Simulation of Daily Pesticide Concentrations from Watershed Characteristics and Monthly Climatic Data

Scientific Investigations Report 2006–5181

Simulation of Daily Pesticide Concentrations from Watershed Characteristics and Monthly Climatic Data

By Aldo V. Vecchia and Charles G. Crawford

Scientific Investigations Report 2006–5181

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, Virginia: 2006

Contents

Figures

Figures—Continued

Figures—Continued

Tables

Conversion Factors and Datum

Multiply	By	To obtain
Area		
square mile (mi^2)	2.590	square kilometer (km^2)

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

°C = (°F − 32) / 1.8

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

Simulation of Daily Pesticide Concentrations from Watershed Characteristics and Monthly Climatic Data

By Aldo V. Vecchia and Charles G. Crawford

Abstract

A time-series model was developed to simulate daily pesticide concentrations for streams in the coterminous United States. The model was based on readily available information on pesticide use, climatic variability, and watershed characteristics and was used to simulate concentrations for four herbicides [atrazine, ethyldipropylthiocarbamate (EPTC), metolachlor, and trifluralin] and three insecticides (carbofuran, ethoprop, and fonofos) that represent a range of physical and chemical properties, application methods, national application amounts, and areas of use in the United States. The time-series model approximates the probability distributions, seasonal variability, and serial correlation characteristics in daily pesticide concentration data from a national network of monitoring stations.

The probability distribution of concentrations for a particular pesticide and station was estimated using the Watershed Regressions for Pesticides (WARP) model. The WARP model, which was developed in previous studies to estimate the probability distribution, was based on selected nationally available watershed-characteristics data, such as pesticide use and soil characteristics. Normality transformations were used to ensure that the annual percentiles for the simulated concentrations agree closely with the percentiles estimated from the WARP model.

Seasonal variability in the transformed concentrations was maintained by relating the transformed concentration to precipitation and temperature data from the United States Historical Climatology Network. The monthly precipitation and temperature values were estimated for the centroids of each watershed. Highly significant relations existed between the transformed concentrations, concurrent monthly precipitation, and concurrent and lagged monthly temperature. The relations were consistent among the different pesticides and indicated the transformed concentrations generally increased as precipitation increased but the rate of increase depended on a temperature-dependent growing-season effect.

Residual variability of the transformed concentrations, after removal of the effects of precipitation and temperature, was partitioned into a signal (systematic variability that is related from one day to the next) and noise (random variability that is not related from one day to the next). Variograms were used to evaluate measurement error, seasonal variability, and serial correlation of the historical data. The variogram analysis indicated substantial noise resulted, at least in part, from measurement errors (the differences between the actual concentrations and the laboratory concentrations). The variogram analysis also indicated the presence of a strongly correlated signal, with an exponentially decaying serial correlation function and a correlation time scale (the time required for the correlation to decay to e^{-1} equals 0.37) that ranged from about 18 to 66 days, depending on the pesticide type.

Simulated daily pesticide concentrations from the time-series model indicated the simulated concentrations for the stations located in the northeastern quadrant of the United States where most of the monitoring stations are located generally were in good agreement with the data. The model neither consistently overestimated or underestimated concentrations for streams that are located in this quadrant and the magnitude and timing of high or low concentrations generally coincided reasonably well with the data. However, further data collection and model development may be necessary to determine whether the model should be used for areas for which few historical data are available.

Introduction

The U.S. Environmental Protection Agency (USEPA) was charged with the establishment of health-based standards for pesticides in drinking water as part of the Food Quality Protection Act of 1996. Pesticide concentrations in surface waters can vary substantially from watershed to watershed because of differences in pesticide use, application practices and timing, and watershed characteristics. To establish the health-based standards for pesticides, concentration data were needed for a large number of surface-water supply locations in the United States. However, because the cost of monitoring pesticide concentrations at frequencies that would be sufficient to assess exposure risks for the large number of surface-water supply locations would be prohibitive, a method was needed to simulate daily pesticide concentrations for those locations for which concentration data were not available. Therefore, the U.S. Geological Survey (USGS), in cooperation with the USEPA, developed a time-series model to simulate daily pesti-

cide concentrations that reproduce the statistical characteristics of historical pesticide concentrations for a national network of monitoring stations. The model was based on readily available information on pesticide use, climatic variability, and watershed characteristics and was used to simulate concentrations for four herbicides [atrazine, ethyldipropylthiocarbamate (EPTC), metolachlor, and trifluralin] and three insecticides (carbofuran, ethoprop, and fonofos) that represent a range of physical and chemical properties, application methods, national application amounts, and areas of use in the United States.

Larson and Gilliom (2001) and Larson and others (2004) previously developed the Watershed Regressions for Pesticides (WARP) model to estimate the probability distribution of pesticide concentrations for unmonitored streams. The model uses empirical relations between historical pesticide concentrations and selected nationally available watershed-characteristics data, such as pesticide use and soil characteristics, to estimate percentiles of the concentrations. This study extended the WARP model to allow simulation of daily pesticide concentrations for randomly generated realizations, or traces, that closely reproduce the estimated percentiles from the earlier WARP model and that realistically reproduce the seasonal variability and serial correlation characteristics in data from the monitoring stations. The time-series characteristics, such as seasonal variability and serial correlation, along with the probability distribution of pesticide concentrations, are important considerations for the assessment of long-term pesticide-exposure risks.

Purpose and Scope

This report describes the time-series model used to simulate daily pesticide concentrations and to evaluate the strengths and limitations of the model for the assessment of long-term pesticide-exposure risks. Although the time-series model uses watershed characteristics and percentiles estimated from the WARP model, new techniques were required to simulate realistic traces, or potential realizations, of the daily concentrations. In particular, monthly time series of precipitation data and of temperature data for each watershed were used to simulate pesticide concentrations that reproduced the seasonal variability in historical data, new data transformation methods were used to ensure that long-term traces for the simulated concentrations closely reproduced the percentiles estimated from the WARP model, and variogram analyses were used to evaluate measurement error, seasonal variability, and serial correlation of the historical data. The time-series model was not intended for use as a tool to predict actual concentration for a specific day or location. Rather, the model was designed to simulate daily pesticide concentrations for which the long-term traces could be used to assess long-term pesticide-exposure risks.

Data Used to Develop Time-Series Model

Concentration data for 112 pesticide-monitoring stations in the coterminous United States (fig. 1, table 1) and annual concentration percentiles estimated from the WARP model for multiple pesticides (C.G. Crawford, S.J. Larson, and R.J. Gilliom, written commun., 2005) were used to develop the time-series model. The stations used are the same as those used to develop the WARP model for atrazine (Larson and others, 2004) and represent a wide variety of environmental settings and watershed characteristics. The concentration data were collected from October 1, 1990, through September 30, 2001 (water years 1991-2001) as part of the U.S. Geological Survey National Water Quality Assessment (NAWQA) and National Stream Quality Accounting Network (NASQAN) programs and are available by accessing the USGS National Water Information System (NWIS) (*http://waterdata.usgs. gov/nwis/qw*). The annual percentiles estimated from the WARP model characterize the probability distribution of daily pesticide concentrations for a randomly selected year. For example, if the 90th percentile for atrazine concentration for a particular station is 0.5 microgram per liter, 90 percent of the daily atrazine concentrations for that station in a typical year would be expected to be less than 0.5 microgram per liter. The annual percentiles do not indicate when during the year the concentrations tend to be high or low nor do they indicate the degree of correlation of neighboring concentrations. However, seasonal and interannual variability in pesticide concentrations generally are a result of seasonal and interannual variability in precipitation, temperature, and application rates. Furthermore, pesticide concentrations generally are correlated from one day to the next because the streamflow volume and pesticide load on any given day represent the integration (both spatially and temporally) of antecedent rainfall-runoff and chemical-transport processes.

For this study, monthly time series of precipitation data and of temperature data were developed for the watershed of each pesticide-monitoring station and used in the time-series model to reproduce seasonal and interannual variability in pesticide concentrations. The monthly time series were developed using monthly total precipitation data and average temperature data from the United States Historical Climatology Network (Karl and others, 1990) for 978 long-term meteorological observation stations in the coterminous United States (fig. 2). The monthly total precipitation data and average temperature data were used to estimate the precipitation and temperature values for the centroid of each watershed for each month. The estimated values were computed using a robust locally weighted regression model (Cleveland and Devlin, 1988) that was fitted to the historical precipitation and temperature data.
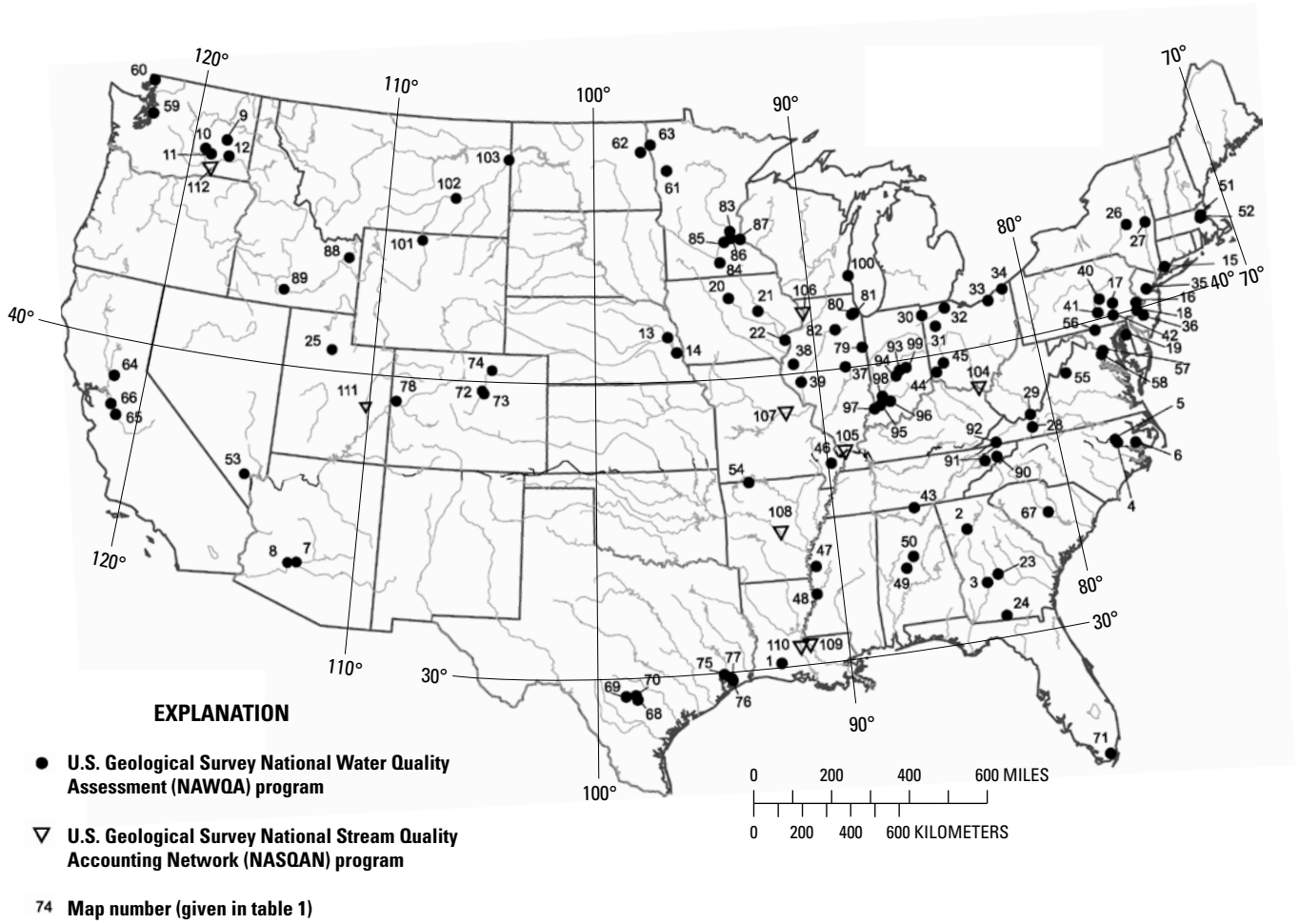
Figure 1.   Locations of pesticide-monitoring stations used to develop time-series model. [From Larson and others, 2004.]

Table 1.   Pesticide-monitoring stations used to develop time-series model.

[Latitude and longitude are for the centroid of the watershed]

Map number (figure 1)	Station name	U.S. Geological Survey station identification number	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square kilometers)	Agricultural land use (percent)
1	Mermentau River at Mermentau, Louisiana	08012150	30.51	92.45	3,576	64.05
2	Sope Creek near Marietta, Georgia	02335870	33.98	84.48	79	.04
3	Lime Creek at County Road near Cobb, Georgia	02350080	32.05	84.09	161	54.45
4	Tar River at Tarboro, North Carolina	02083500	36.16	78.06	5,753	26.37
5	Pete Mitchell Swamp near Penny Hill, North Carolina	02083833	35.83	77.49	44	56.51
6	Albemarle Canal near Swindell, North Carolina	02084558	35.73	76.73	191	60.55
7	Buckeye Canal near Avondale, Arizona	09514000	33.21	110.67	116,964	2.92
8	Hassayampa River near Arlington, Arizona	09517000	33.92	112.70	3,967	4.60
9	Crab Creek near Ritzville, Washington	12464770	47.48	118.04	1,187	23.83
10	Crab Creek Lateral near Othello, Washington	12472380	46.91	119.55	145	92.60
11	E1 68 D Wasteway near Othello, Washington	12473740	46.83	119.03	377	61.16
12	Palouse River at Hooper, Washington	13351000	46.98	117.36	6,380	50.37
13	Maple Creek near Nickerson, Nebraska	06800000	41.66	97.04	955	96.68
14	Platte River at Louisville, Nebraska	06805500	41.37	102.80	221,497	22.45
15	Norwalk River at Winnipauk, Connecticut	01209710	41.24	73.45	85	2.88
16	Little Neshaminy Creek near Neshaminy, Pennsylvania	01464907	40.22	75.18	72	32.29
17	Tulpehocken Creek near Bernville, Pennsylvania	01470779	40.36	76.26	184	82.27
18	Schuylkill River at Philadelphia, Pennsylvania	01474500	40.40	75.78	4,896	38.00
19	Chesterville Branch near Crumpton, Maryland	01493112	39.29	75.92	17	90.76
20	Iowa River near Rowan, Iowa	05449500	43.02	93.64	1,083	89.89
21	Wolf Creek near Dysart, Iowa	05464220	42.24	92.62	775	90.26
22	Iowa River at Wapello, Iowa	05465500	42.46	92.56	32,364	83.74
23	Tucsawhatchee Creek near Hawkinsville, Georgia	02215100	32.29	83.68	420	55.91

Table 1.   Pesticide-monitoring stations used to develop time-series model.—Continued

[Latitude and longitude are for the centroid of the watershed]

Map number (figure 1)	Station name	U.S. Geological Survey station identification number	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square kilometers)	Agricultural land use (percent)
24	Withlacoochee River near Quitman, Georgia	02318500	31.25	83.51	3,863	49.83
25	Jordan River at Salt Lake City, Utah	10171000	40.23	111.66	9,095	7.49
26	Canajoharie Creek near Canajoharie, New York	01349150	42.84	74.65	154	61.11
27	Mohawk River at Cohoes, New York	01357500	42.95	74.67	9,113	27.27
28	Reed Creek at Grahams Forge, Virginia	03167000	36.96	81.13	669	43.58
29	New River at Glen Lyn, Virginia	03176500	36.86	80.96	9,781	29.50
30	St. Joseph River near Newville, Indiana	04178000	41.66	84.70	1,600	78.00
31	Auglaize River near Fort Jennings, Ohio	04186500	40.68	84.20	858	88.87
32	Maumee River at Waterville, Ohio	04193500	41.15	84.41	16,409	87.54
33	Cuyahoga River near Newburgh Heights, Ohio	04208504	41.28	81.42	2,043	27.27
34	Grand River at Harpersfield, Ohio	04211820	41.56	80.88	1,431	41.78
35	Bound Brook at Middlesex, New Jersey	01403900	40.61	74.42	125	1.27
36	Great Egg Harbor River near Sicklerville, New Jersey	01410784	39.77	74.96	39	16.25
37	Sangamon River at Monticello, Illinois	05572000	40.29	88.44	1,426	95.52
38	La Moine River at Colmar, Illinois	05584500	40.50	90.88	1,695	79.69
39	Illinois River at Valley City, Illinois	05586100	40.89	88.85	69,164	78.89
40	East Mahantango Creek at Klingerstown, Pennsylvania	01555400	40.70	76.57	115	55.11
41	Cedar Run at Eberlys Mill, Pennsylvania	01571490	40.22	76.95	32	33.10
42	Mill Creek near Lyndon, Pennsylvania	01576540	40.05	76.15	140	81.12
43	Hester Creek near Plevna, Alabama	0357479650	35.00	86.40	76	66.32
44	Great Miami River at Hamilton, Ohio	03274000	40.01	84.25	9,403	80.35
45	Holes Creek at Kettering, Ohio	393944084120700	39.63	84.19	51	29.42
46	Little River Ditch No. 1 near Morehouse, Missouri	07043500	37.05	89.84	1,143	84.38
47	Bogue Phalia near Leland, Mississippi	07288650	33.72	90.88	1,301	86.10
48	Yazoo River near Long Lake, Mississippi	07288955	33.86	90.10	34,850	55.77

Table 1.   Pesticide-monitoring stations used to develop time-series model.—Continued

[Latitude and longitude are for the centroid of the watershed]

Map number (figure 1)	Station name	U.S. Geological Survey station identification number	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square kilometers)	Agricultural land use (percent)
49	Cahaba Valley Creek at Pelham, Alabama	0242354750	33.36	86.73	65	10.37
50	Cahaba River at Centreville, Alabama	02424000	33.25	86.88	2,656	9.87
51	Aberjona River at Winchester, Massachusetts	01102500	42.49	71.15	59	.01
52	Charles River at Watertown, Massachusetts	01104615	42.23	71.34	694	4.34
53	Las Vegas Wash near Las Vegas, Nevada	094196783	36.32	115.34	2,639	.16
54	Yocum Creek near Oak Grove, Arkansas	07053250	36.40	93.43	134	71.68
55	Muddy Creek at Mount Clinton, Virginia	01621050	38.52	78.95	37	71.37
56	Monocacy River at Bridgeport, Maryland	01639000	39.81	77.25	456	67.83
57	Potomac River at Washington, D.C.	01646580	39.16	78.38	29,975	34.61
58	Accotink Creek near Annandale, Virginia	01654000	38.86	77.26	60	4.56
59	Thornton Creek near Seattle, Washington	12128000	47.72	122.31	29	.02
60	Nooksack River at Brennan, Washington	12213140	48.83	122.08	2,023	12.15
61	Wild Rice River at Twin Valley, Minnesota	05062500	47.30	95.75	2,407	47.96
62	Turtle River near Arvilla, North Dakota	05082625	48.02	97.81	658	93.65
63	Snake River above Alvarado, Minnesota	05085900	48.22	96.53	565	88.16
64	Arcade Creek near Del Paso Heights, California	11447360	38.68	121.30	81	.14
65	Merced River near Newman, California	11273500	37.63	119.96	3,619	13.72
66	San Joaquin River near Vernalis, California	11303500	37.65	120.36	19,023	22.79
67	Gills Creek at Colombia, South Carolina	02169570	34.05	80.92	154	5.30
68	Salado Creek at San Antonio, Texas	08178800	29.58	98.47	505	3.05
69	Medina River at La Coste, Texas	08180640	29.71	99.09	2,100	8.37
70	San Antonio River near Elmendorf, Texas	08181800	29.57	98.82	4,528	16.80
71	C-111 Canal near Homestead, Florida	252414080333200	25.50	80.53	132	72.06

Table 1.   Pesticide-monitoring stations used to develop time-series model.—Continued

[Latitude and longitude are for the centroid of the watershed]

Map number (figure 1)	Station name	U.S. Geological Survey station identification number	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square kilometers)	Agricultural land use (percent)
72	Cherry Creek at Denver, Colorado	06713500	39.67	104.91	61	0.45
73	South Platte River at Denver, Colorado	06714000	39.27	105.40	10,009	2.64
74	Lonetree Creek near Greeley, Colorado	06753990	40.85	104.84	1,472	17.35
75	East Fork Double Bayou near Anahuac, Texas	294349094345999	29.82	94.56	111	85.64
76	Whites Bayou near Anahuac, Texas	295001094384699	29.83	94.62	27	52.69
77	West Prong Old River near Dayton, Texas	295740094542399	30.02	94.95	75	91.51
78	Reed Wash near Mack, Colorado	09153290	39.25	108.79	35	48.03
79	Sugar Creek at Milford, Illinois	05525500	40.57	87.70	1,158	97.30
80	Salt Creek at Western Springs, Illinois	05531500	41.98	88.02	290	2.54
81	Des Plaines River at Riverside, Illinois	05532500	42.22	87.97	1,634	32.75
82	Illinois River at Ottawa, Illinois	05553500	41.54	87.80	28,329	71.32
83	Shingle Creek in Minneapolis, Minnesota	05288705	45.08	93.39	73	2.64
84	Little Cobb River near Beauford, Minnesota	05320270	43.95	93.71	336	92.39
85	Minnesota River near Jordan, Minnesota	05330000	44.68	95.28	42,040	87.47
86	Nine Mile Creek at Bloomington, Minnesota	05330902	44.87	93.38	115	.01
87	Mississippi River at Hastings, Minnesota	05331580	45.54	94.68	95,957	59.29
88	Teton River near St. Anthony, Idaho	13055000	43.78	111.17	2,294	34.08
89	Rock Creek at Twin Falls, Idaho	13092747	42.35	114.35	623	21.95
90	Big Limestone Creek near Limestone, Tennessee	03466208	36.30	82.59	205	59.88
91	Nolichucky River near Lowland, Tennessee	03467609	36.10	82.55	4,373	23.12
92	Copper Creek near Gate City, Virginia	03526000	36.78	82.39	276	29.82
93	Little Buck Creek near Indianapolis, Indiana	03353637	39.67	86.10	44	46.26
94	White River near Centerton, Indiana	03354000	39.99	85.91	6,324	76.84
95	Kessinger Ditch near Monroe City, Indiana	03360895	38.67	87.35	145	92.28
96	East Fork White River at Shoals, Indiana	03373500	39.21	85.87	12,751	67.51

Table 1.   Pesticide-monitoring stations used to develop time-series model.—Continued

[Latitude and longitude are for the centroid of the watershed]

| Map number (figure 1)	Station name	U.S. Geological Survey station identification number	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square kilometers)	Agricultural land use (percent)
97	White River at Hazleton, Indiana	03374100	39.34	86.22	29,290	69.31
98	White River near Elnora, Indiana	3852340870-71801	39.69	86.34	12,434	70.59
99	Sugar Creek at New Palestine, Indiana	3943400855-24601	39.88	85.72	246	92.38
100	Milwaukee River at Milwaukee, Wisconsin	04087000	43.47	88.12	1,803	65.56
101	Bighorn River at Kane, Wyoming	06279500	43.64	108.39	40,824	3.63
102	Yellowstone River at Forsyth, Montana	06295000	44.70	108.71	102,191	5.57
103	Yellowstone River near Sidney, Montana	06329500	45.03	107.50	177,139	5.41
104	Ohio River near Greenup, Kentucky	03216600	39.53	80.79	159,355	23.49
105	Ohio River near Grand Chain, Illinois	03612500	38.23	84.11	527,228	38.70
106	Mississippi River at Clinton, Iowa	05420500	44.83	92.49	238,665	55.03
107	Missouri River at Hermann, Missouri	06934500	43.75	102.89	1,346,053	31.89
108	Arkansas River below Little Rock, Arkansas	07263620	36.86	99.69	408,729	31.07
109	Mississippi River near St. Francisville, Louisiana	07373420	41.26	96.67	2,965,462	39.84
110	Atchafalaya River at Melville, Louisiana	07381495	33.85	96.16	240,627	35.07
111	Green River at Green River, Utah	09315000	40.91	109.28	105,670	2.30
112	Snake River at Burbank, Washington	13353200	44.11	115.17	279,443	11.30
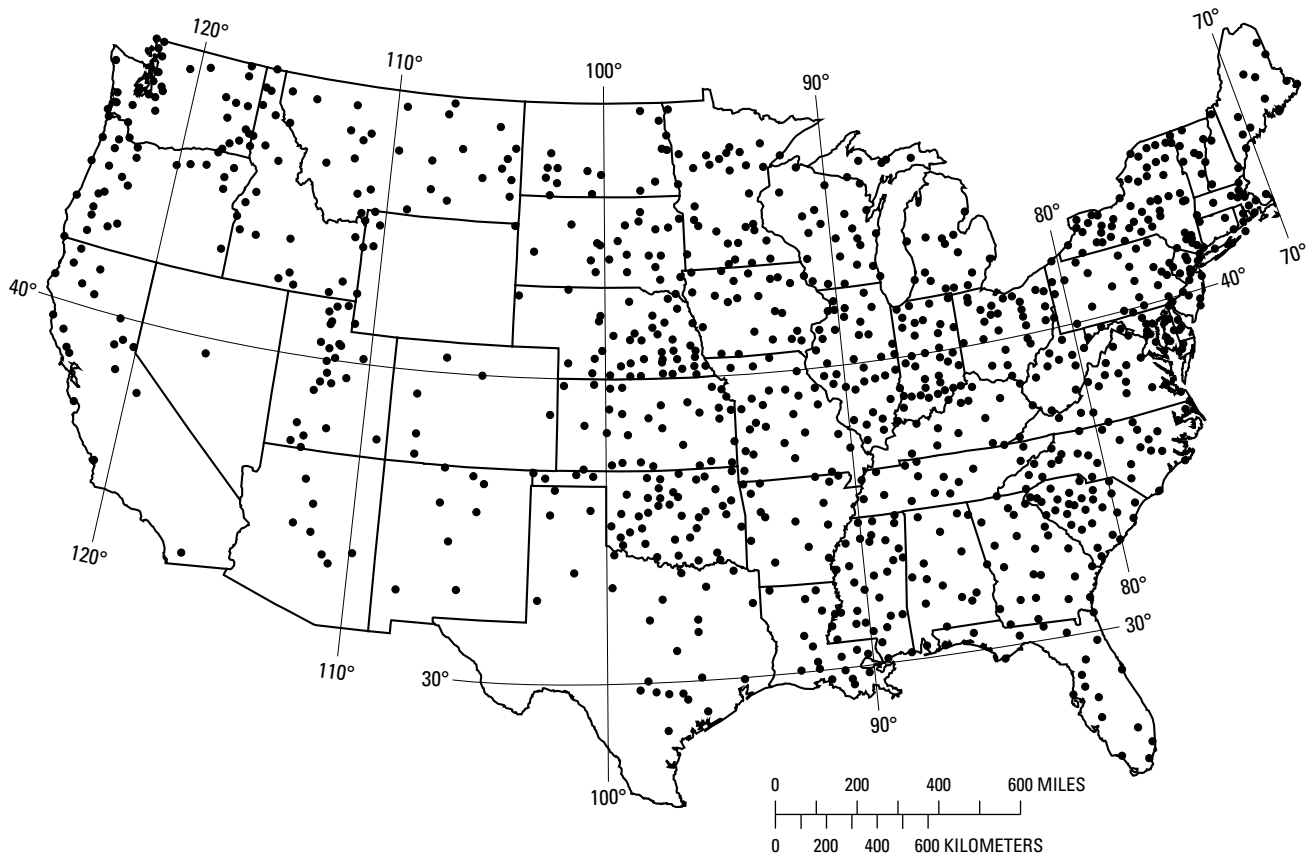
Figure 2.   Locations of selected meteorological observation stations for which data are given in the United States Historical Climatology Network. [From Karl and others, 1990.]

Development of Time-Series Model

Normality Transformations

The first requirement for the time-series model was that the annual percentiles for the simulated concentrations agree closely with the percentiles estimated from the WARP model. To satisfy this requirement, the assumption was made that a transformation existed such that the annual probability distribution of the transformed concentrations was a standard normal distribution. Although such a transformation always exists, a simple transformation that can be applied easily for a wide variety of stations and pesticide species was needed. A normality transformation can be expressed as follows:

$$P\{F(C) < z_p\} = P\{C < F^{-1}(z_p)\} = p \qquad (1)$$

where

$P\{\ \}$	denotes probability;
$F(.)$	is a one-to-one transformation;
C	is the concentration, in micrograms per liter;
z_p	is the pth percentile of a standard normal distribution;
$F^{-1}(.)$	is the inverse transformation; and
p	is a specified value.

