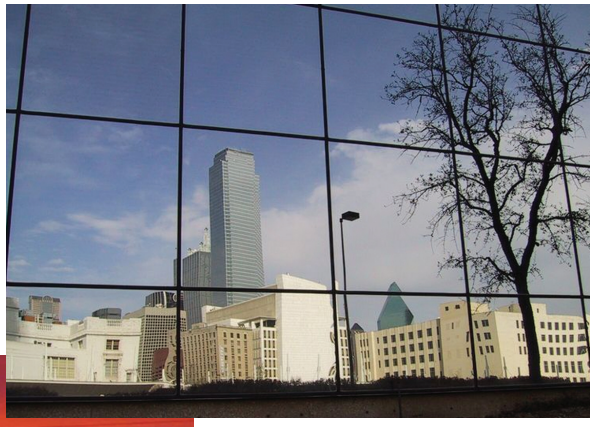
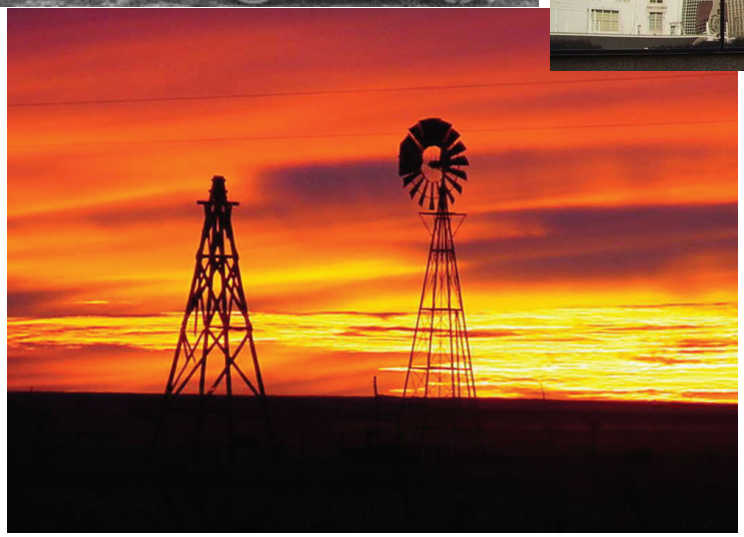
USGS
science for a changing world

Estimating the Susceptibility of Surface Water in Texas to Nonpoint-Source Contamination by Use of Logistic Regression Modeling





Water-Resources Investigations Report 03–4205

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

Estimating the Susceptibility of Surface Water in Texas to Nonpoint-Source Contamination by Use of Logistic Regression Modeling

By William A. Battaglin, Randy L. Ulery, Thomas Winterstein, and Toby Welborn

Water-Resources Investigations Report 03–4205

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

U.S. Department of the Interior
Gale A. Norton, Secretary

U.S. Geological Survey
Charles G. Groat, Director

Contents

Figures

Tables

CONVERSION FACTORS, VERTICAL DATUM, AND ABBREVIATIONS

Multiply	By	To obtain
foot (ft)	0.3048	meter
mile (mi)	1.609	kilometer
square mile (mi^2)	2.590	square kilometer (km^2)

MISCELLANEOUS ABBREVIATIONS AND ABBREVIATED WATER-QUALITY UNITS

meter (m)
microgram per liter (µg/L)
microsiemens per centimeter at 25 degrees Celsius (µS/cm at 25° C)
milligram per liter (mg/L)
milliliter (mL)
picocuries per liter (pCi/L)

MISCELLANEOUS DEFINITIONS

Point source - discharge of a known contaminant from a specific location such as a pipe or smokestack.
Nonpoint source - discharge of a known contaminant from a diffuse source such as a field or urban area.
Relative susceptibility rating - a categorical rating of high, medium, or low that is assigned after analysis of the risk data or other information.

Estimating the Susceptibility of Surface Water in Texas to Nonpoint-Source Contamination by Use of Logistic Regression Modeling

By William A. Battaglin, Randy L. Ulery, Thomas Winterstein, and Toby Welborn

Abstract

In the State of Texas, surface water (streams, canals, and reservoirs) and ground water are used as sources of public water supply. Surface-water sources of public water supply are susceptible to contamination from point and nonpoint sources. To help protect sources of drinking water and to aid water managers in designing protective yet cost-effective and risk-mitigated monitoring strategies, the Texas Commission on Environmental Quality and the U.S. Geological Survey developed procedures to assess the susceptibility of public water-supply source waters in Texas to the occurrence of 227 contaminants. One component of the assessments is the determination of susceptibility of surface-water sources to nonpoint-source contamination. To accomplish this, water-quality data at 323 monitoring sites were matched with geographic information system-derived watershed-characteristic data for the watersheds upstream from the sites. Logistic regression models then were developed to estimate the probability that a particular contaminant will exceed a threshold concentration specified by the Texas Commission on Environmental Quality. Logistic regression models were developed for 63 of the 227 contaminants. Of the remaining contaminants, 106 were not modeled because monitoring data were available at less than 10 percent of the monitoring sites; 29 were not modeled because there were less than 15 percent detections of the contaminant in the monitoring data; 27 were not modeled because of the lack of any monitoring data; and 2 were not modeled because threshold values were not specified.

Introduction

The 1996 Amendment to the Safe Drinking Water Act (U.S. Environmental Protection Agency, 2003) requires that each State prepare a source-water assessment for all public water supplies (PWS). For the assessment, States are required to identify PWS sources, which contaminants are monitored, which contaminants should be monitored, and the intrinsic susceptibility of the source water to contamination. The susceptibility of PWS source water is defined as the potential for a PWS to withdraw water containing a listed contaminant at a concentration that could pose a public health concern. Surface-water sources of public water supply (streams, canals, and reservoirs) are susceptible to contamination from point (discrete) and nonpoint (spatially diffuse) sources.

Texas Source Water Assessment Project

The Texas Commission on Environmental Quality (TCEQ) and the U.S. Geological Survey (USGS) developed procedures to assess the susceptibility of PWS source waters to selected drinking-water contaminants. The procedures as implemented in automated software will better enable TCEQ staff and others to (1) identify the most susceptible PWS's on which to focus source-water protection efforts, (2) reduce monitoring costs associated with ensuring safe drinking water, (3) enhance public understanding of source-water protection, and (4) identify land-management practices needed to protect source waters (Ulery, 2000; http://www.tnrcc.state.tx.us/permitting/waterperm/pdw/swap/, accessed 8/1/2003).

A surface-water susceptibility assessment for a PWS has eight components. Several of the components result in relative susceptibility ratings of high, medium, or low. These ratings are designed to support an overall susceptibility rating for the PWS sources and system. The components are:

1. **Identification Component**—determination of hydrologic attributes (for example, topography and drainage network) used to delineate the area(s) that potentially contribute source water for a PWS.
2. **Delineation Component**—delineation of PWS source area(s) based on attributes determined by the identification component.
3. **Intrinsic Susceptibility Component**—determination of the intrinsic attributes (for example, soil erodibility or mean annual runoff) within PWS source areas and evaluation of their potential to make the PWS more or less susceptible to contamination. This component produces a relative susceptibility rating for all contaminants.

4. **Nonpoint-Source Susceptibility Component**—determination of statistical relations between measured contaminants in surface water and watershed characteristics (for example, land use or pesticide use) within PWS source areas. Statistical model results will determine the probability that a drinking-water contaminant will exceed a TCEQ-specified threshold concentration. In cases of few or no data, a relative susceptibility rating is assigned and is based on the frequency of contaminant detection or level of land-use activities within the PWS source area associated with contaminant production or use.
5. **Point-Source Susceptibility Component—**determination of potential point sources of contaminants to the PWS. A relative susceptibility rating is produced for each contaminant and is based on the permitted discharges from known point-source locations (for example, oil refineries). Estimates of contaminant concentrations at low streamflow will be calculated from permitted releases from point sources within the PWS source area. A relative susceptibility rating is assigned to each contaminant and is based on the ratio of total permitted releases to mean annual streamflow.
6. **Area of Primary Influence Component**—determination of the density and proximity of threatening human activities (for example, pipeline or railroad) and associated contaminants within a limited area (2-hour travel or 1,000-foot buffer) around the PWS surface-water intake (or reservoir). A relative susceptibility rating is produced for each contaminant and is based on number, type, size, and location of threatening activities in the PWS source area.
7. **Contaminant Occurrence Component**—determination of contaminant detections (at or greater than TCEQ-specified threshold concentrations) at water-quality monitoring sites or other nearby PWS. A relative susceptibility rating of "high" is assigned to the contaminant if it has been detected greater than the threshold concentration at a nearby site. No relative susceptibility rating is assigned to the contaminant if it has not been detected greater than the contaminant threshold concentration.
8. **Susceptibility Summary Component**—combination of the seven components into a summary relative susceptibility rating for each PWS source and a single summary rating for the PWS.

Purpose and Scope

The purpose of this report is to describe the data, process, and criteria used to estimate the susceptibility of surface water in Texas to nonpoint-source contamination by use of logistic regression models. The nonpoint-source susceptibility component (component number 4 above) of PWS susceptibility assessments in Texas is addressed by estimating the probability that specified drinking-water contaminants will occur in surface water at or greater than a threshold concentrations specified by the TCEQ. This report focuses on the contaminants for which models are developed and does not discuss in detail how relative susceptibility ratings are assigned to those contaminants that are not modeled. The models developed are specific to Texas, but similar methods could be applied in other locations if sufficient monitoring data are available.

Surface-Water Nonpoint-Source Component

The nonpoint-source susceptibility component (component 4) is addressed statistically by investigating the relations between known contaminant occurrences in surface water and watershed characteristics, either natural or anthropogenic, quantified for the PWS contributing source areas (watersheds). Where monitoring data are sufficient in the State, logistic regression models quantify the relations and provide a means to transfer results statistically from monitored sites to unmonitored PWS intakes. For this analysis, "sufficient" is defined as 50 or more contaminant analyses from 10 percent or more of the monitoring sites with at least 15 percent of sites having contaminant detections at or greater than the TCEQ-specified threshold concentration. The statistical analysis produces logistic regression models that estimate the probability (p) of contaminant detection at or greater than the TCEQ-specified threshold concentration. These probability values are used to assign a relative susceptibility rating as follows:

$$\text{Susceptibility rating is high if } p_{Contaminant} \geq 0.67$$

$$\text{Susceptibility rating is medium if } 0.33 < p_{Contaminant} < 0.67$$

$$\text{Susceptibility rating is low if } p_{Contaminant} \leq 0.33$$

In instances of few or no monitoring data, the nonpoint-source relative susceptibility rating is assigned on the basis of either the frequency of contaminant detections or the level of land-use activities associated with contaminant production or use. When there are less than 50 samples with analyses for the contaminant or those analyses are from less than 10 percent of sites, then a relative susceptibility rating is assigned on the basis of the detection frequency of the contaminant in those samples. When there are few or no monitoring data available for a particular contaminant, a relative susceptibility rating is assigned on the basis of the prevalence of land uses or practices that are likely to be associated with the use or occurrence of the contaminant. This report does not discuss in detail how relative susceptibility ratings are assigned to those contaminants that are not modeled with logistic regression.

Surface-Water Sites

The USGS maintains a surface-water data-collection network in Texas that has more than 500 sites on rivers and reservoirs (Gandara and Barbie, 2001). The majority of data collected at these sites are related to streamflow, but some water-quality data are available at many sites. For the purposes of this report, a set of 323 sites was selected to represent surface-water quality in Texas. Most sites are at USGS gages, but some are PWS reservoirs that were sampled for selected contaminants in 1999 and 2000, and some are sites sampled by the TCEQ. The selected sites are required to have some water-quality data associated with them, and then they are selected to maximize spatial coverage of Texas.

