	—	—	997	0.0	45.6	54.4	0.0	0.0	0.0	0.0	0.0	9.8	86.8	0.0	0.0	3.5
San Antonio Creek Valley	362	—	X	—	—	—	—	—	—	—	76	95.5	0.0	4.5	0.0	0.0	0.0	95.5	0.0	0.0	0.0	4.5	0.0	0.0
San Benito–Upper Pajaro River Basin	371	—	X	—	—	—	—	—	—	—	188	44.4	0.0	40.7	11.1	3.7	0.0	3.7	81.5	14.8	0.0	0.0	0.0	0.0
San Diego Coastal Basins	345	—	X	—	—	—	—	—	—	—	229	77.3	1.5	0.0	0.0	19.7	1.5	18.2	31.8	48.5	1.5	0.0	0.0	0.0
San Emidio Desert	45	—	—	—	—	X	—	—	—	—	132	52.6	0.0	47.4	0.0	0.0	0.0	0.0	0.0	0.0	7.9	2.6	2.6	86.8
San Fernando Valley	356	—	X	—	—	—	—	—	—	—	195	25.0	0.0	0.0	75.0	0.0	0.0	16.1	83.9	0.0	0.0	0.0	0.0	0.0
San Francisco Bay Peninsula Basins	377	—	X	—	—	—	—	—	—	—	139	45.0	0.0	2.5	20.0	30.0	2.5	25.0	75.0	0.0	0.0	0.0	0.0	0.0
San Gabriel Valley	354	—	X	—	—	—	—	—	—	—	208	21.7	0.0	0.0	78.3	0.0	0.0	1.7	98.3	0.0	0.0	0.0	0.0	0.0
San Jacinto Basin	350	—	X	—	—	—	—	—	—	—	288	9.6	0.0	19.3	4.8	48.2	18.1	71.1	24.1	3.6	0.0	1.2	0.0	0.0
San Joaquin Valley	370	—	X	—	—	—	—	—	—	—	12,159	27.1	1.3	12.9	7.9	27.1	23.7	9.5	25.4	31.9	10.4	8.0	7.7	7.2
San Luis Rey-Escondido Coastal Basin	346	—	X	—	—	—	—	—	—	—	90	69.2	11.5	7.7	7.7	3.8	0.0	88.5	11.5	0.0	0.0	0.0	0.0	0.0
San Luis Valley	423	—	—	X	—	—	X	—	—	—	3,534	64.4	26.1	3.3	1.1	1.9	3.2	0.0	4.4	77.7	8.8	2.9	0.0	6.3
San Marcial Basin	274	—	—	—	—	—	X	—	—	—	546	21.0	73.9	1.3	3.8	0.0	0.0	0.6	0.6	0.6	93.0	3.8	1.3	0.0
San Mateo Coastal Basins	375	—	X	—	—	—	—	—	—	—	28	75.0	0.0	0.0	25.0	0.0	0.0	75.0	25.0	0.0	0.0	0.0	0.0	0.0
San Ramon Valley	379	—	X	—	—	—	—	—	—	—	3	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
San Simon Valley	310	X	—	—	—	—	X	—	—	—	1,550	38.6	42.6	17.7	0.2	0.0	0.9	28.7	8.1	4.3	17.3	13.9	6.3	21.5
Santa Ana Coastal Basin	351	—	X	—	—	—	—	—	—	—	365	45.7	1.0	7.6	34.3	11.4	0.0	27.6	44.8	22.9	1.0	0.0	1.0	2.9
Santa Ana Inland Basin	353	—	X	—	—	—	—	—	—	—	695	6.5	1.0	9.5	26.5	37.5	19.0	36.0	63.0	1.0	0.0	0.0	0.0	0.0
Santa Barbara Coastal Basins	360	—	X	—	—	—	—	—	—	—	80	91.3	0.0	8.7	0.0	0.0	0.0	78.3	21.7	0.0	0.0	0.0	0.0	0.0
Santa Clara River Valley	359	—	X	—	—	—	—	—	—	—	87	8.0	0.0	36.0	40.0	8.0	8.0	44.0	32.0	24.0	0.0	0.0	0.0	0.0
Santa Clara Valley	374	—	X	—	—	—	—	—	—	—	288	22.9	2.4	4.8	2.4	67.5	0.0	62.7	32.5	4.8	0.0	0.0	0.0	0.0
Santa Margarita Valley	347	—	X	—	—	—	—	—	—	—	21	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0

Appendix 7. Statistical distribution of predicted nitrate and arsenic concentration classes, by basin, for basin-fill aquifers in the Southwest Principal Aquifers study area.—Continued

[Predictions are for aquifer-penetration depth of 200 feet. States: AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; NV, Nevada; NM, New Mexico; OR, Oregon; TX, Texas; UT, Utah. Nitrate concentration class ranges shown have units of milligrams per liter as nitrogen; arsenic concentration class ranges have units of micrograms per liter. **Abbreviations:** ≥, greater than or equal to; <, less than; X, state the majority of the basin is within; —, state basin is not within]

