Appendix 10. Model Archive Summary for Atrazine Concentration at U.S. Geological Survey Station 07144780, North Fork Ninnescah River above Cheney Reservoir, Kansas, during November 14, 2015, through September 30, 2021

This model archive summary summarizes the atrazine concentration model developed to compute 15-minute, hourly, or daily atrazine concentrations during November 14, 2015, onward. This model supersedes all prior models used during this period. The methods follow U.S. Geological Survey (USGS) guidance as referenced in relevant Office of Surface Water/Office of Water Quality Technical Memoranda and USGS Techniques and Methods, book 3, chapter C4 (Rasmussen and others, 2009; U.S. Geological Survey, 2016).

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

Site and Model Information

Site number: 07144780

Site name: North Fork Ninnescah River above Cheney Reservoir, Kansas

Location: Lat 37°51'45", long 98°00'49" referenced to North American Datum of 1927, in NE 1/4 SE 1/4 NE 1/4 sec.19, T.25 S., R.6 W., Reno County, Kans., hydrologic unit 11030014, on right bank at upstream side of county highway bridge, 10 miles south of Hutchinson, 18.1 miles upstream from Cheney Dam.

Equipment: A YSI, Inc., EXO water-quality monitor (YSI, Inc., 2017) equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, and turbidity was installed November 14, 2015. The EXO monitor was installed in a 4-inch-diameter metal or polyvinyl chloride (or PVC) pipe suspended from the downstream side of the bridge in the deepest fastest flowing water. Measurements from the EXO were recorded every 15 minutes to hourly and transmitted hourly via satellite. Real-time stage was measured using a Design Analysis Water Log H–350/355 nonsubmersible pressure transducer.

Date model was created: August 9, 2022

Model calibration data period: March 2, 2018, through February 1, 2021 (dataset consisted of 22 discrete water-quality samples).

Model application date: November 14, 2015, onward (date of EXO continuous water-quality monitor installation).

Model developed by: Ariele Kramer, USGS, Lawrence, Kans. (akramer@usgs.gov)

Model Calibration Dataset

All data were collected using USGS protocols (U.S. Geological Survey, 2006; Wagner and others, 2006; Bennett and others, 2014) and are stored in the USGS National Water Information System (NWIS) database (https://doi.org/10.5066/F7P55KJN; U.S. Geological Survey, 2022). Potential explanatory variables evaluated individually and in combination were water temperature, specific conductance, pH, dissolved oxygen, turbidity, seasonality (sine and cosine variables), and streamflow.

The regression model is based on 22 concomitant values of discretely collected atrazine concentrations and continuously measured specific conductance and seasonality during March 2, 2018, through February 2, 2021. Discrete samples were collected throughout a range of continuously observed hydrologic conditions. Atrazine concentrations were less than minimum reporting level (less than [<] 0.1 milligram per liter) for 11 samples (50 percent). All potential explanatory variables were time interpolated within the 15-minute to hourly continuous record based on the discrete sample time. The maximum time span between two continuous data points used for interpolation was 4 hours (to preserve the sample dataset, field monitor averages obtained during sample collection were used for model development data if no continuous data were available or if gaps larger than 4 hours in the continuous data record resulted in missing interpolated data). Summary statistics and the complete model-calibration dataset are provided below. Potential outliers were identified using the methods described in Rasmussen and others (2009) and Helsel and others (2020). All potential outliers were investigated by reviewing sample collection information sheets and laboratory reports; if there were no clear issues, explanations, or conditions that would cause a result to be invalid for model calibration, the sample was retained in the dataset. No samples in the model calibration dataset were flagged as outliers.

Atrazine Sampling and Analysis Details

Discrete water-quality samples were collected over a range of hydrologic conditions primarily using a combination of equal depth- and width-integrated and multiple-vertical sample collection techniques (U.S. Geological Survey, 2006). Equal-width-increment and multiple-vertical sample cross sections included five to 12 sampling points with more than 85 percent of samples including 10 or more sampling points. Samples were collected either instream as a wading sample within 300 feet of the bridge or from the downstream side of the bridge using a Federal Interagency Sedimentation Project depth-integrated sampler with a polytetrafluoroethylene bottle, cap, and nozzle. Discrete samples were collected on a semifixed to event-based schedule two to seven times per year.

