Appendix 4. Model Archive Summary for Orthophosphate 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 orthophosphate concentration model developed to compute 15-minute, hourly, or daily orthophosphate 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).

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: April 19, 2016, through August 12, 2021 (dataset consisted of 33 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 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 33 concomitant values of discretely collected orthophosphate concentrations and continuously measured specific conductance during April 19, 2016, through August 12, 2021. Discrete samples were collected throughout the range of continuously observed hydrologic conditions. Orthophosphate concentrations were less than the minimum reporting level (less than [<] 0.04 milligram per liter [mg/L]) for 14 samples (42.4 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. Five samples in the model calibration dataset were flagged as outliers but all were retained in the dataset after further review.

Orthophosphate Sampling 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 5–12 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 one to seven times per year. Samples were analyzed for orthophosphate by the Wichita Municipal Water and Wastewater Laboratory in Wichita, Kans., according to standard methods (Eaton and others, 1995).

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 USGS National Water Information System 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 programming language (R Core Team, 2020) to relate discretely collected orthophosphate concentrations to specific conductance 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 also were strongly considered for continuity.

Censored results (less than the minimum reporting level) made up about 42 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 was selected as a good surrogate for orthophosphate 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 orthophosphate at USGS site 07144780:

Orthophosphate concentration-based model:

$$\log_{10}(OrthoP) = -0.00095 \times SPC - 0.321,$$

where,

OrthoP = orthophosphate concentration, in milligrams per liter as phosphorus;

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

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

The $\log_{10}$-transformed model may be retransformed to the original units so that orthophosphate 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 1.12. The retransformed model, accounting for BCF, is as follows:

$$OrthoP = (10^{-0.00095 \times SPC} \times 10^{-0.321}) \times 1.12.$$

Extrapolation, defined as computation beyond the range of the model calibration dataset, may be 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 orthophosphate concentration using this model is

0.363 milligrams per liter. 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)
OrthoP	Orthophosphate, in milligrams per liter as phosphorus (USGS parameter code 00671)
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

$$\log_{10}(OrthoP) = -0.00095 \times SPC - 0.321$$

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

Variable Summary Statistics

	OrthoP	SPC
Minimum	<0.04	207
1st Quartile	<0.04	576
Median	0.07	1013
Mean	0.07	892
3rd Quartile	0.15	1164
Maximum	0.33	1433
Standard Deviation	0.13	382

Explanatory Variables

```
Coefficients:
              Estimate Standard Error z-score p-value
(Intercept) -0.3209049  0.1056694      -3.037  0.0092
SPC         -0.0009525  0.0001199      -7.943  0.0000
```

Basic Model Statistics

```
Estimated residual standard error (Unbiased)     0.2247
Number of observations                           33
```

```
Number censored                                 14 (42.4 percent)
Log-likelihood (model)                          -6.595
Log-likelihood (intercept only)                 -25.61
      Chi-square         38.03
      Degrees of freedom       1
      p-value                 <0.0001

Pseudo R-squared                                0.7299
Akaike Information Criterion                    19.19
Bayesian Information Criterion                  23.68
```

Outlier Test Criteria

```
leverage   cooksD
 0.09091  0.70838
```

Flagged Observations

```
    Observations exceeding at least one test criterion
   logOrthoP  ycen    yhat   resids leverage   cooksD
1    -1.3979  TRUE -1.6853 -0.04396  0.09308 0.002166
11   -0.4815 FALSE -0.5357  0.05419  0.12544 0.004770
13   -0.5086 FALSE -0.5551  0.04642  0.11971 0.003297
16   -0.6383 FALSE -0.5181 -0.12020  0.13079 0.024780
17   -0.7212 FALSE -0.5233 -0.19794  0.12918 0.066122
```