Aluvial basin		States where basin is									Predicted area (square miles)	Percentage of basin by predicted nitrate class						Percentage of basin by predicted arsenic class						
												1	2	3	4	5	6	1	2	3	4	5	6	7
Name	Number	AZ	CA	CO	ID	NV	NM	OR	TX	UT		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10– 24	≥25
Santa Maria River Valley	364	—	X	—	—	—	—	—	—	—	243	57.1	1.4	12.9	7.1	0.0	21.4	4.3	84.3	10.0	1.4	0.0	0.0	0.0
Santa Rosa Valley	311	X	—	—	—	—	—	—	—	—	657	0.0	0.0	0.0	26.5	73.5	0.0	0.0	0.0	1.1	0.0	0.0	29.6	69.3
Santa Rosa Valley	391	—	X	—	—	X	—	—	—	—	83	62.5	0.0	37.5	0.0	0.0	0.0	0.0	25.0	16.7	0.0	0.0	0.0	58.3
Santa Rosa Vallley	392	—	X	—	—	—	—	—	—	—	90	80.8	3.8	15.4	0.0	0.0	0.0	0.0	80.8	19.2	0.0	0.0	0.0	0.0
Santa Ynez River Valley	361	—	X	—	—	—	—	—	—	—	125	63.9	0.0	5.6	25.0	5.6	0.0	91.7	0.0	0.0	0.0	8.3	0.0	0.0
Sarcobatus Flat	167	—	—	—	—	X	—	—	—	—	532	20.3	0.7	79.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	91.5	7.8
Scott River Basin	414	—	X	—	—	—	—	—	—	—	83	54.2	16.7	29.2	0.0	0.0	0.0	87.5	12.5	0.0	0.0	0.0	0.0	0.0
Searles Valley	202	—	X	—	—	—	—	—	—	—	386	35.1	0.9	60.4	3.6	0.0	0.0	0.0	0.0	0.0	0.0	3.6	26.1	70.3
Sevier Desert	89	—	—	—	—	—	—	—	—	X	3,162	87.5	4.4	7.9	0.1	0.1	0.0	0.7	0.8	0.8	7.6	6.5	52.0	31.8
Shadow Valley	216	—	X	—	—	—	—	—	—	—	254	63.0	0.0	37.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.6	90.4	0.0
Shasta Lake Area	407	—	X	—	—	—	—	—	—	—	184	50.9	49.1	0.0	0.0	0.0	0.0	3.8	45.3	20.8	0.0	0.0	0.0	30.2
Shasta Valley	413	—	X	—	—	—	—	—	—	—	125	44.4	13.9	41.7	0.0	0.0	0.0	0.0	77.8	22.2	0.0	0.0	0.0	0.0
Sierra Valley	396	—	X	—	—	—	—	—	—	—	264	94.7	2.6	2.6	0.0	0.0	0.0	3.9	67.1	3.9	0.0	0.0	0.0	25.0
Silver State and Quinn River valleys	3	—	—	—	—	X	—	X	—	—	976	37.4	0.0	60.9	0.0	0.4	1.4	11.0	2.8	13.2	6.0	3.9	38.8	24.2
Sink Valley	334	—	—	—	—	—	—	—	—	X	163	38.3	0.0	61.7	0.0	0.0	0.0	25.5	44.7	0.0	0.0	0.0	29.8	0.0
Skull Valley	335	—	—	—	—	—	—	—	—	X	615	29.9	16.9	52.5	0.6	0.0	0.0	16.4	66.7	0.0	0.0	0.0	16.9	0.0
Smith Creek Valley	95	—	—	—	—	X	—	—	—	—	351	41.6	3.0	55.4	0.0	0.0	0.0	1.0	0.0	0.0	55.4	0.0	39.6	4.0
Smith River Basin	419	—	X	—	—	—	—	—	—	—	87	80.0	12.0	4.0	4.0	0.0	0.0	84.0	4.0	0.0	12.0	0.0	0.0	0.0
Smith Valley	111	—	X	—	—	X	—	—	—	—	219	38.1	36.5	22.2	1.6	1.6	0.0	27.0	1.6	34.9	23.8	6.3	6.3	0.0
Smoke Creek Desert	34	—	—	—	—	X	—	—	—	—	452	93.8	0.0	6.2	0.0	0.0	0.0	0.0	0.0	20.0	0.0	42.3	2.3	35.4
Snake Valley	84	—	—	—	—	X	—	—	—	X	2,290	95.9	1.7	2.4	0.0	0.0	0.0	10.5	0.2	4.2	75.6	0.9	3.2	5.5
Socorro Basin	268	—	—	—	—	—	X	—	—	—	424	50.0	32.8	10.7	5.7	0.0	0.8	10.7	0.0	0.8	5.7	38.5	18.9	25.4
Soda Lake Valley	228	—	X	—	—	—	—	—	—	—	664	16.8	0.0	16.2	55.5	11.5	0.0	0.0	0.0	0.0	0.0	1.6	11.5	86.9
Sonoma Valley	386	—	X	—	—	—	—	—	—	—	129	97.3	0.0	2.7	0.0	0.0	0.0	0.0	64.9	2.7	32.4	0.0	0.0	0.0
South Butte Valley	74	—	—	—	—	X	—	—	—	—	473	75.7	14.7	9.6	0.0	0.0	0.0	0.0	86.8	0.0	12.5	0.7	0.0	0.0
South Fork area	53	—	—	—	—	X	—	—	—	—	35	80.0	10.0	10.0	0.0	0.0	0.0	40.0	50.0	0.0	0.0	0.0	10.0	0.0
South Ivanpah Valley	219	—	X	—	—	X	—	—	—	—	358	48.5	0.0	46.6	4.9	0.0	0.0	0.0	0.0	0.0	0.0	5.8	88.3	5.8
South Little Smoky Valley	115	—	—	—	—	X	—	—	—	—	323	88.2	0.0	11.8	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.0	95.7	3.2
South Owens Valley	177	—	X	—	—	—	—	—	—	—	646	90.9	8.6	0.5	0.0	0.0	0.0	2.7	0.0	0.0	0.0	0.0	83.3	14.0
South Piute Valley	242	—	X	—	—	—	—	—	—	—	142	0.0	0.0	7.3	92.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
South Railroad Valley	145	—	—	—	—	X	—	—	—	—	400	35.7	24.3	40.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.1	0.9
South Spring Valley	138	—	—	—	—	X	—	—	—	—	83	0.0	16.7	83.3	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
South Three Lakes Valley	188	—	—	—	—	X	—	—	—	—	212	100.0	0.0	0.0	0.0	0.0	0.0	77.0	0.0	0.0	0.0	0.0	23.0	0.0
South Tikapoo Valley	174	—	—	—	—	X	—	—	—	—	205	96.6	0.0	0.0	3.4	0.0	0.0	10.2	0.0	0.0	0.0	0.0	89.8	0.0
Spanish Springs Valley	422	—	—	—	—	X	—	—	—	—	31	33.3	0.0	11.1	22.2	22.2	11.1	0.0	0.0	0.0	11.1	22.2	11.1	55.6
Stanfield area	299	X	—	—	—	—	—	—	—	—	598	0.0	0.0	9.9	9.3	58.1	22.7	0.0	0.0	4.7	23.3	6.4	64.0	1.7
Steptoe Valley	68	—	—	—	—	X	—	—	—	—	1,063	42.5	31.7	25.8	0.0	0.0	0.0	10.8	81.0	0.0	6.5	0.3	1.3	0.0

Appendix 7. Statistical distribution of predicted nitrate and arsenic concentration classes, by basin, for basin-fill aquifers in the Southwest Principal Aquifers study area.—Continued

[Predictions are for aquifer-penetration depth of 200 feet. States: AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; NV, Nevada; NM, New Mexico; OR, Oregon; TX, Texas; UT, Utah. Nitrate concentration class ranges shown have units of milligrams per liter as nitrogen; arsenic concentration class ranges have units of micrograms per liter. **Abbreviations:** ≥, greater than or equal to; <, less than; X, state the majority of the basin is within; —, state basin is not within]

Aluvial basin		States where basin is									Predicted area (square miles)	Percentage of basin by predicted nitrate class						Percentage of basin by predicted arsenic class						
												1	2	3	4	5	6	1	2	3	4	5	6	7
Name	Number	AZ	CA	CO	ID	NV	NM	OR	TX	UT		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10– 24	≥25
Stingaree Valley	100	—	—	—	—	X	—	—	—	—	122	28.6	17.1	54.3	0.0	0.0	0.0	0.0	0.0	0.0	42.9	0.0	57.1	0.0
Stone Cabin Valley	128	—	—	—	—	X	—	—	—	—	570	72.0	12.2	15.9	0.0	0.0	0.0	0.0	0.0	0.0	29.3	0.0	70.7	0.0
Stonewall Flat	157	—	—	—	—	X	—	—	—	—	191	12.7	0.0	87.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	96.4	3.6
Suisun–Fairfield Valley	384	—	X	—	—	—	—	—	—	—	191	72.7	0.0	1.8	23.6	0.0	1.8	0.0	100.0	0.0	0.0	0.0	0.0	0.0
Summit Lake Valley	16	—	—	—	—	X	—	—	—	—	21	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Superior Valley	227	—	X	—	—	—	—	—	—	—	167	35.4	0.0	50.0	4.2	10.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Surprise Valley	410	—	X	—	—	X	—	—	—	—	382	90.0	4.5	5.5	0.0	0.0	0.0	0.0	8.2	50.9	0.0	1.8	2.7	36.4
Susie Creek area	38	—	—	—	—	X	—	—	—	—	7	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0	0.0
Swan Lake Valley	6	—	—	—	—	X	—	—	—	—	3	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Tahoe Valley	389	—	X	—	—	X	—	—	—	—	69	100.0	0.0	0.0	0.0	0.0	0.0	95.0	0.0	0.0	0.0	0.0	0.0	5.0
Teels Marsh Valley	135	—	—	—	—	X	—	—	—	—	76	27.3	0.0	72.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	81.8	18.2
Temecula Valley	348	—	X	—	—	—	—	—	—	—	142	7.3	2.4	68.3	9.8	7.3	4.9	36.6	63.4	0.0	0.0	0.0	0.0	0.0
Thousand Springs Valley–Herrell Siding and Brush Creek area	24	—	—	—	—	X	—	—	—	—	28	25.0	0.0	75.0	0.0	0.0	0.0	12.5	87.5	0.0	0.0	0.0	0.0	0.0
Thousand Springs Valley–Montello and Crittenden area	20	—	—	—	—	X	—	—	—	X	250	56.9	1.4	41.7	0.0	0.0	0.0	0.0	8.3	0.0	4.2	4.2	83.3	0.0
Thousand Springs Valley–Rocky Butte area	22	—	—	—	—	X	—	—	—	—	7	0.0	0.0	100.0	0.0	0.0	0.0	0.0	50.0	0.0	0.0	50.0	0.0	0.0
Thousand Springs Valley–Toano and Rock Spring areas	13	—	—	—	—	X	—	—	—	—	226	16.9	0.0	83.1	0.0	0.0	0.0	0.0	6.2	0.0	1.5	0.0	92.3	0.0
Tippett Valley	82	—	—	—	—	X	—	—	—	—	215	72.6	25.8	1.6	0.0	0.0	0.0	1.6	12.9	0.0	35.5	0.0	50.0	0.0
Tomales-Drakes Basin	383	—	X	—	—	—	—	—	—	—	17	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Tooele Valley	51	—	—	—	—	—	—	—	—	X	257	24.3	31.1	32.4	12.2	0.0	0.0	20.3	62.2	0.0	0.0	0.0	17.6	0.0
Trinity-Salmon Basin	411	—	X	—	—	—	—	—	—	—	7	50.0	50.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Truckee River Basin-Dodge Flat	85	—	—	—	—	X	—	—	—	—	52	93.3	0.0	6.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	40.0	60.0
Truckee River Basin–Reno/Sparks Segment	90	—	—	—	—	X	—	—	—	—	104	16.7	20.0	53.3	3.3	6.7	0.0	0.0	10.0	16.7	6.7	0.0	43.3	23.3
Truckee River Basin–Tracy Segment	93	—	—	—	—	X	—	—	—	—	21	33.3	16.7	50.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	83.3	16.7
Truxton Wash	224	X	—	—	—	—	—	—	—	—	146	0.0	0.0	21.4	78.6	0.0	0.0	4.8	0.0	0.0	0.0	73.8	21.4	0.0
Tularosa Basin	265	—	—	—	—	—	X	—	—	—	4,135	32.4	23.3	42.2	2.1	0.0	0.0	56.0	5.0	20.4	0.9	0.0	17.7	0.0
Tule Valley	97	—	—	—	—	—	—	—	—	X	629	100.0	0.0	0.0	0.0	0.0	0.0	40.9	0.0	0.0	0.0	0.0	43.1	16.0
Twentynine Palms area	253	—	X	—	—	—	—	—	—	—	653	29.3	0.5	62.2	7.4	0.0	0.5	8.0	35.1	1.6	37.8	12.2	2.1	3.2
Ukiah Valley	395	—	X	—	—	—	—	—	—	—	35	50.0	0.0	50.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Upper Cache Creek Basin	394	—	X	—	—	—	—	—	—	—	69	90.0	5.0	0.0	5.0	0.0	0.0	85.0	10.0	0.0	5.0	0.0	0.0	0.0
Upper Hassayampa River Basin	262	X	—	—	—	—	—	—	—	—	236	0.0	0.0	94.1	5.9	0.0	0.0	0.0	16.2	38.2	30.9	14.7	0.0	0.0
Upper Humboldt River Basin	15	—	—	—	—	X	—	—	—	—	306	46.6	2.3	47.7	3.4	0.0	0.0	39.8	55.7	0.0	0.0	0.0	4.5	0.0
Upper Mohave River Valley	243	—	X	—	—	—	—	—	—	—	796	18.3	65.5	14.8	0.9	0.4	0.0	12.7	14.0	25.8	9.6	13.5	23.6	0.9