Thus, for $F(C)$ to have a standard normal distribution, the pth percentile of the annual probability distribution for C must satisfy the following equation:

$$w_p = F^{-1}(z_p) \qquad (2)$$

where

w_p	is the pth percentile, estimated from the WARP model, of the annual probability distribution for C.

The parameters upon which the transformation function generally depends were determined from the percentiles estimated from the WARP model by minimizing the sum of the squared errors between the log-transformed estimated percentiles and the log-transformed normal percentiles as follows:

$$SS = sum\{[\log w_p - \log F^{-1}(z_p)]^2, p = 5, 10, 15, 25, 50, 75, 85, 90, 95\},$$

where

SS	is the sum of the squared deviations,
sum { }	denotes the sum of the values in the braces, and
log	denotes the base-10 logarithm.

The transformation equations given in table 2 then were selected from several transformation equations on the basis of the overall goodness of fit for the stations used in the study. Each transformation depends on empirical constants, or parameters, that need to be estimated for each station-pesticide combination. Transformation LN3 is equivalent to the assumption that the concentration for a randomly selected day of the year has a three-parameter log-normal distribution. Transformations AT1 and AT2 are generalizations of a two-parameter log-normal distribution that allow nonlinearity in the relation between log w_p and z_p. The parameter γ controls the degree of nonlinearity. When $\gamma = 0$, both transformations are equivalent to a two-parameter log-normal distribution in which the relation between log w_p and z_p is linear. When $\gamma \neq 0$ and $z_p < c$, a nonlinear relation exists between log w_p and z_p for both transformations. The parameter constraints ensure that the relation between w_p and z_p is strictly increasing.

The transformation results for atrazine (fig. 3) were similar for all stations. The values for the 50 percent and higher percentiles estimated from the transformations were similar for all three transformations and generally provided a good fit to the percentiles estimated from the WARP model. However, the fitted values for the low percentiles differed among the transformations. Transformation LN3 generally provided a poor fit to the low percentiles, but transformations AT1 and AT2 generally provided a good fit to the low percentiles. The most significant differences among the transformations occurred for percentiles of less than 5 percent—the fitted values for those percentiles generally were highest for transformation LN3 and lowest for transformation AT2. Because the 5th percentile was the lowest value available from the WARP model, the transformations could not be evaluated on the basis of percentiles of less than 5 percent.

The transformation results for EPTC (fig. 4) differed considerably from the results for atrazine. Transformation LN3 generally provided a poor fit to all but the high percentiles estimated from the WARP model. Transformation AT1 provided a reasonable fit to the percentiles for some stations [such as the Milwaukee River at Milwaukee, Wisconsin, station (map number 100)] but a poor fit to the percentiles for other stations [such as the Monocacy River at Bridgeport, Maryland, station (map number 56)]. The fitted curves for transformation AT1 often were convex (curving downward) and the high percentiles were underestimated. Transformation AT2 generally provided a good fit to the percentiles. The fitted curves for that alternative generally were convex for the low percentiles and concave for the high percentiles. Transformation AT2 generally resulted in a higher estimated probability of both extreme high and extreme low concentrations than the other transformations.

The transformation results for fonofos (fig. 5) and the remaining pesticides for which concentrations were simulated (not shown) were similar to the results for EPTC. Transformation LN3 generally provided a poor fit to the moderate and low percentiles estimated from the WARP model, and transforma-

tion AT1 often provided a poor fit to the moderate and high percentiles. Transformation AT2, which hereinafter was used to transform raw concentrations, generally provided a reasonable fit to the percentiles for most stations and pesticides.

Seasonal Structure of Transformed Concentration Data

After raw concentration data were transformed using transformation AT2 (table 2), the transformed concentrations for all stations were combined and used to fit a model to describe seasonal variability in the transformed concentrations. The normality transformations were based on the assumption that the transformed concentration for a randomly selected day and station had a standard normal distribution. However, because spatial and seasonal variability and serial correlation in the time-series data preclude the assumption that daily concentrations are independent and identically distributed, this does not imply that an individual transformed concentration for any given day or station has a standard normal distribution. To reduce the effects of seasonal sampling bias and serial correlation on the fitted model, the following process was used to obtain a random sample of the transformed concentrations. For a particular station and a particular year (y) of the historical record for that station, a uniform random number (u) between zero and one was generated. Then, using $\{y + u_j, j = 1, 2, ..., J\}$ to represent the times, in decimal years, for which concentration data were available during that year, the transformed concentration closest to $y + u$ (the concentration that minimizes $|u_j - u|$) was determined. That concentration then was selected for inclusion in the random sample and another random number was generated for the next year for that station. The process was repeated until a random sample of transformed concentrations was obtained for each station/year pair in the historical record. Although the process did not give a true random sample from the annual distribution because

concentration data were not available for every day of the year, the process did allow for the best selection that could be obtained given the limitations of the historical data.

After the random samples were obtained, the transformed concentrations were related to precipitation and temperature data from the United States Historical Climatology Network (Karl and others, 1990) for various time lags to determine if statistically significant relations existed. The concentrations were related to the historical data by letting $C^*(t) = F(C)$ be the transformed concentration for a particular station and a particular time (t), which was designated by the month and day associated with the historical concentration as follows:

$$t = (m, d)$$

where

$m = 1, 2, ...$ denotes the month, and
$d = 1, 2, ...$ denotes the day of the month.

The monthly precipitation and temperature values estimated as described previously for the centroid of a selected watershed then were used to compute lagged precipitation and temperature values for each historical concentration as follows:

$$P_k(m, d) = \left(1 - \frac{d}{n_m}\right)P(m - k - 1) + \left(\frac{d}{n_m}\right)P(m - k)$$

and

$$T_k(m, d) = \left(1 - \frac{d}{n_m}\right)T(m - k - 1) + \left(\frac{d}{n_m}\right)T(m - k)$$

where

$P_k(m, d)$ is lag-k precipitation, in inches, for month m and day d;
n_m is the number of days in month m;

Table 2.   Transformation equations for percentiles estimated from Watershed Regressions for Pesticides model.

Transformation	Equation[1]	Parameters to be estimated	Parameter constraints
LN3 (three-parameter log-normal)	$\ln (w_p - \zeta) = \mu + \sigma z_p$	ζ, μ, σ	$\zeta > 0, \sigma > 0$
AT1 (alternative transformation 1)	$\ln (w_p) = \alpha + \beta(z_p - c) + \gamma\left\{\dfrac{(z_p - c)^2}{(1 + c - z_p)}\right\} I[z_p < c]$	α, β, c, γ	$\beta > 0, \gamma > -\beta$
AT2 (alternative transformation 2)	$\ln (w_p) = \alpha + \beta(z_p - c) + \gamma\left\{(z_p - c)^2 + \left(\dfrac{1}{6}\right)(z_p - c)^3\right\} I[z_p < c]$	α, β, c, γ	$\beta > 0, \gamma > \dfrac{\beta}{2}$

[1] w_p is the pth percentile, estimated from the Watershed Regressions for Pesticides model, of the annual probability distribution for the concentration; z_p is the pth percentile of a standard normal distribution; c is a specified constant; ln denotes natural logarithm $I[z_p < c] = 1$ if $z_p < c$; and $I[z_p < c] = 0$ if $z_p \geq c$.

$k = 0, 1, 2, ...$ is the time lag, in months;

$P(m)$ is the estimated monthly precipitation value in inches;

$T_k(m, d)$ is lag-k temperature, in degrees Fahrenheit, for month m and day d; and

$T(m)$ is the estimated monthly temperature value, in degrees Fahrenheit.

A weighted average of the values for months $m - k - 1$ and $m - k$, with weights determined by d, was used to ensure a smooth transition from one month to the next. For example, lag-0 precipitation for the middle of month m ($d = 15$) is essentially equal to the average of $P(m)$ and $P(m - 1)$, and lag-0 precipitation for the end of month m ($d = 31$), assuming a month with 31 days, is equal to $P(m)$.

Seasonal variability is particularly difficult to analyze when a large percentage of the data are below the detection limit. For the constituents analyzed for this study, only atrazine concentrations (12 percent censored) and metolachlor concentrations (20 percent censored) routinely exceeded the detection limit. EPTC concentrations (85 percent censored), trifluralin concentrations (81 percent censored), carbofuran concentrations (98 percent censored), ethoprop concentrations (99 percent censored), and fonofos concentrations (96 percent censored) seldom exceeded the detection limit. Therefore, the variability in atrazine and metolachlor concentrations was analyzed before the variability in the remaining pesticide concentrations. The insight gained from the analysis of the atrazine and metolachlor concentrations then was used to extend the model to other pesticide data. The transformed atrazine and metolachlor concentrations are shown in figures 6 through 8 in relation to the square root of lag-0 precipitation ($SRP0$), lag-0 temperature ($T0$), and lag-0 minus lag-1 temperature ($T01$), respectively. $T0$, $SRP0$, and $T01$, in an exploratory analysis, provided the best climatic variables for the model for seasonal variability. The smoothed line in figure 6 indicates the transformed concentrations generally are high for high lag-0 precipitation. The smoothed line in figure 7 indicates an apparent nonmonotonic relation exists between the transformed concentrations and $T0$. The concentrations generally increase for temperatures between about 40 and 55 degrees Fahrenheit and decrease for temperatures of more than 55 degrees Fahrenheit. The smoothed line in figure 8

indicates the transformed concentrations generally increase when $T01$ is positive (that is, when the average monthly temperature is increasing).

The potential interaction between the transformed concentrations and the climatic variables was evaluated using three-dimensional plots for which the transformed concentrations within each temperature or precipitation bin were averaged to smooth the random variability in the concentrations. The highest transformed concentrations generally occurred when both $T0$ and $T01$ were large (fig. 9). However, when $T0$ was more than about 70 degrees Fahrenheit, $T01$ was either negative or slightly positive for many concentrations. The joint occurrence of high $T0$ values and low $T01$ values resulted in the apparent decrease in transformed concentrations shown in figure 7 for high $T0$ values. Likewise, the generally constant transformed concentrations shown in figure 8 for high $T01$ values resulted from the joint occurrence of low $T0$ values and high $T01$ values. When $T0$ was less than about 40 degrees Fahrenheit, the transformed concentrations generally had little sensitivity to $SRP0$ (fig. 10). However, when $T0$ was high, the transformed concentrations generally increased as $SRP0$ increased. The high $T0$ and $SRP0$ values resulted in the apparent increase in transformed concentrations shown in figure 6 for high $SRP0$ values.

Because of the interactions noted between the transformed concentrations and the climatic variables, the following regression model was used to relate the transformed concentrations to the variables:

$$C^* = b_0 + b_1[T0 - 40]^+ + b_2[T01]^+ + b_3[T0 - 40]^+ [T01]^+ + b_4 SRP0 + b_5 SRP0[T0 - 40]^+ + b_6 SRP0[T01]^+ + \varepsilon$$

where

C^* is the actual (uncensored) transformed concentration (dimensionless);

$b_0, ..., b_6$ are regression coefficients;

$T0$ is lag-0 temperature, in degrees Fahrenheit;

$[.]^+$ denotes the maximum of zero and the quantity in brackets;

$T01$ is lag-0 minus lag-1 temperature, in degrees Fahrenheit;

$SRP0$ is the square root of lag-0 precipitation, in inches; and

ε is the error.

The regression model was fitted to the transformed concentrations for all of the pesticides (not just for atrazine and metolachlor). Because of the large percentage of censored historical concentrations used to fit the model, Tobit regression (Tobin, 1958; Cohen, 1976), which uses maximum likelihood estimation (assuming normally distributed errors) to fit a regression model to data with multiple censoring levels, was used rather than ordinary least-squares regression. The survival regression (survreg) function in S-Plus (Insightful Corp., 2002) was used to fit the Tobit regression model. Generalized likelihood-ratio tests were used to determine the best combination of terms for the model and to determine the statistical significance of the coefficients in the model. The following Tobit regression model, with $SRP0[T0 - 40]^+$ and $SRP0[T01]^+$ (in addition to the intercept), was the best model for explaining the variability in the transformed concentrations:

$$C^* = -0.158 + 0.011\ SRP0[T0 - 40]^+ + \\ 0.043\ SRP0[T01]^+ + X \quad (3)$$

where

X is the model residual.

The number of uncensored historical concentrations used to fit the model (eq. 3) was 878, and the number of censored historical concentrations was 1,728. The terms that involve the climatic variables were highly significant and had p-values of less than 10^{-9} for both coefficients.

The goodness of fit of the Tobit regression model was not easily ascertained from a comparison of the fitted and historical values of the transformed concentrations for all pesticides (fig. 11) because of the large percentage of censored values. However, the fitted values for atrazine (fig. 12) and metolachlor (fig. 13) generally were in good agreement with the transformed concentrations (note from equation 3 that - 0.158 is the smallest fitted value for transformed concentration). Although a substantial amount of unexplained variability remained in the transformed concentrations for both atrazine and metolachlor, no significant bias or lack of fit occurred between the fitted values and the transformed concentrations, and the data were consistent with the assumption of normally distributed, homoscedastic (and occasionally censored) errors. Because of the large percentage of censored historical concentrations for the remaining pesticides, the potential bias or lack of fit between the fitted values and the transformed concentra-

tions was difficult to determine. For example, the transformed concentrations for fonofos (fig. 14) and trifluralin (fig. 15) indicated no obvious violations of the model assumptions, but, because of the large percentage of censored historical concentrations, the model could not be validated for either pesticide.

The residuals (X) from the Tobit regression model (eq. 3) were assumed to be normally distributed and involved conditioning on the values of the climatic variables. These values vary from day to day and from location to location. The transformed concentrations from the WARP model (C^*) were assumed to be normally distributed for a randomly selected day and location and, thus, did not involve conditioning on the climatic variables. The simulated values for C^*, computed using equation 3, therefore needed to be approximately normally distributed for a randomly selected day and station for the assumptions of the Tobit regression model and the WARP model to be compatible. Although this assumption was reasonable for the stations analyzed for this report, the assumption may not be reasonable for all locations because the stations analyzed do not represent a random sample of stream locations.

The Tobit regression model (eq. 3) also can be expressed in the following form:

$$C^* = -0.158 + G(T0, T01)SRP0 + X$$

where

$$G(T0, T01) = 0.011[T0 - 40]^+ + 0.043[T01]^+.$$

Thus, the transformed concentrations are linearly related to $SRP0$ with a positive slope $G(T0, T01)$ that can be interpreted as a growing-season effect. If $T0$ is less than 40 degrees Fahrenheit and $T01$ is negative, the data probably were collected past the end of the growing season. In that case, the slope is zero and the transformed concentration is small irrespective of precipitation. The slope remains at zero as long as the average monthly temperature continues to decrease. If $T0$ is more than 40 degrees Fahrenheit and increasing, the data probably were collected near the beginning of the growing season. In that case, the slope is positive, presumably indicating the presence of pesticide residue that can be transported to the stream during precipitation. If $T0$ is more than 40 degrees Fahrenheit and decreasing, the data probably were collected near the end of the growing season. In that case, the slope is positive but decreasing, presumably indicating the declining presence of pesticide residue late in the growing season.
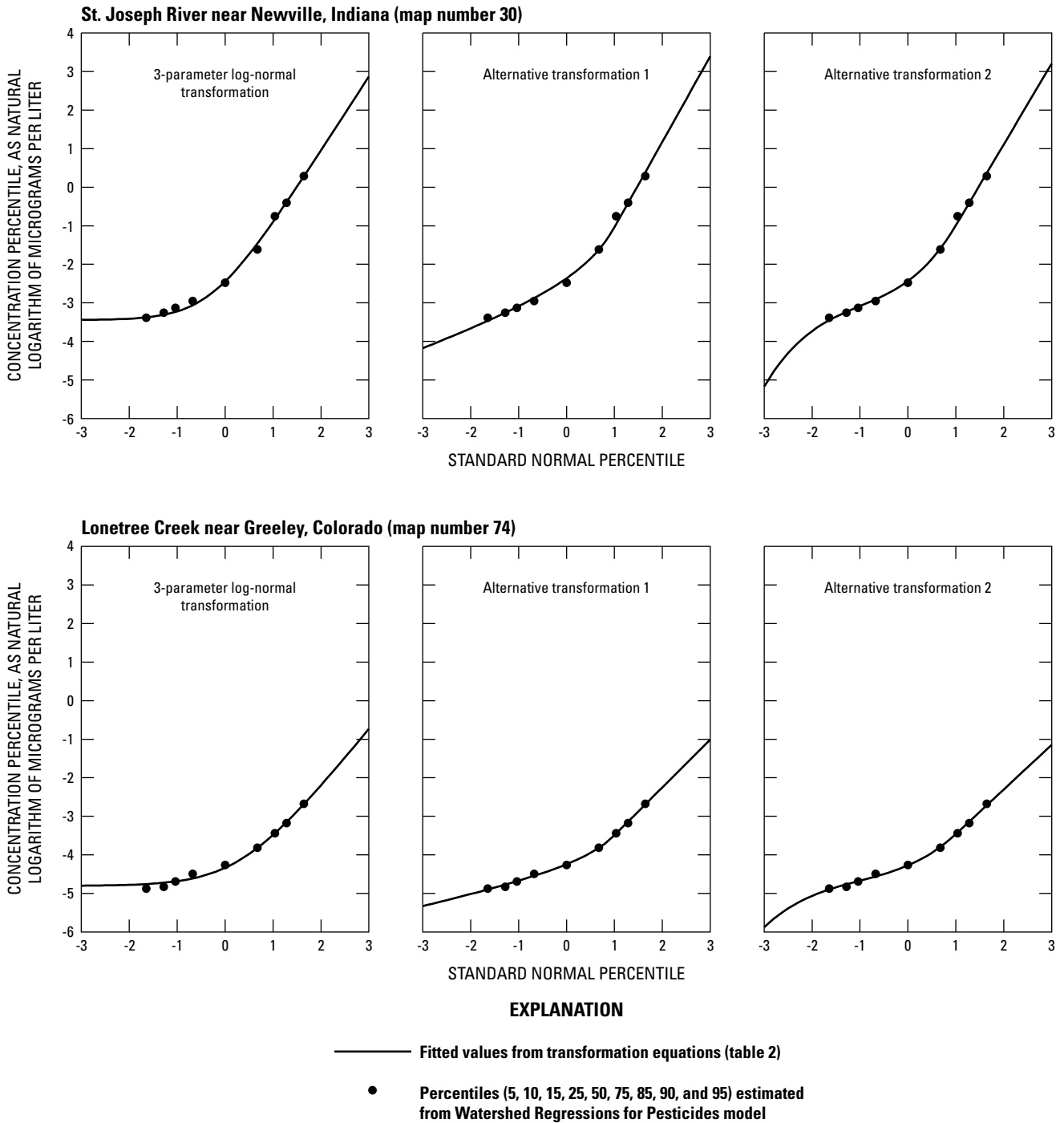
Figure 3.   Log-transformed atrazine concentration percentiles for the St. Joseph River near Newville, Indiana, and Lonetree Creek near Greeley, Colorado, stations.
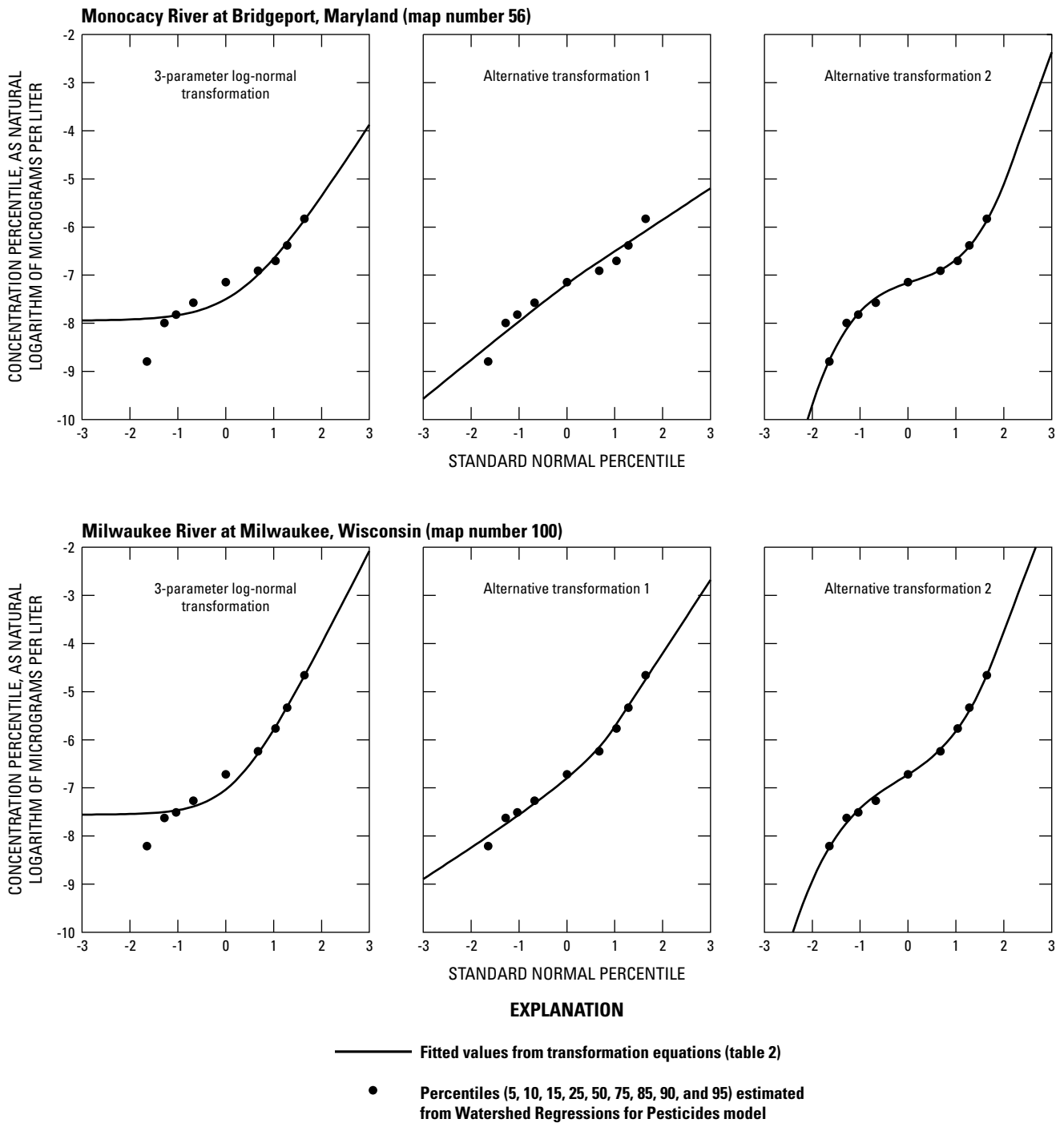
Figure 4. Log-transformed ethyldipropylthiocarbamate concentration percentiles for the Monocacy River at Bridgeport, Maryland, and Milwaukee River at Milwaukee, Wisconsin, stations.
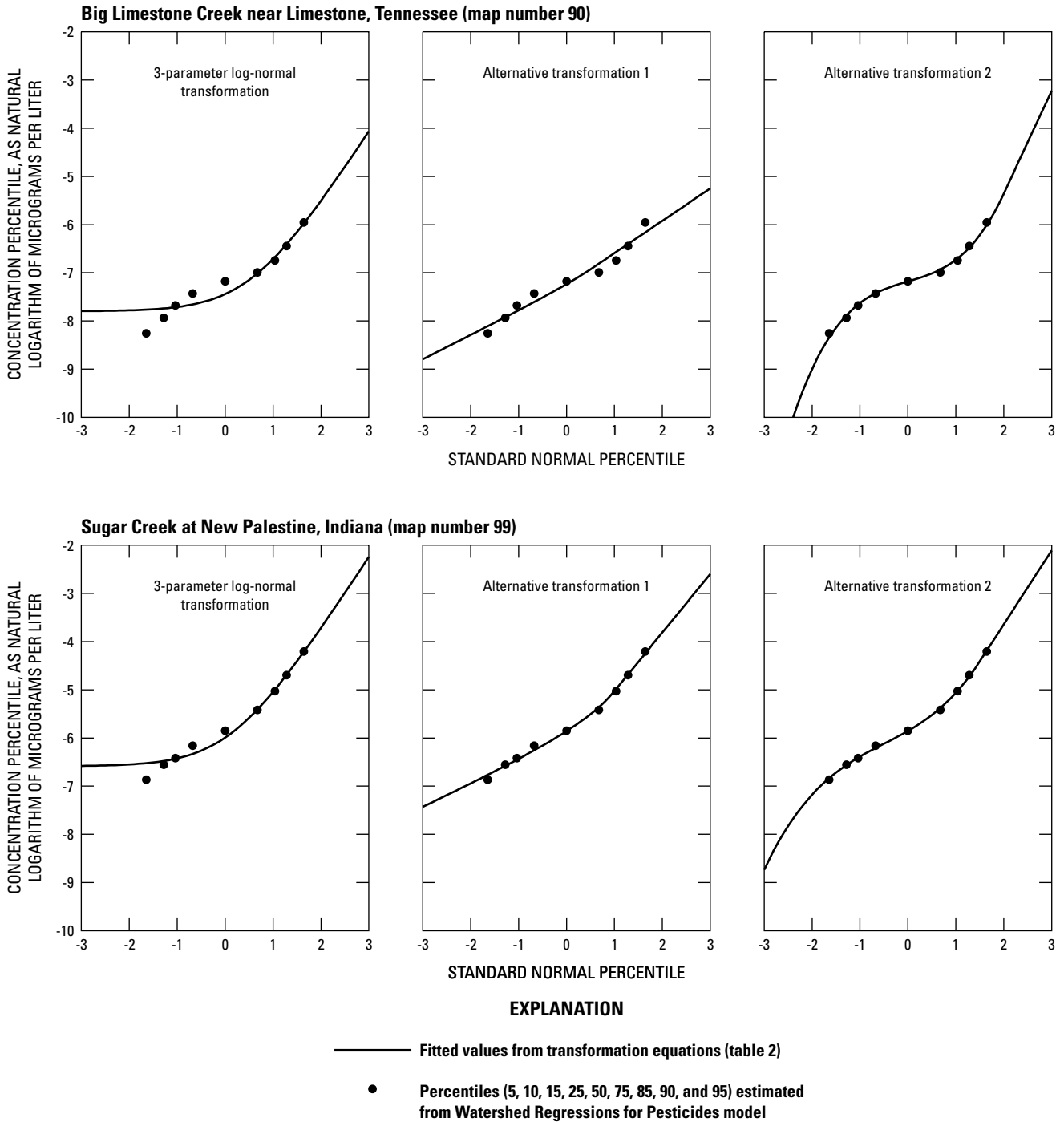
Figure 5. Log-transformed fonofos concentration percentiles for the Big Limestone Creek near Limestone, Tennessee, and Sugar Creek at New Palestine, Indiana, stations.
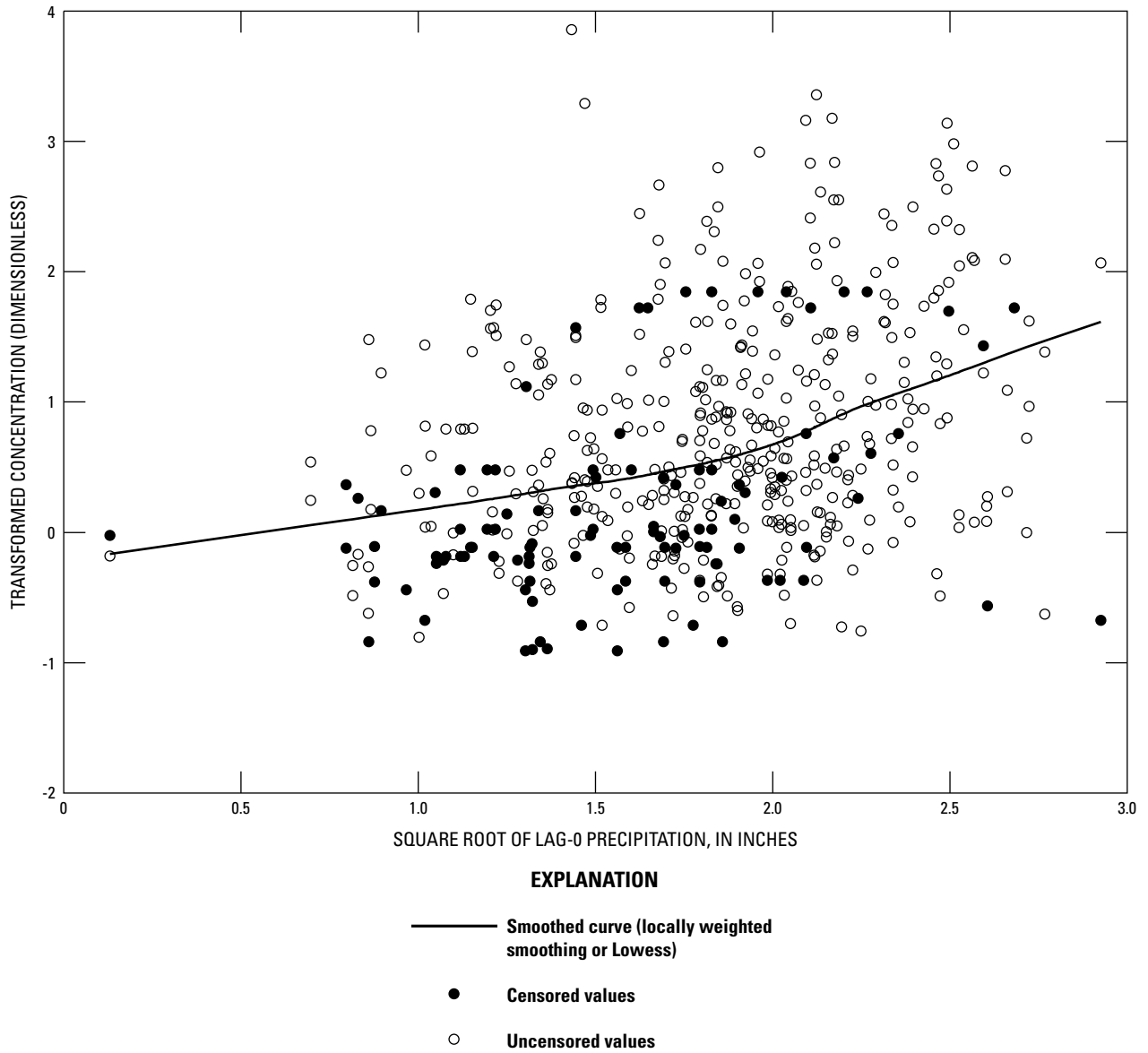
Figure 6.   Transformed atrazine and metolachlor concentrations and lag-0 precipitation.
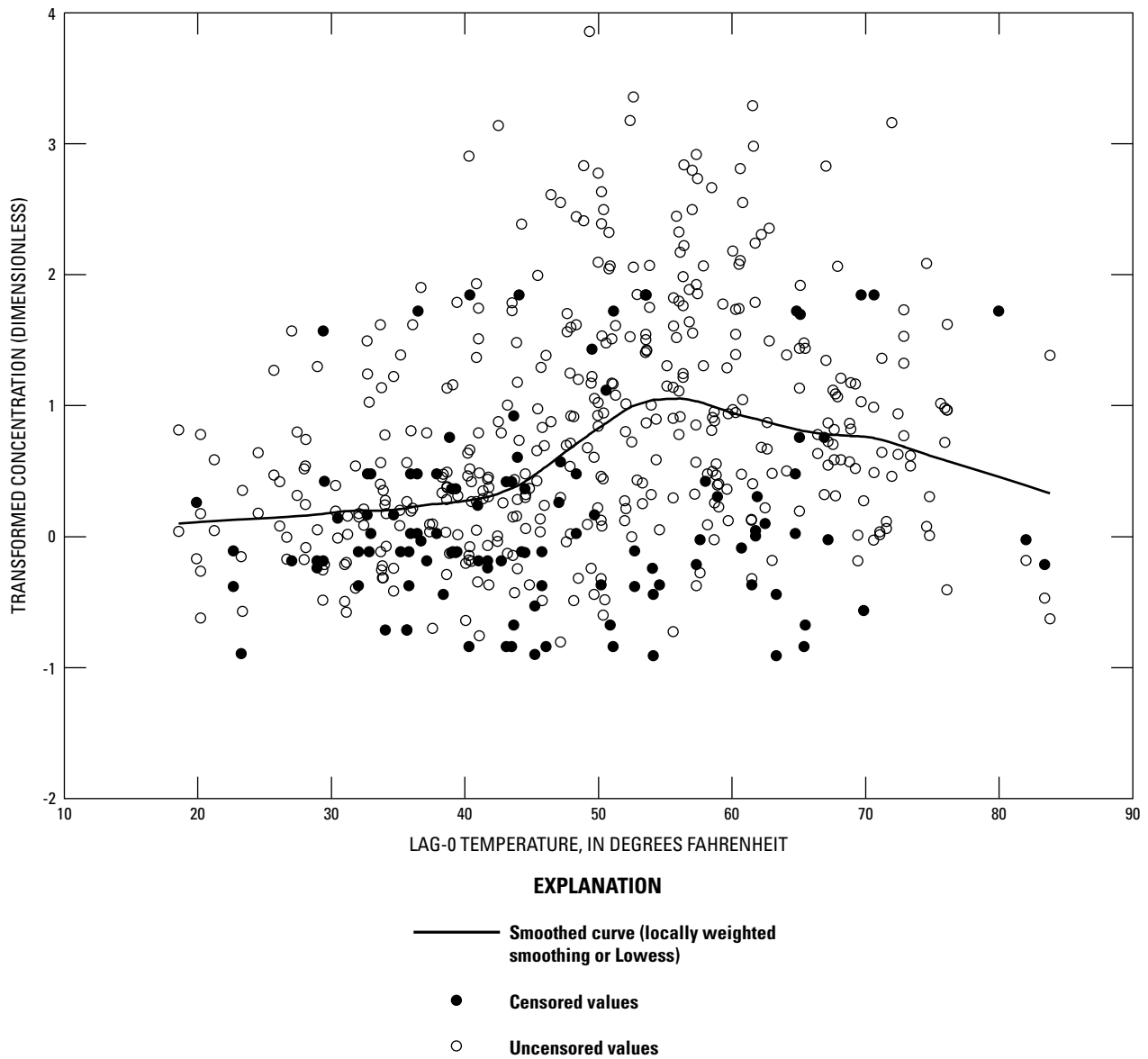
Figure 7.   Transformed atrazine and metolachlor concentrations and lag-0 temperature.
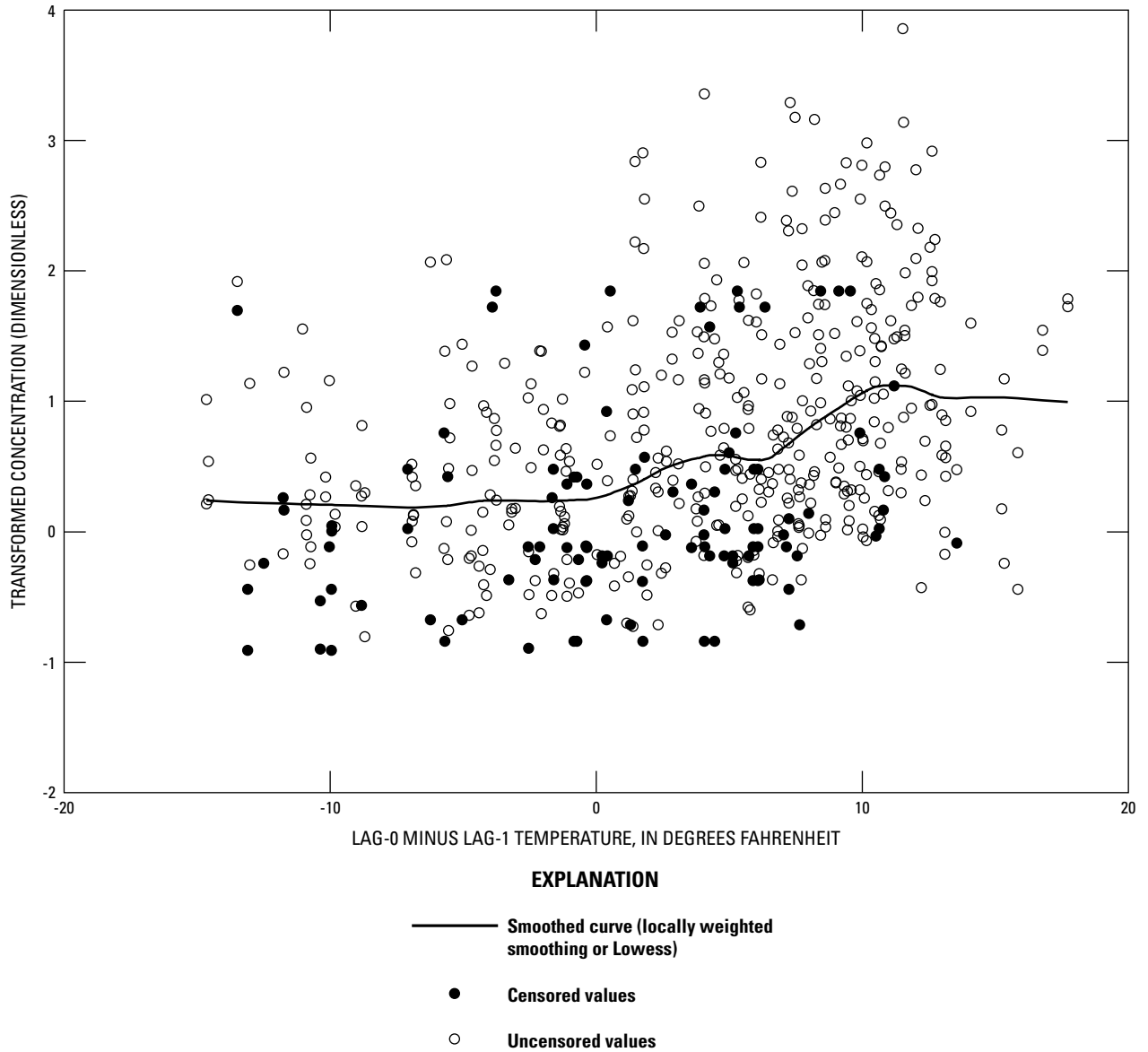
Figure 8. Transformed atrazine and metolachlor concentrations and lag-0 minus lag-1 temperature.