Atrazine analysis (LCAZ) was a direct injection ultra-performance liquid chromatography/ tandem mass spectrometry (UPLC/MS/MS) research method developed at the Kansas Water Science Center, Organic Geochemistry Research Laboratory (OGRL). Surface water samples were analyzed for atrazine with D_5-atrazine as the surrogate internal standard and a reporting level of 0.1 microgram per liter (µg/L). Water samples were filtered in the field and collected in 4-oz (125-milliliter (mL)) amber glass bottles with polytetrafluoroethylene lined caps and shipped overnight or delivered on ice to the OGRL. After arrival, samples were stored at -20 degrees Celsius (°C) until preparation for analysis.

Standard curve, check standards, and laboratory blank samples were prepared in Type I water and treated the same as the environmental water samples. A duplicate sample, spike sample, carryover blank, and check standards were analyzed after every tenth sample. Samples, standards, and blanks were prepared for analysis by spiking with the appropriate amount of surrogate internal standard and analyte mix and filtering with Waters Acrodisc 13-millimeter (mm), 0.2-micron (µm) hydrophilic polypropylene syringe filters into 2-mL glass chromatography vials and stored at -20°C until analysis. Standard curve solutions were prepared at concentrations of 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, and 50 µg/L, check standards at 0.2 and 5 µg/L, surrogate internal standards at 5 µg/L, and the spike samples at 5 µg/L.

During instrumental analysis atrazine was separated with a Waters Acquity H-class Bio UPLC and a Waters Acquity BEH C18, 1.7-µm, 50 x 2.1 mm UPLC analytical column and vanguard pre-column. The mobile phases were (A) 0.3 percent acetic acid in water and (B) 0.3 percent acetic acid in acetonitrile/methanol (50/50 percent). Compounds were detected using a Waters Quattro Micro triple quadrupole mass spectrometer with electrospray ionization (ESI) in positive-ion mode. Multiple reaction monitoring (MRM) with at least two transitions was used to monitored for each compound. The quantitation MRM used for atrazine had a molecular ion of 216 and a fragment ion of 174 and confirmation MRM had a molecular ion of 218 and a fragment ion of 176; ions used for atrazine-d5 were 221/179 and 223/181. To meet quality control specifications environmental samples and duplicates did not exceed a 20% difference, the target matrix spike recovery range was within 80-120 percent, and the check standard accuracy was within 80-120 percent. Data were reported with least significant figure plus 1 (LSD+1) to comply with the 2002 Office of Water Quality Policy for uploading data to NWIS (U.S. Geological Survey, 2002).

Continuous Water-Quality Data

Specific conductance was continuously measured (15 minutes to hourly) using a YSI, Inc., EXO multiparameter sonde (YSI, Inc., 2017). The water-quality monitor was operated and maintained according to standard USGS methods (Wagner and others, 2006; Bennett and others, 2014). All continuous water-quality data at the North Fork Ninnescah River above Cheney Reservoir are available in near-real time (updated hourly) from the NWIS database (https://doi.org/10.5066/F7P55KJN; U.S. Geological Survey, 2022) using the site number 07144780.

Model Development

Stepwise regression analysis was done using R (R Core Team, 2020) to relate discretely collected atrazine concentrations to specific conductance, seasonality, and other continuously measured data. The distribution of residuals was examined for normality and plots of residuals (the difference between the measured and model calculated values) compared to model calculated nitrate plus nitrite were examined for homoscedasticity (departures from zero did not change substantially over the range of model calculated values). Previously published explanatory variables were also strongly considered for continuity.

Censored results (less than the minimum reporting level) made up 50 percent of the model calibration dataset. Tobit regression models were developed using the adjusted maximum likelihood estimation methods using the smwrQW (v0.7.9) package in R programming language (Hald, 1949; Cohen, 1950; Tobin, 1958; Helsel and others, 2020; Lorenz, in press).

Specific conductance and seasonality were selected as a good surrogate for atrazine based on residual plots, a higher pseudocoefficient of determination (pseudo-R^2), and relatively low estimated standard residual error (RSE). Values for the aforementioned statistics were computed and are included below along with all relevant sample data and additional statistical information.

Model Summary

Summary of final Tobit regression analysis for atrazine at USGS site 07144780:

Atrazine concentration-based model:

$$\log_{10}(Atrazine) = (-0.0011 \times SPC) + (0.1864 \times \sin(2\pi D)) - (0.6006 \times \cos(2\pi D)) - 0.4756,$$

where,

Atrazine = atrazine concentration, in micrograms per liter;

SPC = specific conductance, in microsiemens per centimeter at 25 degrees Celsius;

$\log_{10}$ = decimal logarithm; and

D = date in decimal years.