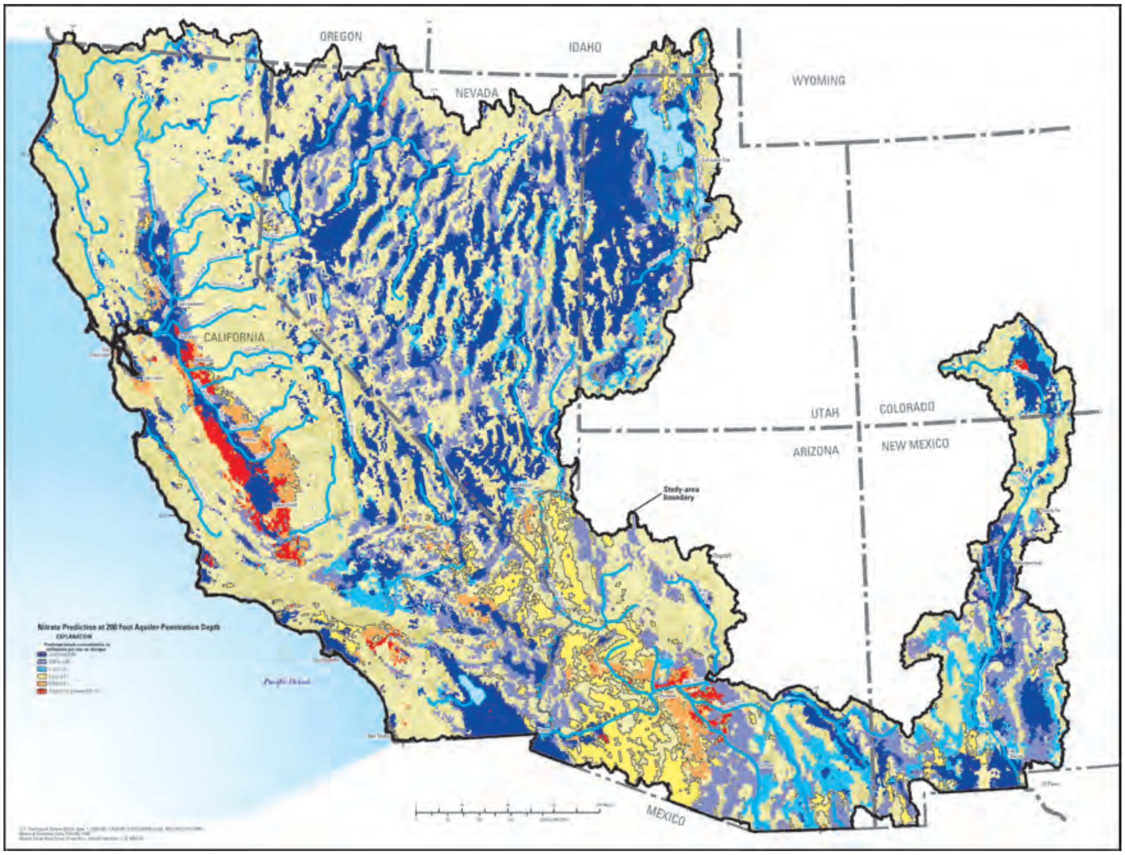
Appendix 7. Statistical distribution of predicted nitrate and arsenic concentration classes, by basin, for basin-fill aquifers in the Southwest Principal Aquifers study area.—Continued

[Predictions are for aquifer-penetration depth of 200 feet. States: AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; NV, Nevada; NM, New Mexico; OR, Oregon; TX, Texas; UT, Utah. Nitrate concentration class ranges shown have units of milligrams per liter as nitrogen; arsenic concentration class ranges have units of micrograms per liter. **Abbreviations:** ≥, greater than or equal to; <, less than; X, state the majority of the basin is within; —, state basin is not within]