EXPLANATION

Concentration for temperature and precipitation bin

Concentration data not available for temperature and precipitation bin
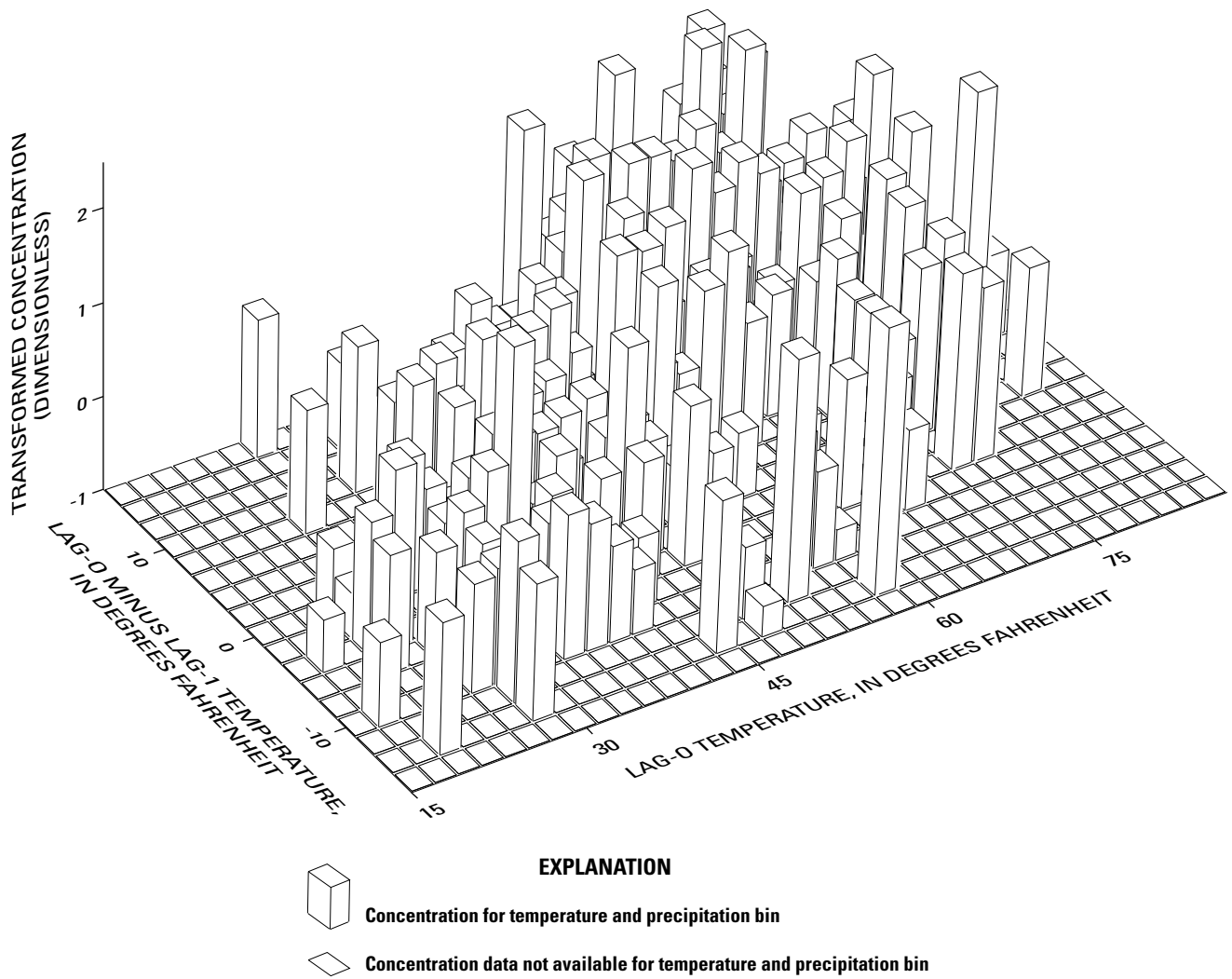
Figure 9. Relation between mean transformed atrazine and metolachlor concentrations and temperature variables.
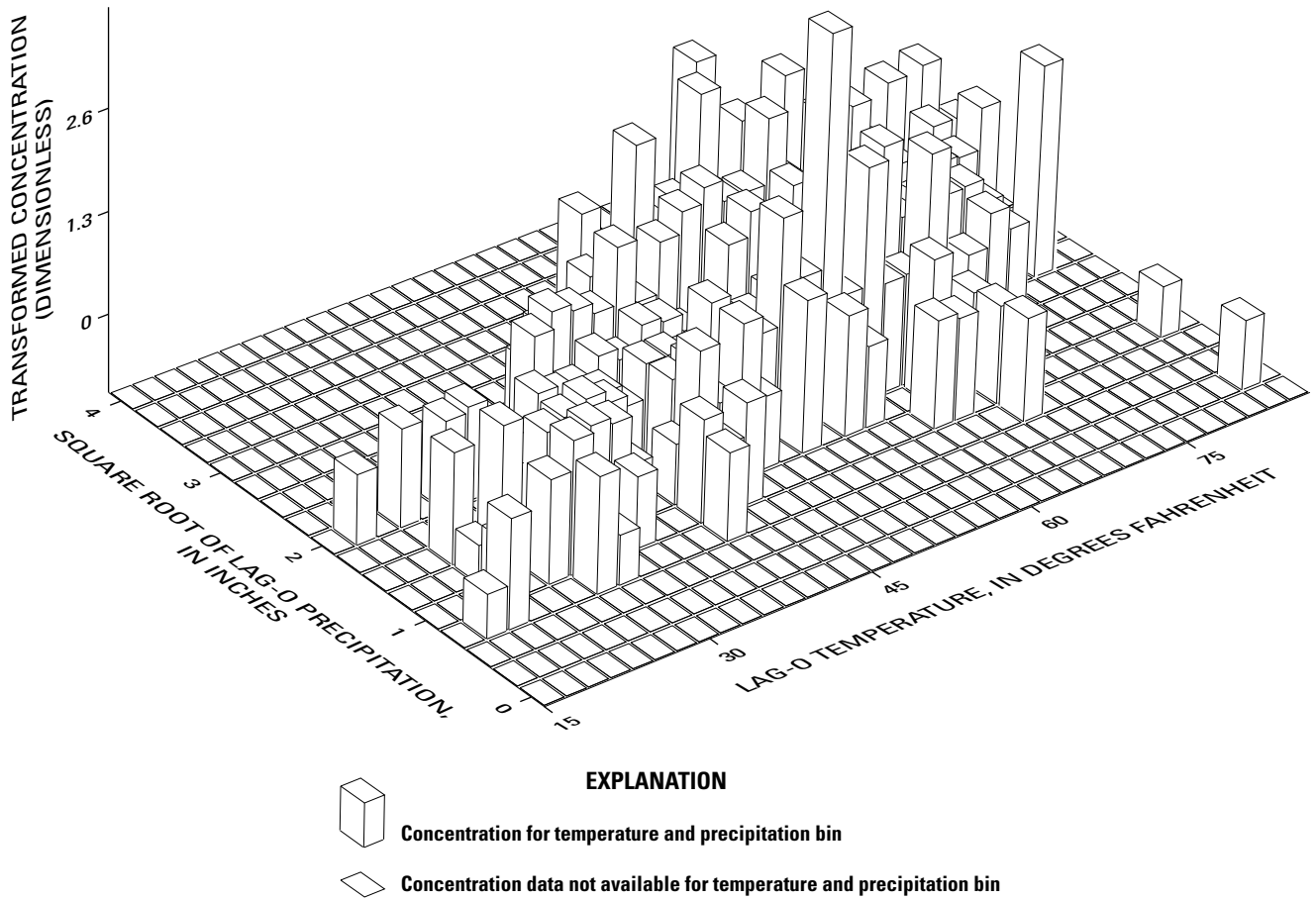
Figure 10. Relation between mean transformed atrazine and metolachlor concentrations and temperature and precipitation variables.
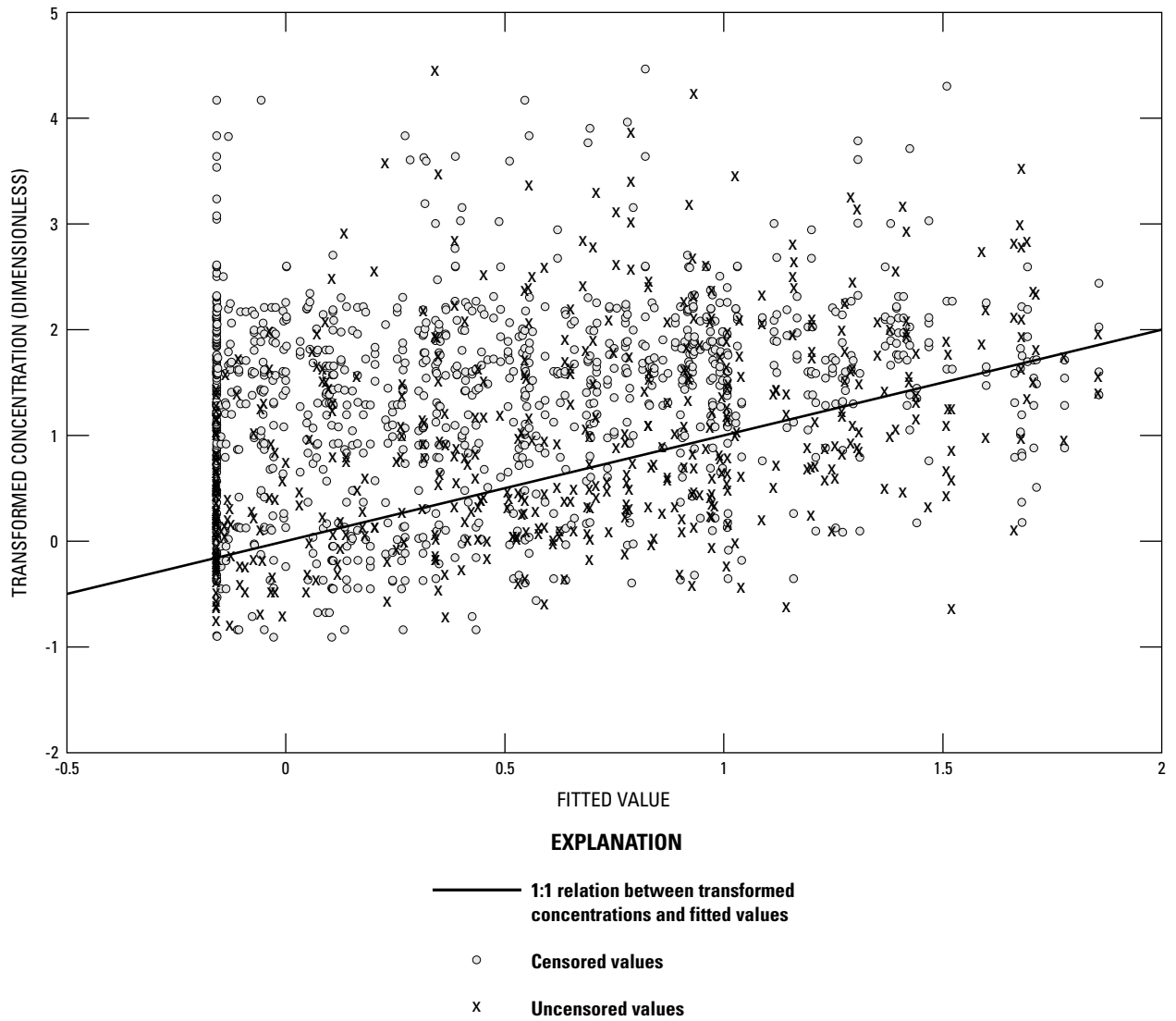
Figure 11.   Transformed concentrations for all pesticides and fitted values from Tobit regression model.
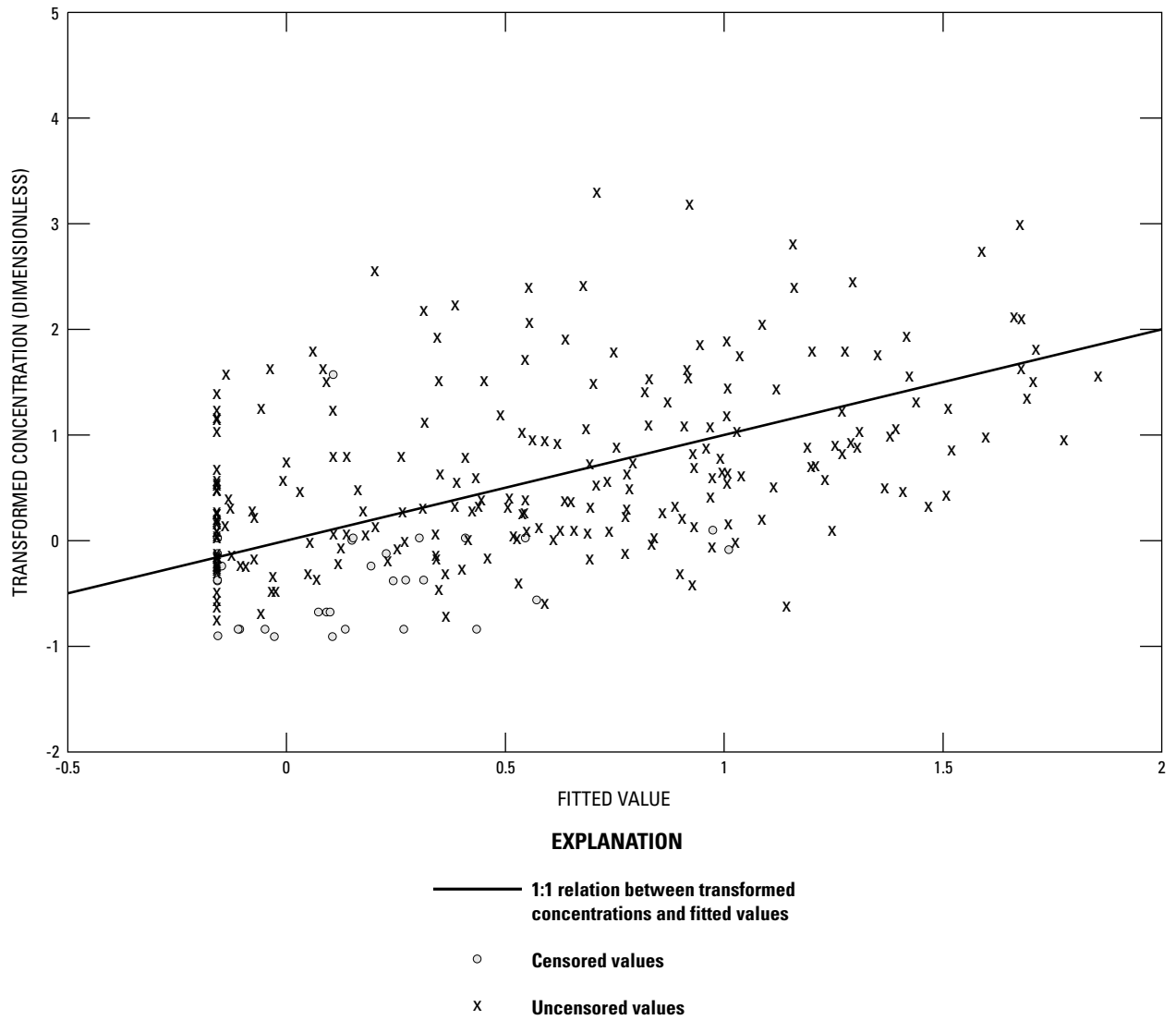
Figure 12.   Transformed atrazine concentrations and fitted values from Tobit regression model.
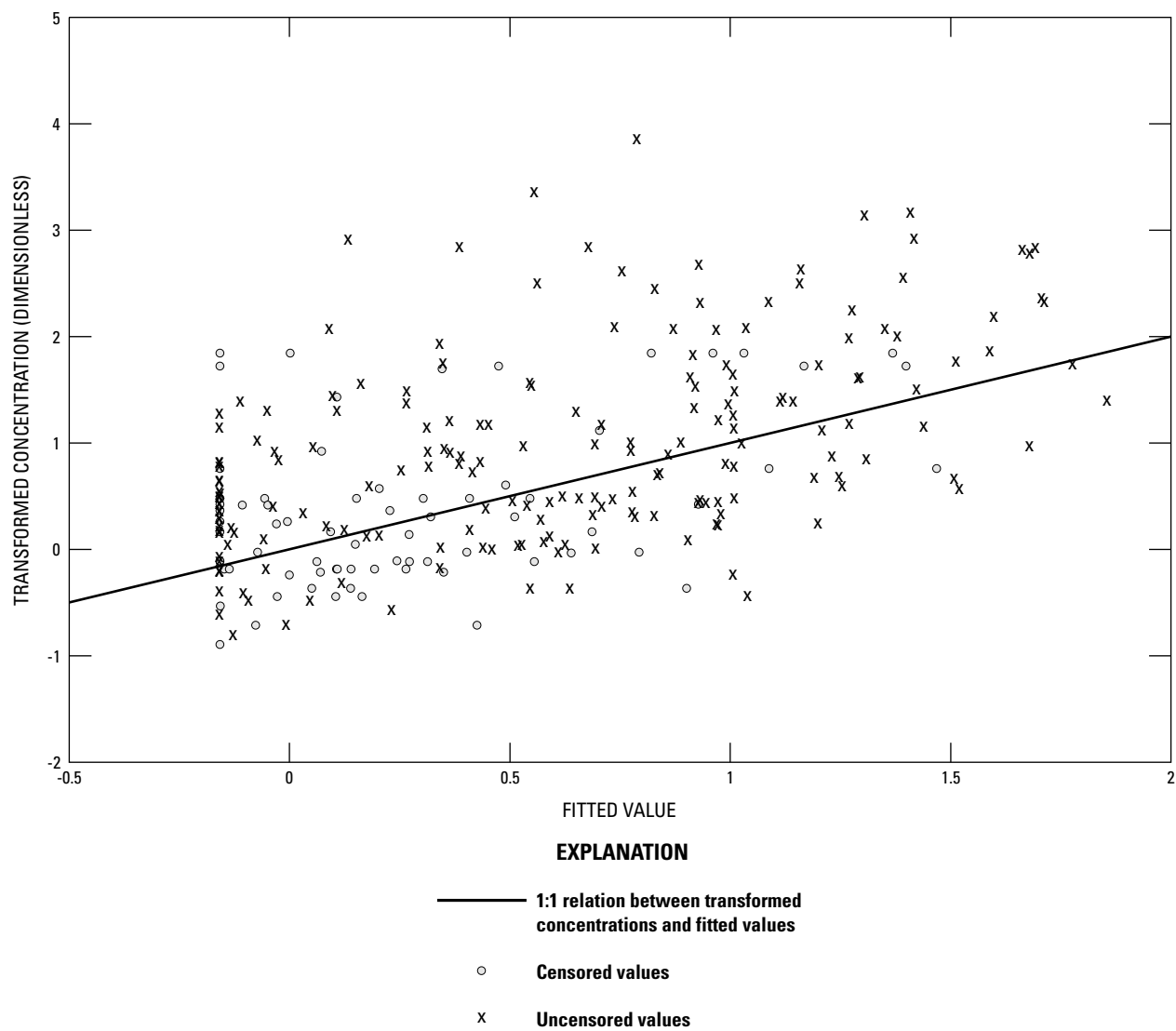
Figure 13. Transformed metolachlor concentrations and fitted values from Tobit regression model.