Water-Quality Data

A total of 227 drinking-water contaminants (for example, lead) and water-quality properties (for example, pH) are identified by the TCEQ as the subjects for this report (table 1). Most of the 227 contaminants are not naturally occurring (for example, pesticides like atrazine) but some (for example, iron) are, and some (for example, pH) are really water-quality properties and not contaminants. The contaminants or properties can be grouped into nine broad classes: inorganics, major ions, metals, microbiologicals, nutrients, organics (other than pesticides), pesticides, physical properties, and radionuclides (table 1).

All water-quality data collected in the past 20 years associated with the 323 surface-water sites from the USGS and TCEQ databases are aggregated into a single database of approximately 2 million records. Duplicate data records and records reported with a remark code that questioned data reliability are eliminated. The data then are used to assign a contaminant occurrence code to each site for each contaminant. The codes are assigned as follows:

–1 = the contaminant was never analyzed for at the site;

0 = the contaminant was analyzed for, but never detected at the site;

1 = the contaminant was analyzed for and detected, but always at a concentration less than the TCEQ-specified threshold; and

2 = the contaminant was analyzed for and detected one or more times at a concentration equal to or greater than the threshold.

This coding process results in a binary response (yes = 2 or no =1 , 0, or –1) to the question, "Has contaminant A occurred at or greater than concentration B at site C?" This binary response is the dependent variable in the logistic regression models. Initial contaminant threshold concentrations were provided by the TCEQ. These initial thresholds occasionally are modified to provide a more suitable data set for logistic regression.

Watershed-Characteristics Data

Data on land use, soil characteristics, population, agricultural chemical use, manure production, and land and stream physiography (Brown and others, 2000) are quantified within the watersheds of the 323 surface-water sites by using a geographic information system (GIS). Most of the resulting watershed characteristics are expressed as either a percentage of the basin areas (for example, percent land use 41 – deciduous forest) or as a basin-area averaged input (for example, atrazine use rate in kilograms per square kilometer). Land-cover characteristics are derived from the 1992 Multi-Resolution Landscape Characterization (MRLC) data set, classified in accordance with the 1992 National Land Cover Dataset (NLCD) land-cover definitions (U.S. Geological Survey, 2000). Average soil characteristics are derived by area weighting soils data (vertically and horizontally) from the U.S. Department of Agriculture (USDA), Natural Resources Conservation Service State Soil Geographic Data Base (STATSGO) (U.S. Department of Agriculture, 1995). Population density estimates are computed from 1990 Census Blocks (U.S. Department of Commerce, 1990), and oil- and gas-well density estimates are obtained from the TCEQ Potential Sources of Contamination (PSOC) database. Agricultural chemical use, fertilizer use, and manure-production estimates are calculated from various sources (Battaglin and Goolsby, 1994; Goolsby and others, 1999). Land and stream physiography are derived from the National Elevation Dataset (NED) digital elevation model (DEM) data (Brown and others, 2000). Watershed characteristics are computed for individual basins by overlaying drainage basin boundaries on all nonpoint-source data sets and computing the spatially averaged values with the GIS.

Subsequently, all GIS-derived watershed characteristics are statistically tested for their ability to predict the occurrence of the 227 contaminants using logistic regression. A total of 25 land use, 6 soils, 2 population, 44 agricultural chemical use, 2 manure production, and 5 land and stream physiography variables are tested for potential use in the logistic regression models. All of the watershed characteristics have approximately normal distributions, so the values are not transformed to improve their statistical properties. The 39 characteristics listed in table 2 are those used in one or more of the logistic regression models developed.

Many of the watershed characteristics are related to or correlated (collinear) with each other. For example, the percentage of high-intensity residential land in the 323 watersheds is correlated with population density in the watersheds, having a Pearson product-moment correlation (Helsel and Hirsch, 1992) of 0.86. Including strongly correlated explanatory variables in multivariate statistical models can have undesirable consequences in regression equations such as coefficients with unrealistic signs or unstable values (Helsel and Hirsch, 1992). For the results presented in this report, if the correlation between two watershed characteristics was 0.7 or greater, those two characteristics were not included in the same regression model.

Table 1. Chemical class, threshold concentration, and model status for 227 surface-water contaminants and water-quality properties in Texas.

[ND, not modeled - no data; LA, not modeled ñ limited analyses; MD, logistic models developed; LD, not modeled ñ limited detections; NT, not modeled ñ no threshold specified; µg/L, micrograms per liter; pCi/L, picocuries per liter; mg/L, milligrams per liter; µS/cm, microsiemens per centimeter at 25 degrees Celsius; mL, milliliter]

Contaminant or property	Threshold concentration	Model status
Inorganics		
Asbestos	1.0 µg/L	ND
Hydrogen sulfide	1.0 mg/L	ND
RDX	0.01 µg/L	ND
Major Ions		
Alkalinity	250 mg/L	MD
Bicarbonate	200 mg/L	MD
Bromide	1.0 µg/L	LA
Calcium	100 mg/L	MD
Carbonate	1.0 mg/L	MD
Chloride	125 mg/L	MD
Cyanide	0.01 µg/L	LA
Fluoride	1.0 mg/L	MD
Hardness	300 mg/L	MD
P-alkalinity	1.0 mg/L	ND
Sodium	75 mg/L	MD
Sulfate	100 mg/L	MD
Metals		
Aluminum	100 µg/L	MD
Antimony	0.01 µg/L	MD
Arsenic	5.0 µg/L	MD
Barium	200 µg/L	MD
Beryllium	1.0 µg/L	MD
Boron	200 µg/L	MD
Cadmium	5.0 µg/L	MD
Chromium	15 µg/L	MD
Copper	20 µg/L	MD
Iron	150 µg/L	MD
Lead	25 µg/L	MD
Magnesium	20 mg/L	MD
Manganese	75 µg/L	MD
Mercury	0.5 µg/L	MD
Nickel	10 µg/L	MD
Selenium	2.0 µg/L	MD
Silver	2.0 µg/L	MD
Thallium	0.1 µg/L	LA
Zinc	50 µg/L	MD
Microbiologicals		
Cryptosporidium parvum	None	ND
Escherichia Coli	None	ND
Fecal virus	None	ND

Table 1. Chemical class, threshold concentration, and model status for 227 surface-water contaminants and water-quality properties in Texas.—Continued

[ND, not modeled - no data; LA, not modeled ñ limited analyses; MD, logistic models developed; LD, not modeled ñ limited detections; NT, not modeled ñ no threshold specified; µg/L, micrograms per liter; pCi/L, picocuries per liter; mg/L, milligrams per liter; µS/cm, microsiemens per centimeter at 25 degrees Celsius; mL, milliliter]

Contaminant or property	Threshold concentration	Model status
Microbiologicals—Continued		
Giardia lamblia	None	ND
Total coliform	20,000/100 mL	MD
Nutrients		
Nitrate	2.0 mg/L	MD
Nitrate + nitrite	2.0 mg/L	MD
Nitrite	0.1 mg/L	MD
Organics		
1,1,1,2-tetrachloroethane	0.1 µg/L	ND
1,1,1-trichloroethane	0.1 µg/L	LA
1,1,2,2-tetrachloroethane	0.1 µg/L	LA
1,1,2-trichloroethane	0.1 µg/L	LA
1,1-dichloroethane	0.1 µg/L	LA
1,1-dichloroethylene	0.1 µg/L	LA
1,1-dichloropropene	0.1 µg/L	LA
1,2,3-trichlorobenzene	0.1 µg/L	LA
1,2,3-trichloropropane	0.1 µg/L	LA
1,2,4-trichlorobenzene	0.1 µg/L	LA
1,2,4-trimethylbenzene	0.1 µg/L	LA
1,2-dichloroethane	0.1 µg/L	LA
1,2-dichloropropane	0.1 µg/L	LA
1,2-diphenylhydrazine	0.1 µg/L	LA
1,3,5-trimethylbenzene	0.1 µg/L	LA
1,3- dichlorobenzene	0.1 µg/L	LA
1,3- dichloropropane	0.1 µg/L	LA
2,2- dichloropropane	0.1 µg/L	LA
2,3,7,8-TCDD (dioxin)	0.1 µg/L	ND
2,4,6- trichlorophenol	0.1 µg/L	LA
2,4-dichlorophenol	0.1 µg/L	LA
2,4-dinitrophenol	0.1 µg/L	LA
2,4-dinitrotoluene	0.1 µg/L	LA
2,6-dinitrotoluene	0.1 µg/L	LA
2-chlorotoluene	0.1 µg/L	LA
2-hexanone	0.1 µg/L	LA
2-methylphenol	0.1 µg/L	ND
4-chlorotoluene	0.1 µg/L	LA
4-isopropyltoluene	0.1 µg/L	LA
4-methyl-2-pentanone	0.1 µg/L	LA
Acenaphthene	0.1 µg/L	LA

Table 1. Chemical class, threshold concentration, and model status for 227 surface-water contaminants and water-quality properties in Texas.—Continued

[ND, not modeled - no data; LA, not modeled ñ limited analyses; MD, logistic models developed; LD, not modeled ñ limited detections; NT, not modeled ñ no threshold specified; µg/L, micrograms per liter; pCi/L, picocuries per liter; mg/L, milligrams per liter; µS/cm, microsiemens per centimeter at 25 degrees Celsius; mL, milliliter]

Contaminant or property	Threshold concentration	Model status
Organics—Continued		
Acenaphthylene	0.1 µg/L	LA
Acetone	0.01 µg/L	LA
Acrylonitrile	0.01 µg/L	LA
Anthracene	0.1 µg/L	LA
Aroclor (pcb)	0.1 µg/L	ND
Benzene	0.01 µg/L	MD
Benzo(A)pyrene	0.01 µg/L	LA
Benzo(A)anthracene	0.01 µg/L	LA
Benzo(B)fluoranthene	0.01 µg/L	LA
Benzo(G,H,I)perylene	0.01 µg/L	LA
Benzo(K)fluoranthene	0.01 µg/L	LA
Bromobenzene	0.01 µg/L	LA
Bromochloromethane	0.01 µg/L	LA
Bromodichloro-methane	0.01 µg/L	LA
Bromoform	0.01 µg/L	LA
Butyl benzyl phthalate	0.01 µg/L	LA
Carbon tetrachloride	0.01 µg/L	LA
Chlorobenzene	0.01 µg/L	LA
Chloroethane	0.01 µg/L	LA
Chloroform	0.1 µg/L	LA
Chloromethane (methyl chloride)	0.01 µg/L	LA
Chrysene	0.1 µg/L	LA
Cis-1,2-dichloroethylene	0.1 µg/L	LA
Cis-1,3-dichloropropene	0.1 µg/L	LA
Di-(2-ethylhexyl) adipate	0.01 µg/L	ND
Di-(2-ethylhexyl) phthalate	0.01 µg/L	ND
Dibenz(A,H)anthracene	0.1 µg/L	LA
Dibromochloro-methane	0.1 µg/L	LA
Dibromochloro-propane	0.1 µg/L	LA
Dibromomethane	0.1 µg/L	LA
Dichlorodifluoro-methane	0.1 µg/L	LA
Diethyl phthalate	0.1 µg/L	LA
Dimethyl phthalate	0.1 µg/L	LA
Di-N-butyl phthalate	0.1 µg/L	LA
Ethyl methacrylate	0.1 µg/L	LA
Ethylbenzene	0.1 µg/L	LD
Ethylene dibromide	0.1 µg/L	LA
Fluorine	0.1 µg/L	LA
Hexachlorobenzene	0.1 µg/L	LA

Table 1. Chemical class, threshold concentration, and model status for 227 surface-water contaminants and water-quality properties in Texas.—Continued

[ND, not modeled - no data; LA, not modeled ñ limited analyses; MD, logistic models developed; LD, not modeled ñ limited detections; NT, not modeled ñ no threshold specified; µg/L, micrograms per liter; pCi/L, picocuries per liter; mg/L, milligrams per liter; µS/cm, microsiemens per centimeter at 25 degrees Celsius; mL, milliliter]