Aluvial basin		States where basin is									Predicted area (square miles)	Percentage of basin by predicted nitrate class						Percentage of basin by predicted arsenic class						
												1	2	3	4	5	6	1	2	3	4	5	6	7
Name	Number	AZ	CA	CO	ID	NV	NM	OR	TX	UT		<0.50	0.50–0.99	1.0–1.9	2.0–4.9	5.0–9.9	≥10	<1.0	1.0–1.9	2.0–2.9	3.0–4.9	5.0–9.9	10– 24	≥25
Upper Pit River Basin	409	—	X	—	—	—	—	—	—	—	177	37.3	0.0	62.7	0.0	0.0	0.0	0.0	39.2	21.6	21.6	0.0	0.0	17.6
Upper Putah Creek Basin	390	—	X	—	—	—	—	—	—	—	24	71.4	0.0	28.6	0.0	0.0	0.0	14.3	71.4	14.3	0.0	0.0	0.0	0.0
Upper Reese River Valley	81	—	—	—	—	X	—	—	—	—	570	86.6	6.1	7.3	0.0	0.0	0.0	21.3	25.6	19.5	26.8	4.9	1.8	0.0
Upper Salinas River Valley	367	—	X	—	—	—	—	—	—	—	108	38.7	0.0	51.6	9.7	0.0	0.0	96.8	0.0	3.2	0.0	0.0	0.0	0.0
Upper San Pedro River Basin	321	X	—	—	—	—	—	—	—	—	1,383	5.0	57.3	20.9	15.8	1.0	0.0	58.5	10.3	21.4	4.5	0.0	4.8	0.5
Upper Santa Cruz River Basin	314	X	—	—	—	—	—	—	—	—	1,466	0.5	55.5	41.9	1.9	0.2	0.0	31.5	0.5	13.0	51.2	3.3	0.5	0.0
Upper Susan River and Eagle Lake Valley	402	—	X	—	—	—	—	—	—	—	24	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	57.1	42.9
Utah Valley	61	—	—	—	—	—	—	—	—	X	546	28.7	10.8	39.5	19.1	0.0	1.9	47.8	16.6	1.9	7.0	19.7	7.0	0.0
Valjean Valley	214	—	X	—	—	—	—	—	—	—	236	39.7	0.0	54.4	2.9	2.9	0.0	0.0	0.0	0.0	0.0	1.5	41.2	57.4
Vallecito, Carrizo, and Coyote Wells valleys	304	—	X	—	—	—	—	—	—	—	268	39.0	0.0	59.7	1.3	0.0	0.0	10.4	0.0	42.9	0.0	11.7	35.1	0.0
Valley of the Ajo	319	X	—	—	—	—	—	—	—	—	462	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5	98.5
Vekol Valley	307	X	—	—	—	—	—	—	—	—	302	2.3	0.0	0.0	55.2	42.5	0.0	0.0	0.0	31.0	18.4	2.3	47.1	1.1
Ventura River Valley	358	—	X	—	—	—	—	—	—	—	28	25.0	0.0	12.5	62.5	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Verde Valley	236	X	—	—	—	—	—	—	—	—	410	34.7	5.9	57.6	1.7	0.0	0.0	0.0	0.0	3.4	0.8	2.5	50.0	43.2
Virgin Valley	4	—	—	—	—	X	—	X	—	—	42	41.7	0.0	58.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Wah Wah Valley	119	—	—	—	—	—	—	—	—	X	368	96.2	0.0	3.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.1	0.9
Walker Lake Valley	106	—	—	—	—	X	—	—	—	—	702	37.1	5.0	56.4	1.5	0.0	0.0	0.0	0.0	0.0	1.0	0.0	96.5	2.5
Ward and Rice valleys	248	—	X	—	—		—	—	—	—	1,102	18.6	0.0	37.9	39.7	3.8	0.0	0.0	0.0	0.0	0.0	36.6	27.1	36.3
Warm Springs Valley	75	—	—	—	—	X	—	—	—	—	69	15.0	5.0	75.0	5.0	0.0	0.0	35.0	0.0	15.0	20.0	15.0	15.0	0.0
Washoe Valley	425	—	—	—	—	X	—	—	—	—	31	88.9	0.0	11.1	0.0	0.0	0.0	44.4	44.4	11.1	0.0	0.0	0.0	0.0
Waterman Wash	296	X	—	—	—	—	—	—	—	—	379	0.0	0.0	0.0	4.6	92.7	2.8	0.0	0.0	22.0	25.7	49.5	2.8	0.0
West Park Valley	19	—	—	—	—	—	—	—	—	X	500	80.6	2.1	17.4	0.0	0.0	0.0	11.1	3.5	0.0	0.0	2.8	44.4	38.2
West Soda Spring Valley	123	—	—	—	—	X	—	—	—	—	66	26.3	0.0	73.7	0.0	0.0	0.0	0.0	0.0	5.3	0.0	0.0	94.7	0.0
White River Valley	108	—	—	—	—	X	—	—	—	—	1,001	48.6	13.2	38.2	0.0	0.0	0.0	1.4	52.8	3.5	38.5	0.0	3.8	0.0
Willcox Basin	312	X	—	—	—	—	—	—	—	—	1,254	6.4	89.8	3.3	0.6	0.0	0.0	36.3	24.7	15.8	4.7	2.5	3.6	12.5
Willow Creek and Horse Lake valleys	401	—	X	—	—	—	—	—	—	—	42	91.7	0.0	8.3	0.0	0.0	0.0	0.0	25.0	16.7	16.7	0.0	0.0	41.7
Willow Creek Valley	25	—	—	—	—	X	—	—	—	—	69	20.0	0.0	80.0	0.0	0.0	0.0	15.0	0.0	0.0	0.0	85.0	0.0	0.0
Winnemucca Lake Valley	57	—	—	—	—	X	—	—	—	—	174	92.0	0.0	8.0	0.0	0.0	0.0	2.0	0.0	2.0	0.0	0.0	52.0	44.0
Yucca Flat	175	—	—	—	—	X	—	—	—	—	174	30.0	0.0	70.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Yuma Basin	306	X	X	—	—	—	—	—	—	—	959	78.6	2.5	17.4	1.4	0.0	0.0	9.1	44.9	1.8	12.3	18.1	9.8	4.0
Yuma Wash	297	X	—	—	—	—	—	—	—	—	361	1.0	0.0	99.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	4.8	94.2

Appendix 8.

Observed and predicted nitrate concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.



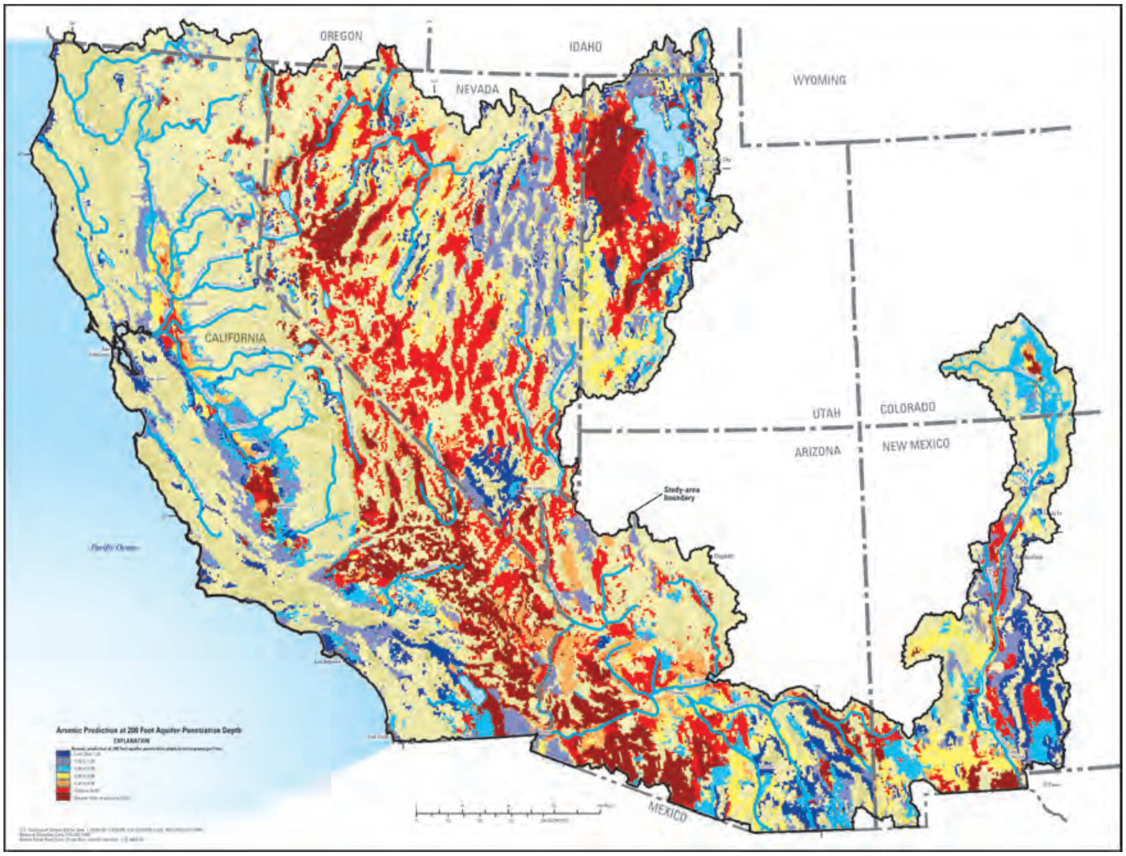
This interactive map has options to display the spatial distribution of observed nitrate concentrations from the training dataset, predicted nitrate concentrations for a 200 foot aquifer-penetration depth, or systematic change in nitrate concentration with aquifer-penetration depth, along with selected hydrologic and geographic referencing information.

Download Nitrate Change Map from the internet.
This link will take you to *http://pubs.usgs.gov/sir/2012/5065/*.

Link to Nitrate Change Map.
In order to link to Nitrate Change Map, the map must have already been downloaded and placed in the same directory on your computer as the main report document.

Appendix 9.

Observed and predicted arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area.