Figure 14. Transformed fonofos concentrations and fitted values from Tobit regression model.

Figure 15.   Transformed trifluralin concentrations and fitted values from Tobit regression model.

Serial Correlation Structure of Transformed Concentration Data

As indicated in the previous section, a substantial part of the variability in transformed pesticide concentrations can be explained using monthly precipitation and temperature data. However, the unexplained variability in the residuals from the Tobit regression model probably results from other factors, including spatial and temporal variability of climatic data within 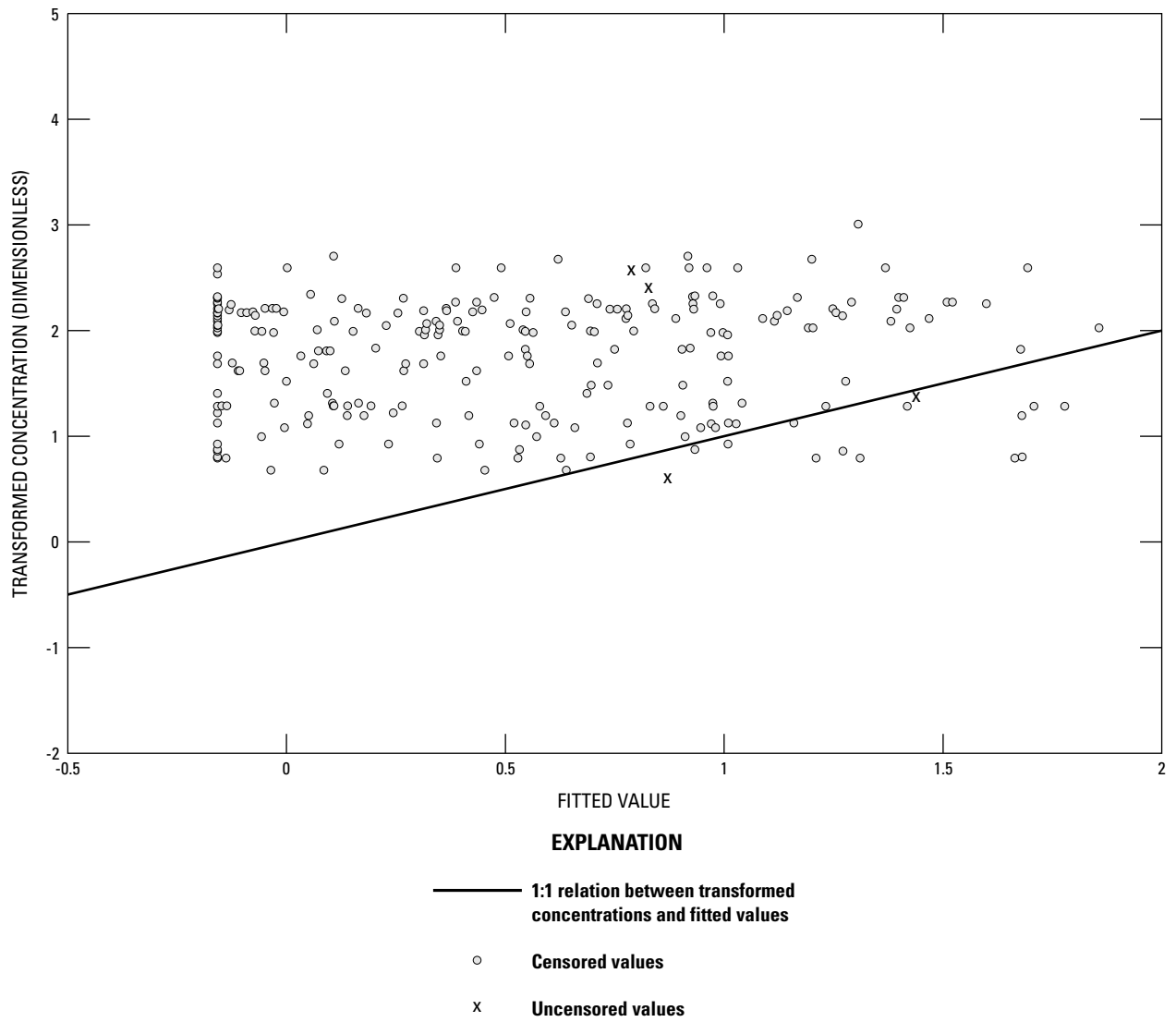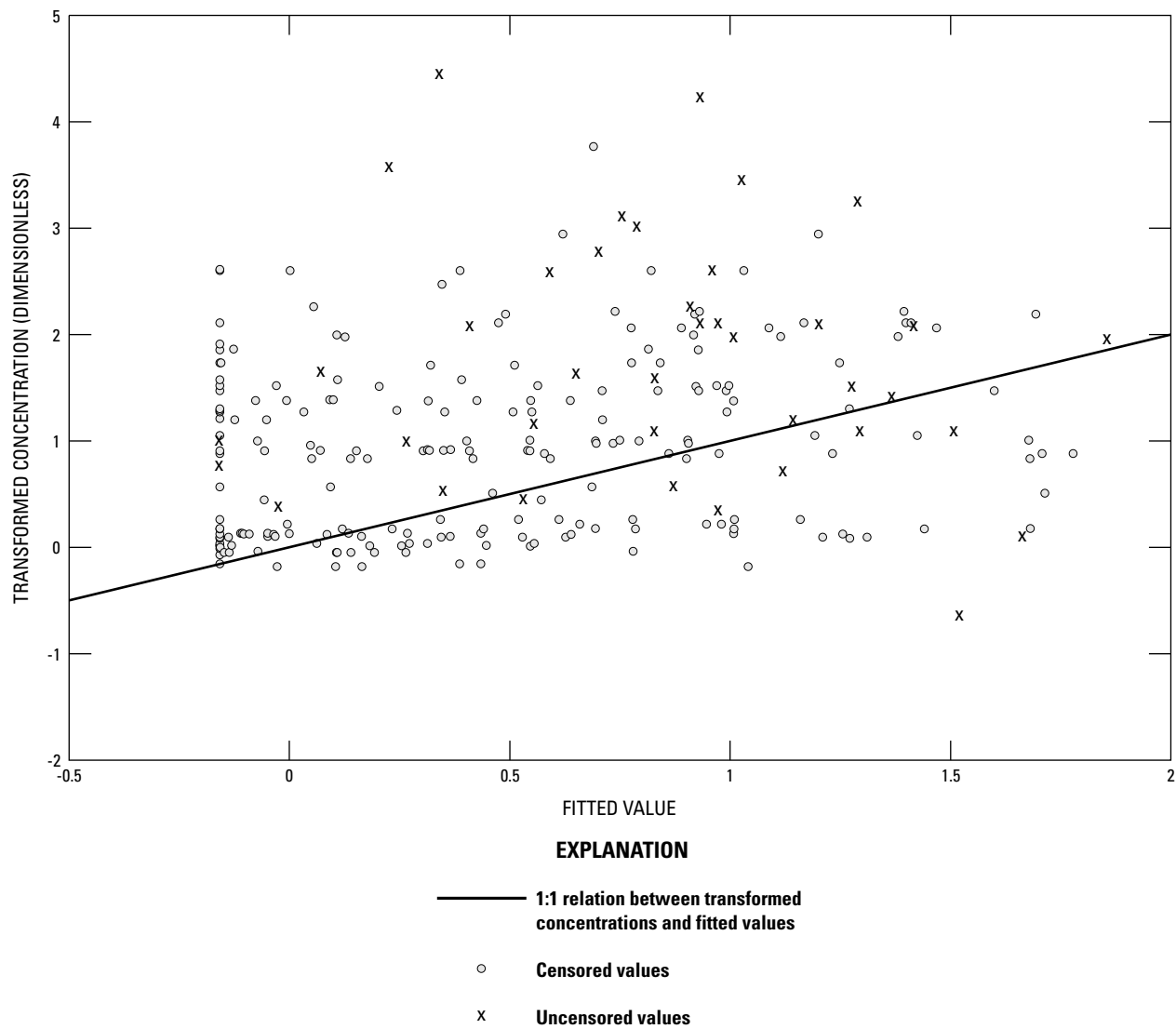watersheds and for time scales smaller than monthly, spatial and temporal variability of pesticide application rates, inability of the Tobit regression model to account for complex source and transport mechanisms, and differences between actual (instream) concentrations and measured concentrations in water samples. A detailed analysis of model errors was beyond the scope of this report. However, a statistical analysis of the errors was used to partition the unexplained variability into two parts: "signal" and "noise." The signal is the systematic variability in concentration that is related consistently from one day to the next, and the noise is the seemingly random variability that is not related from one day to the next.

The residuals from the Tobit regression model for a given time (t) can be expressed as follows:

$$X(t) = Y(t) + N(t) \qquad (4)$$

where

$\qquad X(t) \qquad$ is the model residual for time t,
$\qquad Y(t) \qquad$ is the signal, and
$\qquad N(t) \qquad$ is the noise.

The signal was assumed to be related from one day to the next through its serial correlation function,

$$\rho(k) = \left(\frac{1}{V_Y}\right) E[Y(t)Y(t-k)],$$

where

$\qquad \rho(k) \qquad$ is the serial correlation for lag k, in days;
$\qquad V_y = E[Y(t)^2] \qquad$ is the variance of $Y(t)$; and
$\qquad E[\,.\,] \qquad$ denotes the expected value.

The model was assumed to be unbiased ($E[Y(t)] = 0$), and $Y(t)$ was assumed to be stationary [that is, $Y(t)$ had a constant variance and serial correlation that depended only on the time lag]. The noise was assumed to be uncorrelated with $Y(t)$, to be uncorrelated with itself from one day to the next, and to have constant variance: $E[N(t)^2] = V_N$, $E[N(t)Y(t-k)] = 0$ for all k; and $E[N(t)N(t-k)] = 0$ for all k not equal to zero. Because concentration data were not available for every day of the year, variograms, rather than the serial correlation function, were used to evaluate measurement error, seasonal variability, and serial correlation of the historical data. Separate variograms were computed for atrazine and metolachlor, but the variograms for the remaining pesticides were combined into a single variogram to obtain a sufficient number of pairs of uncensored historical concentrations. Given equation 4 and the associated assumptions, the variogram (referred to as the semi-variogram in some publications) of $X(t)$ can be expressed as follows (Cressie, 1991):

$$\gamma(k) = \left(\frac{1}{2}\right) E[\{X(t) - X(t-k)\}^2] = V_N + V_Y - [1 - \rho(k)]. \qquad (5)$$

The variogram (eq. 5) was estimated using the residuals from the Tobit regression model (eq. 3) according to methods established by Cressie (1991). Whereas the Tobit regression model was fitted using a subsample of the concentration data as described previously, the variogram was estimated using residuals for all of the nonmissing concentration data to increase the number of pairs of observations for small time lags. Because methods were not readily available to fit the variograms with censored historical concentrations, only uncensored historical concentrations were used in the estimations.

The estimated variograms for atrazine and metolachlor and the variogram for the remaining pesticides are shown in figure 16. The points in the figure show estimates of $\gamma(k)$ for lags of 2.5, 7.5, …, and 97.5 days. The graph of these points in relation to the time lag is called the empirical variogram. The empirical variogram ordinate for a particular lag (k) was estimated using the squared differences between residuals separated by an interval of $k - 2.5$ to $k + 2.5$ days. The variograms do not approach zero as the time lag becomes small, thus confirming the presence of noise in the data (as indicated by the positive noise variance in equation 5). The discontinuity of the variogram as the time lag approaches zero is called the nugget effect. The nonzero nugget probably results, at least in part, from measurement errors (the differences between the actual concentrations and the laboratory concentrations). For positive time lags, the empirical variograms generally increase as the time lag increases, thus confirming the presence of serial correlation in the signal, where the serial correlation decreases as the lag increases (eq. 5). The serial correlation in the signal was approximated using the exponential correlation function, $\rho(k) = r^k$, where r is the lag-1 serial correlation. The following variogram, called an exponential variogram with nugget, was obtained by substituting r^k for $\rho(k)$ in equation 5:

$$\gamma(k) = V_N + V_Y[1 - r^k].$$

The variogram shown by equation 6 depends on three parameters—the nugget (V_N), the variance of the signal (V_Y), and the lag-1 serial correlation (r). These parameters were estimated by minimizing the sum of the squared errors between the fitted variograms and the empirical variograms. The fitted variograms shown in figure 16 and given in table 3 were in close agreement with the empirical variograms. The fitted variograms given in table 3 were used to generate the long-term traces of daily pesticide concentrations. The cor-

relation time scales given in table 3 indicate the signals were strongly correlated even for substantial time lags. For example, the correlation time scale for metolachlor was 66 days, which indicates that a time lag of 66 days was required for the correlation function to decay to $e^{-1} = 0.37$.

Simulation of Daily Pesticide Concentrations

The Tobit regression model (eq. 3), the residual variograms (table 3), and the transformation equations (table 2) were combined to develop a time-series model to simulate daily pesticide concentrations that reproduced the probability distributions, seasonal variability, and serial correlation characteristics in data from the monitoring stations. The procedure used to simulate the concentrations is described in this section, and a comparison to historical concentrations is given for selected stations. The potential strengths and limitations of the model for the assessment of long-term pesticide-exposure risks also is discussed.

The following procedure was used to simulate the daily concentrations for a given pesticide, station, and representative climatic period (designated by the historical months used to select the precipitation and temperature data):

1. Select the representative climatic period by designating the beginning ($m1$) and ending ($m2$) months of the simulation period. Use climatic data from the United States Historical Climatology Network (Karl and others, 1990) to compute estimated monthly precipitation, $P(m)$, and monthly temperature, $T(m)$, for the centroid of the watershed for each month, m, with $m = m1 - 1, m1, ..., m2$. For this report, estimated values were computed using a robust locally weighted regression model (Cleveland and Devlin, 1988).

2. For each day of the simulation period ($t = 1, 2, ..., T$), generate a correlated time series of random variables, $\{Y(t)\}$, as follows:

$$Y(1) = (V_Y)^{\frac{1}{2}} Z(1)$$

$$Y(t) = rY(t-1) + (1 - r^2)^{\frac{1}{2}} (V_Y)^{\frac{1}{2}} Z(t), t = 2, ..., T$$

where V_Y and r are given in table 3 and $\{Z(1), ..., Z(T)\}$ is a sequence of independent, standard normal random variables.

3. For each day of the simulation period, generate the simulated value of the transformed concentration using equation 3 with Y replacing X. Let the resulting time series of transformed concentrations be denoted by $U(t)$ to distinguish from $C^*(t)$, which includes the noise $X(t) = Y(t) + N(t)$. The noise was excluded when generating the transformed concentrations because it was expected to consist primarily of measurement error and the objective of the model was to reproduce the statistical characteristics of actual concentrations rather than measured concentrations.

4. Reverse the normality transformation (see table 2, alternative transformation 2) to obtain a time series of simulated concentrations

$$C(t) = \exp\left\{\alpha + \beta[U(t) - c] + \gamma\left\{[U(t) - c]^2 + \left[\frac{1}{6}\right][U(t) - c]^3\right\}I[U(t) < c]\right\}$$

where the transformation parameters depend on the pesticide and station being considered and are determined from the percentiles estimated from the WARP model as described previously.

Table 3. Fitted variograms for residuals from Tobit regression model for transformed concentrations.

[The correlation time scale (τ) satisfies $r = e^{\frac{1}{\tau}}$; γ is the variogram; k is the time lag, in days; V_N is the nugget; V_Y is the variance of the signal; r is the lag-1 serial correlation]

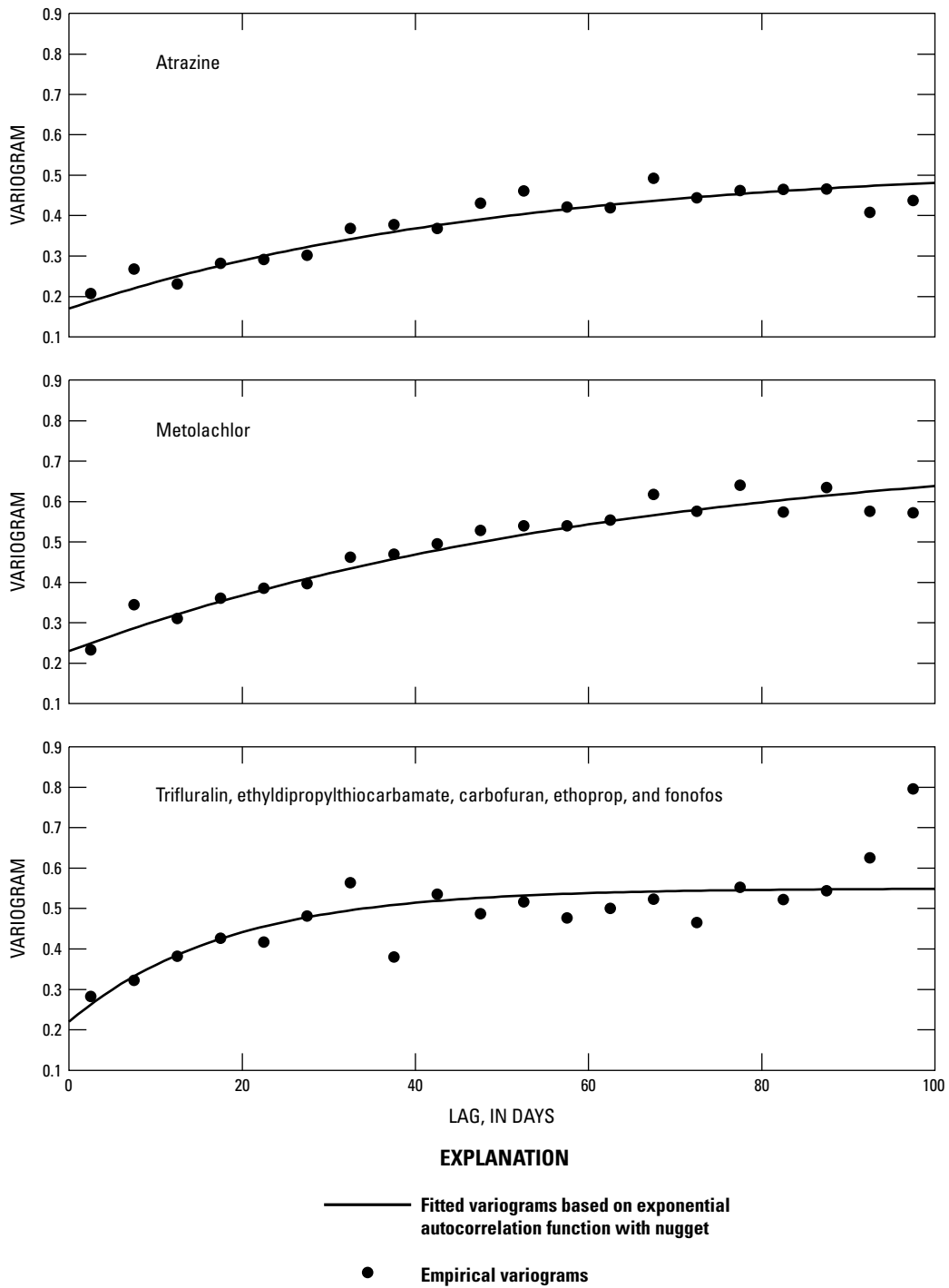| Pesticide	Fitted variogram $\{\gamma(k) = V_N + V_Y[1 - r^k]\}$	Correlation time scale (days)
Atrazine	$V_N = 0.17; V_Y = 0.37; r = 0.981$	52
Metolachlor	$V_N = 0.23; V_Y = 0.52; r = 0.985$	66
Trifluralin, ethyldipropylthiocarbamate, carbofuran, ethoprop, and fonofos	$V_N = 0.22; V_Y = 0.35; r = 0.946$	18

Figure 16. Fitted variograms for residuals from Tobit regression model.

The previous four-step procedure was used to generate long-term traces of daily concentrations for each pesticide and each monitoring station. The representative climatic period for the simulations was October 1, 1990, through September 30, 2001. For each pesticide and station, the procedure was replicated 100 times [with a different generated time series for $Y(t)$ each time]. Thus, 100 potential realizations (traces) of daily concentrations were generated for comparison with the historical data. The following examples illustrate some general observations regarding the ability of the model to reproduce seasonal variability in the concentrations for the various pesticides. An effort was made to choose representative stations for the examples rather than biasing the discussion toward stations for which the model worked particularly well. The examples emphasize stations that are located in areas that have high agricultural land use. Those stations typically corresponded with stations that are located in areas in which the various pesticides are used extensively. The traces for atrazine were similar to the traces for metolachlor and, thus, were not included in the examples. The traces for ethoprop were not included because the simulated concentrations for that pesticide generally were small (less than 0.01 microgram per liter) and almost all of the historical concentrations were censored.

The first few examples are typical of stations located in the northeastern quadrant of the United States where most of the monitoring stations are located (fig. 1). The representative stations for this quadrant are East Mahantango Creek at Klingerstown, Pennsylvania (map number 40; fig. 1; drainage area 115 square kilometers; about 55 percent agricultural land use), Sugar Creek at New Palestine, Indiana (map number 99; 246 square kilometers; about 92 percent agricultural land use), and Iowa River at Wapello, Iowa (map number 22; 32,364 square kilometers; about 84 percent agricultural land use).

The traces for the East Mahantango Creek station are shown in figures 17 through 21. The top graph in each figure shows a single trace from the time-series model, and the bottom graph shows the 5th, 50th, and 95th percentiles for each day of the year. The percentiles were computed from 100 traces or 1,000 values (100 traces times 10 years per trace). The model reasonably reproduced the seasonal and interannual variability of the metolachlor concentrations for this station (fig. 17) but may have slightly overestimated historical concentrations during March and April. The model also reasonably reproduced the timing and magnitude of the high trifluralin, EPTC, carbofuran, and fonofos concentrations for this station (figs. 18 through 21). The simulated concentrations for those pesticides have less serial correlation and, thus, generally fluctuate more rapidly from day to day than those for metolachlor. The simulated trifluralin concentrations (fig. 18) were less variable than the other pesticide concentrations, and the censored historical concentrations occurred more frequently than indicated by the simulations. The simulated EPTC (fig. 19), carbofuran (fig. 20), and fonofos (fig. 21) concentrations generally were near or below the censored historical concentrations from September through March and increased from April through July. The timing and magni-

tude of the occasionally uncensored historical concentrations seem to support the model although the censored historical concentrations occurred more frequently than indicated by the simulations.

The traces for the Sugar Creek station (figs. 22 through 26) were similar to those for the East Mahantango Creek station. However, for metolachlor (fig. 22), the simulated concentrations generally were slightly higher than the historical concentrations from October through March and slightly lower than the historical concentrations during May and June. The simulated trifluralin concentrations (fig. 23), as for the East Mahantango Creek station (fig. 18), generally exceeded the censored historical concentrations. However, the timing and magnitude of the uncensored historical concentrations seem to support the model. The timing and magnitude of the high uncensored historical EPTC concentrations (fig. 24) seem to support the model although the censored historical concentrations occurred more frequently than indicated by the simulations. The simulated carbofuran concentrations (fig. 25) were similar to those for the East Mahantango Creek station (fig. 20). The censored historical fonofos concentrations (fig. 26), as for the East Mahantango Creek station (fig. 21), occurred more frequently than indicated by the simulations. The traces for other stations for which a large number of fonofos concentrations were available indicated the model generally did not overestimate historical fonofos concentrations. For example, the simulated fonofos concentrations for the White River at Hazelton, Indiana, station (map number 97) (fig. 27) agree closely with the historical concentrations for that station.

The traces for the Iowa River station are shown in figures 28 through 30 (the traces for trifluralin and carbofuran were similar to those for the East Mahantango Creek and Sugar Creek stations and, thus, are not shown). The model reasonably reproduced the seasonal and interannual variability of the metolachlor, EPTC, and fonofos concentrations for the Iowa River station.

The traces for the remaining stations in the northeastern quadrant generally supported the results of the previous examples and indicated the model should be useful as a screening tool to assess long-term pesticide-exposure risks for streams in the northeastern United States. Except for trifluralin concentrations, the model generally did not overestimate or underestimate concentrations, and the timing and magnitude of the high and low historical concentrations seem to support the model. The good agreement between the simulated concentrations and the historical concentrations was expected because most of the historical concentrations used to fit the model were for streams that are located in this quadrant.

For stations that are not located in the northeastern quadrant (fig. 1), the adequacy of the model was difficult to determine. For certain stations and/or pesticides, the agreement between the simulated concentrations and the historical concentrations was good. However, in many cases, too few data were available to determine the adequacy of the model and, in some cases, the agreement between the simulated concentra-

tions and the historical concentrations was poor. The following examples illustrate model results for some of the remaining stations, but further data collection and model development may be necessary to determine whether the model should be used for areas for which few historical data are available.

The traces for the Bogue Phalia near Leland, Mississippi, station (map number 47) are shown in figures 31 through 33 for metolachlor, trifluralin, and fonofos, respectively. For this station, the model generally underestimated the historical metolachlor concentrations, especially from April through June (fig. 31). The model closely reproduced the historical trifluralin concentrations during the early spring but generally overestimated the historical concentrations during the rest of the year (fig. 32). Most of the simulated and all of the historical fonofos concentrations were below detection limits (fig. 33), but, because of a lack of uncensored historical concentrations, a comparison could not be made between those concentrations and the simulated concentrations.

The simulated metolachlor concentrations for the Withlacoochee River near Quitman, Georgia, station (map number 24) (fig. 34), in contrast to those for the Bogue Phalia station, generally were higher than the historical concentrations during most of the year. The simulated trifluralin concentrations (fig. 35) generally were higher than the historical concentrations, but the timing and magnitude of the few uncensored concentrations seem to support the model. The simulated fonofos concentrations (fig. 36) generally were higher than the historical concentrations.

The simulated EPTC concentrations for the Palouse River at Hooper, Washington, station (map number 12) (fig. 37) were in good agreement with the historical concentrations. The simulated fonofos concentrations (fig. 38) generally were less than the detection limit as were the historical concentrations. The simulated fonofos concentrations for the Rock Creek at Twin Falls, Idaho, station (map number 89) (fig. 39) were less than the detection limit, with occasional values above the limit during March through August. Although only one uncensored historical concentration occurred, the concentrations on days when samples were not collected also may have been high.

The simulated metolachlor concentrations for the San Joaquin River near Vernalis, California, station (map number 66) (fig. 40) had little, if any, seasonal variability. Although the simulated concentrations were in general agreement with the historical concentrations during some months, the simulated concentrations were less than the historical concentrations from May through August. Apparently, the climatic variables used in the model to reproduce seasonal variability did not correlate well with the growing season for this station. The simulated EPTC concentrations (fig. 41) were in poor agreement with the historical concentrations. However, the simulated trifluralin and fonofos concentrations (figs. 42 and 43) were in close agreement with the historical concentrations. The historical concentrations did not increase from May through August, in contrast to those for metolachlor and EPTC, and the timing and magnitude of the high historical concentrations supported the model.

Implications and Limitations of Model

The time-series model developed in this report to simulate daily pesticide concentrations is intended for use as a preliminary screening tool to assess long-term pesticide-exposure risks for streams in the coterminous United States. The primary strength of the model is its relative simplicity. In contrast to physically based, chemical source and transport models, the time-series model developed in this report does not require high-resolution rainfall, runoff, and pesticide source and transport data to simulate the daily pesticide concentrations. Rather, to apply the model to a particular watershed or pesticide, only readily available data, such as the watershed location (as specified by the latitude and longitude for the centroid of the watershed), monthly precipitation and temperature data from the United States Historical Climatology Network (Karl and others, 1990), and watershed averages of readily available watershed characteristics (as required by the WARP model), are needed. The most noticeable variable not included in the model is daily streamflow. Although daily streamflow data were available for most of the monitoring stations used in the study, those data were not included so the model could be applied easily to ungaged watersheds for which daily streamflow would need to be estimated. The ability of the time-series model to simulate pesticide concentrations probably would be enhanced if daily streamflow was included as a predictor variable.