The $\log_{10}$-transformed model may be retransformed to the original units so that atrazine concentrations can be calculated directly. The retransformation introduces a negative bias in the retransformed calculated constituent (Helsel and others, 2020). This bias may be corrected using Duan's bias correction factor (BCF; Duan, 1983; Helsel and others, 2020). For this model, the calculated BCF was 1.50. The retransformed model, accounting for BCF, is as follows:

$$Atrazine = (10^{-0.0011 \times SPC} \times 10^{0.1864 \times \sin(2\pi D)} \times 10^{-0.6006 \times \cos(2\pi D)} \times 10^{-0.4756}) \times 1.50.$$

Previous studies by Christensen and Ziegler (1998), Ziegler and others (1999), and Rasmussen and others (2016) observed inverse relationships between atrazine and chloride, thus indicating an inverse relation to specific conductance as well. Additionally, the largest concentrations of atrazine were observed in the spring and summer when atrazine is most likely applied to crops and when higher rainfall amounts are common, indicating a seasonal pattern. In this atrazine model, a seasonal component (periodic function including the day of year) was included in the final linear regression equation in addition to specific conductance as explanatory variables (table 3; Helsel and others, 2020).

Extrapolation, defined as computation beyond the range of the model calibration dataset, may be used to extrapolate no more than 10 percent outside the range of the calibration data used to fit the model and is therefore limited. The extrapolation limit for atrazine concentration using this model is 3.85 μg/L. Computed estimates outside that limit are not supported by the current model calibration dataset.

Model statistics, data, and plots

Definitions

Variable	Explanation
Cook's D	Cook's distance, a measure of influence (Helsel and others, 2020)
Leverage	An outlier's measure in the x direction (Helsel and others, 2020)
Atrazine	Atrazine, in micrograms per liter (μg/L) (USGS parameter code 69362)
p-value	The probability that the independent variable has no effect on the dependent variable (Helsel and others, 2020)
pseudo-R^2	Pseudocoefficient of determination. An estimation of the proportion of variance in the response variable explained by the model (McKelvey and Zavoina, 1975)
SPC	Specific conductance, in microsiemens per centimeter at 25 degrees Celsius (μS/cm at 25°C) (USGS parameter code 00095)
z-score	The estimated coefficient divided by its associated standard error (Helsel and others, 2020)

Model Information

```
Atrazine = -0.0011 * SPC + 0.1864 * sin(2piD) – 0.6006 * cos(2piD) – 0.4756

Computation Method: Adjusted Maximum Likelihood Estimation (AMLE)
```

Variable Summary Statistics

	Atrazine	SPC	sin2piD	cos2piD
Minimum	<0.1	207	-0.9899	-0.9955
1st Quartile	0.05	439	-0.3126	-0.9282
Median	0.07	1081	0.1924	-0.5558
Mean	0.28	887	0.1527	-0.3452
3rd Quartile	0.24	1170	0.8065	0.1029
Maximum	3.50	1402	0.9999	0.9787
Standard Deviation	0.64	426	0.6743	0.6661

Explanatory Variables

```
Coefficients:
             Estimate Standard Error z-score p-value
```

```
(Intercept) -0.475584  0.4164373     -1.1420  0.1199
SPC         -0.001073  0.0004062     -2.6414  0.0022
sin2piD      0.186402  0.2374299      0.7851  0.3699
cos2piD     -0.600618  0.4602108     -1.3051  0.0269
```

Basic Model Statistics

```
Estimated residual standard error (Unbiased)          0.5484
Number of observations                                22
Number censored                                       11 (50 percent)
Log-likelihood (model)                                -10.48
Log-likelihood (intercept only)                       -18.82
      Chi-square                16.67
      degrees of freedom        3
      p-value                   0.0008

Pseudo R-squared                                      0.6809
Akaike Information Criterion                          30.97
Bayesian Information Criterion                        36.42
Variance inflation factors
    SPC          1.44
    sin2piD      1.16
    cos2piD      1.26
```