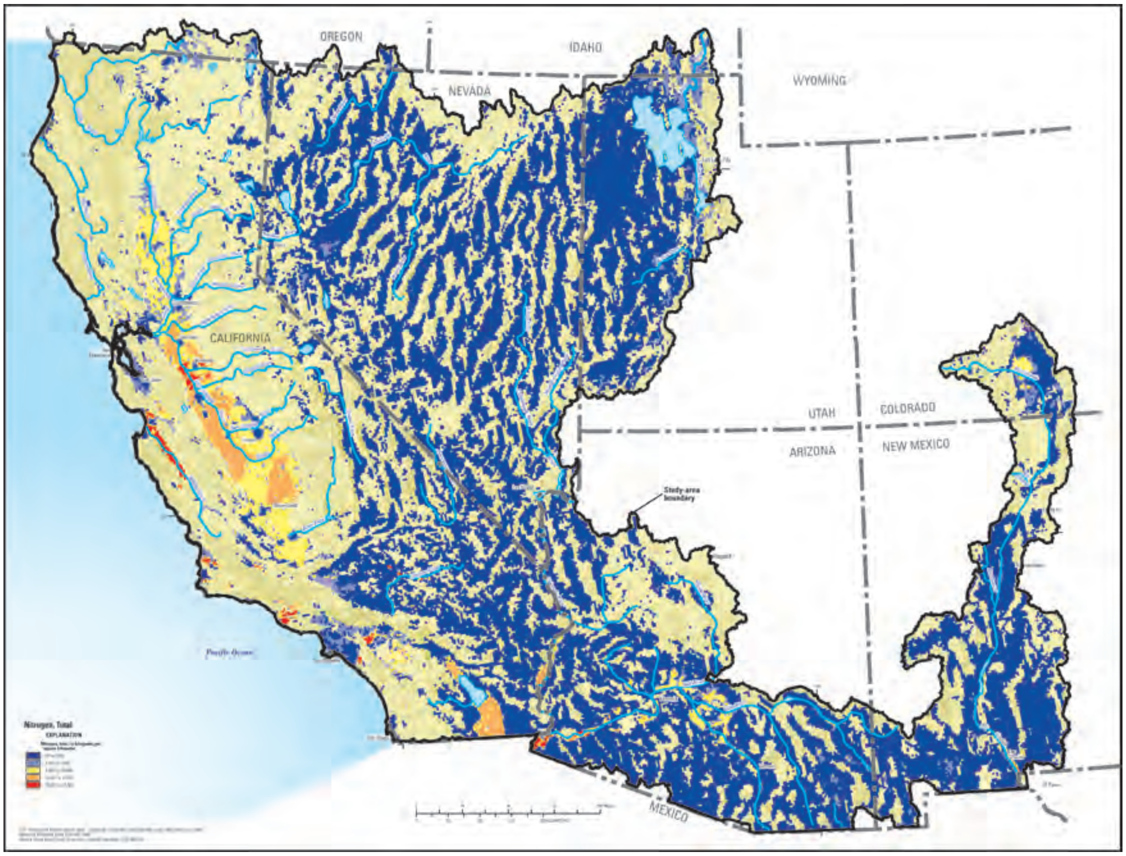
This interactive map has options to display the spatial distribution of observed arsenic concentrations from the training dataset, predicted arsenic concentrations for a 200 foot aquifer-penetration depth, or systematic change in arsenic concentration with aquifer-penetration depth, along with selected hydrologic and geographic referencing information.

Download Arsenic Concentrations Map from the internet.
This link will take you to *http://pubs.usgs.gov/sir/2012/5065/*.

Link to Arsenic Concentrations Map.
In order to link to Arsenic Concentrations Map, the map must have already been downloaded and placed in the same directory on your computer as the main report document.

Appendix 10.

Nitrogen loading variables used in the random forest classifiers of nitrate concentrations in the Southwest Principal Aquifers study area.



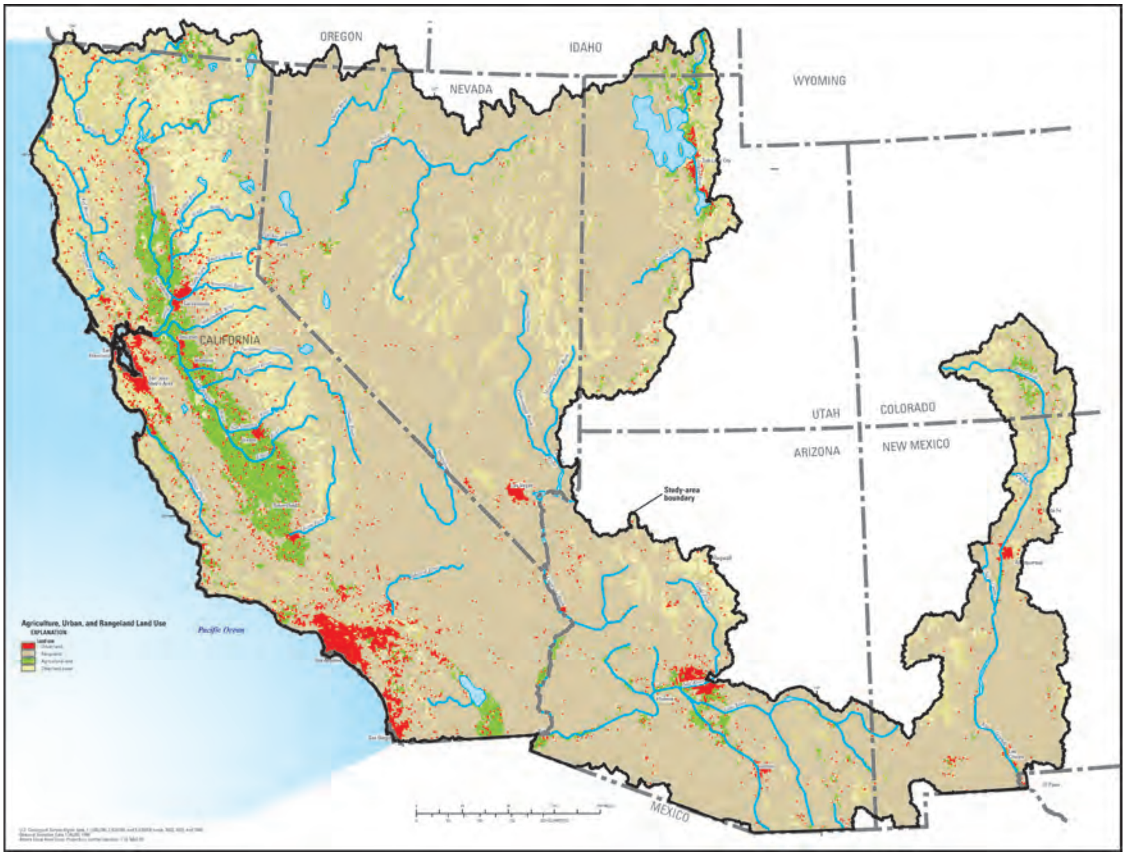
This interactive map has options to display the spatial distribution of nitrogen loading from atmospheric, confined manure, unconfined manure, farm fertilizer, or nonfarm fertilizer sources, along with the total loading from these sources. In addition, these sources can be displayed with selected hydrologic and geographic referencing information.

Download Nitrogen Loading Map from the internet.
This link will take you to *http://pubs.usgs.gov/sir/2012/5065/*.

Link to Nitrogen Loading Map.
In order to link to Nitrogen Loading Map, the map must have already been downloaded and placed in the same directory on your computer as the main report document.

Appendix 11.

Landcover variables used in the random forest classifiers of nitrate and arsenic concentrations in the Southwest Principal Aquifers study area.