Because of the relative simplicity of the time-series model and because of the inherent noise and unpredictability of pesticide concentrations, many limitations of the model need to be considered before the model can be used to assess long-term pesticide-exposure risks. Also, because the model uses watershed characteristics and percentiles estimated from the WARP model for each particular watershed and pesticide, the limitations of the WARP model, as described by Larson and others (2004) and C.G. Crawford, S.J. Larson, and R.J. Gilliom (written commun., 2005), need to be considered. Although the WARP model was developed to provide unbiased estimates of the percentiles of the concentrations when averaged for many locations that have similar watershed characteristics and pesticide application rates, the model may contain local bias for certain regions, especially those for which few historical data are available.

Another limitation of the time-series model is that the regression equation used to relate transformed concentrations to climatic variables (monthly precipitation and temperature) did not account for differences among the geographic locations of the monitoring stations or differences among watershed characteristics. However, even with this limitation, the model reasonably reproduced the spatial differences in the timing and magnitude of the pesticide concentrations. An evaluation of the relations between residuals from the regression model, watershed characteristics, and station locations indicated the regression equation could not be improved substantially given the available data. However, as more data

become available, significant relations (from both a practical and statistical standpoint) could be exploited to improve the ability of the model to simulate daily pesticide concentrations.

The regression model used to relate transformed concentrations to the climatic variables was assumed to be the same for all of the pesticides for which concentrations were simulated. Although the herbicides and insecticides were not necessarily expected to respond to the climatic variables in the same way because of the basic differences in chemical properties, application methods, and application amounts, attempts to fit a separate model for the insecticides were not successful because of the small percentage (generally less than 2 percent) of uncensored, in relation to censored, historical concentrations. Because of this small percentage, the maximum likelihood function for the Tobit regression model fitted to insecticide data alone was not well defined and the parameters could not be determined. The fitted model for herbicides was essentially the same as the regression model for herbicides and insecticides combined. The similarity of the two models may indicate the insecticides react in a manner consistent with the herbicides or, more likely, that the herbicide data dominate the likelihood function. A separate model for insecticides could be developed in the future as laboratory methods used to detect low insecticide concentrations are improved.

The time-series model developed in this report, despite its limitations, should be valuable for the simulation of daily pesticide concentrations needed for a preliminary assessment of long-term pesticide-exposure risks for unmonitored streams, particularly in the northeastern quadrant of the United States. Because of the statistically rigorous representation of seasonal variability and serial correlation of pesticide concentrations provided by the model and the probability distributions provided by the WARP model, the simulated concentrations are a good alternative to concentrations generated using a deterministic model with errors in the input data.

Summary

To simulate daily pesticide concentrations that reproduce the statistical characteristics of historical pesticide concentrations for a national network of monitoring stations, the U.S. Geological Survey, in cooperation with the U.S. Environmental Protection Agency, developed a time-series model that was based on readily available information on pesticide use, climatic variability, and watershed characteristics. The model was used to simulate concentrations for four herbicides [atrazine, ethyldipropylthiocarbamate (EPTC), metolachlor, and trifluralin] and three insecticides (carbofuran, ethoprop, and fonofos) that represent a range of physical and chemical properties, application methods, national application amounts, and areas of use in the United States.

Concentration data for 112 pesticide-monitoring stations in the coterminous United States were used to develop the time-series model. The stations used are the same as those used to develop the Watershed Regressions for Pesticides (WARP) model for atrazine and represent a wide variety of environmental settings and watershed characteristics. The concentration data were collected from October 1, 1990, through September 30, 2001 (water years 1991-2001) as part of the U.S. Geological Survey National Water Quality Assessment (NAWQA) and National Stream Quality Accounting Network (NASQAN) programs.

Monthly time series of precipitation data and of temperature data were developed for the watershed of each pesticide-monitoring station and used in the time-series model to reproduce seasonal and interannual variability in pesticide concentrations.

The first requirement for the time-series model was that the annual percentiles for the simulated concentrations agree closely with the percentiles estimated from the WARP model. To satisfy this requirement, the assumption was made that a transformation existed such that the annual probability distribution of the transformed concentrations was a standard normal distribution. The parameters upon which the transformation function generally depends were determined from the percentiles estimated from the WARP model by minimizing the sum of the squared errors between the log-transformed estimated percentiles and the log-transformed normal percentiles. After raw concentration data were transformed, the transformed concentrations for all stations were combined and used to fit a model to describe seasonal variability in the transformed concentrations. The transformed concentrations were related to precipitation and temperature data from the United States Historical Climatology Network for various time lags to determine if statistically significant relations existed. Because of the large percentage of censored historical concentrations used to fit the model, Tobit regression (Tobin, 1958; Cohen, 1976), which uses maximum likelihood estimation (assuming normally distributed errors) to fit a regression model to data with multiple censoring levels, was used rather than ordinary least-squares regression. The terms that involve the climatic variables were highly significant and had p-values of less than 10^{-9}.

The goodness of fit of the Tobit regression model was not easily ascertained from a comparison of the fitted and historical values of the transformed concentrations for all pesticides because of the large percentage of censored values. However, the fitted values for atrazine and metolachlor generally were in good agreement with the transformed concentrations. Although a substantial amount of unexplained variability remained in the transformed concentrations for both atrazine and metolachlor, no significant bias or lack of fit occurred between the fitted values and the transformed concentrations. Because of the large percentage of censored historical concen-

trations for the remaining pesticides, the potential bias or lack of fit between the fitted values and the transformed concentrations was difficult to determine.

A substantial part of the variability in transformed pesticide concentrations can be explained using monthly precipitation and temperature data. However, the unexplained variability in the residuals from the Tobit regression model probably results from other factors, including spatial and temporal variability of climatic data within watersheds and for time scales smaller than monthly, spatial and temporal variability of pesticide application rates, inability of the Tobit regression model to account for complex source and transport mechanisms, and differences between actual (instream) concentrations and measured concentrations in water samples. A statistical analysis of the errors was used to partition the unexplained variability into two parts: "signal" and "noise."

Variograms were used to evaluate measurement error, seasonal variability, and serial correlation of the historical data. Separate variograms were computed for atrazine and metolachlor, but the variograms for the remaining pesticides were combined into a single variogram to obtain a sufficient number of pairs of uncensored historical concentrations.

The Tobit regression model, the residual variograms, and the transformation equations were combined to develop a time-series model to simulate daily pesticide concentrations that reproduced the probability distributions, seasonal variability, and serial correlation characteristics in data from the monitoring stations.

The representative climatic period for the simulations was October 1, 1990, through September 30, 2001. For each pesticide and station, 100 potential realizations (traces) of daily concentrations were generated for comparison with the historical data.

The traces indicated the model should be useful as a screening tool to assess long-term pesticide-exposure risks for streams in the northeastern United States. Except for trifluralin concentrations, the model generally did not overestimate or underestimate concentrations, and the timing and magnitude of the high and low historical concentrations seem to support the model.

For stations that are not located in the northeastern quadrant, the adequacy of the model was difficult to determine. For certain stations and/or pesticides, the agreement between the simulated concentrations and the historical concentrations was good. However, in many cases, too few data were available to determine the adequacy of the model and, in some cases, the agreement between the simulated concentrations and the historical concentrations was poor. Further data collection and model development may be necessary to determine whether the model should be used for areas for which few historical data are available.

Because of the relative simplicity of the time-series model and because of the inherent noise and unpredictability of pesticide concentrations, many limitations of the model need to be considered before the model can be used to assess long-term pesticide-exposure risks.

However, the model reasonably reproduced the spatial differences in the timing and magnitude of the pesticide concentrations. An evaluation of the relations between residuals from the regression model, watershed characteristics, and station locations indicated the regression equation could not be improved substantially given the available data. However, as more data become available, significant relations (from both a practical and statistical standpoint) could be exploited to improve the ability of the model to simulate daily pesticide concentrations.

References

Cleveland, W.S., and Devlin, S.J., 1988, Locally weighted regression—An approach to regression analysis by local fitting: Journal of the American Statistical Association, v. 83, p. 596-610.

Cohen, A.C., 1976, Progressively censored sampling in the three-parameter log-normal distribution: Technometrics, v. 18, p. 99-103.

Cressie, N.A.C., 1991, Statistics for spatial data: Wiley, New York, 900 p.

Insightful Corp., 2002, S-PLUS version 6.1.2, release 1, for Microsoft Windows: accessed January 25, 2006, at *www.insightful.com*

Karl, T.R., Williams, C.N., Quinlan, F.T., and Boden, T.A., 1990, United States Historical Climatology Network (HCN) serial temperature and precipitation data: Publication No. 3404, Environmental Sciences Division, Oak Ridge National Laboratory, 83 p.

Larson, S.J., and Gilliom, R.J., 2001, Regression models for estimating herbicide concentrations in U.S. streams from watershed characteristics: Journal of the American Water Resources Association, v. 37, no. 5, p. 1349-1368.

Larson, S.J., Crawford, C.G., and Gilliom, R.J., 2004, Development and application of watershed regressions for pesticides (WARP) for estimating atrazine concentration distributions in streams: U.S. Geological Survey Water-Resources Investigations Report 03-4047, 68 p.

Tobin, J., 1958, Estimation of relationships for limited dependent variables: Econometrica, v. 31, p. 24-36.
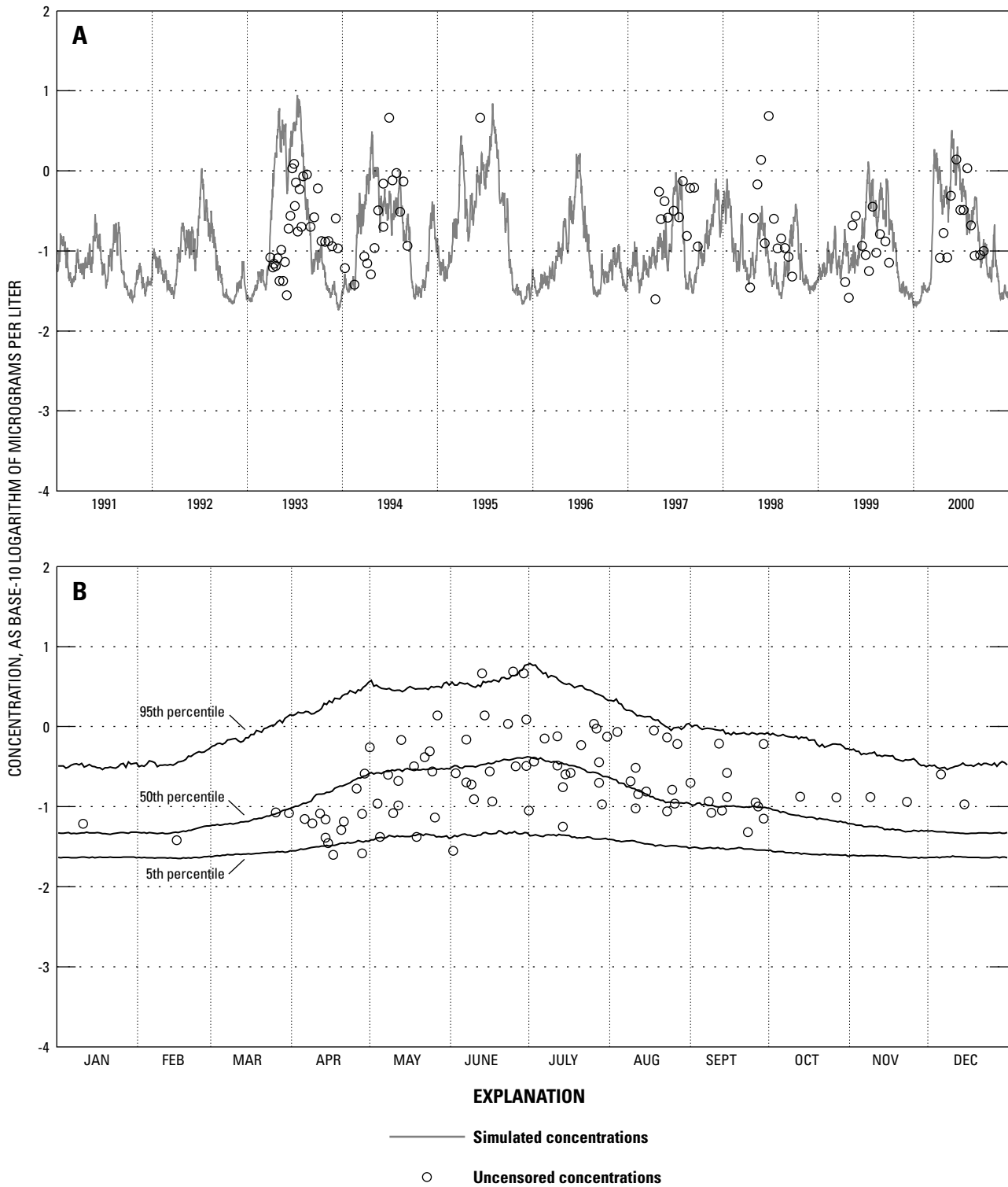
Figure 17.   Generated trace of daily metolachlor concentrations for 1991-2000 for the East Mahantango Creek at Klingerstown, Pennsylvania, station (map number 40) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
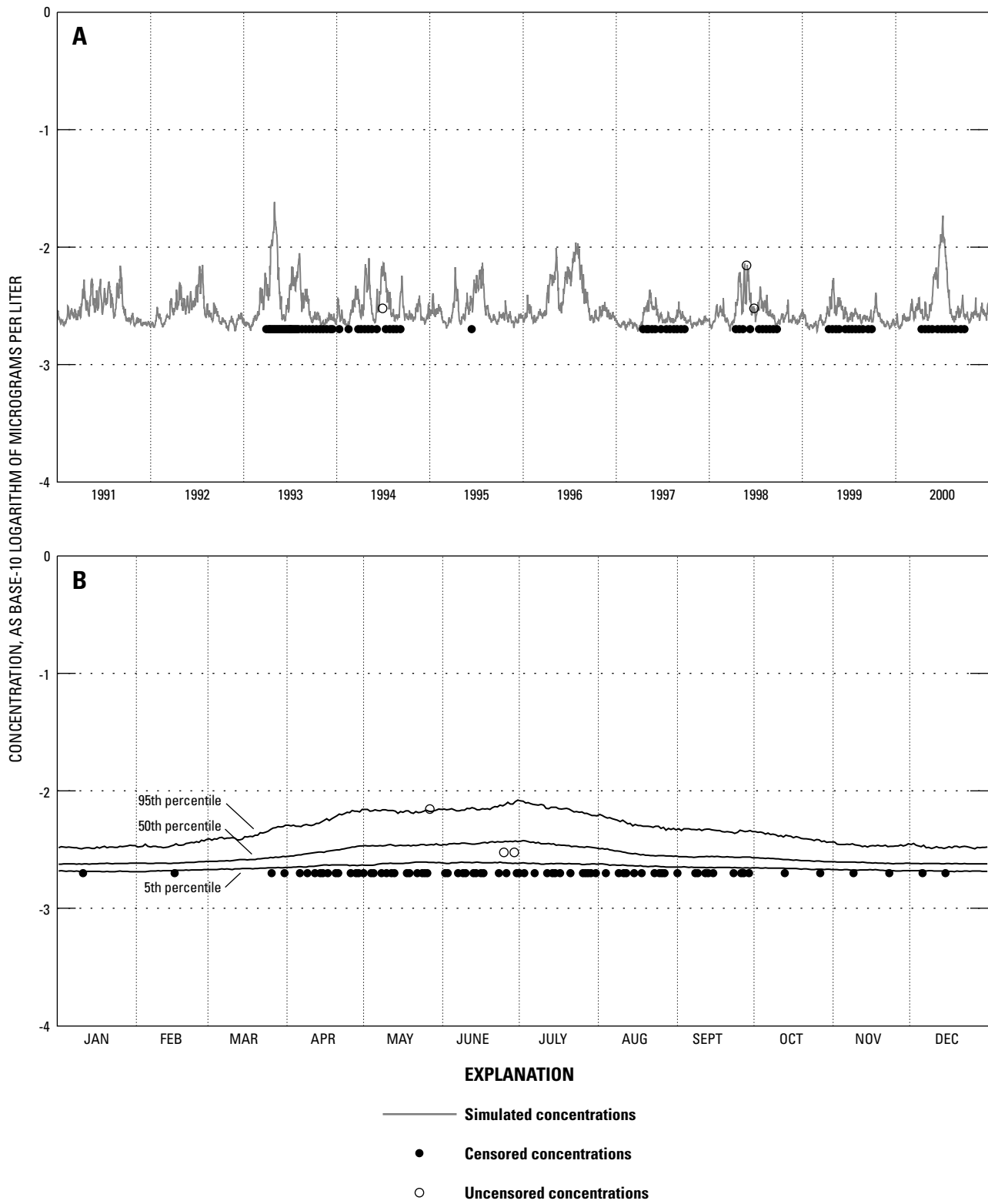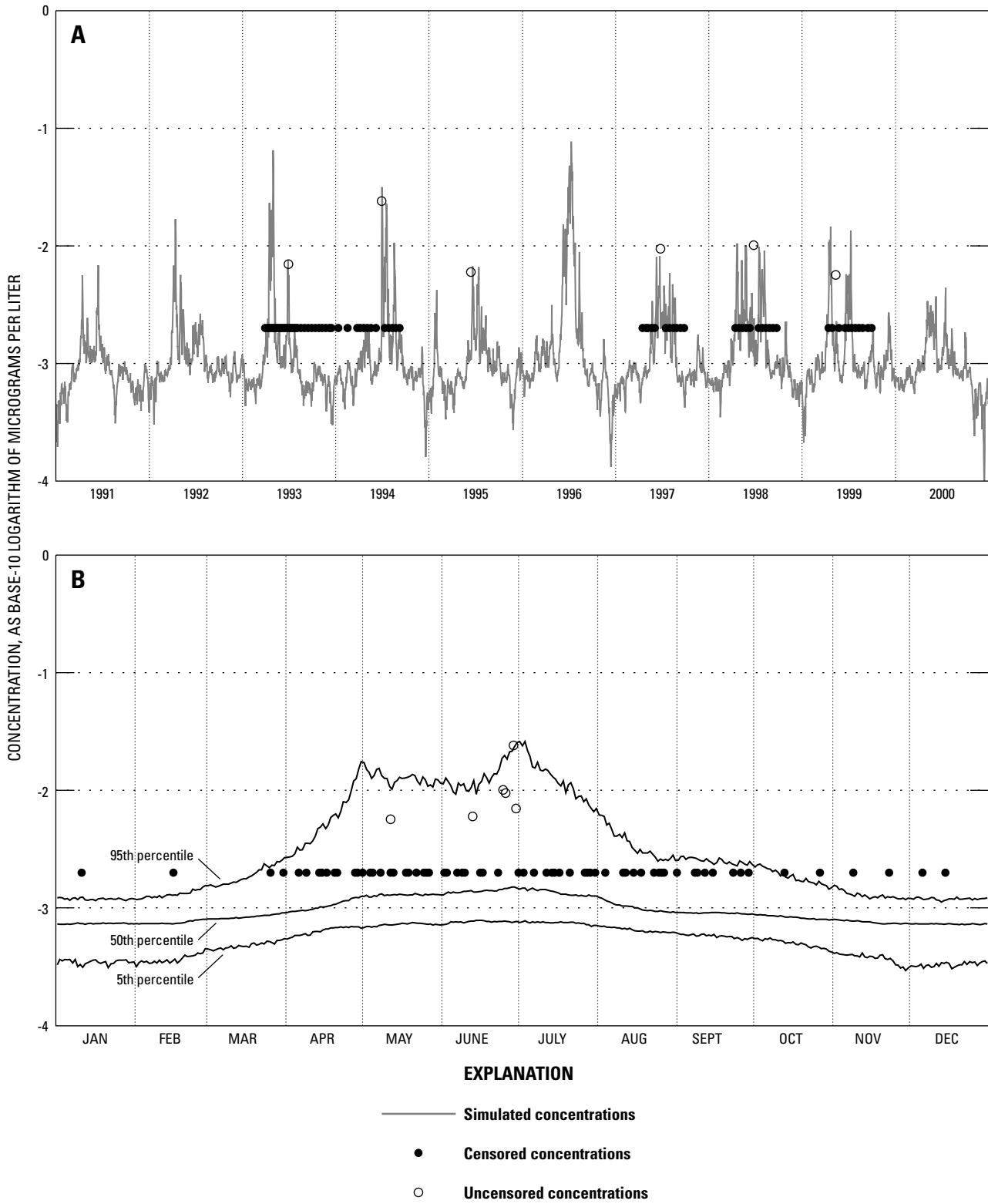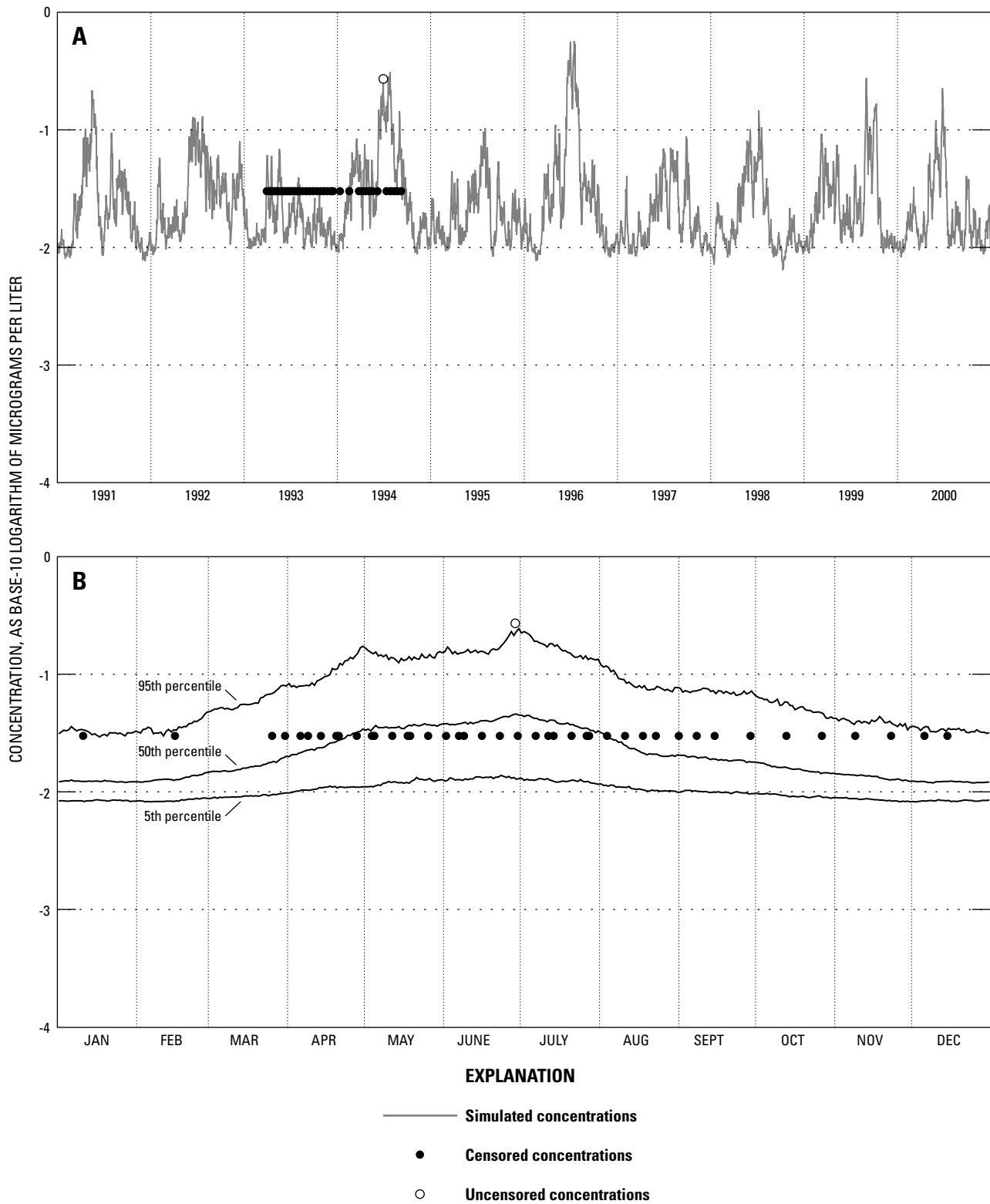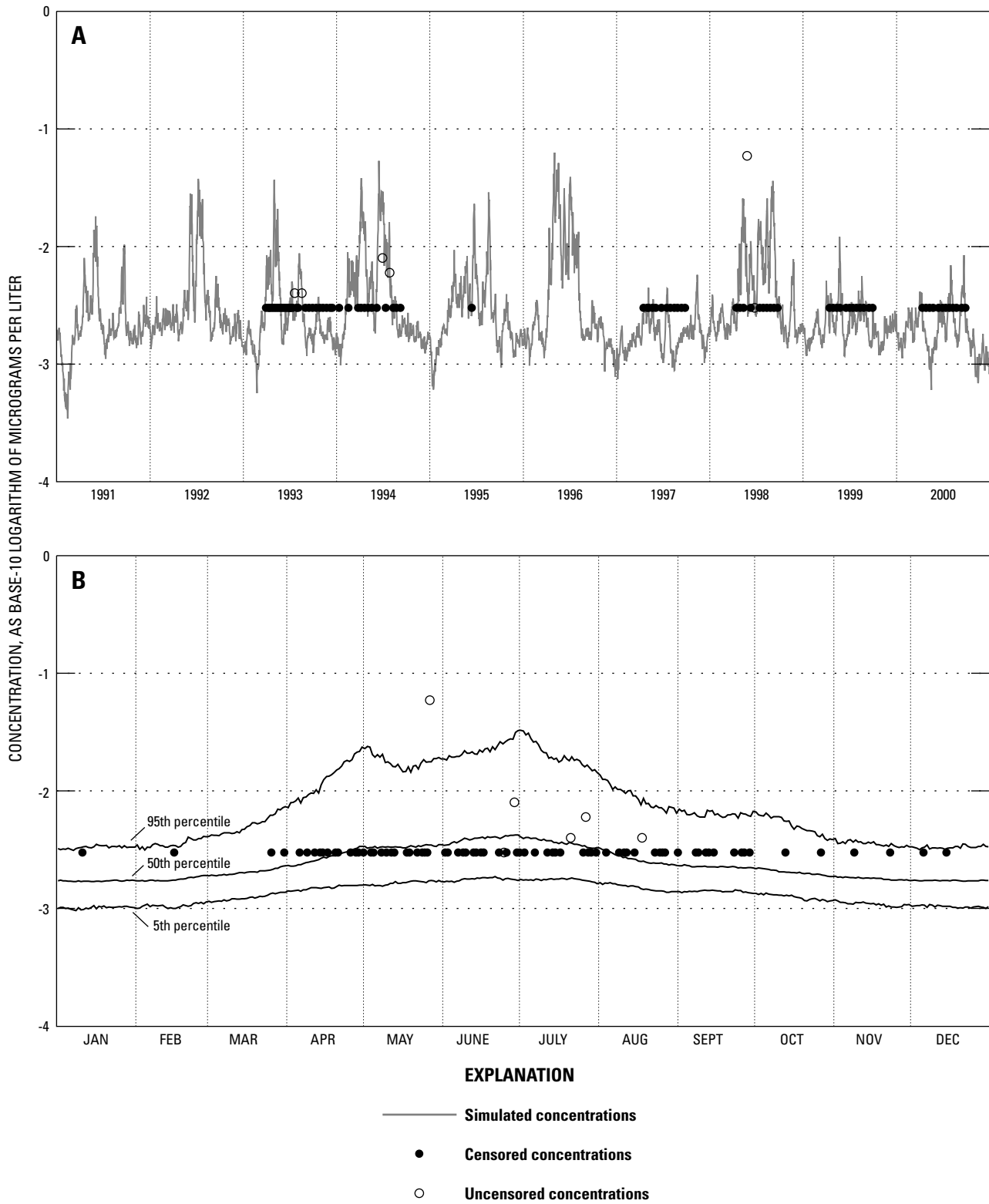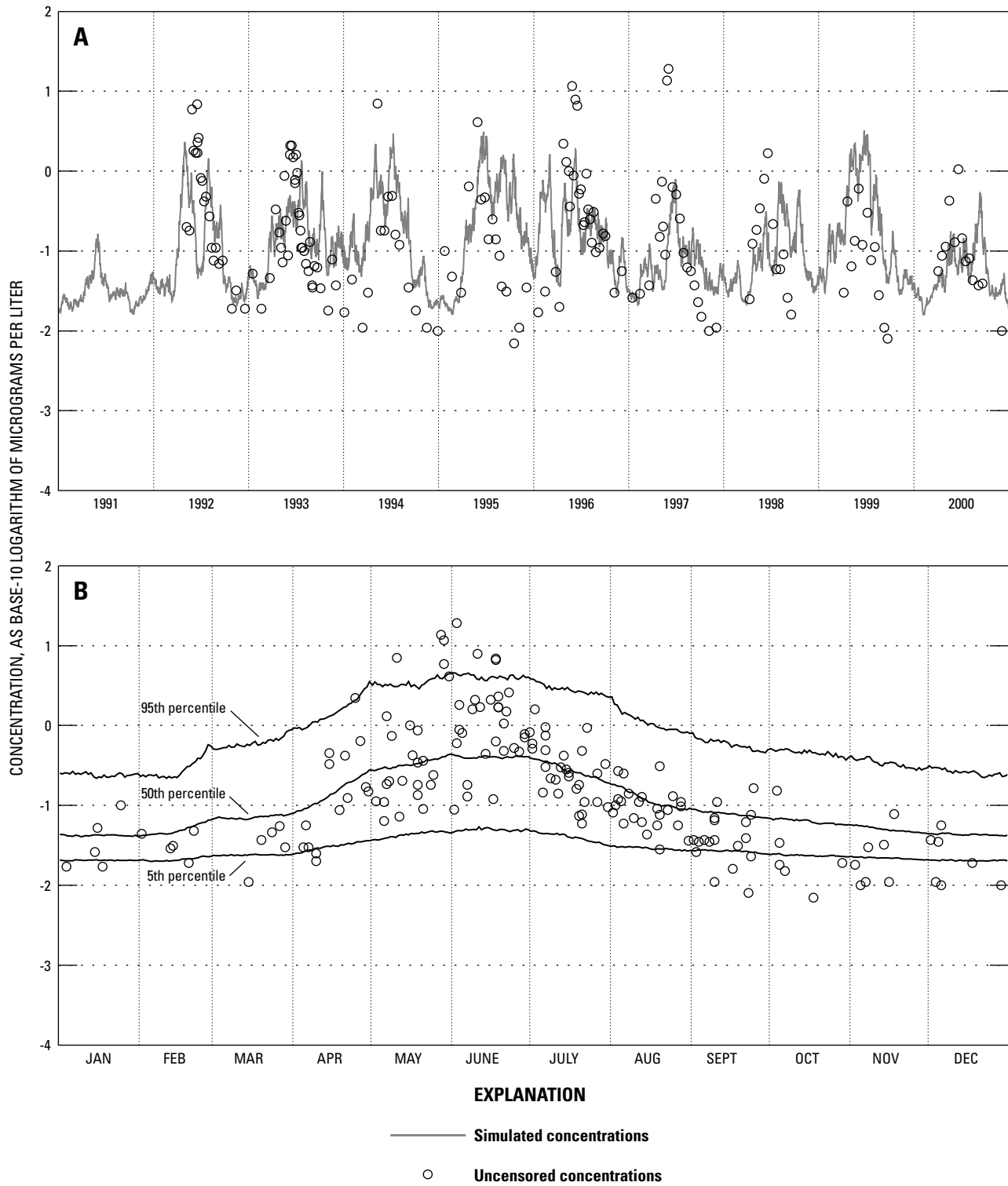
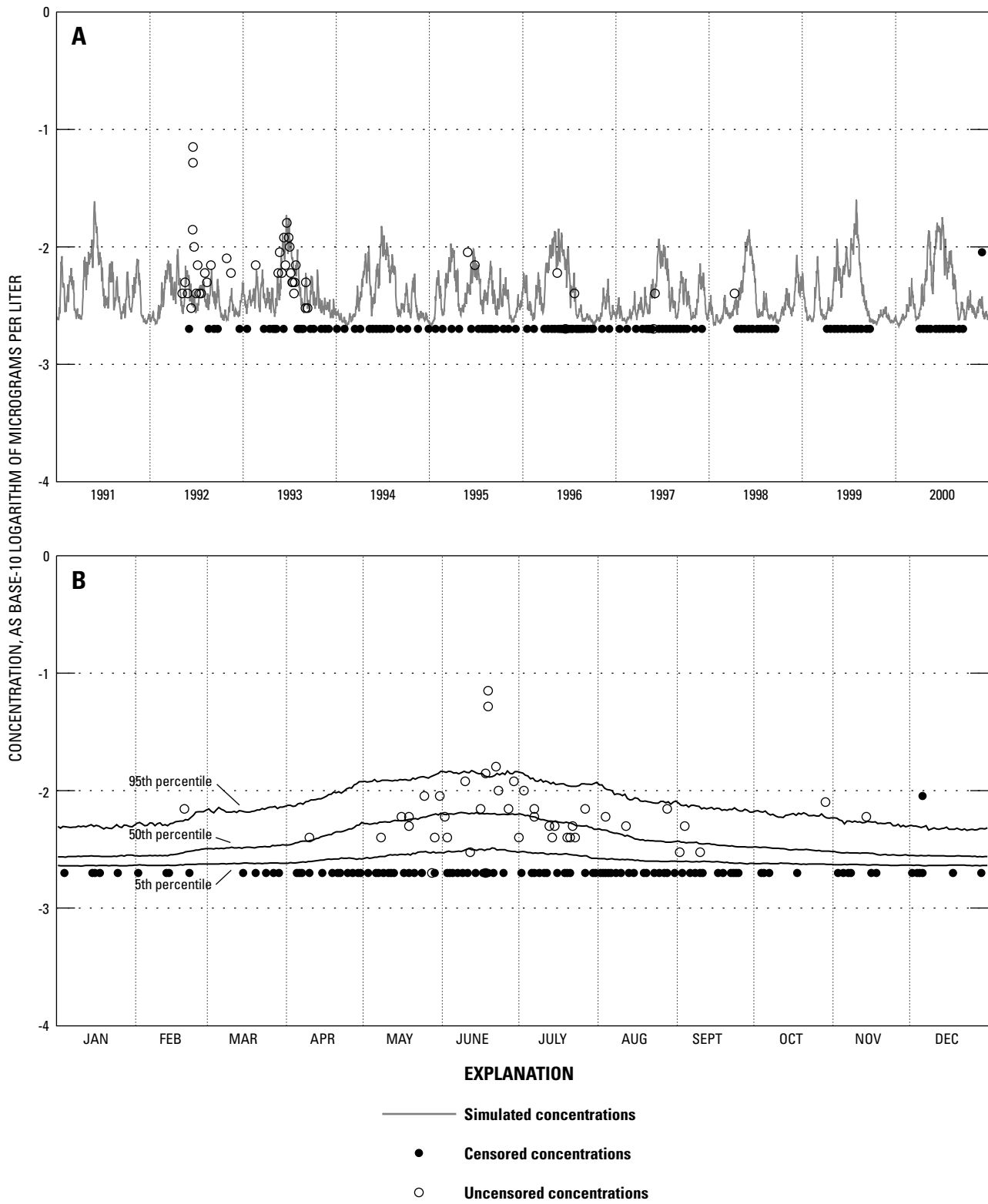**Figure 18.** Generated trace of daily trifluralin concentrations for 1991-2000 for the East Mahantango Creek at Klingerstown, Pennsylvania, station (map number 40) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
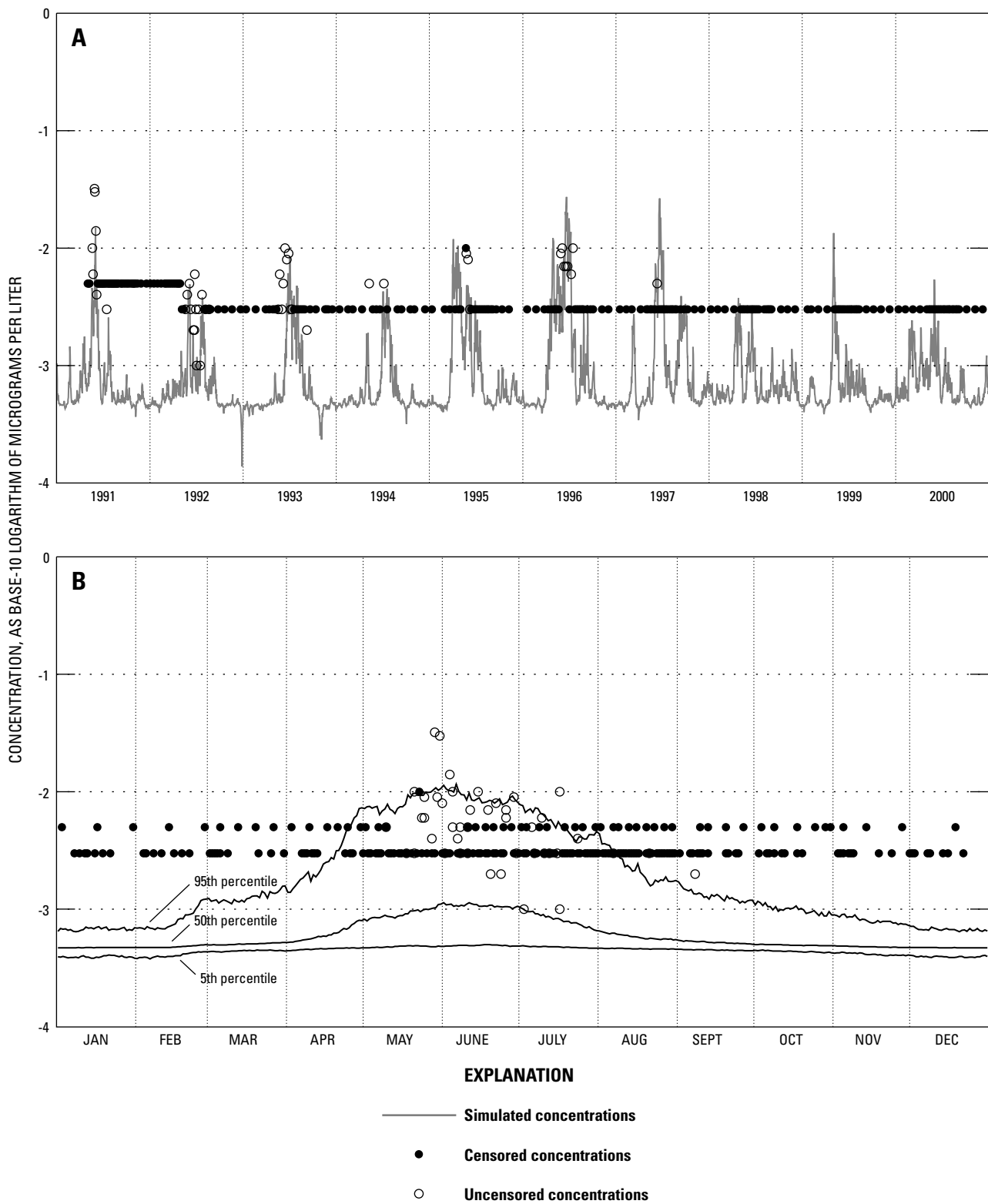
Figure 19. Generated trace of daily ethyldipropylthiocarbamate concentrations for 1991-2000 for the East Mahantango Creek at Klingerstown, Pennsylvania, station (map number 40) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).