Contaminant or property	Threshold concentration	Model status
Organics—Continued		
Hexachlorobutadiene	0.1 µg/L	LA
Hexachlorocyclo-pentadiene	0.1 µg/L	LA
Indeno(1,2,3,CD)pyrene	0.1 µg/L	LA
Iodomethane (methyl iodide)	0.1 µg/L	LA
Isopropylbenzene	0.1 µg/L	LA
M+P xylene	0.1 µg/L	MD
Methyl ethyl ketone	0.1 µg/L	LA
Methyl methacrylate	0.1 µg/L	LA
Methyl-tert-butyl ether (MTBE)	0.2 µg/L	MD
Monochlorobenzene	0.01 µg/L	LA
M-xylene	0.1 µg/L	MD
Naphthalene	0.1 µg/L	LA
N-butylbenzene	0.1 µg/L	LA
Nitrobenzene	0.1 µg/L	LA
N-propylbenzene	0.1 µg/L	LA
Orth-1,2-dichlorobenzene	0.1 µg/L	LA
O-xylene	0.01 µg/L	MD
Para-1,4-dichlorobenzene	0.1 µg/L	LA
PCBs	0.01 µg/L	MD
Pentachlorophenol	0.1 µg/L	LA
Perchlorate	0.01 µg/L	ND
Phenanthrene	0.01 µg/L	LA
P-xylene	0.1 µg/L	MD
Pyrene	0.1 µg/L	LA
S-butylbenzene	0.01 µg/L	LA
Styrene	0.01 µg/L	LA
T-butylbenzene	0.1 µg/L	LA
Tetrachloro-ethylene	0.1 µg/L	LA
Tetrahydrofuran	0.1 µg/L	LA
Toluene	0.01 µg/L	MD
Total trihalomethane	0.01 µg/L	ND
Trans-1,2-dichloroethylene	0.1 µg/L	LA
Trans-1,3-dichloropropene	0.1 µg/L	LA
Trichloroethylene	0.1 µg/L	LA
Trichlorofluoro-methane	0.1 µg/L	LA
Vinyl acetate	0.1 µg/L	LA
Vinyl chloride	0.1 µg/L	LA
Xylenes (total)	0.1 µg/L	MD

Table 1. Chemical class, threshold concentration, and model status for 227 surface-water contaminants and water-quality properties in Texas.—Continued

[ND, not modeled - no data; LA, not modeled ñ limited analyses; MD, logistic models developed; LD, not modeled ñ limited detections; NT, not modeled ñ no threshold specified; µg/L, micrograms per liter; pCi/L, picocuries per liter; mg/L, milligrams per liter; µS/cm, microsiemens per centimeter at 25 degrees Celsius; mL, milliliter]

Contaminant or property	Threshold concentration	Model status
Pesticides		
1,3- dichloropropene	0.1 µg/L	LA
2,4,5-T	0.01 µg/L	MD
2,4,5-TP (silvex)	0.01 µg/L	MD
2,4-D	0.01 µg/L	MD
3-hydroxycarbofuran	0.1 µg/L	LD
Acetochlor	0.01 µg/L	LD
Alachlor	0.01 µg/L	MD
Aldicarb	0.01 µg/L	LD
Aldicarb sulfone	0.01 µg/L	LD
Aldicarb sulfoxide	0.01 µg/L	LD
Aldrin	0.01 µg/L	MD
Atrazine	0.1 µg/L	MD
Bentazon	0.01 µg/L	LD
Bromacil	0.01 µg/L	LD
Bromomethane (methyl bromide)	0.01 µg/L	LA
Butachlor	0.01 µg/L	LA
Carbaryl	0.01 µg/L	MD
Carbofuran	0.01 µg/L	LD
Carbon disulfide	0.01 µg/L	LA
Chlordane	0.01 µg/L	MD
Cis-chlordane (alpha-chlordane)	0.01 µg/L	LA
Trans-chlordane (gamma-chlordane)	0.01 µg/L	LA
Chlordane (trans-nonachlor)	0.01 µg/L	ND
Cyanazine	0.01 µg/L	LD
Dalapon	0.01 µg/L	ND
DCPA di-acid degradate	0.001 µg/L	MD
DCPA mono-acid degradate	0.01 µg/L	LD
DDE	0.001 µg/L	MD
Diazinon	0.01 µg/L	MD
Dicamba	0.01 µg/L	LD
Dichloromethane (methylene chloride)	0.1 µg/L	LA
Dieldrin	0.01 µg/L	MD
Dinoseb	0.01 µg/L	LD
Diquat	0.01 µg/L	ND
Disulfoton	0.01 µg/L	LD
Diuron	0.01 µg/L	MD
Endothall	0.01 µg/L	ND
Endrin	0.01 µg/L	LD
EPTC	0.01 µg/L	LD

Table 1. Chemical class, threshold concentration, and model status for 227 surface-water contaminants and water-quality properties in Texas.—Continued

[ND, not modeled - no data; LA, not modeled ñ limited analyses; MD, logistic models developed; LD, not modeled ñ limited detections; NT, not modeled ñ no threshold specified; µg/L, micrograms per liter; pCi/L, picocuries per liter; mg/L, milligrams per liter; µS/cm, microsiemens per centimeter at 25 degrees Celsius; mL, milliliter]

Contaminant or property	Threshold concentration	Model status
Pesticides—Continued		
Fonofos	0.01 µg/L	LD
Glyphosate	0.01 µg/L	LA
Heptachlor	0.01 µg/L	MD
Heptachlor epoxide	0.01 µg/L	MD
Lambast	0.01 µg/L	LA
Lindane	0.01 µg/L	MD
Linuron	0.01 µg/L	LD
Methiocarb	0.01 µg/L	LD
Methomyl	0.01 µg/L	LD
Methoxychlor	0.01 µg/L	LD
Metolachlor	0.01 µg/L	MD
Metribuzin	0.01 µg/L	MD
Molinate	0.01 µg/L	LD
Organotins	0.1 µg/L	ND
Oxamyl	0.1 µg/L	LD
Picloram	0.01 µg/L	LD
Prometon	0.01 µg/L	MD
Propachlor	0.01 µg/L	LA
Propazine	0.01 µg/L	LD
Simazine	0.01 µg/L	MD
Terbacil	0.01 µg/L	LD
Terbufos	0.01 µg/L	LD
Toxaphene	0.01 µg/L	LD
Triazines	None	NT
Trifluralin	0.001 µg/L	LD
Physical Properties		
pH	None	NT
Specific conductance	1,500 µS/cm at 25^{o}C	MD
Total dissolved solids	500 mg/L	MD
Radionuclides		
Gross Alpha	1.0 pCi/L	LA
Gross Beta	1.0 pCi/L	LA
Radium-226	0.1 pCi/L	LA
Radium-228	0.1 pCi/L	ND
Radon	0.1 pCi/L	ND
Strontium-89	1.0 pCi/L	ND
Strontium-90	1.0 pCi/L	ND
Total alpha-emitting radium	1.0 pCi/L	ND
Tritium	1.0 pCi/L	ND
Uranium	2.0 µg/L	MD

Table 2. Watershed characteristics used in surface-water contaminant occurrence logistic regression models.

Variable	Definition
p21	Percent land use 21 – low intensity residential
p22	Percent land use 22 – high intensity residential
p23	Percent land use 23 – commercial/industrial/transportation
p41	Percent land use 41 – deciduous forest
p42	Percent land use 42 – evergreen forest
p43	Percent land use 43 – mixed forest
p51	Percent land use 51 – shrubland
p61	Percent land use 61 – orchards/vineyards
p71	Percent land use 71 – grasslands/herbaceous
p81	Percent land use 81 – pasture/hay
p82	Percent land use 82 – row crops
p83	Percent land use 83 – small grains
p85	Percent land use 85 – urban/recreational grasses
p91	Percent land use 91 – woody wetlands
p92	Percent land use 92 – emergent herbaceous wetlands
p_ag	Percent agricultural land – sum of p81, p82, p83, p84, p85, and p61
p_clay	Total clay content of soil expressed as percent clay in material less than 2 millimeters in size.
p_drain	Percent artificially drained agricultural land
p_forest	Percent forest land – sum of p41, p42, and p43
p_org	Total soil organic matter in percent by weight.
p_urban	Percent urban land – sum of p21, p22, and p23
p_wet	Percent wetlands – sum of p91 and p92
ala_use	Alachlor use rate in kilograms per square kilometer
atr_use	Atrazine use rate in kilograms per square kilometer
BS	Average basin slope in feet per mile
bulk_d	Average soil bulk density in grams per cubic centimeter
DCPA_use	DCPA use rate in kilograms per square kilometer
diuron_use	Diuron use rate in kilograms per square kilometer
fert_n	Fertilizer nitrogen use in metric tons per square kilometer
hydgrp	Average soil hydrologic group
kfact	Average soil total K factor
manure_n	Manure N production in metric tons per square kilometer
manure_p	Manure P production in metric tons per square kilometer
MCS	Main channel slope in feet per mile
perm	Average soil permeability in inches per hour
popdn	Basin population density in people per square mile
sod-chl_use	Sodium-chloride use rate in kilograms per square kilometer
TDA	Total drainage area in square miles
welldn	Density of oil- and gas-wells in number per square kilometer

Logistic Regression Models

Logistic regression is a statistical technique that uses one or more independent explanatory variables to predict the probability of a categorical response. Logistic regression is well suited for modeling the probability or likelihood of the occurrence or nonoccurrence of some event (Hosmer and Lemeshow, 2000). Used primarily in medical research, logistic regression is gaining popularity in environmental investigations (Battaglin and Goolsby, 1998; Davidson and others, 2001; Nolan, 2001; Rupert, 2003; Smith and others, 2001; Mineau, 2002).

The response or dependent variable in logistic regression is the logarithm of the odds ratio p/(1-p), where p is the probability of a data value being in one of the categories (Helsel and Hirsch, 1992). The logit transformation converts a variable constrained between zero and one into a continuous variable that is linear with respect to the vector of the explanatory variables, X:

$$Y = \text{Log}\,[p/(1-p)] = b_0 + b_1X_1 + \ldots b_kX_k \quad (1)$$

where

Y is the response variable; [p/(1-p)] is the odds ratio;

b_0 is the intercept;

b_1 is the slope coefficient for the first explanatory variable X_1; and

b_k is the slope coefficient for the kth explanatory variable X_k.

Slope coefficients for logistic regression models are fit to the categorical response variable by using the method of maximum likelihood, which optimizes the probability that the observed data will be estimated from the set of slope coefficients. To calculate the predicted value of the response variable (a probability), the logistic transformation (the inverse of the logit transformation) is used. The resulting equation has the form:

$$p = \frac{(\exp(b_0 + b_1X_1 + ..b_kX_k))}{(1+(\exp(b_0 + b_1X_1 + ..b_kX_k)))} \quad (2)$$

The SAS statistical software system (SAS Institute, 1990) is used to develop the logistic regression models. The "Best" and the "Stepwise" model-selection procedures are used to screen potential models. The "Best" procedure tests all possible combinations of model variables and lists the 10 best 1 through 5 variable models. The "Stepwise" procedure starts with one variable and then adds in the next most significant variable. After each "Step," the remaining potential variables are tested for their ability to explain the remaining model error. The variable that explains the most of the remaining model error is the next to be added. Neither the "Best" nor the "Stepwise" model-selection procedures will prevent correlated explanatory variables from entering the models, so supervision of the process is required.