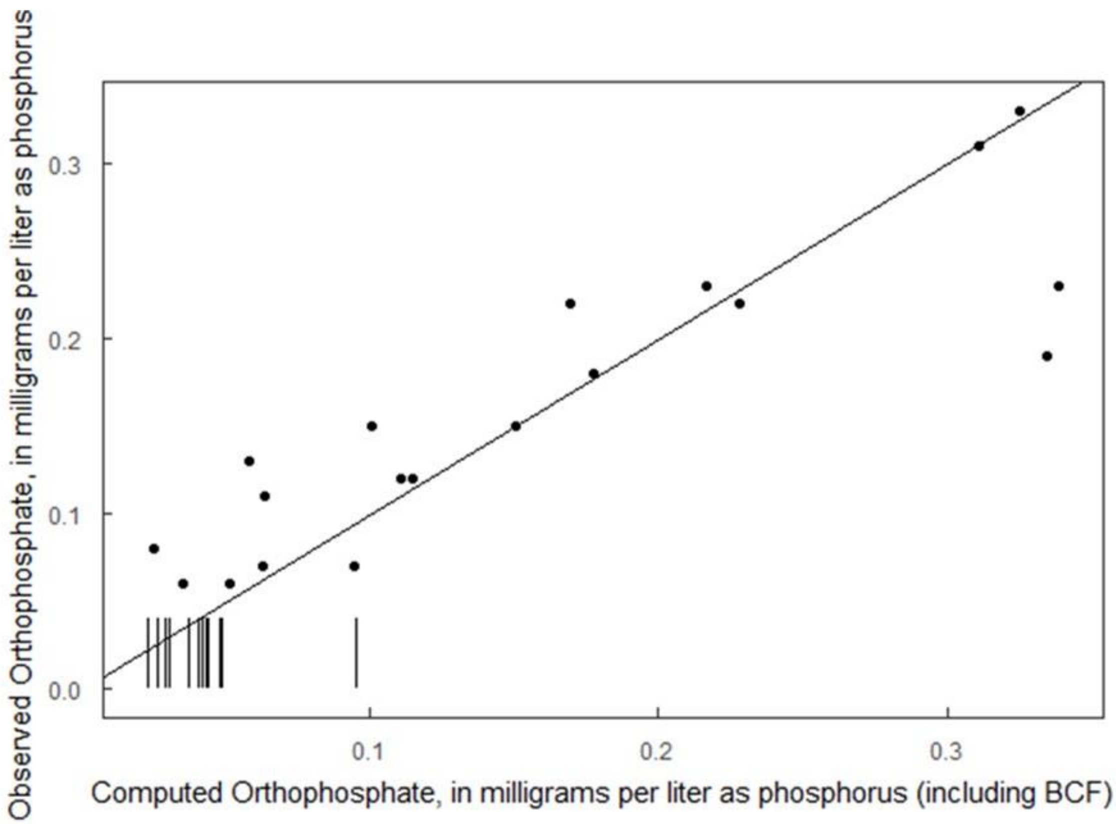
Bias correction factor

```
1.116237
```