Figure 20. Generated trace of daily carbofuran concentrations for 1991-2000 for the East Mahantango Creek at Klingerstown, Pennsylvania, station (map number 40) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).

Figure 21. Generated trace of daily fonofos concentrations for 1991-2000 for the East Mahantango Creek at Klingerstown, Pennsylvania, station (map number 40) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).

Figure 22. Generated trace of daily metolachlor concentrations for 1991-2000 for the Sugar Creek at New Palestine, Indiana, station (map number 99) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).

Figure 23.   Generated trace of daily trifluralin concentrations for 1991-2000 for the Sugar Creek at New Palestine, Indiana, station (map number 99) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
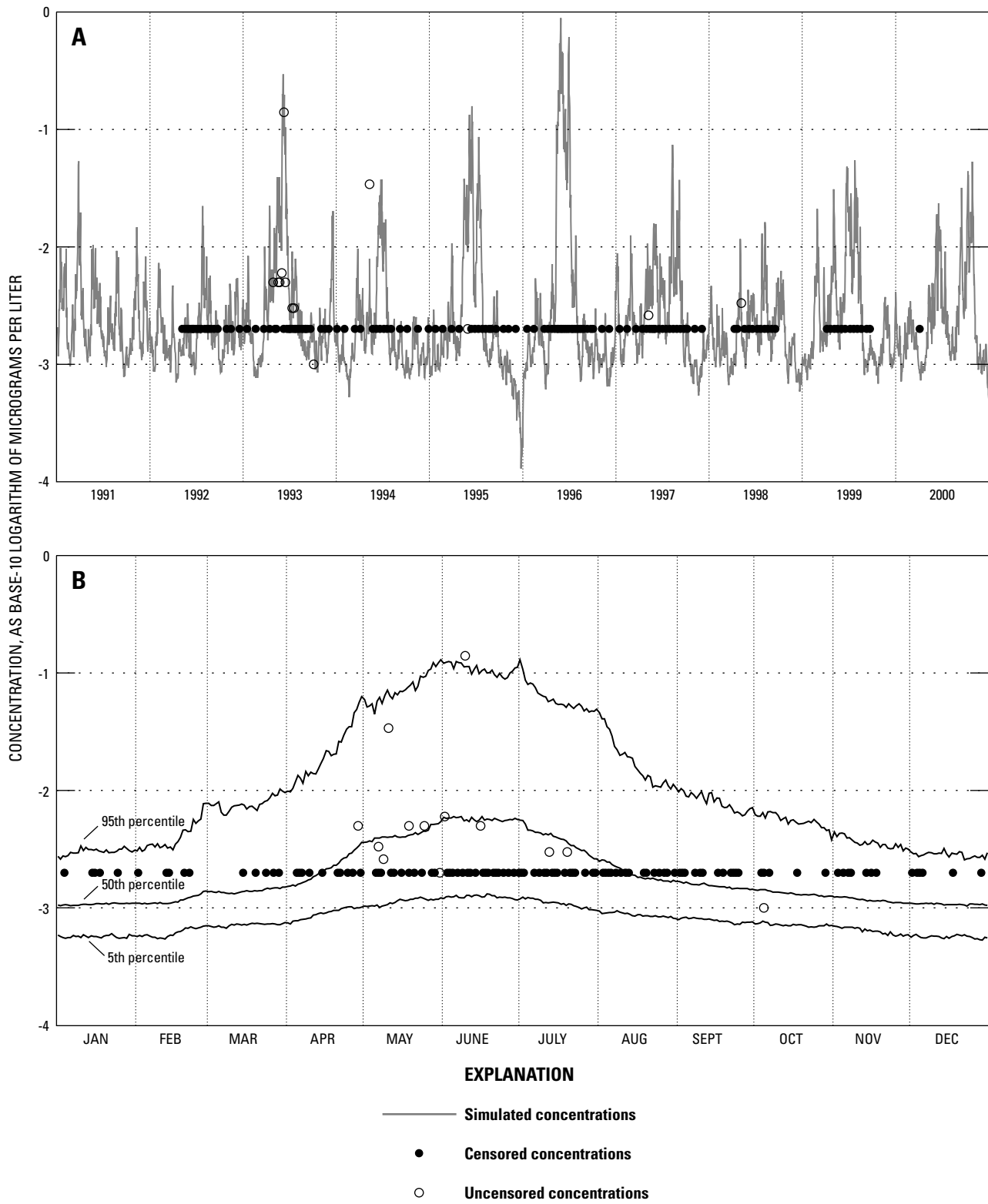
Figure 24. Generated trace of daily ethyldipropylthiocarbamate concentrations for 1991-2000 for the Sugar Creek at New Palestine, Indiana, station (map number 99) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
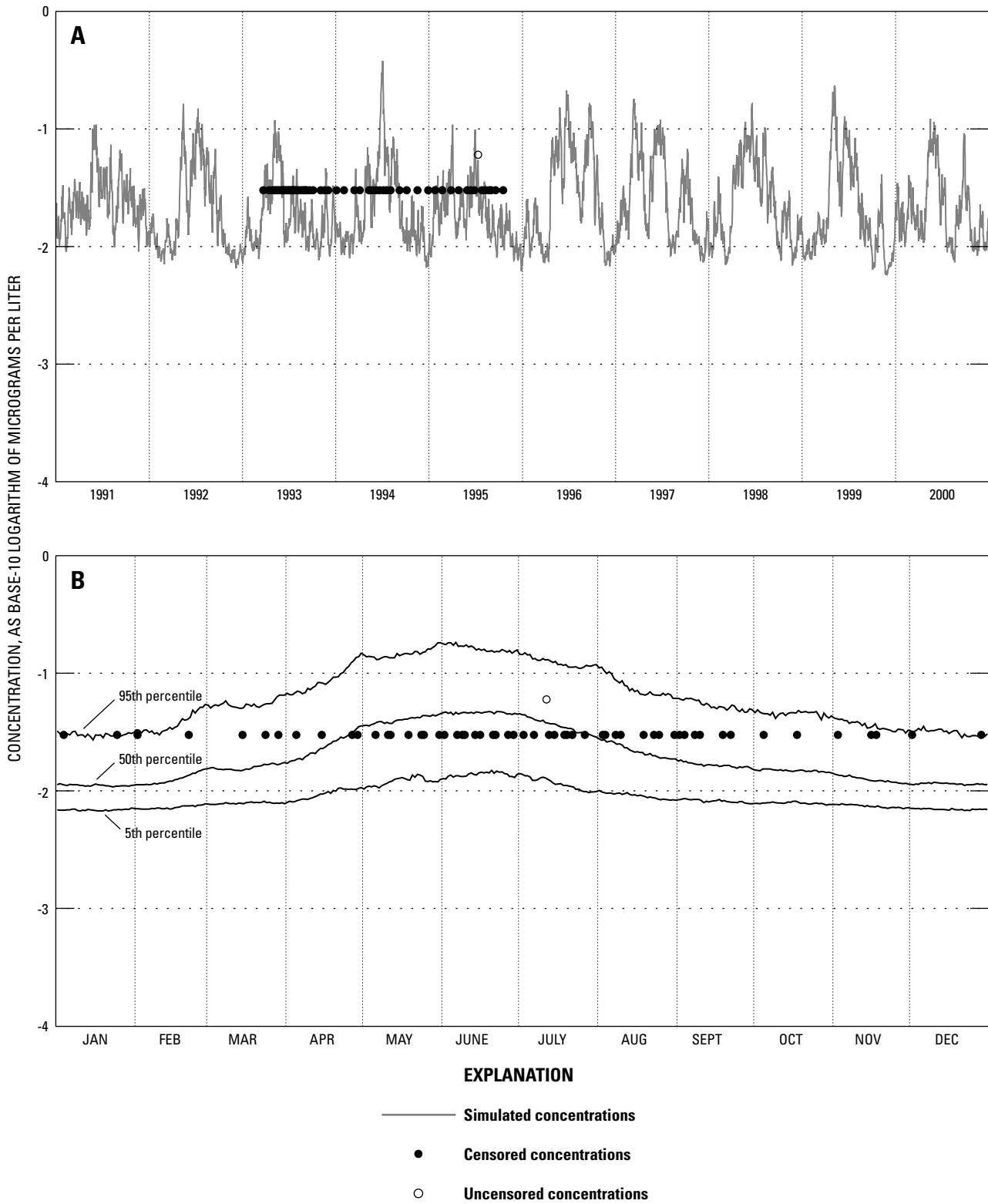
Figure 25. Generated trace of daily carbofuran concentrations for 1991-2000 for the Sugar Creek at New Palestine, Indiana, station (map number 99) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
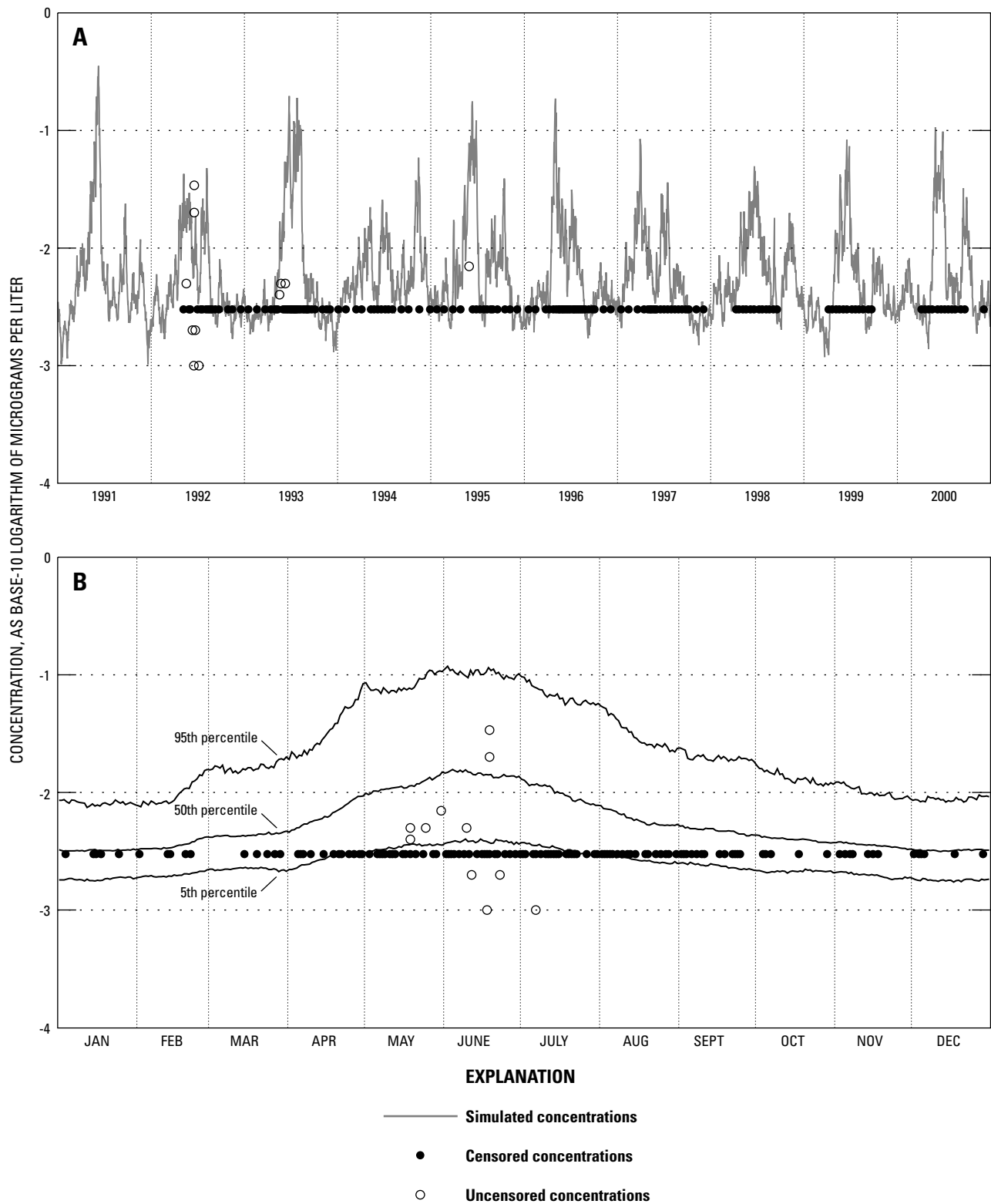
Figure 26.   Generated trace of daily fonofos concentrations for 1991-2000 for the Sugar Creek at New Palestine, Indiana, station (map number 99) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).