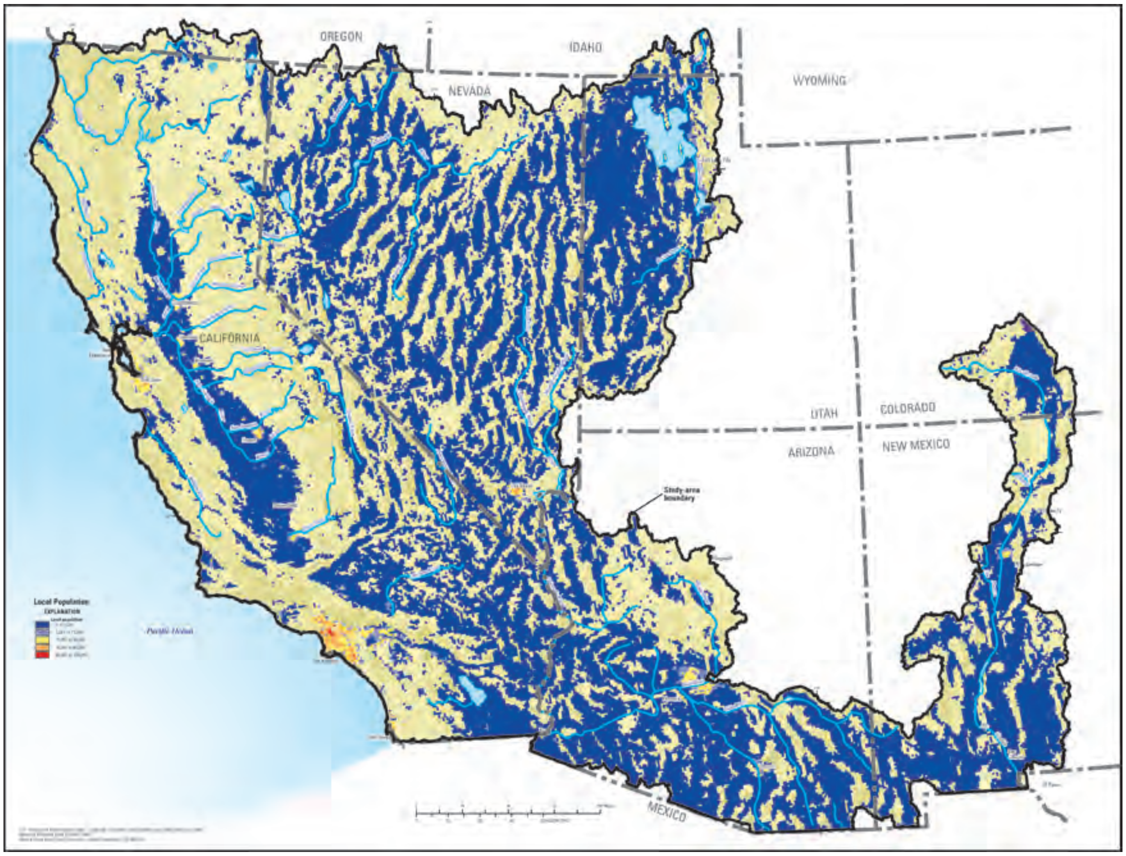
This interactive map has options to display the spatial distribution of agricultural land, urban land, rangeland, and other landcover at local and basin scales, along with selected hydrologic and geographic referencing information.

Download Landcover Variables Map from the internet.
This link will take you to *http://pubs.usgs.gov/sir/2012/5065/*.

Link to Landcover Variables Map.
In order to link to Landcover Variables Map, the map must have already been downloaded and placed in the same directory on your computer as the main report document.

Appendix 12.

Population variables used in the random forest classifiers of nitrate and arsenic concentrations in the Southwest Principal Aquifers study area.



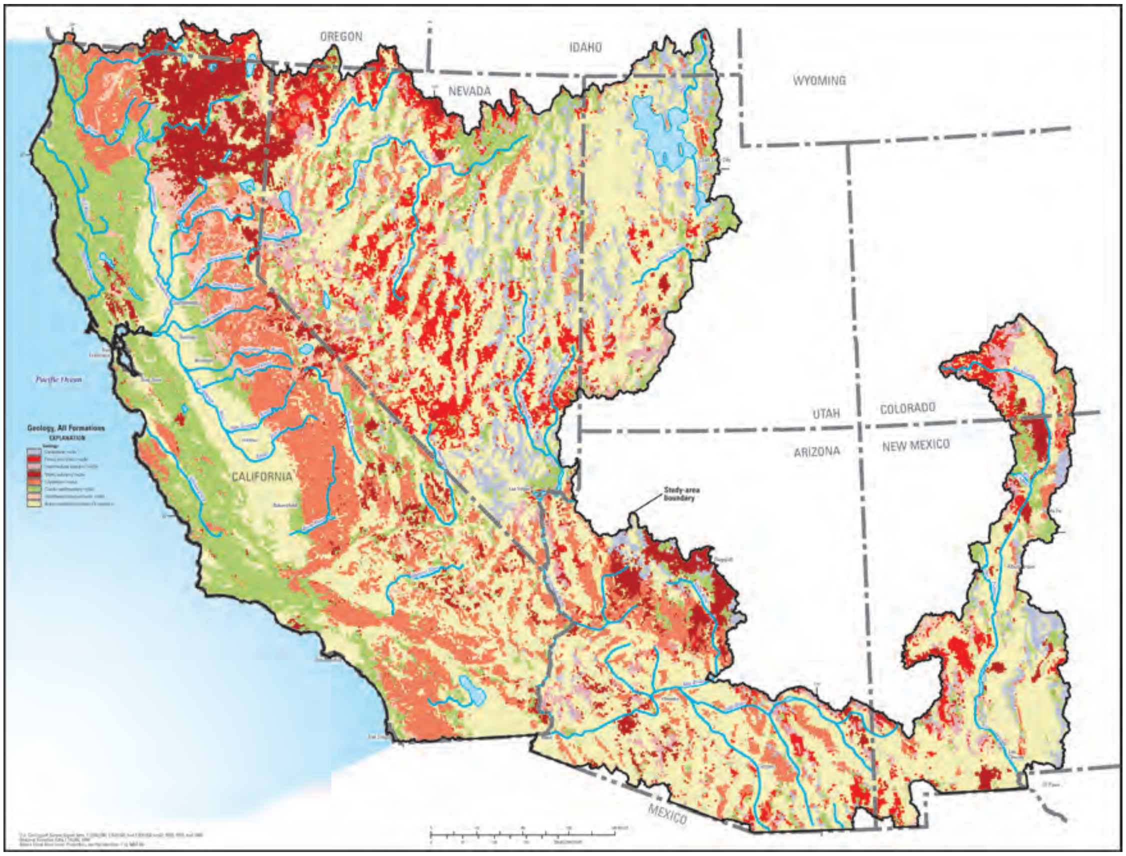
This interactive map has options to display the spatial distribution of local and basin population and population density information, as well as the ratio of population on septic systems to those on sewer systems. In addition, these population variables can be displayed with selected hydrologic and geographic referencing information.

Download Population Variables Map from the internet.
This link will take you to *http://pubs.usgs.gov/sir/2012/5065/*.

Link to Population Variables Map.
In order to link to Population Variables Map, the map must have already been downloaded and placed in the same directory on your computer as the main report document.

Appendix 13.

Geologic source variables used in the random forest classifiers of nitrate and arsenic concentrations in the Southwest Principal Aquifers study area.



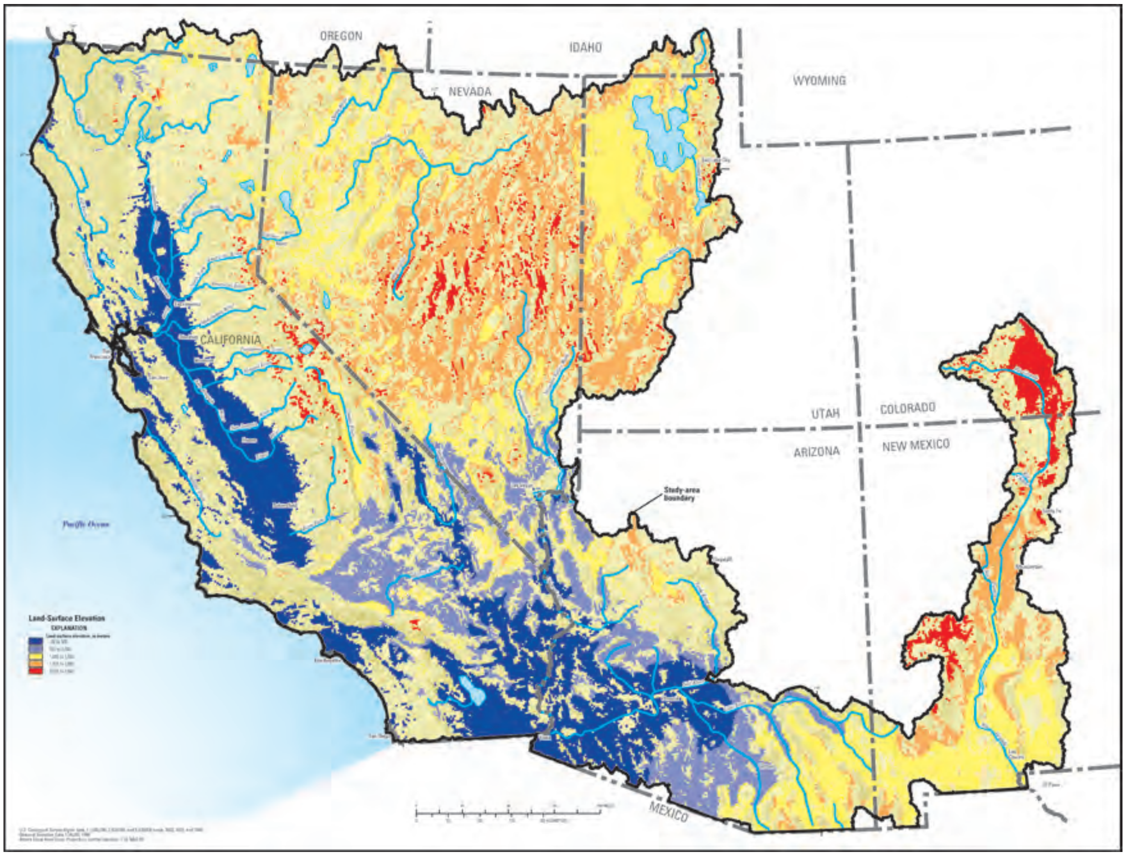
This interactive map has options to display the spatial distribution of selected rock types and their percent area in the bedrock surrounding the alluvial basins, along with selected hydrologic and geographic referencing information.