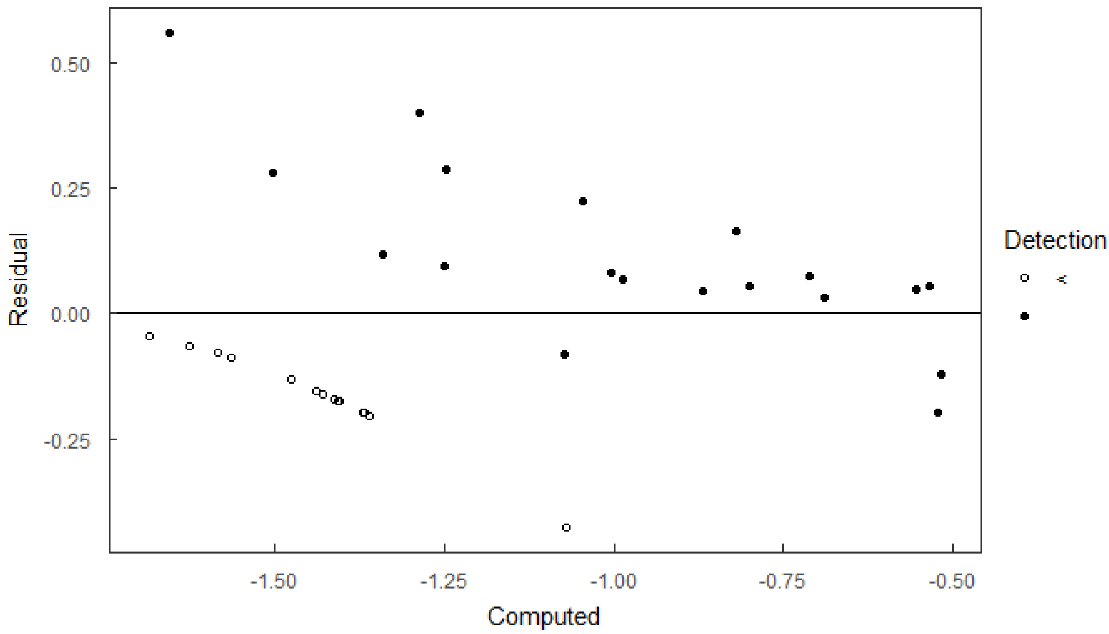
95% Confidence Intervals

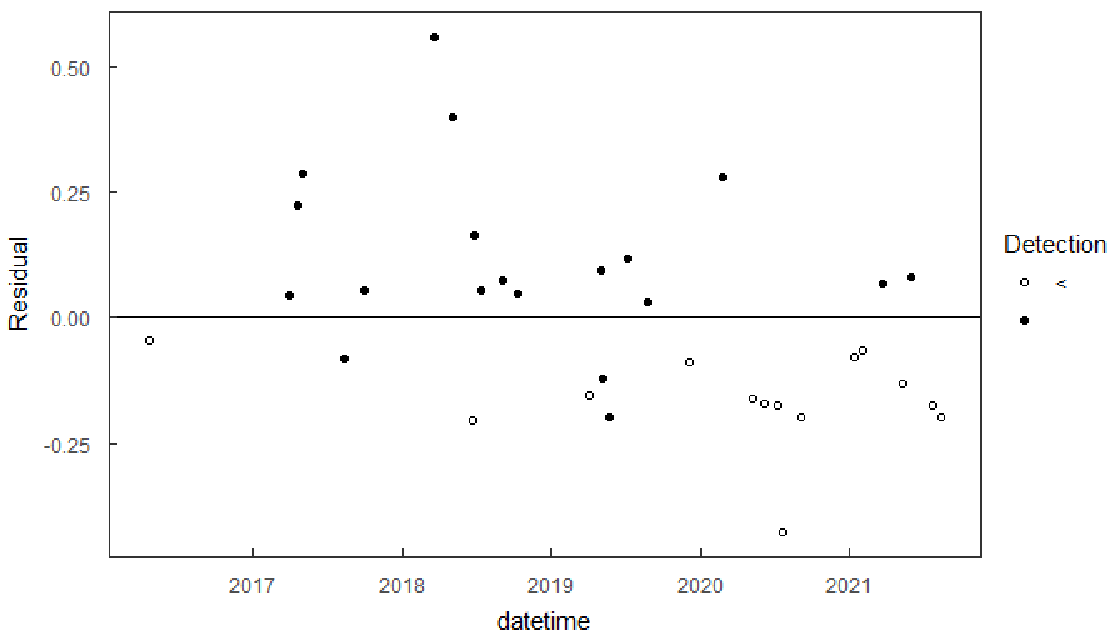
```
                   2.5 %        97.5 %
(Intercept) -0.528013052 -0.1137966496
SPC         -0.001187498 -0.0007174526
```