Figure 27. Generated trace of daily fonofos concentrations for 1991-2000 for the White River at Hazelton, Indiana, station (map number 97) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
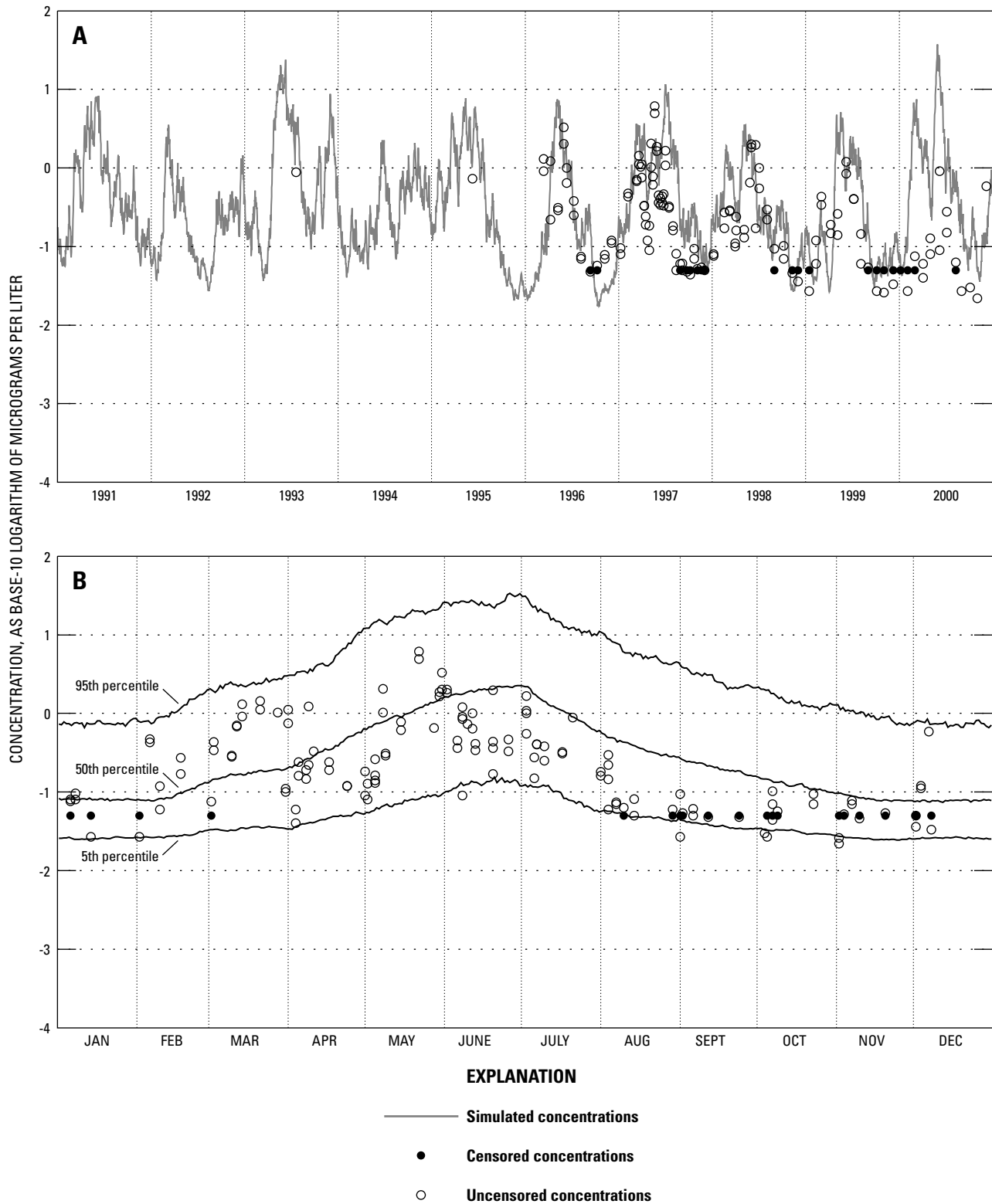
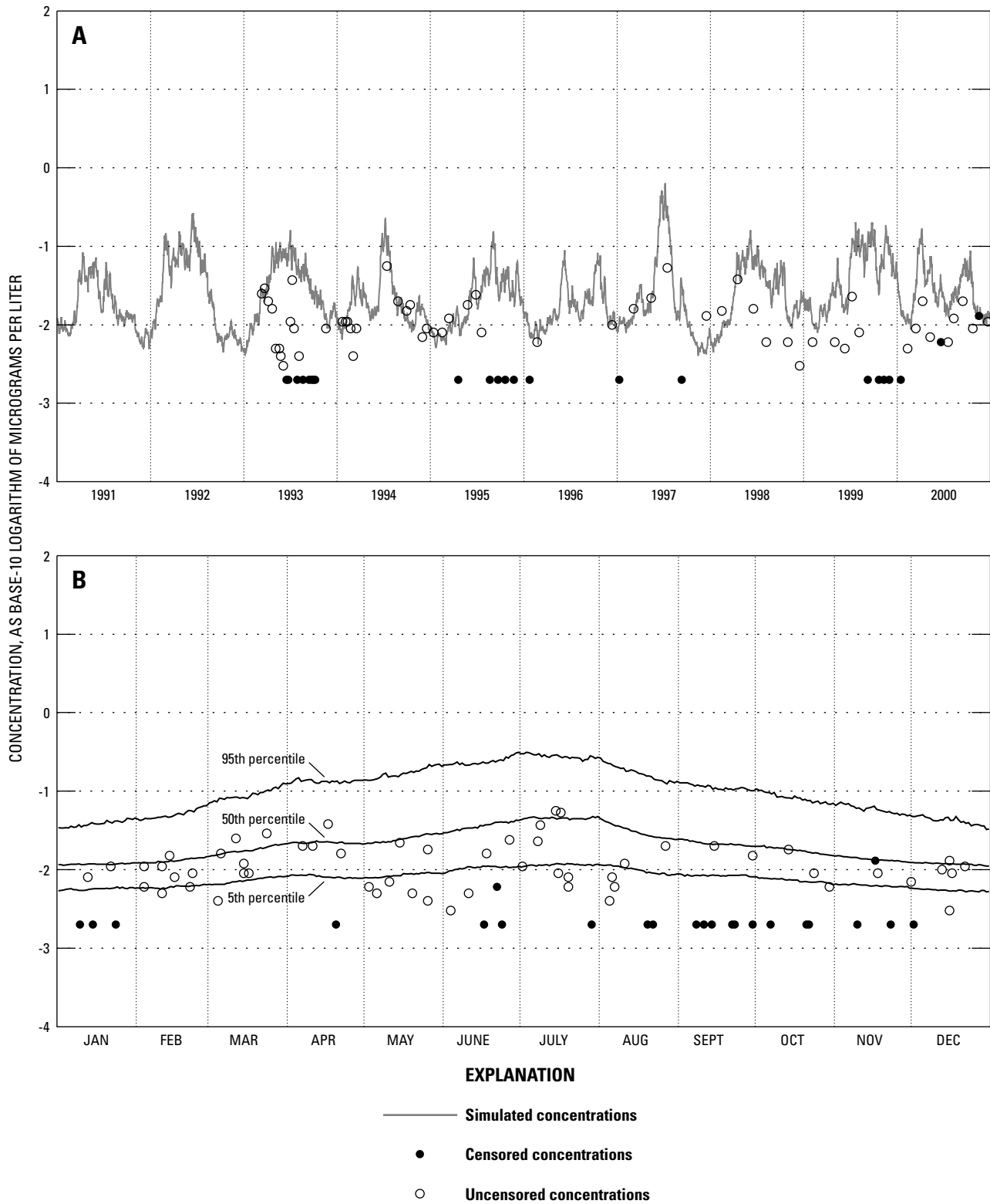**Figure 28.** Generated trace of daily metolachlor concentrations for 1991-2000 for the Iowa River at Wapello, Iowa, station (map number 22) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
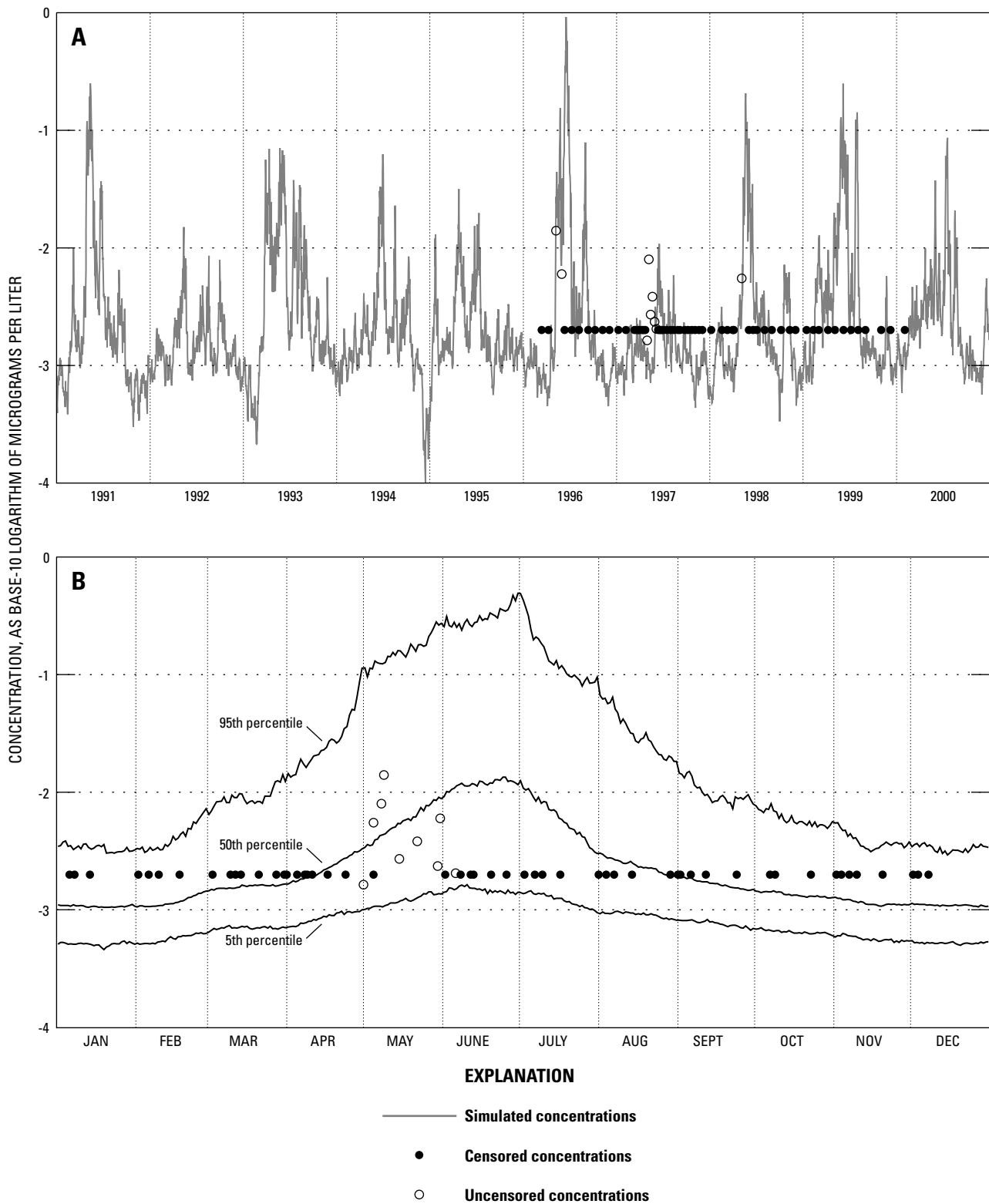
Figure 29.   Generated trace of daily ethyldipropylthiocarbamate concentrations for 1991-2000 for the Iowa River at Wapello, Iowa, station (map number 22) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
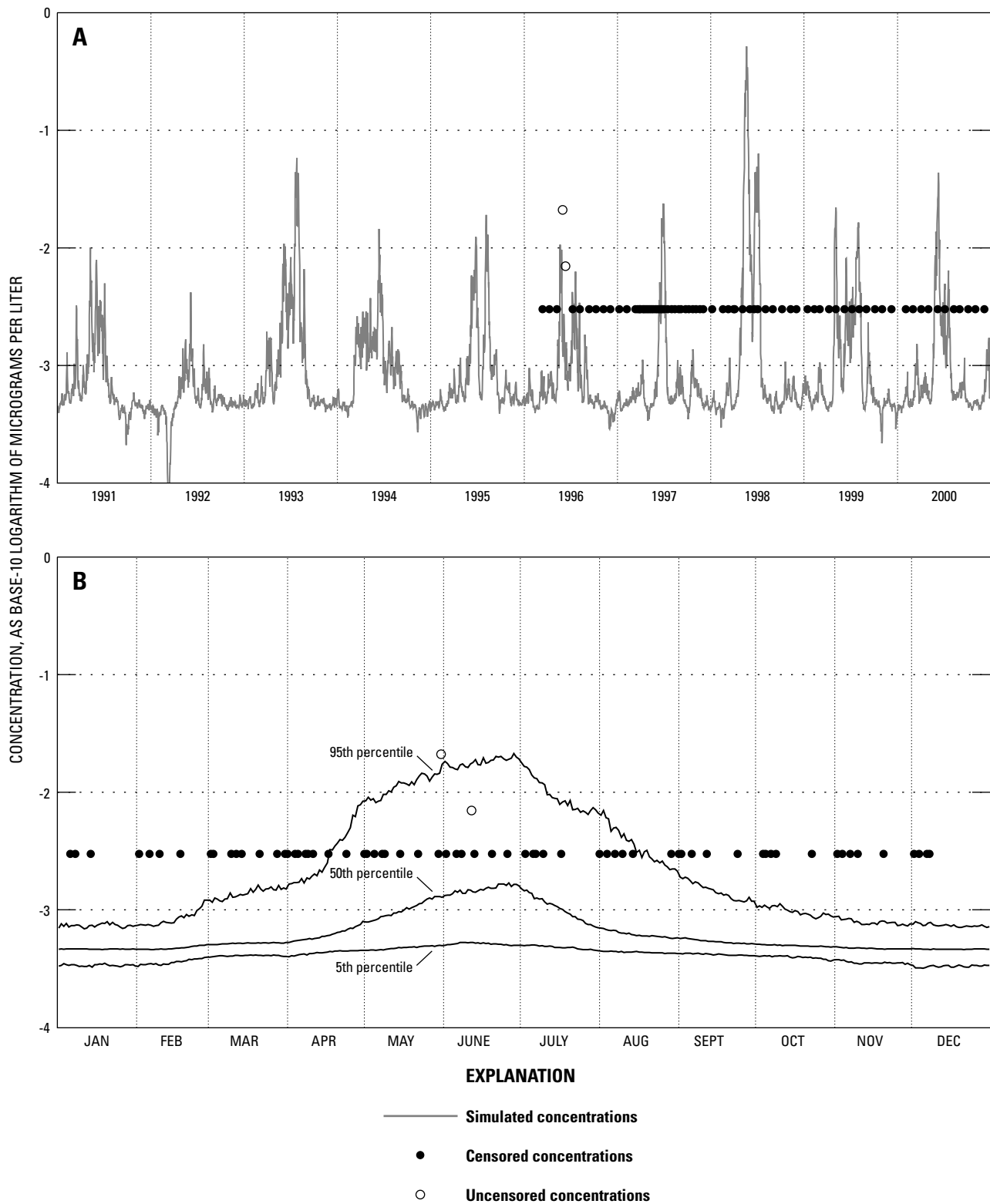
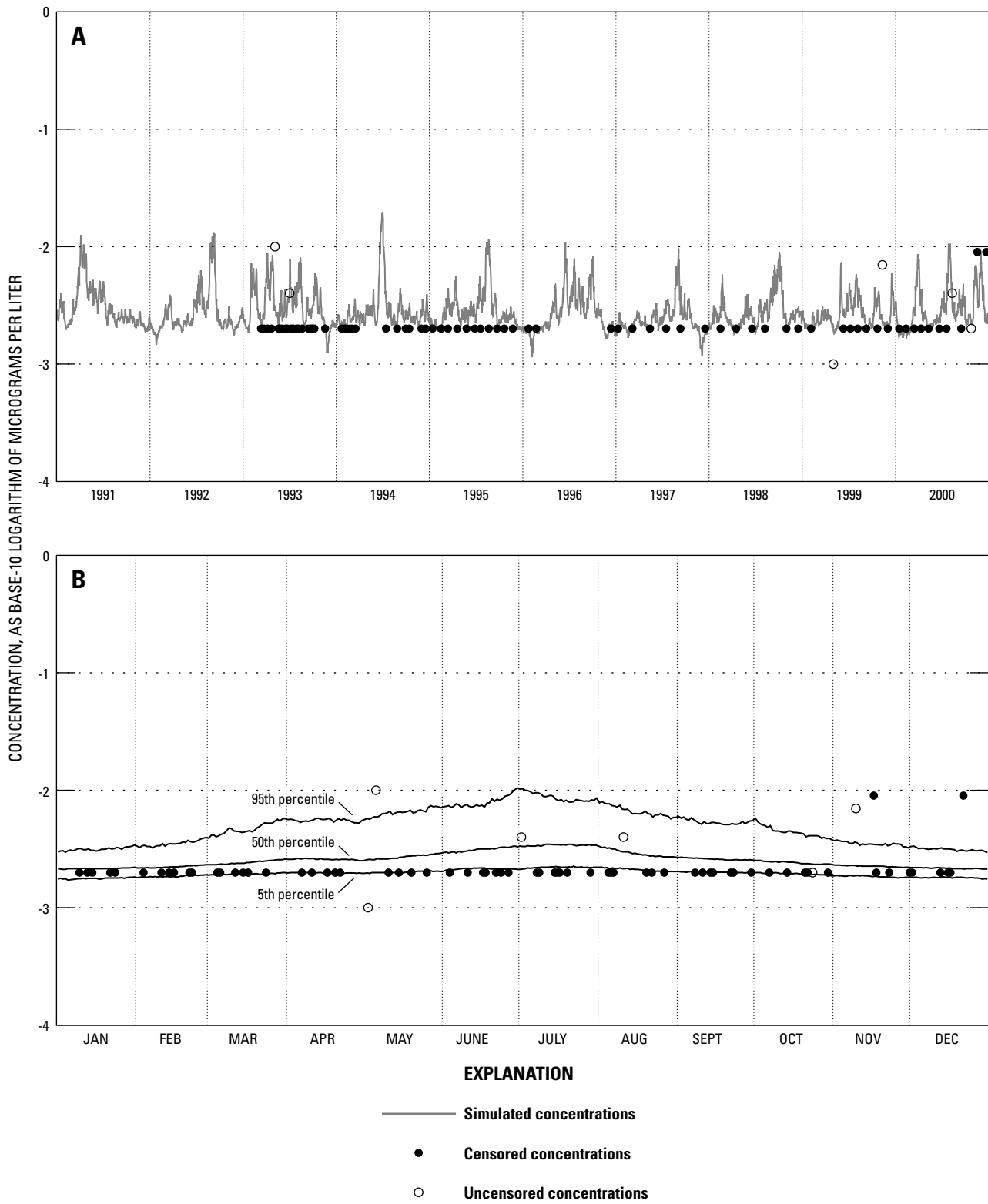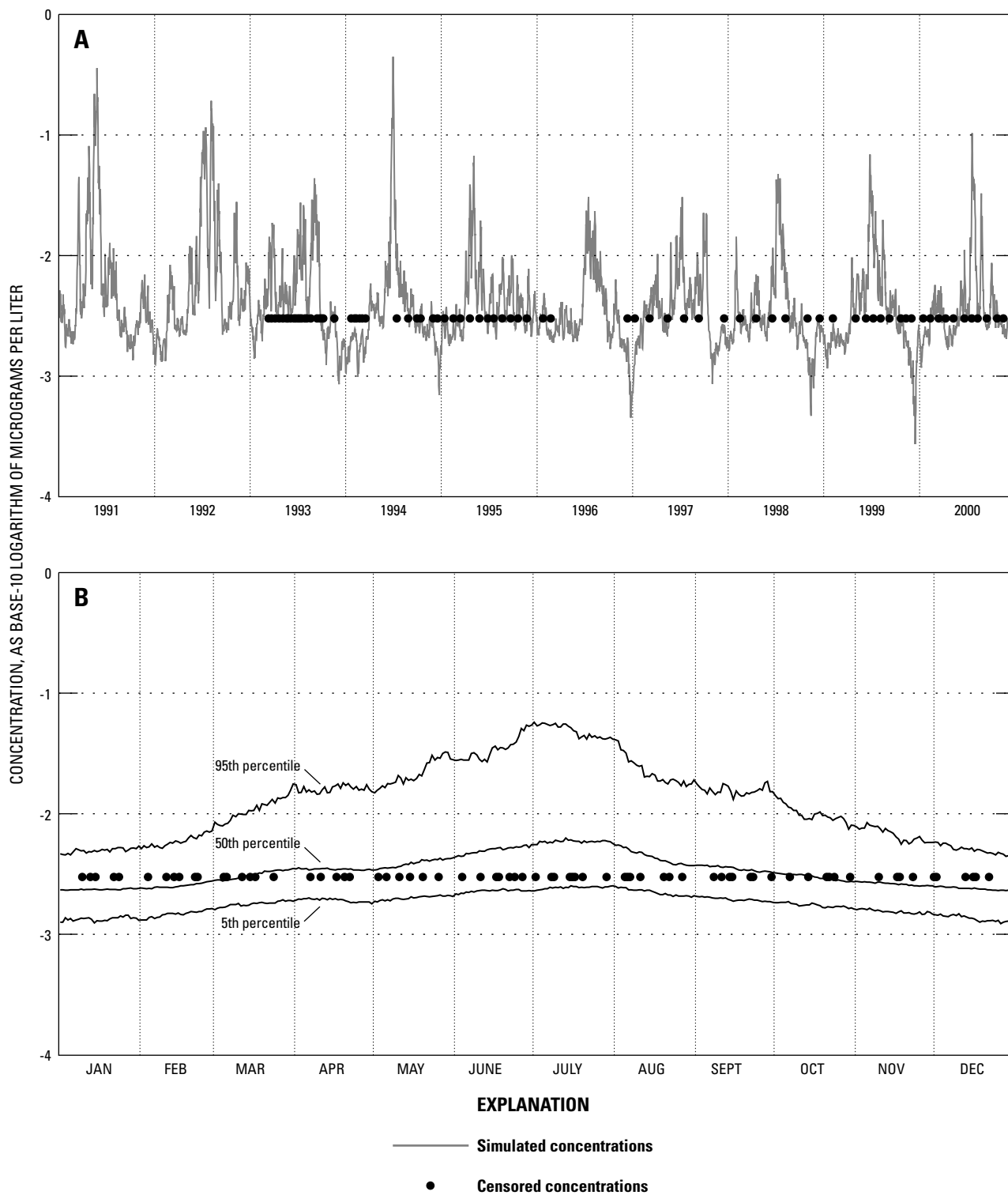**Figure 30.**   Generated trace of daily fonofos concentrations for 1991-2000 for the Iowa River at Wapello, Iowa, station (map number 22) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
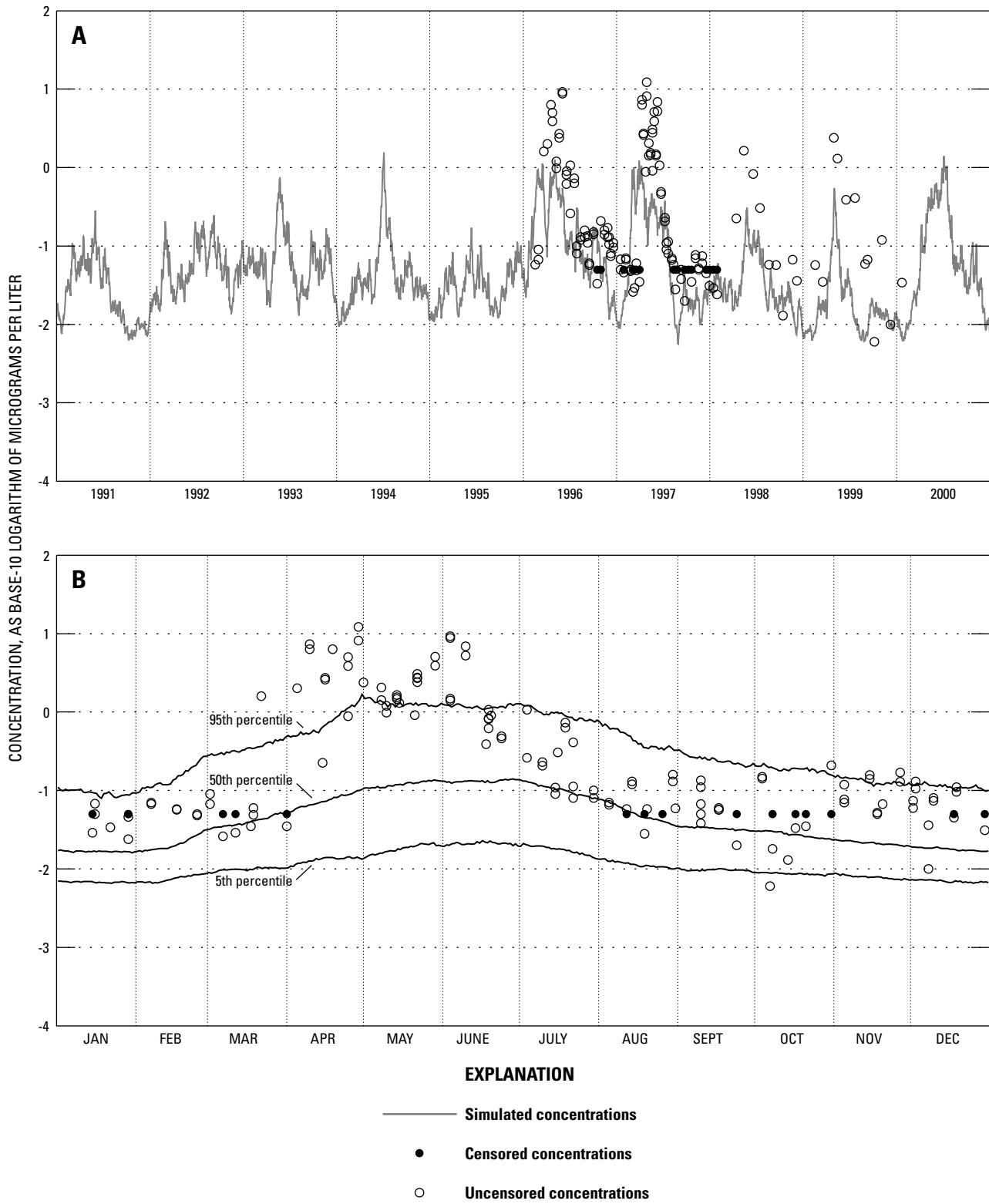
Figure 31.   Generated trace of daily metolachlor concentrations for 1991-2000 for the Bogue Phalia near Leland, Mississippi, station (map number 47) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
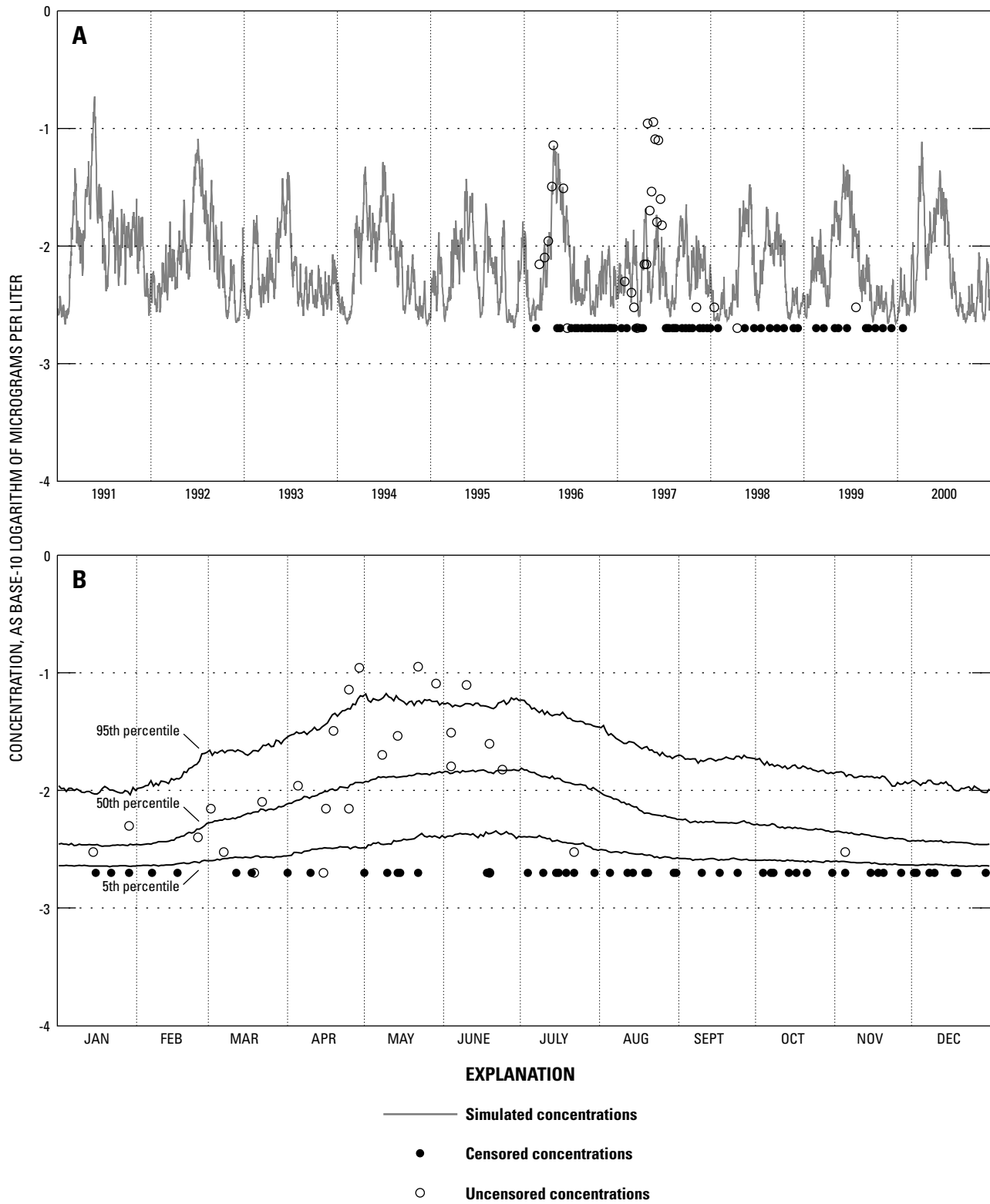
Figure 32.   Generated trace of daily trifluralin concentrations for 1991-2000 for the Bogue Phalia near Leland, Mississippi, station (map number 47) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
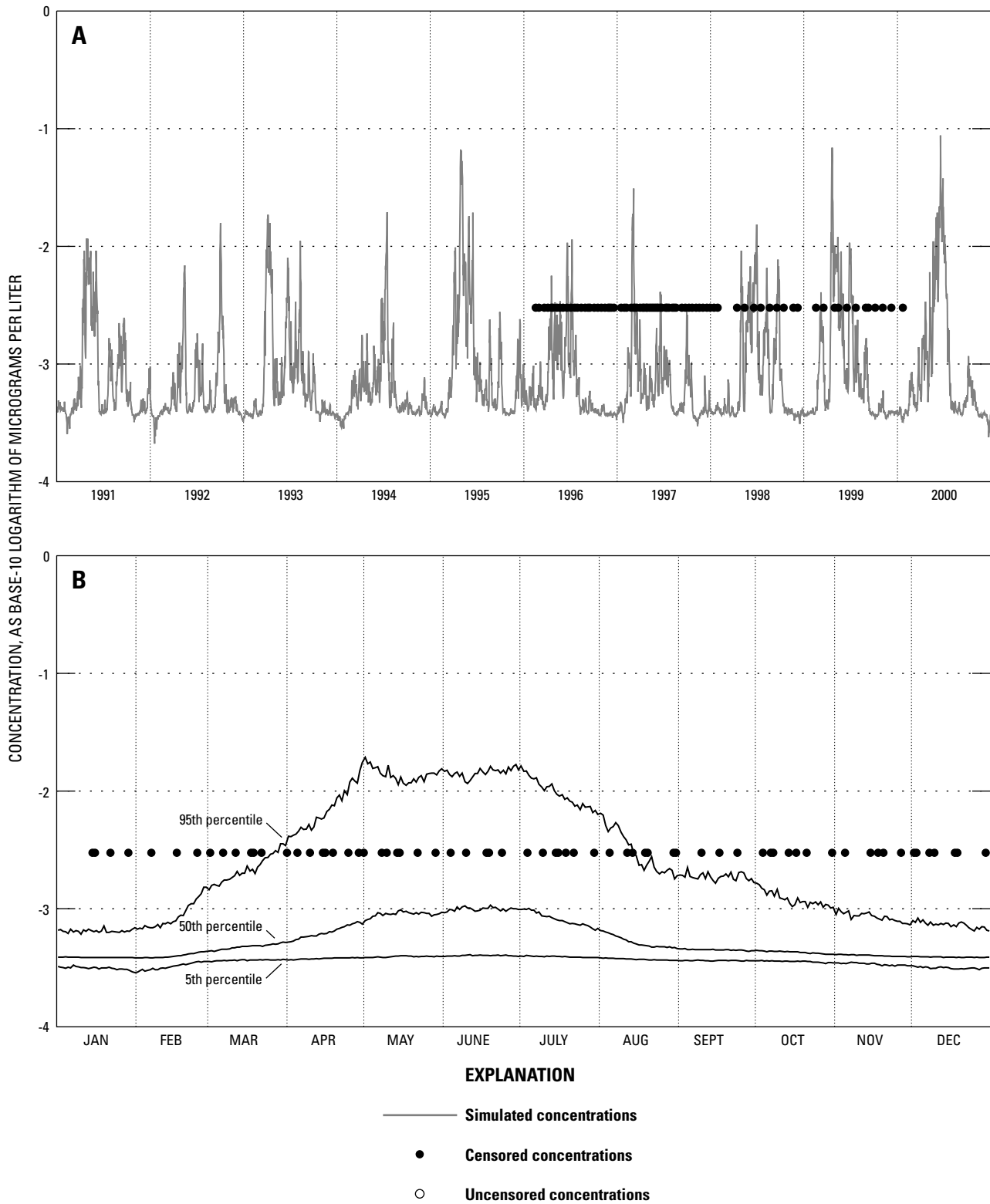
Figure 33. Generated trace of daily fonofos concentrations for 1991-2000 for the Bogue Phalia near Leland, Mississippi, station (map number 47) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).