Outlier Test Criteria

```
leverage   cooksD
  0.4091   0.8699
```

Flagged Observations

```
    No observations exceeded any test criteria
```

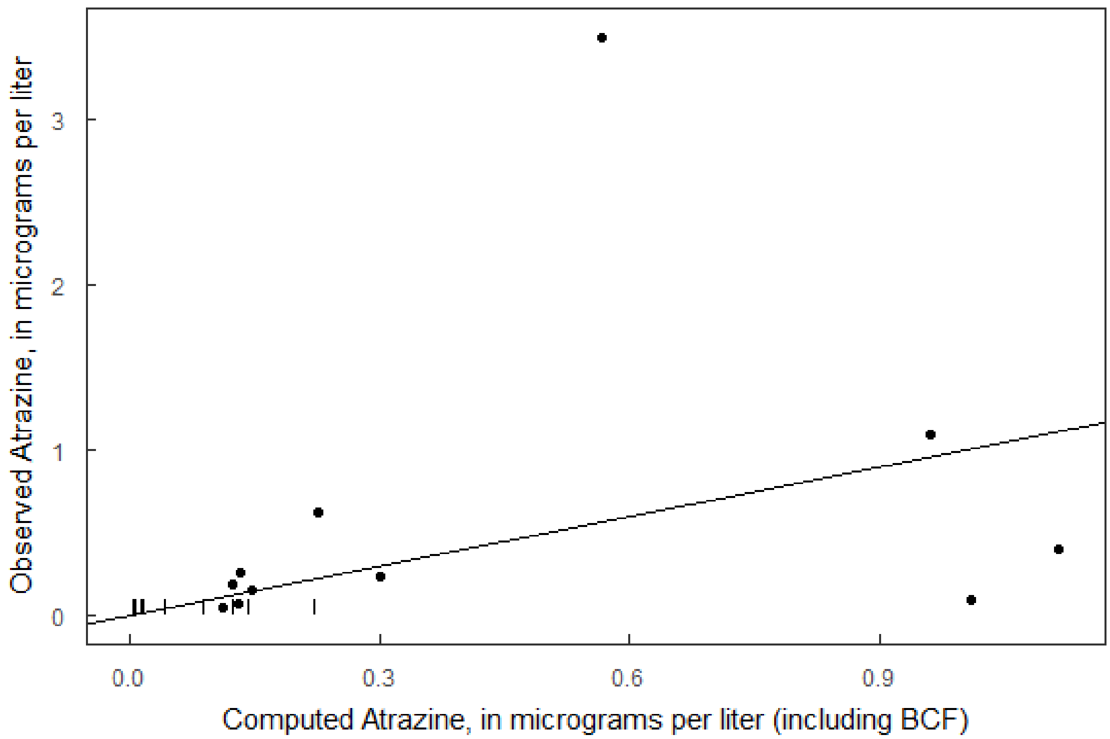
Bias correction factor

```
1.495246
```