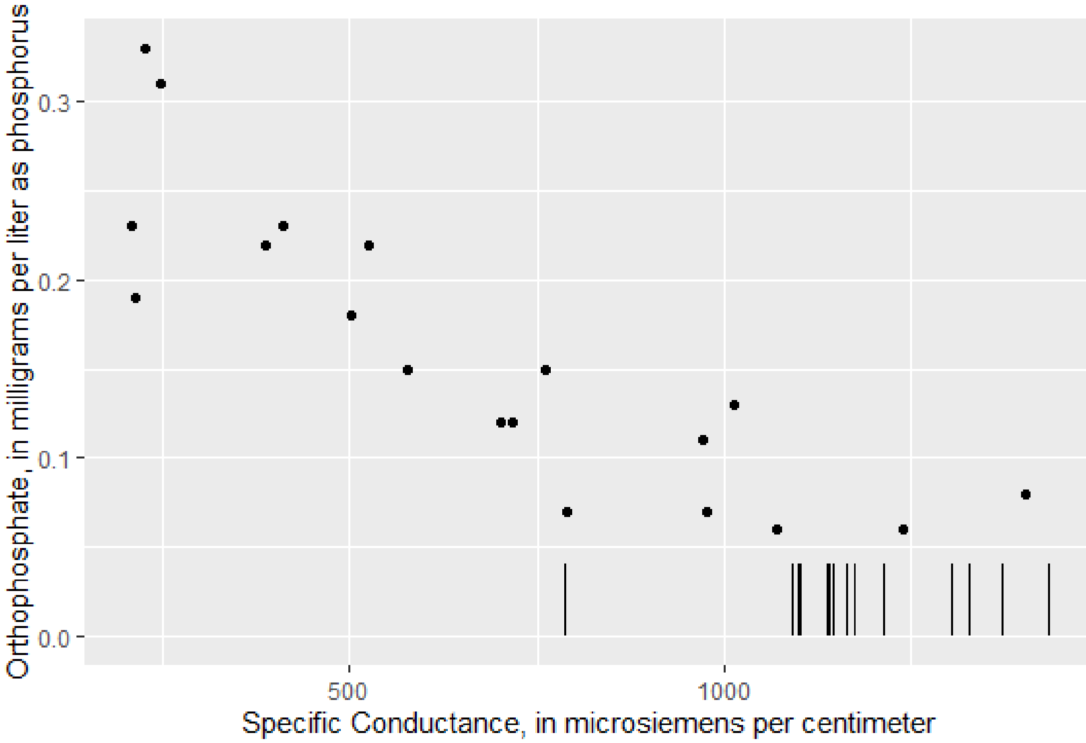
Plots



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 represent observations. The black line represents the 1:1 line.


Residual
0.50
0.25
0.00
-0.25
-1.50
-1.25
-1.00
-0.75
-0.50
Computed
Detection
<


Residual
0.50
0.25
0.00
-0.25
2017
2018
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.

Model Calibration Dataset

	datetime	logOrthoP	OrthoP	SPC	Computed_logOrthoP	Computed_OrthoP
1	2016-04-19 10:25:00	<-1.4	<0.04	1432	-1.685	0.0230
2	2017-03-29 10:45:00	-0.824	0.15	576	-0.870	0.1507
3	2017-04-20 12:00:00	-0.824	0.15	762	-1.047	0.1002
4	2017-05-02 09:50:00	-0.959	0.11	972	-1.246	0.0633
5	2017-08-11 11:00:00	-1.15	0.07	789	-1.072	0.0945
6	2017-09-28 10:30:00	-0.745	0.18	502	-0.799	0.1772
7	2018-03-20 10:30:00	-1.1	0.08	1402	-1.656	0.0246
8	2018-05-04 10:00:00	-0.886	0.13	1013	-1.286	0.0578
9	2018-06-21 10:10:00	<-1.4	<0.04	1092	-1.361	0.0486
10	2018-06-26 13:20:00	-0.658	0.22	524	-0.820	0.1691
11	2018-07-14 12:00:00	-0.481	0.33	225	-0.536	0.3251
12	2018-09-05 09:55:00	-0.638	0.23	411	-0.713	0.2163
13	2018-10-09 10:10:00	-0.509	0.31	246	-0.555	0.3110
14	2019-04-02 10:50:00	<-1.4	<0.04	1175	-1.440	0.0405
15	2019-05-02 11:20:00	-1.15	0.07	975	-1.250	0.0628
16	2019-05-08 12:00:00	-0.638	0.23	207	-0.518	0.3386
17	2019-05-21 12:30:00	-0.721	0.19	212	-0.523	0.3345
18	2019-07-08 11:30:00	-1.22	0.06	1070	-1.341	0.0510
19	2019-08-26 11:30:00	-0.658	0.22	387	-0.690	0.2282
20	2019-12-03 10:20:00	<-1.4	<0.04	1305	-1.564	0.0305
21	2020-02-26 10:30:00	-1.22	0.06	1240	-1.502	0.0351
22	2020-05-07 10:30:00	<-1.4	<0.04	1164	-1.430	0.0415
23	2020-06-04 10:20:00	<-1.4	<0.04	1146	-1.412	0.0432
24	2020-07-08 11:00:00	<-1.4	<0.04	1137	-1.404	0.0440
25	2020-07-21 10:10:00	<-1.4	<0.04	788	-1.071	0.0948
26	2020-09-03 10:20:00	<-1.4	<0.04	1099	-1.367	0.0479
27	2021-01-12 10:10:00	<-1.4	<0.04	1327	-1.585	0.0290
28	2021-02-01 11:00:00	<-1.4	<0.04	1370	-1.626	0.0264
29	2021-03-23 11:40:00	-0.921	0.12	700	-0.988	0.1148
30	2021-05-10 10:50:00	<-1.4	<0.04	1213	-1.476	0.0373
31	2021-06-01 10:40:00	-0.921	0.12	716	-1.003	0.1108
32	2021-07-22 10:40:00	<-1.4	<0.04	1139	-1.406	0.0438
33	2021-08-12 11:00:00	<-1.4	<0.04	1102	-1.371	0.0476

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.

Cohen, A.C., Jr., 1950, Estimating the mean and variance of normal populations from singly truncated 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. [Also available at https://doi.org/10.1080/01621459.1983.10478017.]

Eaton, A.D., Clesceri, L.S., and Greenberg, A.E., eds., 1995, Standard methods for the examination of water and wastewater (19th ed.): New York, American Public Health Association, 905 p.

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.]

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, 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.