Download Geologic Source Map from the internet.
This link will take you to *http://pubs.usgs.gov/sir/2012/5065/*.

Link to Geologic Source Map.
In order to link to Geologic Source Map, the map must have already been downloaded and placed in the same directory on your computer as the main report document.

Appendix 14.

Flow path variables used in the random forest classifiers of nitrate and arsenic concentrations in the Southwest Principal Aquifers study area.



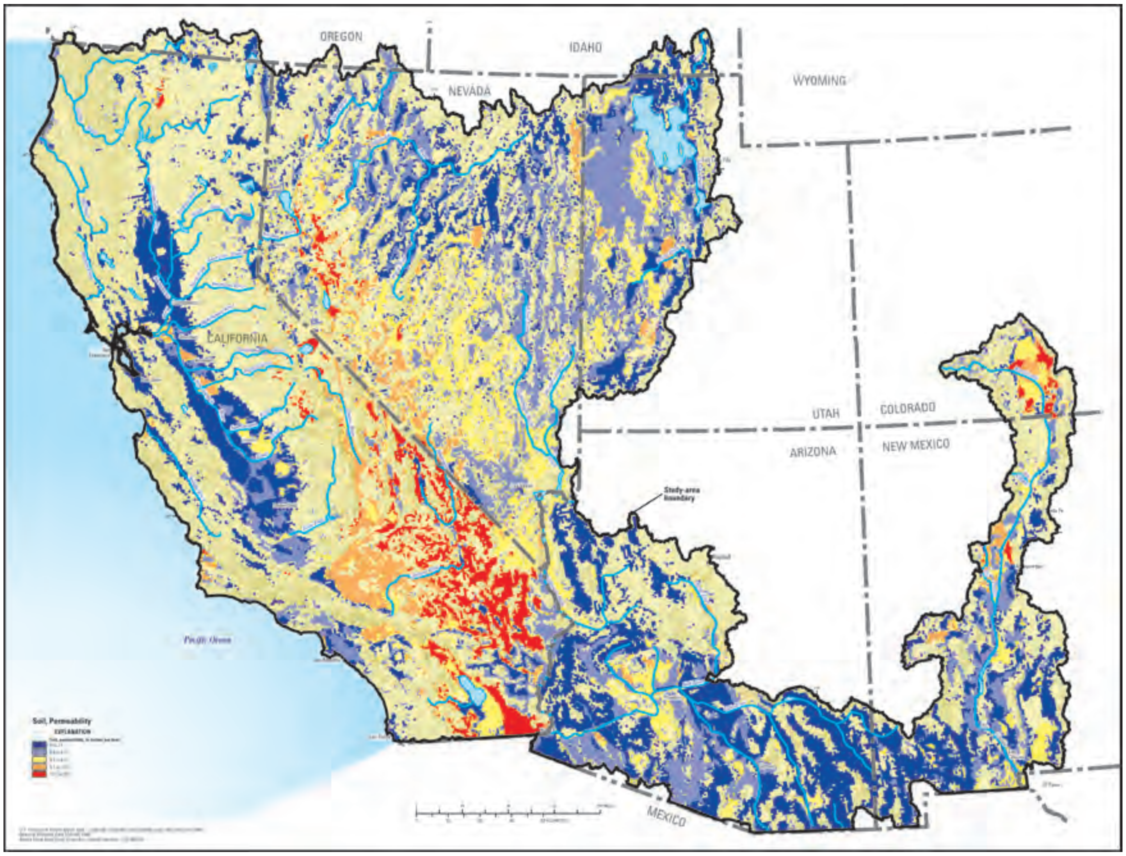
This interactive map has options to display the spatial distribution of land-surface elevation, land-surface elevation percentile, and land-surface slope, along with other selected hydrologic and geographic referencing information.

Download Flow Path Variables Map from the internet.
This link will take you to *http://pubs.usgs.gov/sir/2012/5065/.*

Link to Flow Path Variables Map.
In order to link to Flow Path Variables Map, the map must have already been downloaded and placed in the same directory on your computer as the main report document.

Appendix 15.

Soil property variables used in the random forest classifiers of nitrate and arsenic concentrations in the Southwest Principal Aquifers study area.



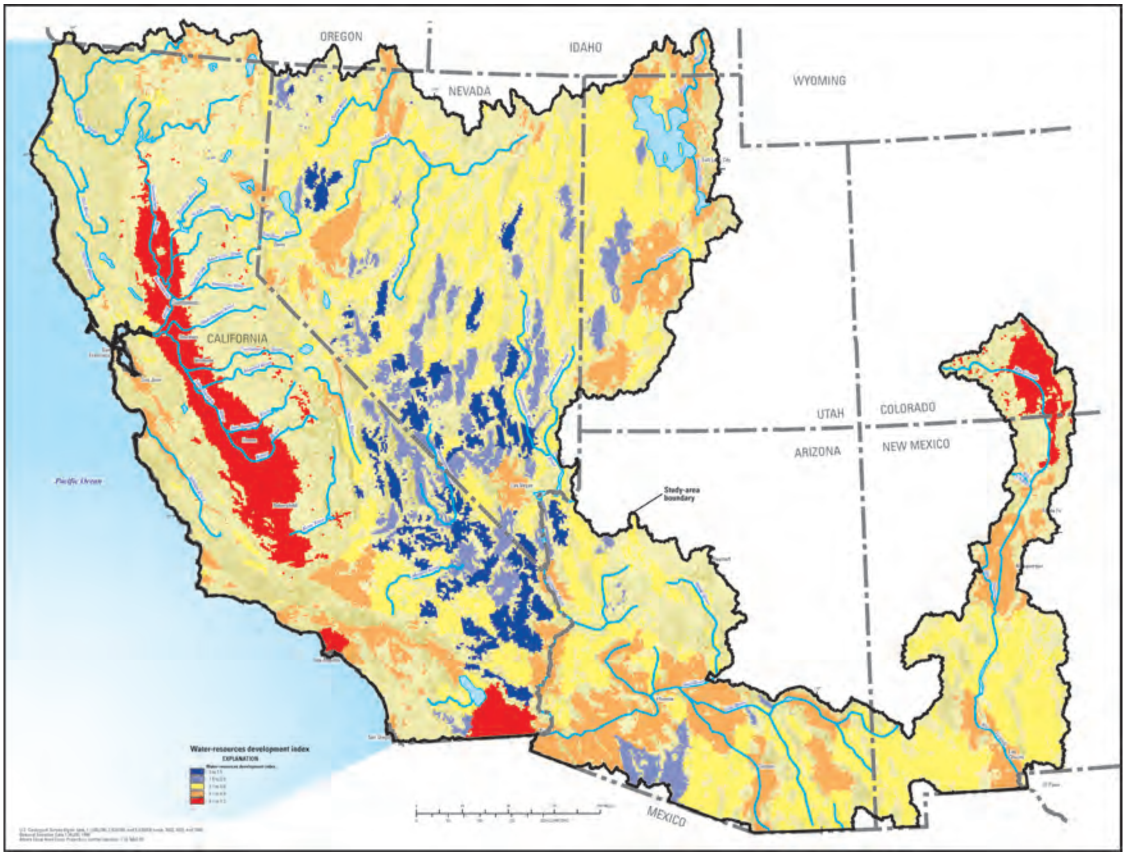
This interactive map has options to display the spatial distribution of several soil properties, including permeability, seasonally high water depth, hydric fraction, and fraction in soil group A, B, C, or D. In addition, these soil property variables can be displayed with selected hydrologic and geographic referencing information.