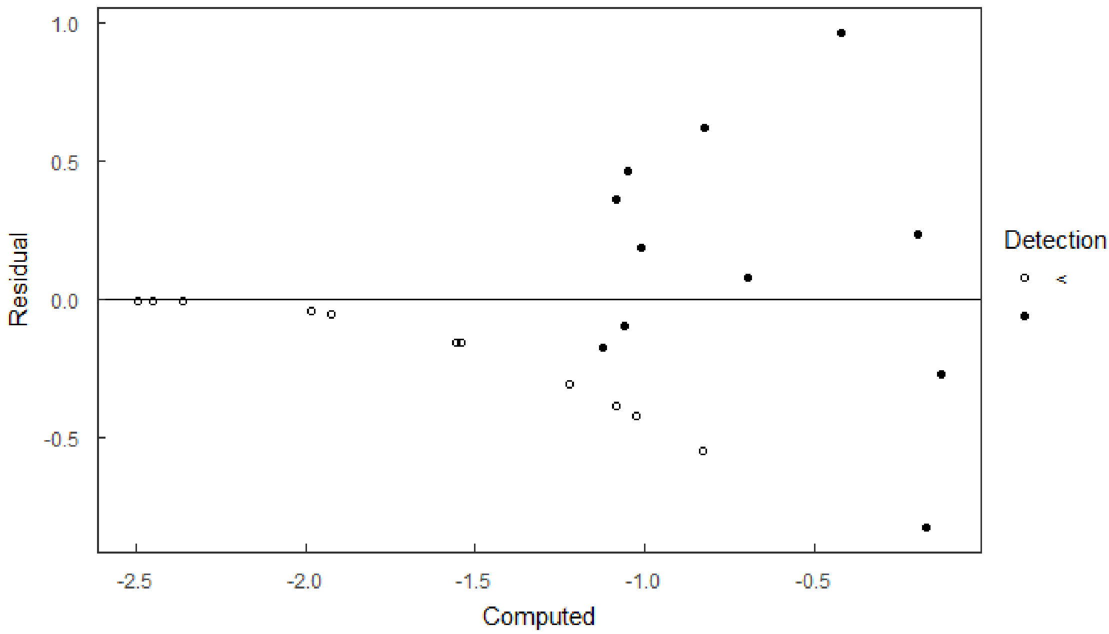
95% Confidence Intervals

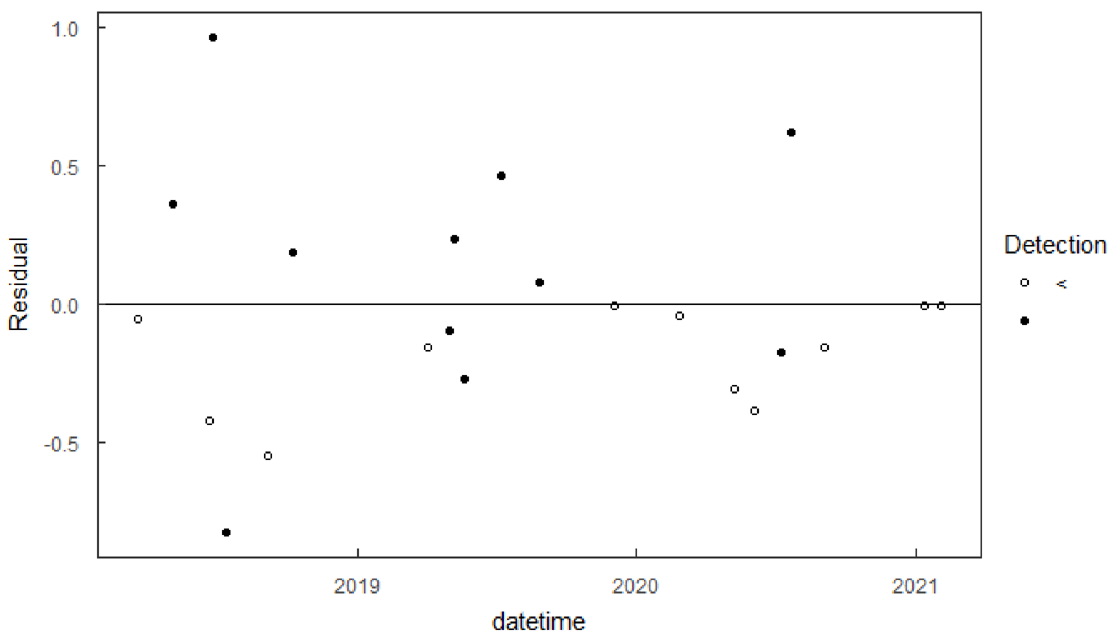
```
                  2.5 %        97.5 %
(Intercept) -1.291786579  0.3406176785
SPC         -0.001868932 -0.0002767673
sin2piD     -0.278952580  0.6517556737
cos2piD     -1.502614833  0.3013783063
```