Logistic regression models for individual contaminants are compared on the basis of model fit statistics such as the Akaike information criterion (AIC) and the Schwartz criterion (SAS Institute, 1990). The AIC is analogous to Mallow's Cp and includes a measure of model error and a penalty for extra variables (Helsel and Hirsch, 1992). The better models have lower AIC values, but AIC values for models of different contaminants cannot be compared unless they have identical source data. The predictive abilities of the models are compared on the basis of the Somers' D (SAS Institute, 1990) measure of rank correlation between the observed and predicted responses, and by contingency table analysis of predicted compared to observed responses. Somers' D values range from –1 to 1; the higher values indicate better model predictive abilities. Model fit and predictive ability are used to select the best two models for each contaminant. The final logistic regression models should not be affected by collinearity (correlation) among the explanatory variables because variables that are highly correlated (Pearson's correlation coefficient of 0.7 or greater) are not included in the same model.

Optimal performance of logistic regression models is obtained when there are nearly equal numbers of responses for the dichotomous response variable. For instances when there are either few detections greater than the threshold concentration or few nondetections, the resulting logistic regression models are subject to fitting problems, and the likelihood of developing spurious models increases (Hosmer and Lemeshow, 2000).

Logistic regression models are developed that predict the probability of occurrence at or greater than the TCEQ-specified threshold concentration for 63 of the 227 contaminants. For the remaining 164 contaminants, logistic regression models for susceptibility of occurrence in surface water from nonpoint sources cannot be produced. These contaminants are not modeled for one of the following mutually exclusive reasons:

1. There are fewer than 50 sample analyses total at the 323 surface-water sites; hence, there is not enough analytical data to justify model development.
2. There are less than 10 percent (~32) of the 323 surface-water sites with analytical data; hence, there is not enough spatial coverage of data to justify model development.
3. There are more than 50 samples total, at more than 32 sites, but the target contaminant is detected at or greater than the threshold concentration at less than 15 percent of those sites; hence, there is insufficient variability in the contaminant occurrence to produce a model.
4. Threshold concentrations are not specified. Two contaminants, pH and total triazines, are not modeled for this reason.

Estimating the Susceptibility of Surface Water to Nonpoint-Source Contamination

The result of the nonpoint-source susceptibility component for each of the 227 contaminants is either two logistic regression models that estimate the probability of contaminant occurrence at or greater than a threshold concentration at surface-water PWS intakes in Texas, or an assignment of a relative susceptibility rating based on frequency of contaminant detection or level of land-use activities associated with the contaminant.

Summary of Modeled Contaminants

The modeling status of each of the 227 contaminants is listed in table 1. Models were developed for 63 contaminants. No analytical data were available for 27 contaminants. All other contaminants had at least 50 sample analyses total, but 106 contaminants were analyzed at fewer than 32 sites. Monitoring data were sufficient for 29 contaminants but they were detected greater than the threshold concentration too infrequently to develop a model. Threshold concentrations were not specified for two contaminants.

The percentage of contaminants that were modeled varies considerably by contaminant type. Logistic regression models are developed for 100 percent (3 of 3) of nutrients, 95 percent (18 of 19) of metals, 75 percent (9 of 12) of major ions, 67 percent (2 of 3) of physical properties, 32 percent (20 of 63) of pesticides, 20 percent (1 of 5) of microbiologicals, 10 percent (1 of 10) of radionuclides, 8 percent (9 of 109) of organics, and 0 percent (0 of 3) of inorganics (table 1). Pesticide occurrence is not modeled primarily because of a lack of detections in the monitoring data (28 of 43 that were not modeled); organics are not modeled primarily because of limitations in the extent of monitoring data (91 of 100 not modeled); and microbiologicals, inorganics, and radionuclides are all not modeled primarily because of the complete absence of monitoring data.

Examples of Modeling Results

Examples of the monitoring data and modeling results are shown in figures 1 and 2. Figures 1(A) and 2(A) show the monitoring data that are available for model development, whereas figures 1(B) and 2(B) show model results. In both figures, areas shaded as "not in study area" are the parts of Texas that are not covered by the watershed of 1 or more of the 323 surface-water sites. The 323 watersheds are shaded according to the results of the monitoring data analysis or logistic regression model results. Figure 1(A) shows where dissolved manganese occurred at least once at a concentration of 75 µg/L (the TCEQ-specified threshold concentration) or greater, occurred but always at a concentration less than 75 µg/L, was analyzed for but not detected, and was not analyzed for. In figure 1(B) the modeled probabilities that manganese will occur at a concentration of 75 µg/L or greater are shown as high (probability of 0.67 or greater), medium (probability of 0.34 to 0.66), or low (probability of less than 0.34). The monitoring data for manganese were available at 232 of the 323 sites with about 14,000 analyses. Equation 3 is the model used to predict the probability of manganese occurrence in figure 1. This model indicates that the presence of evergreen and mixed forestland and shrubland is inversely related to manganese occurrence.

(3)

$$p = \frac{(\exp(5.35 - 0.069*p42 + 0.058*p43 - 0.020*p51 - 11.3*kfact))}{(1 + (\exp(5.35 - 0.069*p42 + 0.058*p43 - 0.020*p51 - 11.3*kfact)))}$$

Figure 2(A) shows where dissolved simazine occurred at least once at a concentration of 0.01 µg/L (the TCEQ-specified threshold concentration) or greater, occurred but always at a concentration less than 0.01 µg/L, was analyzed for but not detected, and was not analyzed for. In figure 1(B) the modeled probabilities that simazine will occur at a concentration of 0.01 µg/L or greater are shown as high, medium, or low. The monitoring data for simazine were available at 129 of the 323 sites with about 1,300 analyses. Equation 4 is the model used to predict the probability of simazine occurrence in figure 2. This model indicates that the presence of low-intensity residential land and pasture/hay is positively related to simazine occurrence, and evergreen forestland and woody wetlands are inversely related to simazine occurrence.

(4)

$$p = \frac{(\exp(-0.719 + 0.098*p21 + 0.052*p81 - 0.028*p42 - 0.463*p91))}{(1 + (\exp(-0.719 + 0.098*p21 + 0.052*p81 - 0.028*p42 - 0.463*p91)))}$$

The logistic regression models developed for the 63 contaminants and selected model fit statistics are listed in table 3. This table and figures similar to 1 and 2 for the other 61 contaminants for which logistic regression models were developed are available on the Internet at http://co.water.usgs.gov/midconherb/html/texas.html.

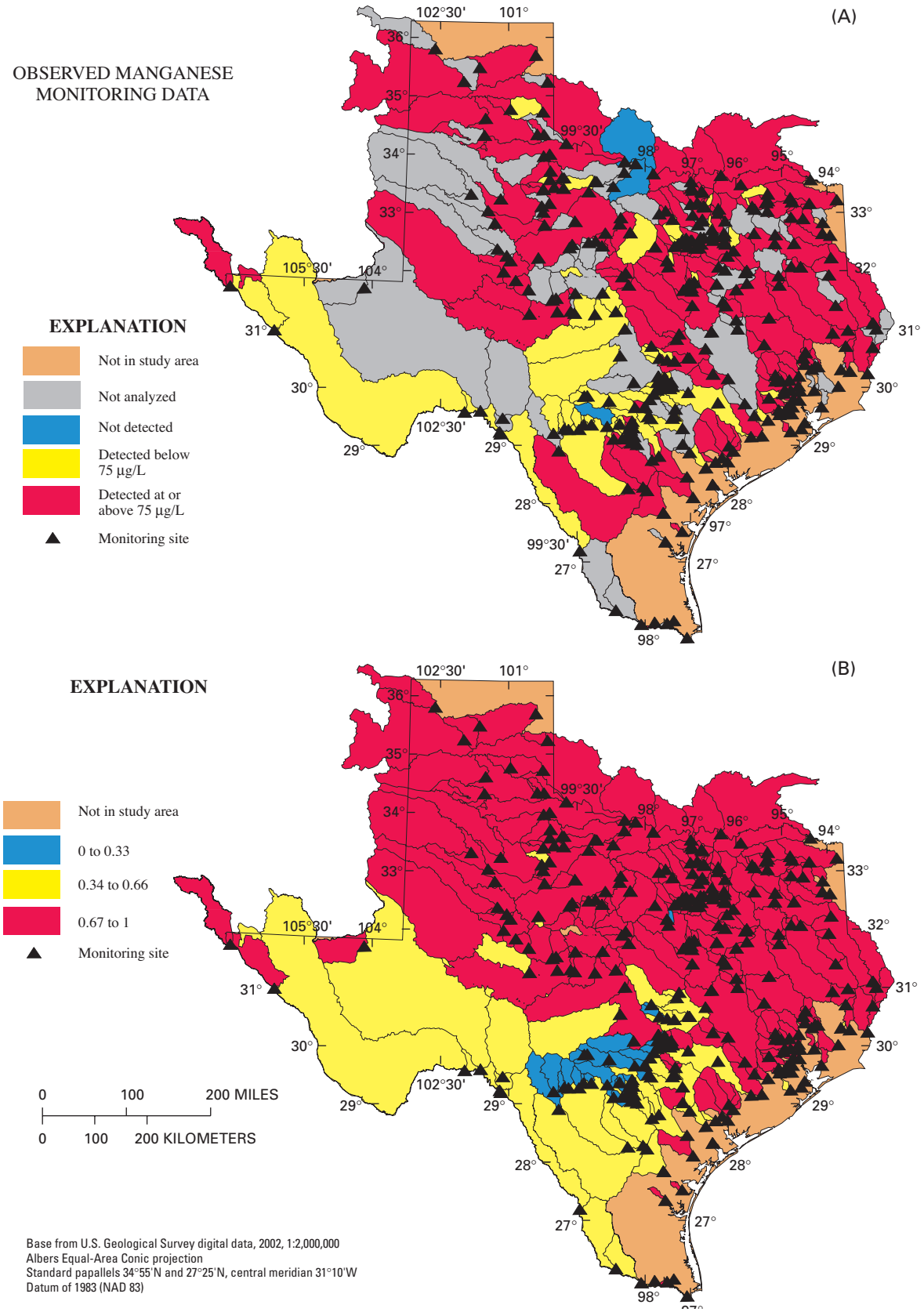


Figure 1. Summary of (A) dissolved manganese monitoring data and (B) the results of a logistic regression model predicting the probability of manganese occurrence at or greater than 75 micrograms per liter in Texas surface water.