Download Soil Property Variables Map from the internet.
This link will take you to *http://pubs.usgs.gov/sir/2012/5065/*.

Link to Soil Property Variables Map.
In order to link to Soil Property Variables Map, the map must have already been downloaded and placed in the same directory on your computer as the main report document.

Appendix 16.

Water use variables used in the random forest classifiers of nitrate and arsenic concentrations in the Southwest Principal Aquifers study area.



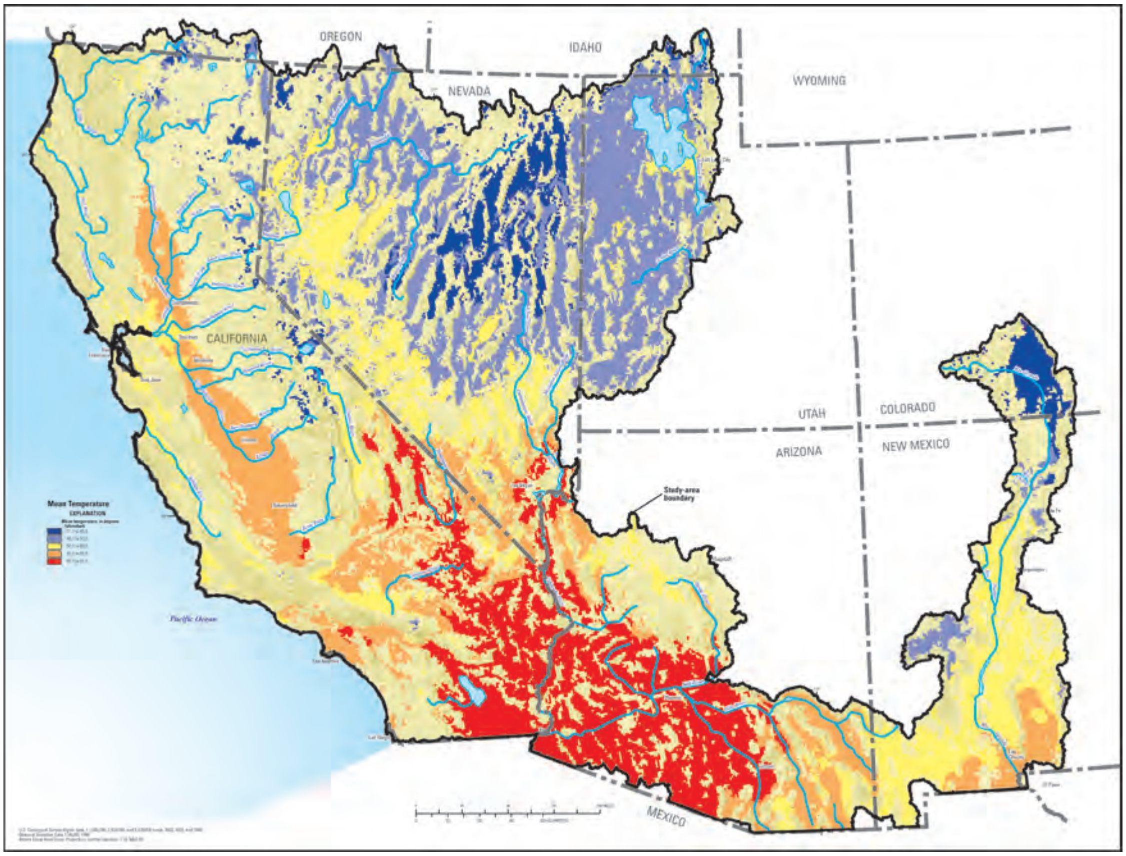
This interactive map has options to display the spatial distribution of groundwater or surface water use for irrigated agriculture or for public water supply, along with the water-resources development index. In addition, these water-use variables can be displayed with other selected hydrologic and geographic referencing information.

Download Water Use Variables Map from the internet.
This link will take you to *http://pubs.usgs.gov/sir/2012/5065/*.

Link to Water Use Variables Map.
In order to link to Water Use Variables Map, the map must have already been downloaded and placed in the same directory on your computer as the main report document.

Appendix 17.

Hydroclimatic variables used in the random forest classifiers of nitrate and arsenic concentrations in the Southwest Principal Aquifers study area.

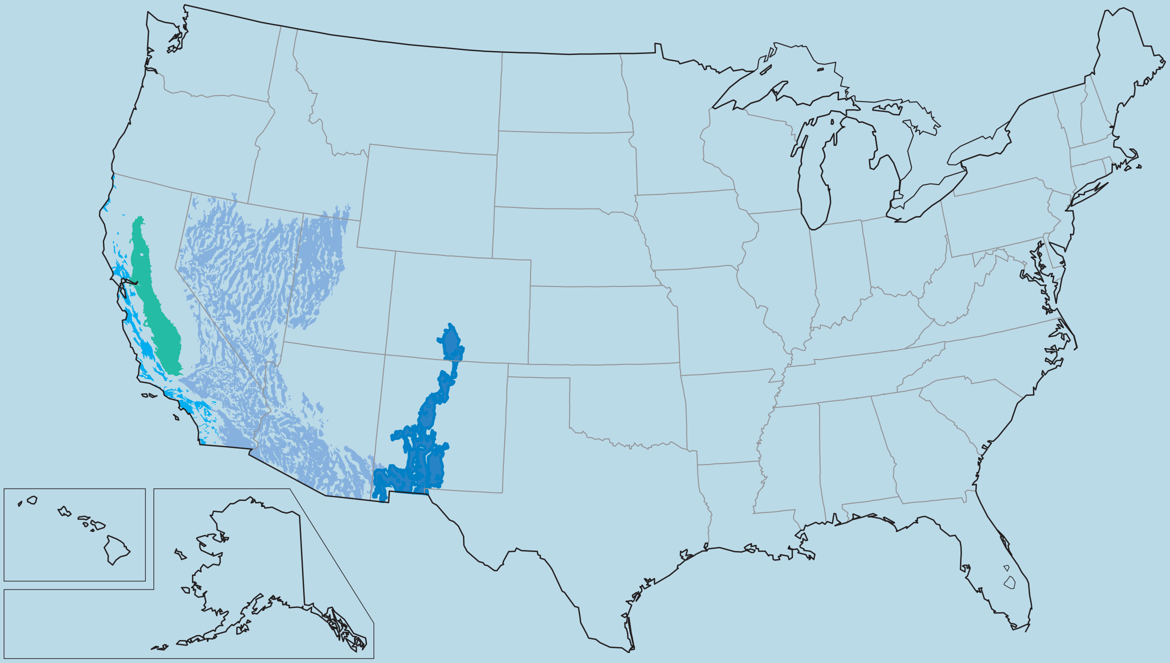


This interactive map has options to display the spatial distribution of mean air temperature, potential evapotranspiration, basin recharge and contributing area recharge, as well as other selected hydrologic and geographic referencing information.

Download Hydroclimatic Variables Map from the internet.
This link will take you to *http://pubs.usgs.gov/sir/2012/5065/*.

Link to Hydroclimatic Variables Map.
In order to link to Hydroclimatic Variables Map, the map must have already been downloaded and placed in the same directory on your computer as the main report document.

Anning and others—**Predicted Nitrate and Arsenic Concentrations in Basin-Fill Aquifers of the Southwestern United States**—Scientific Investigations Report 2012–5065



Southwest Principal Aquifers—Includes (from left to right) California Coastal Basin aquifers, Central Valley aquifer system, Basin and Range basin-fill aquifers, and Rio Grande aquifer system