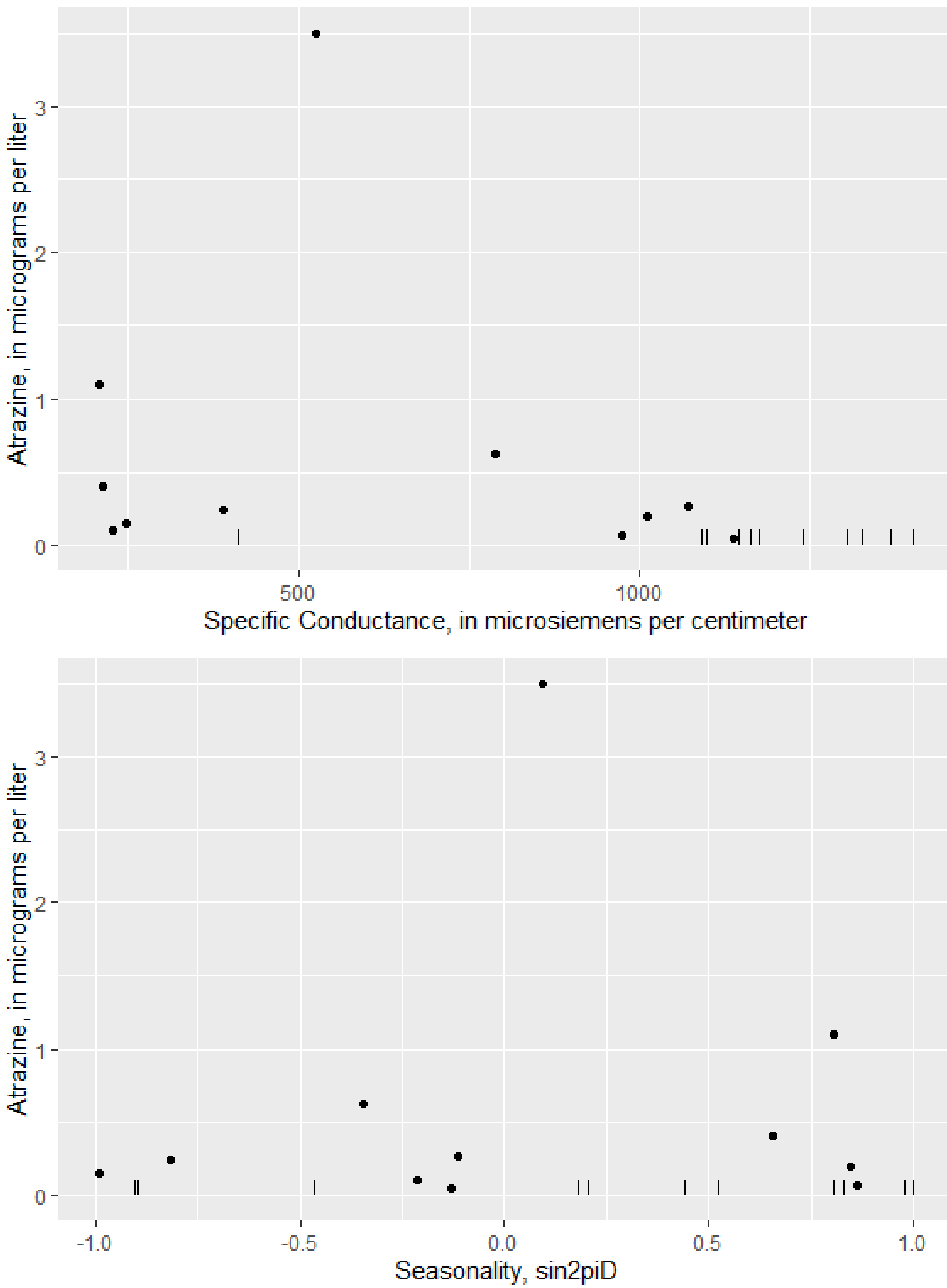
Plots



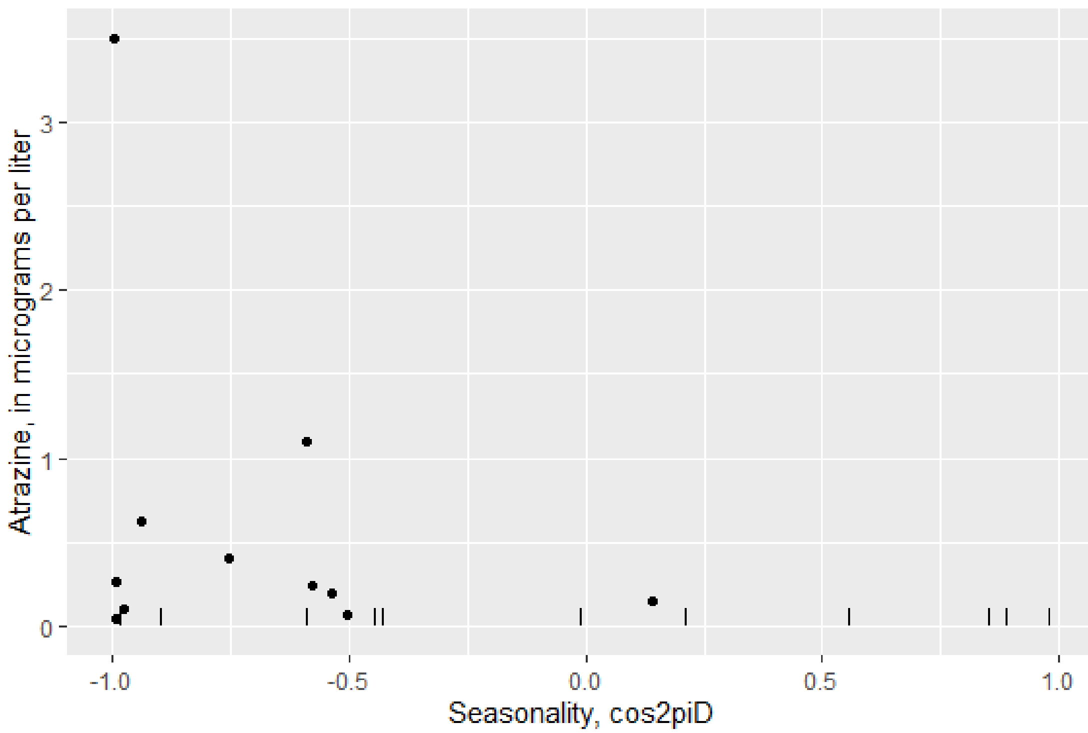
The black vertical lines correspond to the censored results in the model calibration dataset as they are distributed in the model computations. The black dots correspond to observations. The trend line represents the 1:1 line.