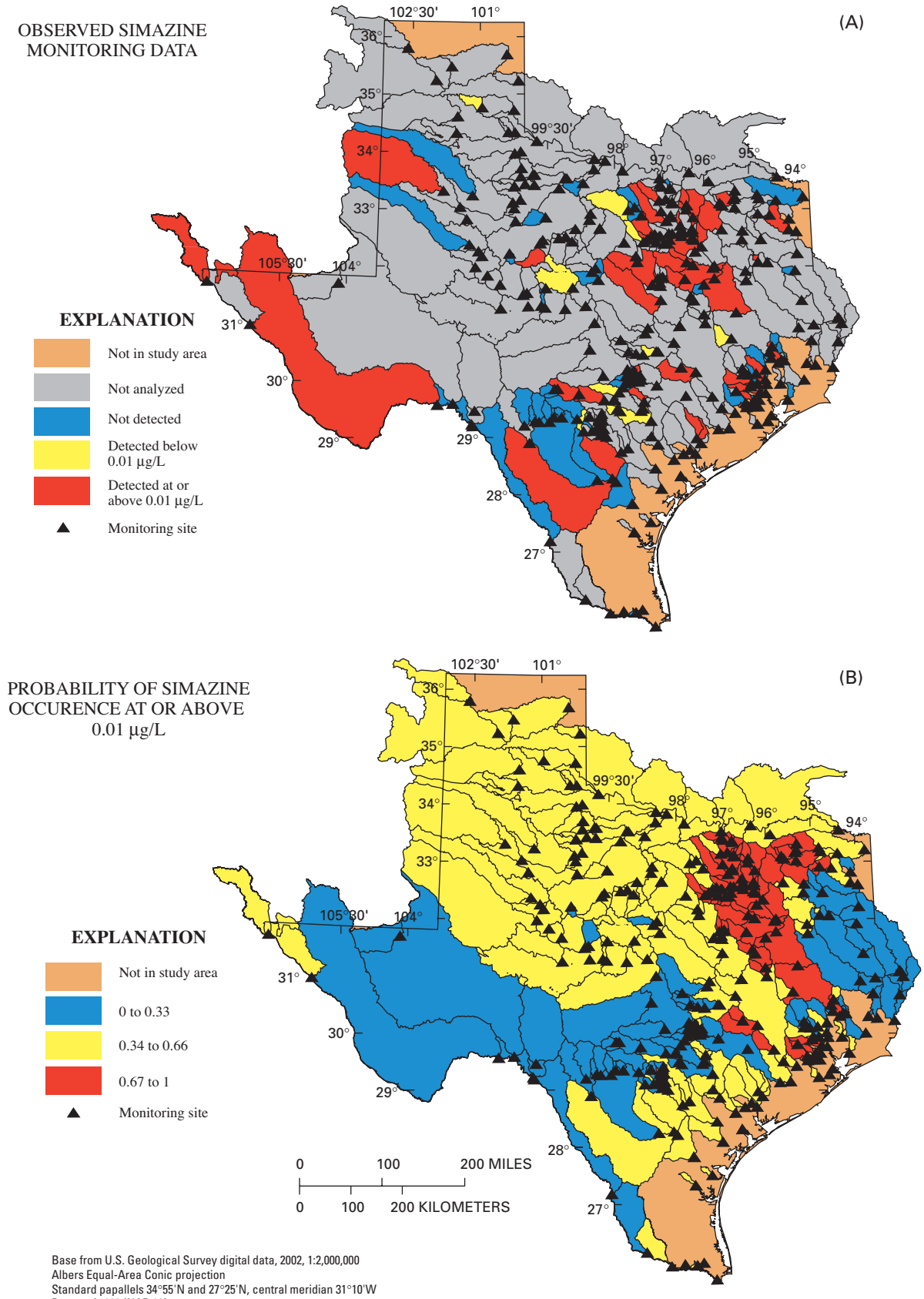


Figure 2. Summary of (A) dissolved simazine monitoring data and (B) the results of a logistic regression model predicting the probability of simazine occurrence at or greater than 0.01 microgram per liter in Texas surface water.

Table 3. Logistic regression-model equations and model-fit statistics for selected surface-water nonpoint-source contaminants in Texas.

[AIC, Akaike information criterion; TCEQ, Texas Commission on Environmental Quality; see table 2 for abbreviation definitions of model variables]

Contaminant and (TCEQ contaminant number)	Water-quality parameter code(s)	Threshold concentration	Model comments	
First logistic regression model			**AIC**	**Somers' D**
Second logistic regression model			**AIC**	**Somers' D**
2,4,5-T (21)	39742, 39740	0.01 µg/L		
p=(exp(-6.44 + 0.085*p22 + 0.0001*TDA - 0.034*p51 + 1.77*hydgrp))/ (1+(exp(-6.44 + 0.085*p22 + 0.0001*TDA - 0.034*p51 + 1.77*hydgrp)))			178.5	0.618
p=(exp(-6.28 + 0.031*p_urb + 0.0001*TDA - 0.033*p51 + 1.71*hydgrp))/ (1+(exp(-6.28 + 0.031*p_urb + 0.0001*TDA - 0.033*p51 + 1.71*hydgrp)))			180.5	0.613
2,4,5-TP (SILVEX)(22)	39762, 39760	0.01 µg/L		
p=(exp(-2.38 + 0.056*p_urb + 118.5*perm - 0.066*MCS + 0.122*p_wet))/ (1+(exp(-2.38 + 0.056*p_urb + 118.5*perm - 0.066*MCS + 0.122*p_wet)))			148.6	0.631
p=(exp(-2.94 + 0.849*p85 + 74.0*perm + 0.000035*TDA + 0.388*p92))/ (1+(exp(-2.94 + 0.849*p85 + 74.0*perm + 0.000035*TDA + 0.388*p92)))			145.3	0.607
2,4-D (24)	39732, 39730	0.01 µg/L		
p=(exp(-3.42 + 0.151*p21 + 0.889*hydgrp + 0.00005*TDA + 0.325*p91))/ (1+(exp(-3.42 + 0.151*p21 + 0.889*hydgrp + 0.00005*TDA + 0.325*p91)))			194.9	0.531
p=(exp(-3.35 + 0.055*p_urb + 0.852*hydgrp + 0.00006*TDA + 0.241*p_wet))/ (1+(exp(-3.35 + 0.055*p_urb + 0.852*hydgrp + 0.00006*TDA + 0.241*p_wet)))			193.3	0.551
Alachlor (41)	46342, 77825	0.01 µg/L		
p=(exp(-12.9 + 6.66*ala_use + 0.146*p22 + 0.059*p81 + 2.32*hydgrp + 0.066*p71))/ (1+(exp(-12.9 + 6.66*ala_use + 0.146*p22 + 0.059*p81 + 2.32*hydgrp + 0.066*p71)))			88.2	0.754
p=(exp(-9.09 + 3.97*ala_use + 0.032*p_urban + 0.026*p_ag + 1.79*hydgrp))/ (1+(exp(-9.09 + 3.97*ala_use + 0.032*p_urban + 0.026*p_ag + 1.79*hydgrp)))			91.9	0.705
Aldrin (45)	39330, 39331, 39332	0.01 µg/L		
p=(exp(-3.52 + 0.166*P23 + 0.042*P81 + 0.491*welldn))/ (1+(exp(-3.52 + 0.166*P23 + 0.042*P81 + 0.491*welldn)))			81.8	0.646
p=(exp(-3.51 + 0.043*p_urb + 0.025*p_ag + 0.487*welldn))/ (1+(exp(-3.51 + 0.043*p_urb + 0.025*p_ag + 0.487*welldn)))			85.7	0.624
Alkalinity (46)	39036, 39086, 39087	250 mg/L	Poor models	
p=(exp(-0.169 + 0.058*p41 - 0.224*p43 - 0.025*p81 ñ 0.0003*popdn ñ 0.0001*TDA))/ (1+(exp(-0.169 + 0.058*p41 - 0.224*p43 - 0.025*p81 ñ 0.0003*popdn ñ 0.0001*TDA)))			247.0	0.515
p=(exp(0.811 - 0.045*p22 - 0.190*p43 - 0.026*p81 ñ 0.002*BS – 0.0001*TDA))/ (1+(exp(0.811 - 0.045*p22 - 0.190*p43 - 0.026*p81 – 0.002*BS – 0.0001*TDA)))			247.9	0.491

Table 3. Logistic regression-model equations and model-fit statistics for selected surface-water nonpoint-source contaminants in Texas.—Continued

[AIC, Akaike information criterion; TCEQ, Texas Commission on Environmental Quality; see table 2 for abbreviation definitions of model variables]

Contaminant and (TCEQ contaminant number)	Water-quality parameter code(s)	Threshold concentration	Model comments	
Aluminum (47)	01106	100 µg/L	Poor models	
p=(exp(-1.12 + 0.549*fert_n + 0.046*MCS + 0.304*p91 + 0.269*p85))/ (1+(exp(-1.12 + 0.549*fert_n + 0.046*MCS + 0.304*p91 + 0.269*p85)))			154.4	0.427
p=(exp(-1.23 + 0.445*fert_n + 0.053*MCS + 0.199*p_wet + 0.513*welldn))/ (1+(exp(-1.23 + 0.445*fert_n + 0.053*MCS + 0.199*p_wet + 0.513*welldn)))			155.6	0.435
Antimony (49)	01095, 01097, 99897	0.01 µg/L		
p=(exp(-2.16 - 0.349*P41 + 0.376*P81 + 0.00006*TDA))/ (1+(exp(-2.16 - 0.349*P41 + 0.376*P81 + 0.00006*TDA)))			43.8	0.879
p=(exp(3.40 - 0.172*P22 - 0.123*P71 - 0.079*p42))/ (1+(exp(3.40 - 0.172*P22 - 0.123*P71 - 0.079*p42)))			50.5	0.812
Arsenic (51)	01000	5.0 µg/L		
p=(exp(1.40 - 0.048*p_forest + 0.111*p_urb + 0.105*p_drain ñ 0.032*MCS))/ (1+(exp(1.40 - 0.048*p_forest + 0.111*p_urb + 0.105*p_drain ñ 0.032*MCS)))			202.9	0.634
p=(exp(-0.193 - 0.049*p_forest + 0.098*p_urb + 0.624*hydgrp ñ 0.038*MCS))/ (1+(exp(-0.193 - 0.049*p_forest + 0.098*p_urb + 0.624*hydgrp ñ 0.038*MCS)))			209.3	0.599
Atrazine (53)	39632, 39630	0.1 µg/L		
p=(exp(-5.02 + 0.135*atr_use + 0.250*P21 + 0.193*p_clay))/ (1+(exp(-5.02 + 0.135*atr_use + 0.250*P21 + 0.193*p_clay)))			132.9	0.746
p=(exp(-8.77 + 0.057*p_ag + 0.119*p_urban + 0.035*p71 + 1.75*hydgrp))/ (1+(exp(-8.77 + 0.057*p_ag + 0.119*p_urban + 0.035*p71 + 1.75*hydgrp)))			132.6	0.740
Barium (54)	01005	200 µg/L		
p=(exp(6.76 + 0.244*P22 + 0.921*P92 - 0.061*p43 – 2.30*hydgrp))/ (1+(exp(6.76 + 0.244*P22 + 0.921*P92 - 0.061*P43 – 2.30*hydgrp)))			219.7	0.596
p=(exp(3.33 + 0.269*P22 + 1.03*P92 + 0.040*p71 – 1.53*hydgrp))/ (1+(exp(3.33 + 0.269*P22 + 1.03*P92 + 0.040*P71 – 1.53*hydgrp)))			220.9	0.579
Benzene (56)	34030	0.01 µg/L		
p=(exp(-4.33 + 0.125*p21 – 3.18*fert_n + 0.282*p_clay + 184*p61))/ (1+(exp(-4.33 + 0.125*p21 – 3.18*fert_n + 0.282*p_clay + 184*p61)))			48.6	0.657
p=(exp(-3.14 + 0.027*p_urban – 1.86*fert_n + 0.178*p_clay))/ (1+(exp(-3.14 + 0.027*p_urban – 1.86*fert_n + 0.178*p_clay)))			51.0	0.585
Beryllium (62)	01010	1.0 µg/L		
p=(exp(-2.61 - 0.209*p22 + 5.05*bulk_d + 0.650*p85 - 0.096*p_clay))/ (1+(exp(-2.61 - 0.209*p22 + 5.05*bulk_d + 0.650*p85 - 0.096*p_clay)))			165.2	0.505
p=(exp(-3.23 - 0.061*p22 + 6.20*bulk_d - 0.430*fert_n - 0.078*p_clay))/ (1+(exp(-3.23 - 0.061*p22 + 6.20*bulk_d - 0.430*fert_n - 0.078*p_clay)))			167.3	0.505

Table 3. Logistic regression-model equations and model-fit statistics for selected surface-water nonpoint-source contaminants in Texas.—Continued

[AIC, Akaike information criterion; TCEQ, Texas Commission on Environmental Quality; see table 2 for abbreviation definitions of model variables]