Figure 34. Generated trace of daily metolachlor concentrations for 1991-2000 for the Withlacoochee River near Quitman, Georgia, station (map number 24) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).

Figure 35.   Generated trace of daily trifluralin concentrations for 1991-2000 for the Withlacoochee River near Quitman, Georgia, station (map number 24) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).

Figure 36. Generated trace of daily fonofos concentrations for 1991-2000 for the Withlacoochee River near Quitman, Georgia, station (map number 24) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
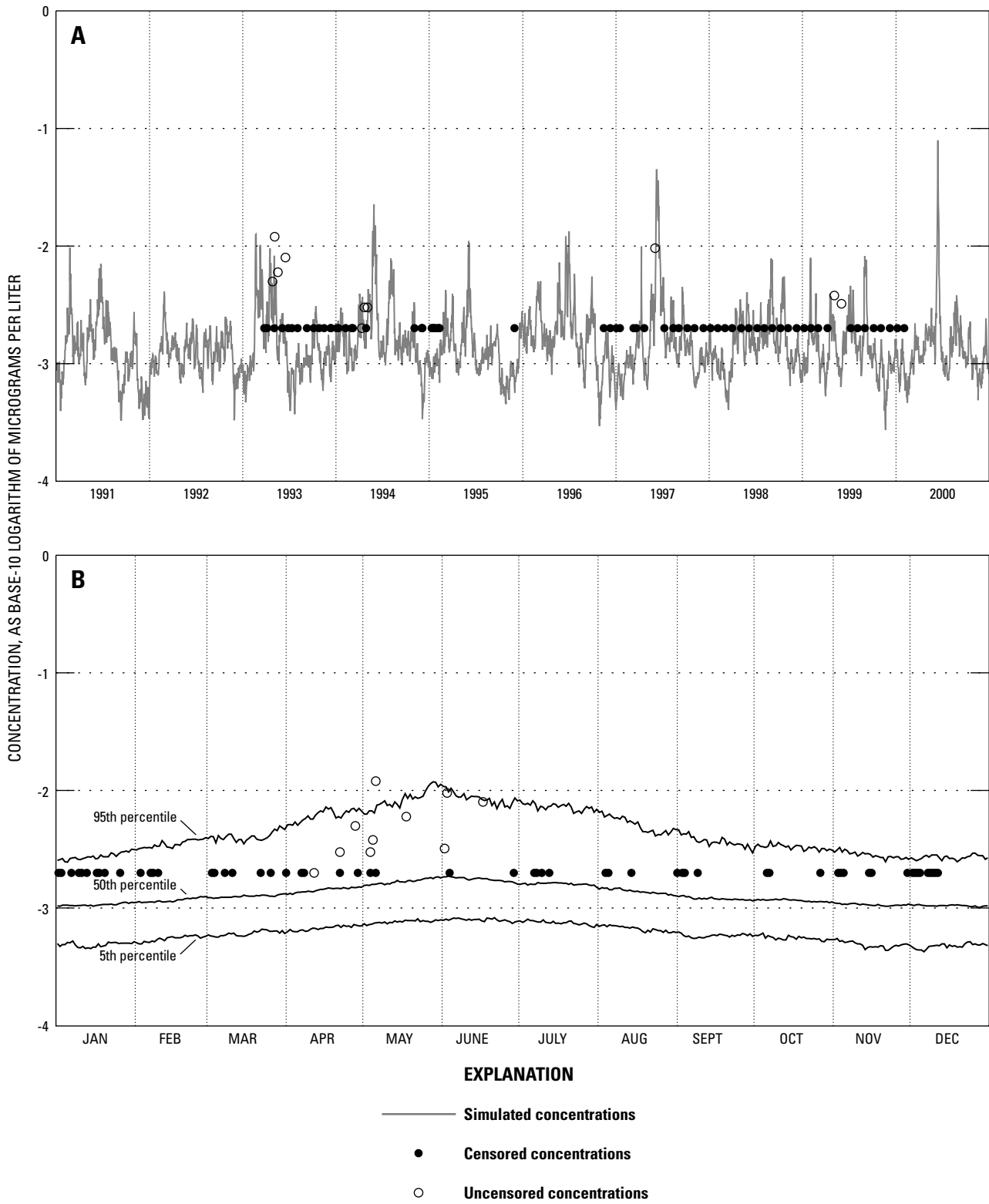
Figure 37. Generated trace of daily ethyldipropylthiocarbamate concentrations for 1991-2000 for the Palouse River at Hooper, Washington, station (map number 12) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
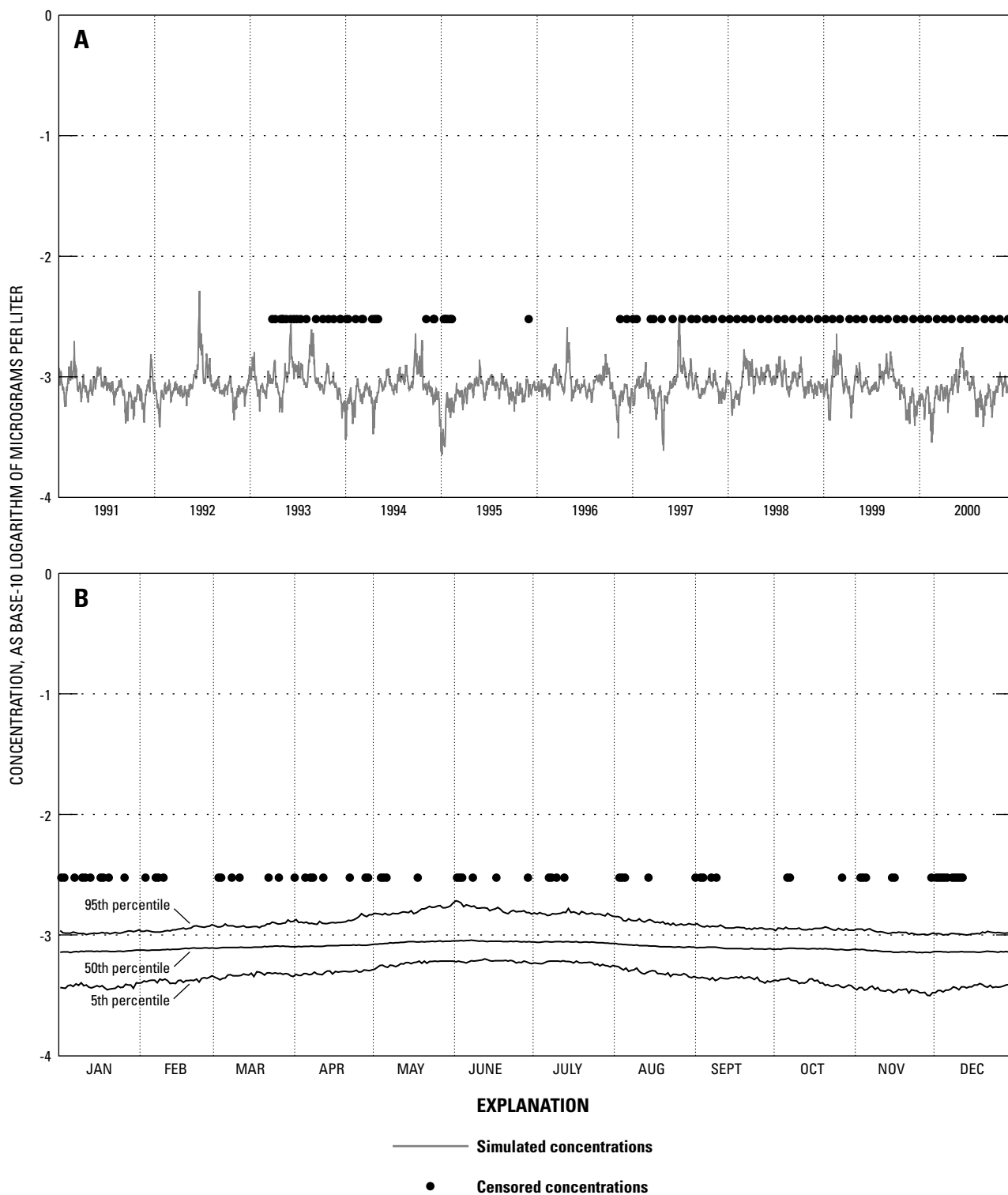
Figure 38.   Generated trace of daily fonofos concentrations for 1991-2000 for the Palouse River at Hooper, Washington, station (map number 12) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
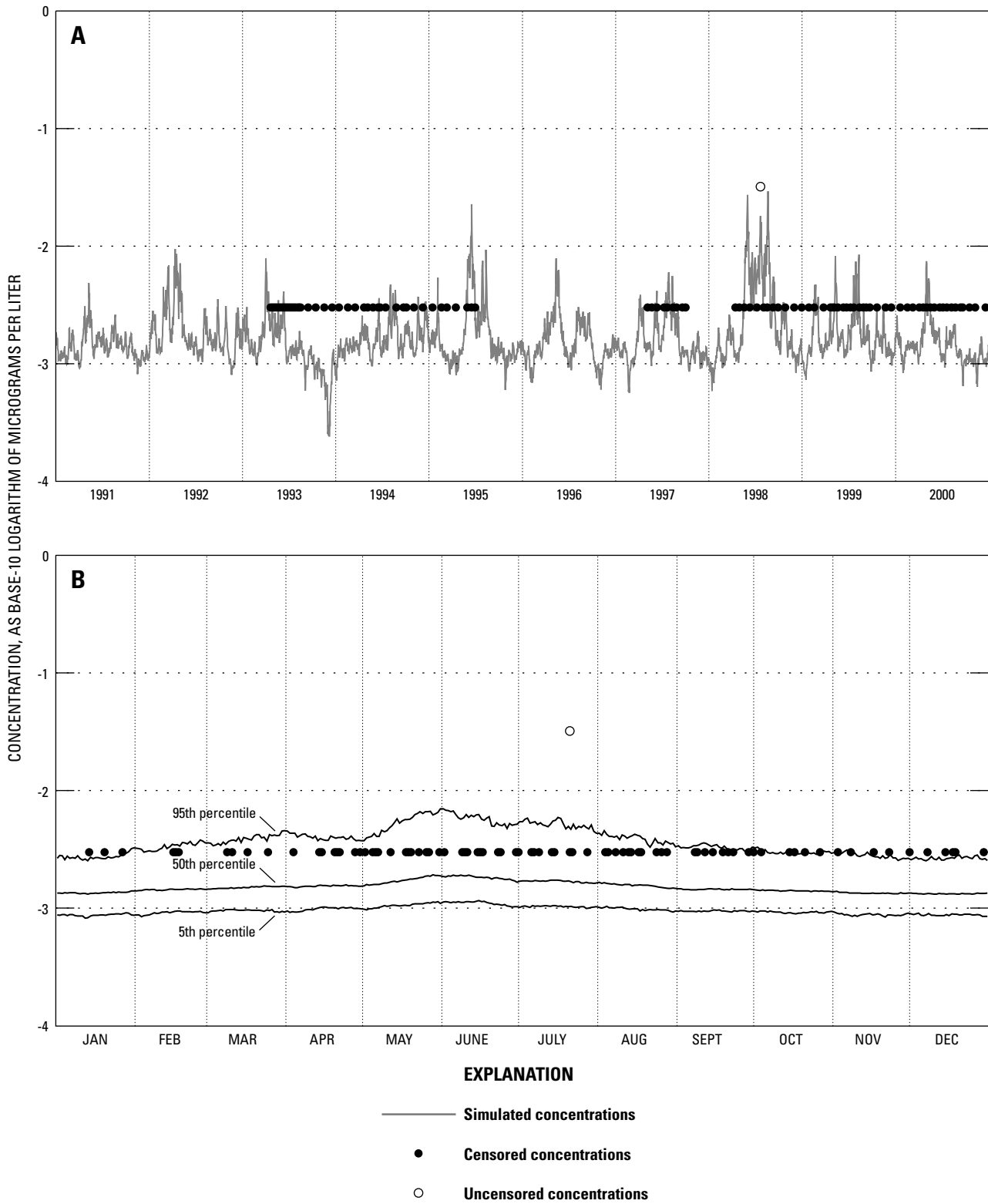
Figure 39.   Generated trace of daily fonofos concentrations for 1991-2000 for the Rock Creek at Twin Falls, Idaho, station (map number 89) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
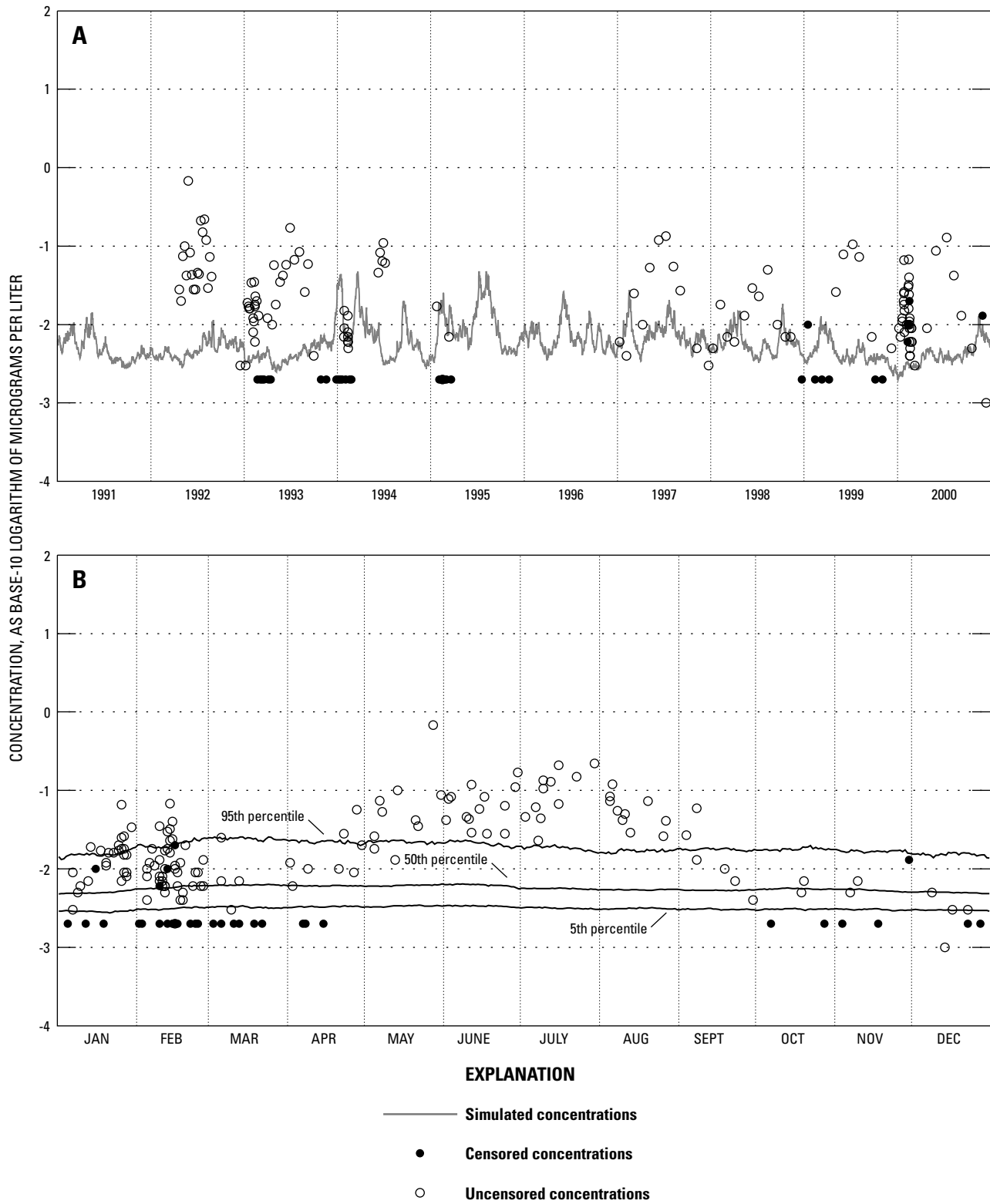
Figure 40. Generated trace of daily metolachlor concentrations for 1991-2000 for the San Joaquin River near Vernalis, California, station (map number 66) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
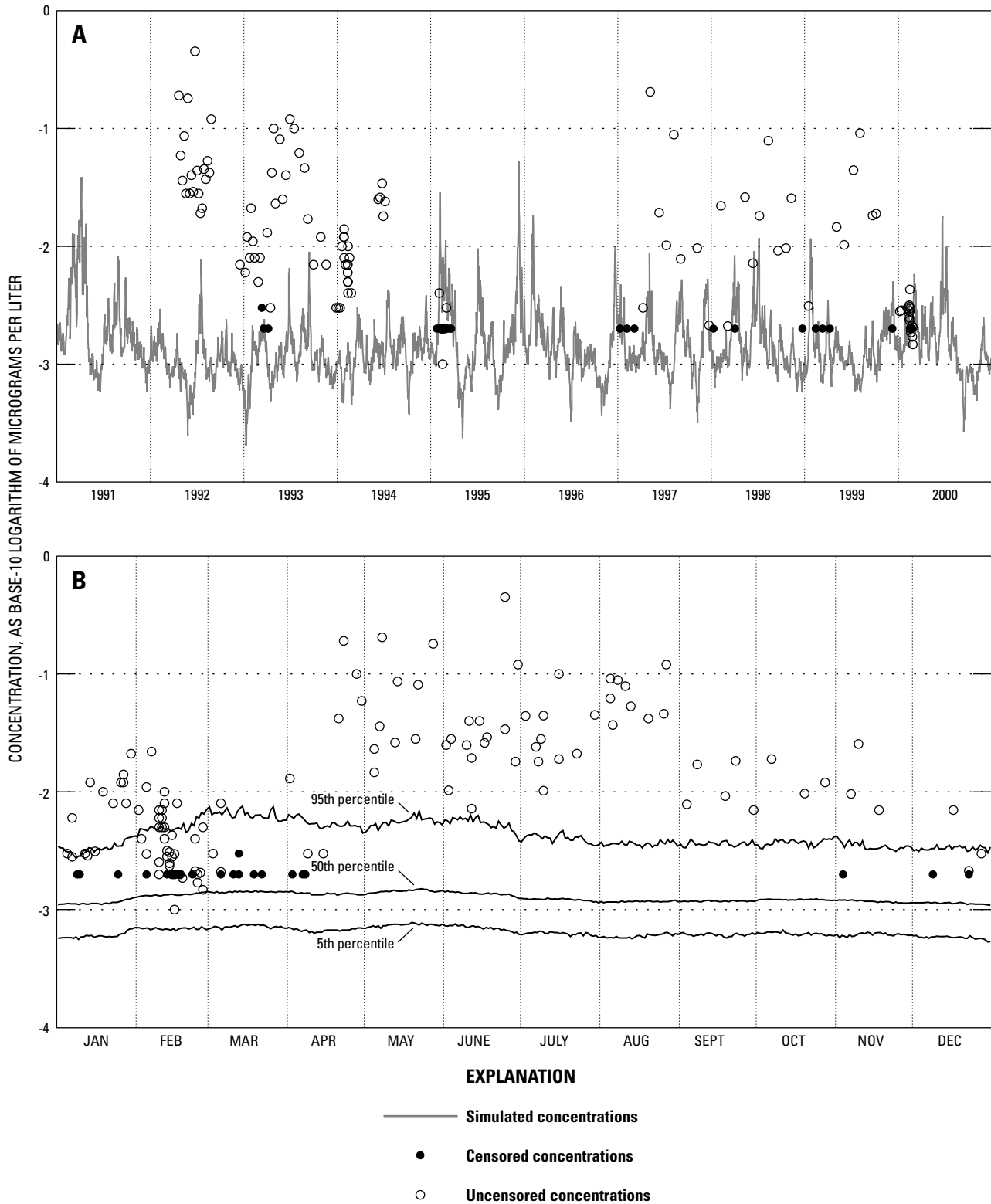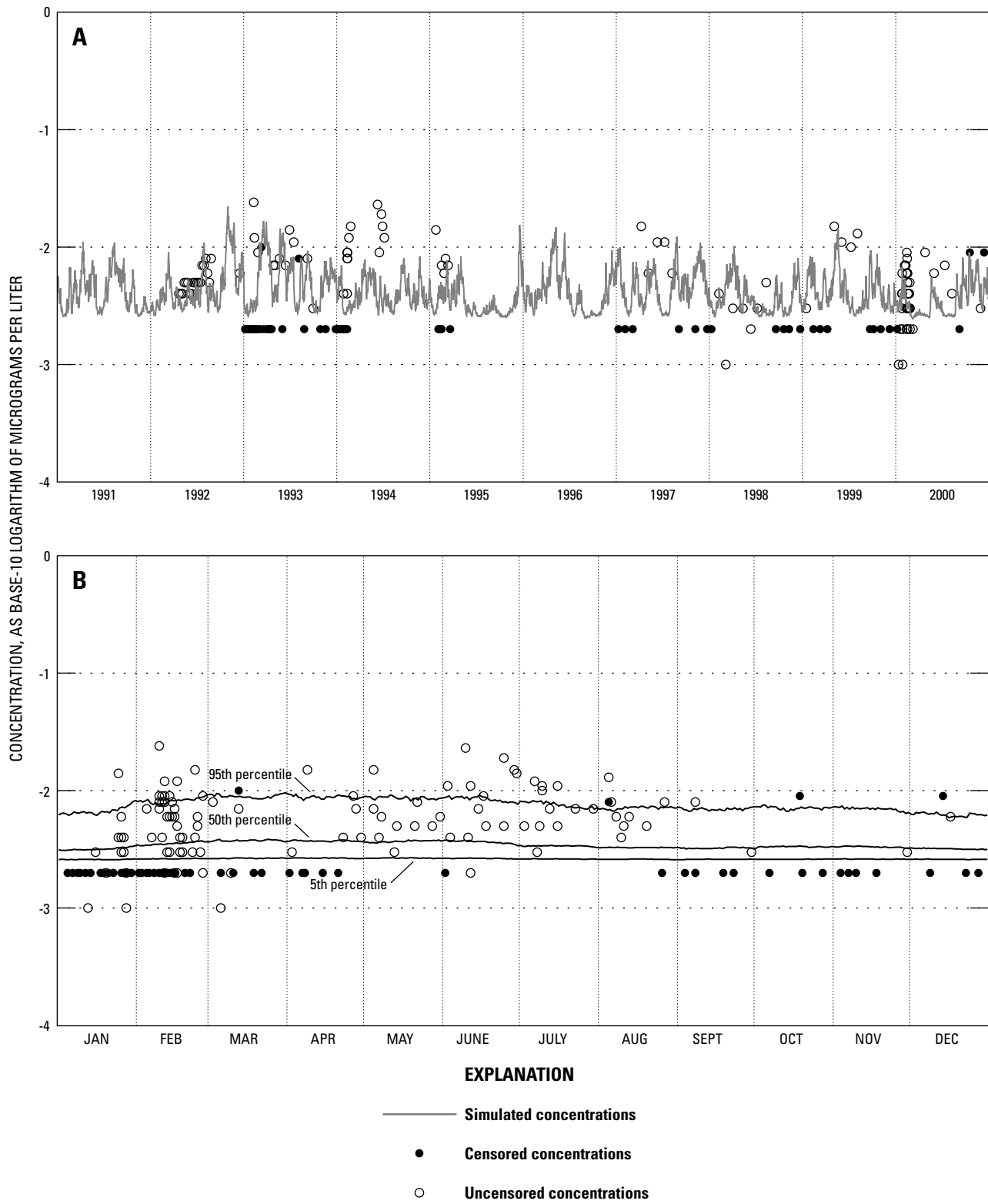
Figure 41.   Generated trace of daily ethyldipropylthiocarbamate concentrations for 1991-2000 for the San Joaquin River near Vernalis, California, station (map number 66) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).

Figure 42.   Generated trace of daily trifluralin concentrations for 1991-2000 for the San Joaquin River near Vernalis, California, station (map number 66) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).
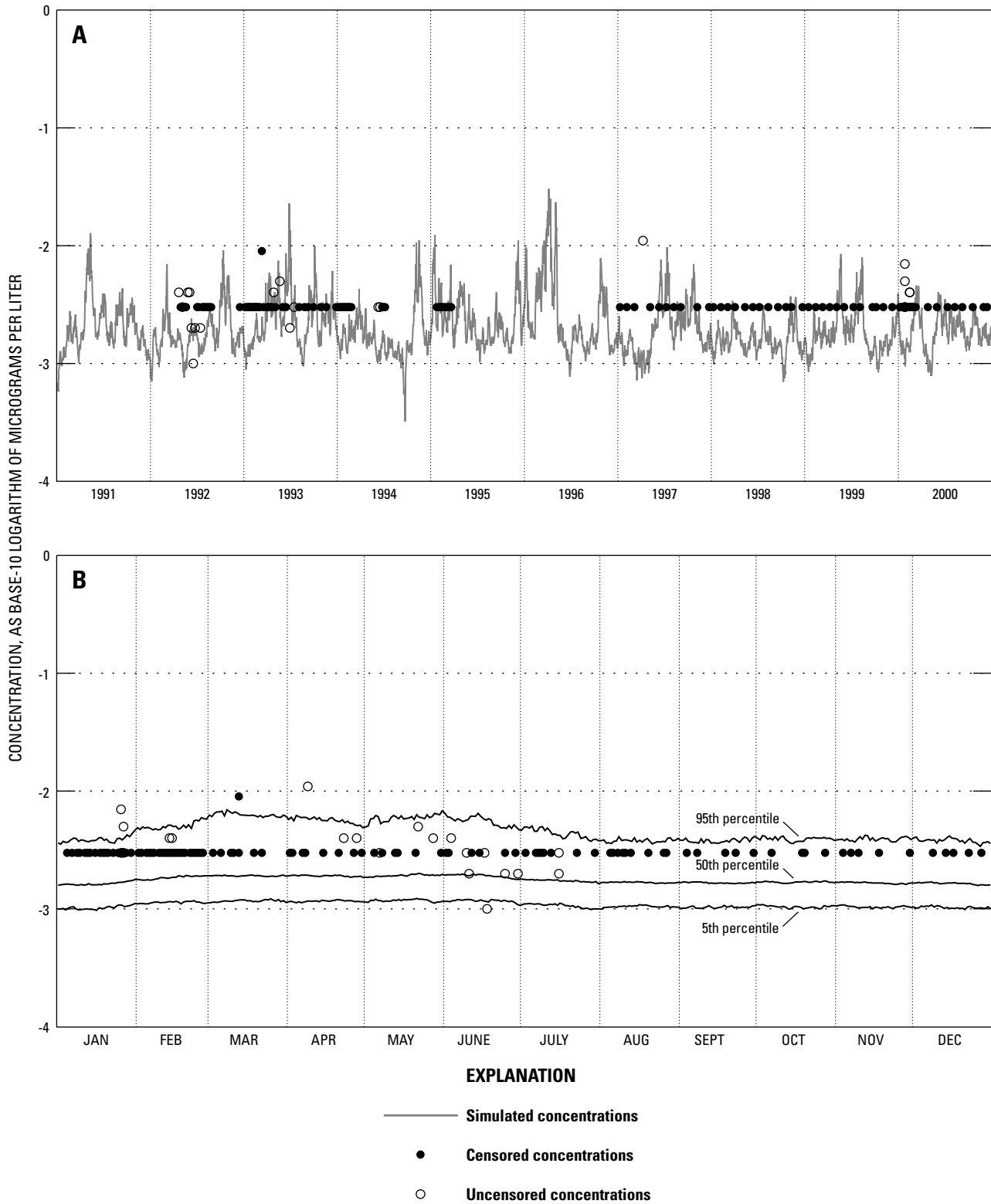
Figure 43. Generated trace of daily fonofos concentrations for 1991-2000 for the San Joaquin River near Vernalis, California, station (map number 66) (A) and 5th, 50th, and 95th percentiles computed from 100 generated traces (B).

USGS