1.0
0.5
0.0
-0.5
Residual
-2.5
-2.0
-1.5
-1.0
-0.5
Computed
Detection
<


1.0
0.5
0.0
-0.5
Residual
2019
2020
2021
datetime
Detection
<



The black vertical lines correspond to the censored results in the model calibration dataset as they are distributed in the model computations.



The black vertical lines correspond to the censored results in the model calibration dataset as they are distributed in the model computations.

Model Calibration Dataset

	datetime	logAtrazine	Atrazine	SPC	Computed_logAtrazine	Computed Atrazine	sin2piD	cos2piD
1	2018-03-20 10:30:00	<-1	<0.1	1402	-1.923	0.01785	0.9778	0.2093
2	2018-05-04 10:00:00	-0.721	0.19	1013	-1.084	0.12331	0.8452	-0.5344
3	2018-06-21 10:10:00	<-1	<0.1	1092	-1.023	0.14194	0.1798	-0.9837
4	2018-06-26 13:20:00	0.544	3.5	524	-0.422	0.14194	0.1798	-0.9837
5	2018-07-14 12:00:00	-1	0.1	225	-0.171	1.00960	-0.2135	-0.9769
6	2018-09-05 09:55:00	<-1	<0.1	411	-0.828	0.22232	-0.9034	-0.4289
7	2018-10-09 10:10:00	-0.824	0.15	246	-1.009	0.14649	-0.9899	0.1415
8	2019-04-02 10:50:00	<-1	<0.1	1175	-1.542	0.04290	0.9999	-0.0129
9	2019-05-02 11:20:00	-1.15	0.07	975	-1.058	0.13085	0.8631	-0.5050
10	2019-05-08 12:00:00	0.0414	1.1	207	-0.192	0.96044	0.8065	-0.5913
11	2019-05-21 12:30:00	-0.398	0.4	212	-0.128	1.11430	0.6552	-0.7555
12	2019-07-08 11:30:00	-0.585	0.26	1070	-1.048	0.13382	-0.1117	-0.9937
13	2019-08-26 11:30:00	-0.62	0.24	387	-0.696	0.30089	-0.8165	-0.5773
14	2019-12-03 10:20:00	<-1	<0.1	1305	-2.494	0.00479	-0.4636	0.8861
15	2020-02-26 10:30:00	<-1	<0.1	1240	-1.985	0.01547	0.8312	0.5560
16	2020-05-07 10:30:00	<-1	<0.1	1164	-1.219	0.09028	0.8065	-0.5913
17	2020-06-04 10:20:00	<-1	<0.1	1146	-1.084	0.12335	0.4405	-0.8977
18	2020-07-08 11:00:00	-1.3	0.05	1137	-1.124	0.11236	-0.1287	-0.9917
19	2020-07-21 10:10:00	-0.201	0.63	788	-0.822	0.22550	-0.3456	-0.9384
20	2020-09-03 10:20:00	<-1	<0.1	1099	-1.555	0.04170	-0.8958	-0.4444
21	2021-01-12 10:10:00	<-1	<0.1	1327	-2.449	0.00532	0.2051	0.9787
22	2021-02-01 11:00:00	<-1	<0.1	1370	-2.360	0.00653	0.5234	0.8521

References Cited

Bennett, T.J., Graham, J.L., Foster, G.M., Stone, M.L., Juracek, K.E., Rasmussen, T.J., and Putnam, J.E., 2014, U.S. Geological Survey quality-assurance plan for continuous water-quality monitoring in Kansas, 2014: U.S. Geological Survey Open-File Report 2014–1151, 34 p. plus appendixes, accessed September 7, 2022, at https://doi.org/10.3133/ofr20141151.