Contaminant and (TCEQ contaminant number)	Water-quality parameter code(s)	Threshold concentration	Model comments	
Bicarbonate (63)	00453	200 mg/L		
p=(exp(-4.39 + 0.038*p_forest ñ 2.11*p_wet + 0.053*p51 + 0.186*p_clay))/ (1+(exp(-4.39 + 0.038*p_forest ñ 2.11*p_wet + 0.053*p51 + 0.186*p_clay)))			96.3	0.635
p=(exp(-2.22 + 0.127*p41 ñ 3.22*p92 + 0.057*p51 + 0.872*fert_n))/ (1+(exp(-2.22 + 0.127*p41 – 3.22*p92 + 0.057*p51 + 0.872*fert_n)))			100.8	0.632
Boron (64)	01020	200 µg/L		
p=(exp(2.61 - 0.056*p_forest - 0.045*MCS - 0.253*welldn))/ (1+(exp(2.61 - 0.056*p_forest - 0.045*MCS - 0.253*welldn)))			95.2	0.675
p=(exp(1.84 - 0.021*p42 - 0.437*p91 - 0.005*BS))/ (1+(exp(1.84 - 0.021*p42 - 0.437*p91 - 0.005*BS)))			100.4	0.591
Cadmium (74)	01025	5.0 µg/L	Poor models	
p=(exp(-0.424 - 0.025*p51 + 0.027*p71 + 0.455*fert_n + 0.002*BS))/ (1+(exp(-0.424 - 0.025*p51 + 0.027*p71 + 0.455*fert_n + 0.002*BS)))			245.3	0.371
p=(exp(2.03 - 0.052*p21 - 0.053*p41 - 0.034*p43 - 0.043*p51))/ (1+(exp(2.03 - 0.052*p21 - 0.053*p41 - 0.034*p43 - 0.043*p51)))			244.1	0.360
Calcium (75)	00915	100 mg/L		
p=(exp(0.642 - 0.055*p42 + 0.059*p82 - 0.032*p81 ñ 0.401*p91))/ (1+(exp(0.642 - 0.055*p42 + 0.059*p82 - 0.032*p81 ñ 0.401*p91)))			299.7	0.650
p=(exp(-1.33 - 0.040*p_forest - 0.431*p_wet + 2.71*bulk_d))/ (1+(exp(-1.33 - 0.040*p_forest - 0.431*p_wet + 2.71*bulk_d)))			312.3	0.627
Carbaryl (76)	39750, 49310, 82680	0.01 µg/L		
p=(exp(-1.18 + 0.118*p23 + 21.6*p61 - 0.061*MCS))/ (1+(exp(-1.18 + 0.118*p23 + 21.6*p61 - 0.061*MCS)))			119.4	0.600
p=(exp(-1.46 + 0.032*p_urban + 0.030*p51 - 0.065*MCS))/ (1+(exp(-1.46 + 0.032*p_urban + 0.030*p51 - 0.065*MCS)))			120.8	0.570
Carbonate (80)	00452	1.0 mg/L		
p=(exp(1.16 ñ 0.140*p41 + 0.078*p51 + 0.072*p21 ñ 0.064*MCS))/ (1+(exp(1.16 - 0.140*p41 + 0.078*p51 + 0.072*p21 ñ 0.064*MCS)))			92.9	0.714
p=(exp(6.00 ñ 0.089*p_forest ñ 0.060*p_ag - 0.070*MCS ñ 1.16*welldn))/ (1+(exp(6.00 - 0.089*p_forest ñ 0.060*p_ag - 0.070*MCS ñ 1.16*welldn)))			92.3	0.699
Chlordane (81)	39350, 39352	0.01 µg/L		
p=(exp(12.3 + 0.49*p21 – 49.7*Kfact + 0.00004*TDA))/ (1+(exp(12.3 + 0.49*p21 – 49.7*Kfact + 0.00004*TDA)))			85.7	0.848
p=(exp(9.19 + 0.148*p_urban – 38.1*Kfact + 0.00003*TDA))/ (1+(exp(9.19 + 0.148*p_urban – 38.1*Kfact + 0.00003*TDA)))			88.8	0.820

Table 3. Logistic regression-model equations and model-fit statistics for selected surface-water nonpoint-source contaminants in Texas.—Continued

[AIC, Akaike information criterion; TCEQ, Texas Commission on Environmental Quality; see table 2 for abbreviation definitions of model variables]

Contaminant and (TCEQ contaminant number)	Water-quality parameter code(s)	Threshold concentration	Model comments	
Chloride (85)	00940	125 mg/L		
p=(exp(-6.42 + 0.066*p51 - 0.034*p_forest + 6.63*bulk_d + 0.646*welldn))/ (1+(exp(-6.42 + 0.066*p51 - 0.034*p_forest + 6.63*bulk_d + 0.646*welldn)))			262.8	0.766
p=(exp(-7.85 + 0.067*p51 + 0.034*p71 + 6.75*bulk_d + 0.048*p82))/ (1+(exp(-7.85 + 0.067*p51 + 0.034*p71 + 6.75*bulk_d + 0.048*p82)))			273.1	0.728
Chromium (90)	01030, 01032	15.0 µg/L	Poor models	
p=(exp(-0.979 - 0.062*p51 - 0.042*p81 + 0.845*hydgrp + 0.00007*TDA))/ (1+(exp(-0.979 - 0.062*p51 - 0.042*p81 + 0.845*hydgrp + 0.00007*TDA)))			240.0	0.514
p=(exp(-2.01 + 0.038*p_urban + 0.032*p_forest + 0.693*fert_n + 0.00006*TDA))/ (1+(exp(-2.01 + 0.038*p_urban + 0.032*p_forest + 0.693*fert_n + 0.00006*TDA)))			246.0	0.456
Copper (94)	01040	20 µg/L	Poor models	
p=(exp(-1.71 - 0.028*p_forest + 0.160*p_wet + 0.020*p51 + 2.20*bulk_d))/ (1+(exp(-1.71 - 0.028*p_forest + 0.160*p_wet - 0.020*p51 + 2.20*bulk_d)))			220.7	0.461
p=(exp(-1.18 - 0.041*p42 - 0.302*p85 + 0.021*p82 + 0.019*p71))/ (1+(exp(-1.18 - 0.041*p42 - 0.302*p85 + 0.021*p82 + 0.019*p71)))			219.0	0.468
DCPA di-acid degradate (99)	82682	0.001 µg/L		
p=(exp(-4.03 + 0.197*p21 + 0.043*p71 + 2.48*DCPA_use + 0.004*BS + 0.0001*TDA))/ (1+(exp(-4.03 + 0.197*p21 + 0.043*p71 + 2.48*DCPA_use + 0.004*BS + 0.0001*TDA)))			91.5	0.713
p=(exp(-0.837 + 0.040*p_urban ñ 1.10*p_wet - 0.070*MCS + 0.0001*TDA))/ (1+(exp(-0.837 + 0.040*p_urban ñ 1.10*p_wet - 0.070*MCS + 0.0001*TDA)))			87.0	0.666
DDE (101)	346853, 39366	0.001 µg/L		
p=(exp(-4.30 + 0.035*p_urban + 0.055*p_drain + 0.00008*TDA + 0.008*BS))/ (1+(exp(-4.30 + 0.035*p_urban + 0.055*p_drain + 0.00008*TDA + 0.008*BS)))			75.6	0.663
p=(exp(-12.0 + 0.954*p85 + 0.008*BS + 25.1*Kfact + 0.00008*TDA))/ (1+(exp(-12.0 + 0.954*p85 + 0.008*BS + 25.1*Kfact + 0.00008*TDA)))			71.4	0.665
Diazinon (104)	39570, 39572, 39573	0.01 µg/L		
p=(exp(-1.24 + 0.402*p21 + 0.00007*TDA + 0.027*p81))/ (1+(exp(-1.24 + 0.402*p21 + 0.00007*TDA + 0.027*p81)))			188.3	0.664
p=(exp(-0.866 + 0.078*p_urban + 0.00008*TDA + 1.60*p_org – 0.027*p51))/ (1+(exp(-0.866 + 0.078*p_urban + 0.00008*TDA + 1.60*p_org – 0.027*p51)))			202.5	0.599
Dieldrin (112)	39380, 39381	0.01 µg/L		
p=(exp(-8.02 + 0.048*p_urban - 0.080*p51 + 2.02*hydgrp + 0.0001*TDA))/ (1+(exp(-8.02 + 0.048*p_urban - 0.080*p51 + 2.02*hydgrp + 0.0001*TDA)))			141.0	0.745
p=(exp(-7.59 + 0.313*p23 + 0.499*p92 + 1.52*hydgrp + 0.00005*TDA))/ (1+(exp(-7.59 + 0.313*p23 + 0.499*p92 + 1.52*hydgrp + 0.00005*TDA)))			140.5	0.705
Diuron (119)	49300	0.01 µg/L		

Table 3. Logistic regression-model equations and model-fit statistics for selected surface-water nonpoint-source contaminants in Texas.—Continued

[AIC, Akaike information criterion; TCEQ, Texas Commission on Environmental Quality; see table 2 for abbreviation definitions of model variables]

Contaminant and (TCEQ contaminant number)	Water-quality parameter code(s)	Threshold concentration	Model comments	
p=(exp(-1.81 + 0.980*diuron_use + 0.127*p21 + 0.544*welldn))/ (1+(exp(-1.81 + 0.980*diuron_use + 0.127*p21 + 0.544*welldn)))			95.3	0.543
p=(exp(-1.75 + 0.868*diuron_use + 0.051*p_urban + 0.542*welldn))/ (1+(exp(-1.75 + 0.868*diuron_use + 0.051*p_urban + 0.542*welldn)))			96.8	0.536
Fluoride (129)	00950, 00951	1.0 mg/L		
p=(exp(-1.68 - 0.018*p_ag - 0.044*p_forest + 80.3*perm + 0.096*p_clay))/ (1+(exp(-1.68 - 0.018*p_ag - 0.044*p_forest + 80.3*perm + 0.096*p_clay)))			335.5	0.507
p=(exp(-1.10 - 0.024*p81 - 0.040*p42 ñ 0.079*p43 + 1.86*bulk_d))/ (1+(exp(-1.10 - 0.024*p81 - 0.040*p42 ñ 0.079*p43 + 1.86*bulk_d)))			335.3	0.506
Hardness (135)	00900	300 mg/L		
p=(exp(-4.39 + 0.111*p22 + 0.046*p51 + 0.034*p71 + 0.073*p82 + 2.66*bulk_d))/ (1+(exp(-4.39 - 0.111*p22 ñ 0.046*p51 + 0.034*p71 + 0.073*p82 + 2.66*bulk_d)))			305.7	0.642
p=(exp(-3.75 - 0.014*p_ag ñ 0.055*p_forest + 206*perm + 1.23*hydgrp + 0.139*welldn))/ (1+(exp(-3.75 - 0.014*p_ag ñ 0.055*p_forest + 206*perm + 1.23*hydgrp + 1.39*welldn)))			323.9	0.575
Heptachlor (136)	39410, 39411	0.01 µg/L		
p=(exp(-4.01 + 0.031*p_ag + 0.058*p_urban + 0.045*p_drain))/ (1+(exp(-4.01 + 0.031*p_ag + 0.058*p_urban + 0.045*p_drain)))			69.6	0.772
p=(exp(-3.87 + 0.029*p_ag + 0.0012*popdn + 0.043*p_drain))/ (1+(exp(-3.87 + 0.029*p_ag + 0.0012*popdn + 0.043*p_drain)))			68.9	0.769
Heptachlor epoxide(137)	39420, 39421	0.01 µg/L		
p=(exp(-3.75 + 0.027*p_ag + 0.071*p_urban + 0.375*welldn))/ (1+(exp(-3.75 + 0.028*p_ag + 0.071*p_urban + 0.375*welldn)))			77.1	0.820
p=(exp(-3.39 + 0.088*p81 + 0.198*p21 – 1.39*p91))/ (1+(exp(-3.39 + 0.088*p81 + 0.198*p21 – 1.39*p91)))			71.2	0.818
Iron (144)	01046	150 µg/L	Poor models	
p=(exp(2.42 – 0.048*p42 - 0.031*p51 – 0.020*p71 - 0.030*P82))/ (1+(exp(2.42 – 0.048*p42 - 0.031*p51 – 0.020*p71 – 0.030*p82)))			269.3	0.504
p=(exp(-1.54 - 0.019*p51 + 2.31*bulk_d + 0.509*manure_n))/ (1+(exp(-1.54 - 0.019*p51 + 2.31*bulk_d + 0.509*manure_n)))			275.0	0.472
Lead (147)	01049	25 µg/L	Poor models	
p=(exp(0.185 – 0.050*p43 – 0.026*p51 + 0.155*p85 + 0.002*BS))/ (1+(exp(0.185 – 0.050*p43 - 0.026*p51 + 0.155*p85 + 0.002*BS)))			260.7	0.318
p=(exp(-0.315 + 0.029*p22 – 0.028*p51 + 0.027*p71 + 0.002*BS))/ (1+(exp(-0.315 + 0.029*p22 - 0.028*p51 + 0.021*p71 + 0.002*BS)))			266.2	0.282
Lindane (148)	39340, 39341	0.01 µg/L		