Christensen, V.G., and Ziegler, A.C., 1998, Atrazine in source water intended for artificial ground-water recharge, south-central Kansas: U.S. Geological Survey Fact Sheet 074–98, 4 p.

Cohen, A.C., Jr., 1950, Estimating the mean and variance of normal populations from singly and doubly truncated samples: Annals of Mathematical Statistics, v. 21, no. 4, p. 557–569, accessed October 2019 at https://doi.org/10.1214/aoms/1177729751.

Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: Journal of the American Statistical Association, v. 78, n. 383 p. 605–610.

Hald, A., 1949, Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point: Scandinavian Actuarial Journal, v. 1949, no. 1, p. 119–134. [Also available at https://doi.org/10.1080/03461238.1949.10419767.]

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J., 2020, Statistical methods in water resources: U.S. Geological Survey Techniques and Methods, book 4, chap. A3, 458 p. [Also available at https://doi.org/10.3133/tm4A3.] [Supersedes USGS Techniques of Water-Resources Investigations, book 4, chap. A3, version 1.1.]

Lorenz, D.L., [in press], smwrQW—An R package for managing and analyzing water-quality data, version 0.7.9: R Foundation for Statistical Computing software release.

R Core Team, 2020, R—A language and environment for statistical computing: R Foundation for Statistical Computing software release (version 4.0.2), accessed September 7, 2022, at https://www.R-project.org/.

McKelvey, R.D., and Zavoina, W., 1975, A statistical model for the analysis of ordinal level dependent variables: The Journal of Mathematical Sociology, v. 4, no. 1, p. 103–120. [Also available at https://doi.org/10.1080/0022250X.1975.9989847.]

Rasmussen, P.P., Gray, J.R., Glysson, G.D., and Ziegler, A.C., 2009, Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data: U.S. Geological Survey Techniques and Methods, book 3, chap. C4, 52 p. [Also available at https://doi.org/10.3133/tm3C4.]

Rasmussen, P.P., Eslick, P.J., and Ziegler, A.C., 2016, Relations between continuous real-time physical properties and discrete water-quality constituents in the Little Arkansas River, south-central Kansas, 1998-2014: U.S. Geological Survey Open-File Report 2016–1057, 20 p., https://dx.doi.org/10.3133/ofr20161057.

Tobin, J., 1958, Estimation of relationships for limited dependent variables: Econometrica, v. 26, no. 1, p. 24– 36. [Also available at https://doi.org/10.2307/1907382.]

U.S. Geological Survey, 2002, Policy for Storing and Reporting Significant Figures for Chemical Data: U.S. Geological Survey Office of Surface Water Technical Memorandum 2002.11, Office of Water Quality Technical Memorandum 2002.11, accessed September 30, 2022, at https://water.usgs.gov/admin/memo/QW/qw02.11.html.

U.S. Geological Survey, 2006, Collection of water samples (ver. 2.0, September 2006): U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A4 [variously paged]. [Also available at https://doi.org/10.3133/twri09A4.]

U.S. Geological Survey, 2016, Policy and guidance for approval of surrogate regression models for computation of time series suspended-sediment concentration and loads: U.S. Geological Survey Office of Surface Water Technical Memorandum 2016.07, Office of Water Quality Technical Memorandum 2016.10, 40 p., accessed September 7, 2022, at https://water.usgs.gov/water-resources/memos/memo.php?id=467.

U.S. Geological Survey, 2022, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed September 7, 2022, at https://doi.org/10.5066/F7P55KJN.

Wagner, R.J., Boulger, R.W., Jr., Oblinger, C.J., and Smith, B.A., 2006, Guidelines and standard procedures for continuous water-quality monitors—Station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods, book 1, chap. D3, 51 p. plus 8 attachments. [Also available at https://doi.org/10.3133/tm1D3.]

YSI, Inc., 2017, EXO user manual—Advanced water quality monitoring platform (rev. G): Yellow Springs, Ohio, YSI, Inc., 154 p., accessed September 7, 2022, at https://www.ysi.com/file%20library/documents/manuals/exo-user-manual-web.pdf.

Ziegler, A.C., Christensen, V.G., and Ross, H.C., 1999, Baseline water-quality and preliminary effects of artificial recharge on groundwater, south-central Kansas, 1995-98: U.S. Geological Survey Water Resources Investigations Report 99–4250, 74 p. [Also available at https://doi.org/10.3133/wri994250.]