Table 3. Logistic regression-model equations and model-fit statistics for selected surface-water nonpoint-source contaminants in Texas.—Continued

[AIC, Akaike information criterion; TCEQ, Texas Commission on Environmental Quality; see table 2 for abbreviation definitions of model variables]

Contaminant and (TCEQ contaminant number)	Water-quality parameter code(s)	Threshold concentration	Model comments	
p=(exp(-5.03 + 0.199*p23 – 0.040*p51 + 1.11*hydgrp + 0.0001*TDA))/ (1+(exp(-5.03 + 0.199*p23 - 0.040*p51 + 1.11*hydgrp + 0.0001*TDA)))			175.4	0.645
p=(exp(-4.80 + 0.043*p_urban ñ 0.039*p51 + 1.06*hydgrp + 0.0001*TDA))/ (1+(exp(-4.80 + 0.043*p_urban - 0.039*p51 + 1.06*hydgrp + 0.0001*TDA)))			179.1	0.641
M+P xylene (150)	85795	0.1 µg/L	Poor models	
p=(exp(-3.79 + 0.047*p_urban + 0.005*BS + 171.5*perm))/ (1+(exp(-3.79 + 0.047*p_urban + 0.005*BS + 171.5*perm)))			47.3	0.475
p=(exp(-3.19 + 0.100*p22 + 0.052*p42 + 157.2*perm))/ (1+(exp(-3.19 + 0.100*p22 + 0.052*p42 + 157.2*perm)))			46.7	0.496
Magnesium (151)	00925	20.0 mg/L		
p=(exp(-3.04 - 0.068*p22 - 0.132*p43 - 0.089*p81 + 6.49*bulk_d))/ (1+(exp(-3.04 - 0.068*p22 - 0.132*p43 - 0.089*p81 + 6.49*bulk_d)))			267.8	0.747
p=(exp(-2.23 + 0.067*p51 + 0.040*p71 + 0.068*p82 + 0.0006*popdn))/ (1+(exp(-2.23 + 0.067*p51 + 0.040*p71 + 0.068*p82 + 0.0006*popdn)))			284.0	0.680
Manganese (152)	01056	75 µg/L		
p=(exp(5.35 - 0.069*p42 + 0.058*p43 - 0.020*p51 – 11.3*kfact))/ (1+(exp(5.35 - 0.069*p42 + 0.058*p43 - 0.020*p51 – 11.3*kfact)))			225.1	0.613
p=(exp(7.59 + 0.013*p_ag – 1.81*hydgrp - 0.007*BS – 0.00004*TDA))/ (1+(exp(7.59 + 0.013*p_ag – 1.81*hydgrp - 0.007*BS – 0.00004*TDA)))			233.8	0.559
Mercury (153)	71890	0.5 µg/L	Poor models	
p=(exp(-4.11 + 0.121*p23 + 0.026*p82 + 1.29*hydgrp + 0.00004*TDA))/ (1+(exp(-4.11 + 0.121*p23 + 0.026*p82 + 1.29*hydgrp + 0.00004*TDA)))			230.3	0.391
p=(exp(-3.20 + 0.017*p_urban + 0.225*atr_use + 0.967*hydgrp + 0.00004*TDA))/ (1+(exp(-3.20 + 0.017*p_urban + 0.225*atr_use + 0.967*hydgrp + 0.00004*TDA)))			228.6	0.418
Methyl-t-butyle ether (MTBE)(159)	78032	0.2 µg/L	Poor models	
p=(exp(-1.48 + 0.041*p81 – 2.01*p91 + 0.106*p43 + 1.69*manure_n))/ (1+(exp(-1.48 + 0.041*p81 ñ 2.01*p91 + 0.106*p43 + 1.69*manure_n)))			62.2	0.568
p=(exp(2.75 + 0.037*p_ag ñ 5.77*bulk_d + 1.06*manure_n))/ (1+(exp(2.75 + 0.037*p_ag ñ 5.77*bulk_d + 1.06*manure_n)))			66.3	0.421
Metolachlor (160)	39415, 82612	0.01 µg/L		
p=(exp(-4.17 + 0.048*p82 + 0.032*p81 + 6.20*p_org + 70.0*p61))/ (1+(exp(-4.17 + 0.048*p82 + 0.032*p81 + 6.20*p_org + 70.0*p61)))			122.6	0.660
p=(exp(-2.58 + 0.028*p_ag - 0.066*MCS + 5.29*p_org + 0.00007*TDA))/ (1+(exp(-2.58 + 0.028*p_ag - 0.066*MCS + 5.29*p_org + 0.00007*TDA)))			119.7	0.655
Metribuzin (161)	82611, 82612	0.01 µg/L		

Table 3. Logistic regression-model equations and model-fit statistics for selected surface-water nonpoint-source contaminants in Texas.—Continued

[AIC, Akaike information criterion; TCEQ, Texas Commission on Environmental Quality; see table 2 for abbreviation definitions of model variables]

Contaminant and (TCEQ contaminant number)	Water-quality parameter code(s)	Threshold concentration	Model comments	
p=(exp(-1.61 + 0.040*p21 - 0.057*p42 + 12.4*p61))/ (1+(exp(-1.61 + 0.040*p21 - 0.057*p42 + 12.4*p61)))			87.2	0.533
p=(exp(-0.541 + 0.013*p_urban - 0.033*p_forest – 0.075*MCS))/ (1+(exp(-0.541 + 0.013*p_urban - 0.033*p_forest – 0.075*MCS)))			87.8	0.474
M-xylene (164)	No codes	0.1 µg/L	Same models as for M+P xylene	
p=(exp(-3.79 + 0.047*p_urban + 0.005*BS + 171.5*perm))/ (1+(exp(-3.79 + 0.047*p_urban + 0.005*BS + 171.5*perm)))			47.3	0.475
p=(exp(-3.19 + 0.100*p22 + 0.052*p42 + 157.2*perm))/ (1+(exp(-3.19 + 0.100*p22 + 0.052*p42 + 157.2*perm)))			46.7	0.496
Nickel (167)	01065	10 µg/L	Poor models	
p=(exp(-4.17 + 0.077*p22 + 0.023*p71 + 0.238*p83 + 0.539*p91 + 1.00*hydgrp))/ (1+(exp(-4.17 + 0.077*p22 + 0.023*p71 + 0.238*p83 + 0.539*p91 + 1.00*hydgrp)))			205.5	0.434
p=(exp(-0.185 + 0.056*p22 - 0.022*p82 + 0.253*p83 + 0.377*p91))/ (1+(exp(-0.185 + 0.056*p22 - 0.022*p82 + 0.253*p83 + 0.377*p91)))			206.0	0.414
Nitrate (168) and nitrate + nitrite (169)	00630, 00631, 00618, 00620	2.0 mg/L	Poor models	
p=(exp(-6.10 + 0.078*p82 + 0.244*p85 + 1.75*hydgrp ñ 0.018*MCS)))/ (1+(exp(-6.10 + 0.078*p82 + 0.244*p85 + 1.75*hydgrp ñ 0.018*MCS)))			336.3	0.494
p=(exp(-6.02 + 0.079*p82 + 0.046*p22 + 1.74*hydgrp ñ 0.021*MCS))/ (1+(exp(-6.02 + 0.079*p82 + 0.046*p22 + 1.74*hydgrp ñ 0.021*MCS)))			336.8	0.496
Nitrite (170)	00613, 00615	0.1 mg/L	Poor models	
p=(exp(-2.56 + 0.020*p_ag + 0.019*p_urban + 0.608*hydgrp ñ 0.036*p_drain))/ (1+(exp(-2.56 + 0.020*p_ag + 0.019*p_urban + 0.608*hydgrp ñ 0.036*p_drain)))			316.1	0.394
p=(exp(-2.51 + 0.692*fert_n + 0.077*p22 + 0.487*hydgrp ñ 0.039*p_drain))/ (1+(exp(-2.51 + 0.692*fert_n + 0.077*p22 + 0.487*hydgrp ñ 0.039*p_drain)))			306.4	0.424
O-xylene (176)	77135	0.01 µg/L		
p=(exp(-6.24 + 0.492*p23 + 0.051*p71 + 0.014*BS))/ (1+(exp(-6.24 + 0.492*p23 + 0.051*p71 + 0.014*BS)))			29.4	0.764
p=(exp(-4.49 + 0.104*p_urban + 0.012*BS))/ (1+(exp(-4.49 + 0.104*p_urban + 0.012*BS)))			29.3	0.694
PCBs (179)	39516, 39517	0.01 µg/L		
p=(exp(-1.14 + 0.059*p_urban - 0.088*MCS + 0.393*welldn))/ (1+(exp(-1.14 + 0.059*p_urban - 0.088*MCS + 0.393*welldn)))			85.8	0.653
p=(exp(-0.900 + 0.178*p22 - 0.128*MCS + 0.034*p82))/ (1+(exp(-0.900 + 0.178*p22 - 0.128*MCS + 0.034*p82)))			82.7	0.638
Prometon (185)	04037, 39056	0.01 µg/L		

Table 3. Logistic regression-model equations and model-fit statistics for selected surface-water nonpoint-source contaminants in Texas.—Continued

[AIC, Akaike information criterion; TCEQ, Texas Commission on Environmental Quality; see table 2 for abbreviation definitions of model variables]

Contaminant and (TCEQ contaminant number)	Water-quality parameter code(s)	Threshold concentration	Model comments	
p=(exp(0.870 + 0.080*p_urban - 0.043*p_forest))/ (1+(exp(0.870 + 0.080*p_urban - 0.043*p_forest)))			139.1	0.611
p=(exp(0.762 + 0.139*p22 - 0.075*p43 - 0.003*BS))/ (1+(exp(0.762 + 0.139*p22 - 0.075*p43 - 0.003*BS)))			150.1	0.509
P-xylene (188)	No codes or data	0.1 µg/L	Same models as for M+P xylene	
p=(exp(-3.79 + 0.047*p_urban + 0.005*BS + 171.5*perm))/ (1+(exp(-3.79 + 0.047*p_urban + 0.005*BS + 171.5*perm)))			47.3	0.475
p=(exp(-3.19 + 0.100*p22 + 0.052*p42 + 157.2*perm))/ (1+(exp(-3.19 + 0.100*p22 + 0.052*p42 + 157.2*perm)))			46.7	0.496
Selenium (195)	01145	2.0 µg/L		
p=(exp(-1.58 - 0.061*p43 + 0.028*p82 + 0.259*p85 + 146*perm))/ (1+(exp(-1.58 - 0.061*p43 + 0.028*p82 + 0.259*p85 + 146*perm)))			245.9	0.426
p=(exp(-2.49 - 0.076*p43 - 0.027*p81 + 3.58*bulk_d))/ (1+(exp(-2.49 - 0.076*p43 - 0.027*p81 + 3.58*bulk_d)))			247.1	0.411
Silver (196)	01075	2.0 µg/L	Poor models	
p=(exp(0.703 + 0.107*p83 + 0.230*p91 + 0.338*welldn - 122*perm))/ (1+(exp(0.703 + 0.107*p83 + 0.230*p91 + 0.338*welldn – 122*perm)))			258.6	0.298
p=(exp(1.67 + 0.016*p_ag + 0.342*welldn – 2.06*p_org - 130*perm))/ (1+(exp(1.67 + 0.016*p_ag + 0.342*welldn – 2.06*p_org - 130*perm)))			260.7	0.259
Simazine (197)	04035, 39055	0.01 µg/L		
p=(exp(-0.719 + 0.098*p21 + 0.052*p81 – 0.028*p42 - 0.463*p91))/ (1+(exp(-0.719 + 0.098*p21 + 0.052*p81 ñ 0.028*p42 - 0.463*p91)))			146.8	0.640
p=(exp(-2.97 + 0.045*p_urban + 0.038*p_ag + 2.61*p_org + 0.00004*TDA))/ (1+(exp(-2.97 + 0.045*p_urban + 0.038*p_ag + 2.61*p_org + 0.00004*TDA)))			147.8	0.615
Sodium (198)	00930	75 mg/L		
p=(exp(-3.65 - 0.043*p_forest + 0.037*p51 + 4.76*bulk_d + 0.821*welldn))/ (1+(exp(-3.65 - 0.043*p_forest + 0.037*p51 + 4.76*bulk_d + 0.821*welldn)))			254.4	0.773
p=(exp(-6.13 - 0.074*p43 + 0.041*p51 + 6.48*bulk_d + 0.790*welldn + 0.872*sod-chl_use))/ (1+(exp(-6.13 - 0.074*p43 + 0.041*p51 + 6.48*bulk_d + 0.790*welldn + 0.872*sod-chl_use)))			253.1	0.771
Specific conductance (199)	00095	1,500 µS/cm		
p=(exp(-2.02 – 0.066*p42 + 0.090*p82 + 0.048*p51 + 0.208*welldn))/ (1+(exp(-2.02 - 0.066*p42 + 0.090*p82 + 0.048*p51 + 0.208*welldn)))			253.2	0.757
p=(exp(-0.871 ñ 0.067*p_forest ñ 0.018*p_urban + 173*perm + 0.187*welldn))/ (1+(exp(-0.871 - 0.067*p_forest ñ 0.018*p_urban + 173*perm + 0.187*welldn)))			266.6	0.720
Sulfate (203)	00945	100 mg/L		

Table 3. Logistic regression-model equations and model-fit statistics for selected surface-water nonpoint-source contaminants in Texas.—Continued

[AIC, Akaike information criterion; TCEQ, Texas Commission on Environmental Quality; see table 2 for abbreviation definitions of model variables]

Contaminant and (TCEQ contaminant number)	Water-quality parameter code(s)	Threshold concentration	Model comments	
p=(exp(-3.47 - 0.089*p41 + 5.76*bulk_d ñ 0.182*p43 - 0.053*p_drain))/ (1+(exp(-3.47 - 0.089*p41 + 5.76*bulk_d ñ 0.182*p43 - 0.053*p_drain)))			293.4	0.676
p=(exp(0.575 ñ 0.065*p_forest + 124*perm + 0.015*p_urban ñ 0.035*p_drain))/ (1+(exp(0.575 ñ 0.065*p_forest + 124*perm + 0.015*p_urban ñ 0.035*p_drain)))			292.9	0.681
Total dissolved solids (TDS)(205)	70301	500 mg/L		
p=(exp(1.11 - 0.061*p42 ñ 0.051*p43 - 0.033*p81 + 0.048*p82 + 0.75*welldn))/ (1+(exp(1.11 - 0.061*p42 ñ 0.051*p43 - 0.033*p81 + 0.048*p82 + 0.75*welldn)))			273.1	0.717
p=(exp(-0.532 - 0.063*p_forest - 0.024*p_ag + 3.28*bulk_d + 0.653*welldn))/ (1+(exp(-0.532 - 0.063*p_forest - 0.024*p_ag + 3.28*bulk_d + 0.653*welldn)))			273.8	0.710
Toluene (211)	34010	0.01 µg/L		
p=(exp(-1.58 + 0.120*p21 – 0.222*p43 - 0.079*MCS + 301*perm))/ (1+(exp(-1.58 + 0.120*p21 – 0.222*p43 - 0.079*MCS + 301*perm)))			66.5	0.615
p=(exp(9.69 - 0.269*p43 – 0.050*p71 – 2.73*hydgrp – 0.048*MCS))/ (1+(exp(9.69 - 0.269*p43 – 0.050*p71 – 2.73*hydgrp ñ 0.048*MCS)))			69.0	0.607
Total coliform (213)	31501	20,000 counts/100 mL		
p=(exp(-0.266 + 0.085*p_urban ñ 0.028*p_forest + 4.09*manure_p))/ (1+(exp(-0.266 + 0.085*p_urban ñ 0.028*p_forest + 4.09*manure_p)))			54.5	0.801
p=(exp(-1.77 + 1.09*p21 ñ 0.057*p42 + 7.54*p_org))/ (1+(exp(-1.77 + 1.09*p21 ñ 0.057*p42 + 7.54*p_org)))			52.1	0.786
Uranium (223)	22703, 75990, 80020	2.0 µg/L		
p=(exp(-8.84 + 0.117*p81 + 0.122*p82 + 1.11*p85 + 0.138*p51))/ (1+(exp(-8.84 + 0.117*p81 + 0.122*p82 + 1.11*p85 + 0.138*p51)))			44.1	0.856
p=(exp(-13.1 + 0.157*p_ag + 0.114*p_urban + 185*perm + 0.162*p51))/ (1+(exp(-13.1 + 0.157*p_ag + 0.114*p_urban + 185*perm + 0.162*p51)))			45.0	0.877
Xylenes (total) (226)	81551	0.1 µg/L		
p=(exp(-2.43 + 0.00013*TDA – 1.88*welldn + 0.625*p92 + 127*perm))/ (1+(exp(-2.43 + 0.00013*TDA – 1.88*welldn + 0.625*p92 + 127*perm)))			51.6	0.533
p=(exp(1.56 + 0.00011*TDA – 1.34*welldn - 0.934*hydgrp))/ (1+(exp(1.56 + 0.00011*TDA – 1.34*welldn - 0.934*hydgrp)))			52.1	0.563
Zinc (227)	01090	50 µg/L		
p=(exp(-0.401 + 0.0005*popdn + 0.146*p83 + 0.384*p92))/ (1+(exp(-0.401 + 0.0005*popdn + 0.146*p83 + 0.384*p92)))			262.4	0.342
p=(exp(-0.446 + 0.075*p22 + 0.152*p83 + 0.410*p92))/ (1+(exp(-0.446 + 0.075*p22 + 0.152*p83 + 0.410*p92)))			259.7	0.373

Summary

The Texas Commission on Environmental Quality (TCEQ) and the U.S. Geological Survey (USGS) developed procedures for assessing the susceptibility of public water-supply (PWS) source waters to the occurrence of 227 drinking-water contaminants identified by the TCEQ. These procedures are being used by TCEQ to protect sources of drinking water and to develop protective and cost-effective water-supply monitoring and management strategies. To assess the susceptibility of surface waters used as PWS to contamination from nonpoint sources, water-quality data at 323 USGS and TCEQ sampling sites were matched with geographic information system-derived watershed characteristics and contaminant source information for the drainage basins associated with these sites. Logistic regression models were developed to estimate the probability that a contaminant will occur at or greater than a TCEQ-specified threshold. Statistically significant regression models were developed for 63 of the 227 contaminants. Of the remaining 164 contaminants, 106 were not modeled because monitoring data were available at less than 10 percent of the 323 sites; 29 were not modeled because of few detections of the contaminant in the otherwise adequate monitoring data; 27 were not modeled because of the absence of monitoring data; and 2 were not modeled because thresholds were not specified. The logistic regression models in this report were developed using the best available data. Many of the models could be improved with additional data, in particular from locations that were lacking data. Likewise, many of the 164 unmodeled contaminants could potentially be modeled if additional monitoring data are collected that target gaps in the current information.

References

Battaglin, W.A., and Goolsby, D.A., 1994, Spatial data in geographic information system format on agricultural chemical use, land use, and cropping practices in the United States: U.S. Geological Survey Water-Resources Investigations Report 94–4176, 87 p.

Battaglin, W.A., and Goolsby, D.A., 1998, Regression models of herbicide concentrations in outflow from reservoirs in the midwestern USA, 1992–1993: Journal of the American Water Resources Association, v. 34, no. 6, p. 1369–1390.

Brown, J.R., Ulery, R.L., and Parcher, J.W., 2000, Creating a standardized watersheds database for the Lower Rio Grande/Rio Bravo, Texas: U.S. Geological Survey Open-File Report 00–065, 18 p.

Davidson, Carlos., Shaffer, H.B., and Jennings, M.R., 2001, Declines of the California red-legged frog—Climate, UV–B, habitat, and pesticides hypotheses: Ecological Applications, v. 11, no. 2, p. 464–479.

Gandara, S.C., and Barbie, D.L., 2001, Index of stations—Surface-water data-collection network of Texas, September 1999: U.S. Geological Survey Open-File Report 01–103, 17 p.

Goolsby, D.A., Battaglin, W.A., Lawrence, G.B., Atrz, R.S., Aulenback, B.T., and Hooper, R.P., 1999, Flux and sources of nutrients in the Mississippi-Atchafalaya River Basin—Topic 3 Report for the Integrated Assessment on Hypoxia in the Gulf of Mexico: Silver Spring, Md., National Oceanic and Atmospheric Administration Coastal Decision Analysis Series No. 17, NOAA Coastal Ocean Program, 130 p.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: Amsterdam, The Netherlands, Elsevier, 522 p.

Hosmer, D.W., and Lemeshow, S., 2000, Applied logistic regression: New York, John Wiley, 375 p.

Mineau, P., 2002, Estimating the probability of bird mortality from pesticide sprays on the basis of the field study record: Environmental Toxicology and Chemistry, v. 21, no. 7, p. 1497–1506.

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

Rupert, M.G., 2003, Probability of detecting atrazine/desethyl-atrazine and elevated concentrations of nitrate in ground water in Colorado: U.S. Geological Survey Water-Resources Investigations Report 02–4269, 35 p.

SAS Institute, 1990, SAS/STAT User's guide, Version 6, 4th ed.: Cary, North Carolina, 1,686 p.

Smith, J.H., Wickham, J.D., Norton, D., Wade, T.G., and Jones, K.B., 2001, Utilization of landscape indicators to model potential pathogen impaired waters: Journal of the American Water Resources Association, v. 37, no. 4, p. 805–814.

Ulery, R.L., 2000, Overview of the Texas Source Water Assessment Project: U.S. Geological Survey Fact Sheet 101–00, 6 p.

U.S. Department of Agriculture, 1995, STATSGO—U.S. coverage by States including Puerto Rico (excludes Alaska): U.S. Department of Agriculture, Natural Resources Conservation Service, National Cooperative Soil Survey, State Soil Survey Geographic Data Base, CD–ROM.

U.S. Department of Commerce, 1990, 1990 Census of population and housing: U.S. Department of Commerce, U.S. Census Bureau, accessed June 15, 2000, at URL http://www.census.gov/geo/www/tiger/index.html

U.S. Environmental Protection Agency, 2003, The Safe Drinking Water Act, Public Law 104–182, 104th Congress: accessed 2/2003 at http://www.epa.gov/safewater/sdwa/text.html

U.S. Geological Survey, 2000, National Land Cover Dataset: U.S. Geological Survey Fact Sheet 108–00, 2 p.