Appendix 15. Model Archive Summary for Nitrate Plus Nitrite Concentration at U.S. Geological Survey Station 07144790, Cheney Reservoir near Cheney, Kansas, during October 1, 2014, through September 30, 2021

This model archive summary summarizes the nitrate plus nitrite (NO_3NO_2) concentration model developed to compute 15-minute, hourly, or daily nitrate plus nitrite concentrations during October 1, 2014, 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: 07144790

Site name: Cheney Reservoir near Cheney, Kansas

Location: Lat 37°43'34", long 97°47'38" referenced to North American Datum of 1927, in SE 1/4 NE 1/4 NW 1/4 sec.06, T.27 S., R.4 W., Sedgwick County, Kans., hydrologic unit 11030014, in control house structure at outlet works of Cheney Dam on North Fork Ninnescah River, 6.0 mi north of Cheney, and at mile 15.9.

Equipment: A YSI, Inc., EXO water-quality monitor (YSI, Inc., 2017) equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, turbidity, chlorophyll fluorescence, and phycocyanin fluorescence was installed November 14, 2015. The EXO monitor is suspended from the dam intake tower walkway. The monitor is at a depth that fluctuates between three to six feet depending on reservoir elevation. Measurements from the EXO were recorded every 15 minutes to hourly and transmitted hourly via satellite. Reservoir elevation was measured using a Design Analysis H–350 nonsubmersible pressure transducer and H–355 gas system.

Date model was created: August 9, 2022

Model calibration data period: February 7, 2016, through August 31, 2021 (dataset consisted of 45 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, chlorophyll fluorescence, phycocyanin fluorescence, seasonality (sine and cosine variables), and reservoir elevation.

The regression model is based on 45 concomitant values of discretely collected nitrate plus nitrite concentration and continuously measured specific conductance and temperature during February 7, 2016, through August 31, 2021. Discrete samples were collected throughout the range of continuously observed hydrologic conditions. Nitrate plus nitrate concentrations were less than minimum reporting level (less than [<] 0.02 milligram per liter) in 15 samples (33.3 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. Six samples in the model calibration dataset were flagged as outliers but all were retained in the dataset after further review.

Nitrate plus Nitrite Sampling Details

Discrete water-quality samples were collected primarily by depth-integrating through the photic-zone (depth at which light is approximately 1 percent of that at the surface) using a double check-valve bailer (Lane and others, 2003). Vertical water-quality profiles collected during sampling indicated that thermal stratification rarely occurs, and water-quality conditions are typically uniform throughout the water column. Samples were collected from the walkway on the dam intake tower. Discrete samples were collected on a semifixed to event-based schedule seven to eight times per year. All samples were collected between 9:15 a.m. and 12:20 p.m. Samples were analyzed for nitrate plus nitrite 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 and temperature were 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 Cheney Reservoir near Cheney, Kans. 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 07144790.

Model Development

Stepwise regression analysis was done using the R programming language (R Core Team, 2020) to relate discretely collected nitrate plus nitrite concentrations to specific conductance, temperature, 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) represented 33 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 temperature were selected as good surrogates for nitrate plus nitrite 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 nitrate plus nitrite (NO_3NO_2) at USGS site 07144790:

NO_3NO_2 concentration-based model:

$$\log_{10}(NO3NO2) = -0.004 \times SPC - 0.043 \times TEMP + 2.936,$$

where,

NO3NO2 = nitrate plus nitrite concentration, in milligrams per liter as nitrogen;

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

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

TEMP = water temperature, in degrees Celsius.

The $\log_{10}$-transformed model may be retransformed to the original units so that nitrate plus nitrite 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:

$$NO3NO2 = (10^{-0.004 \times SPC} \times 10^{-0.043 \times TEMP} \times 10^{2.936}) \times 1.50.$$

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 nitrate plus nitrite concentration using this model is 0.847 milligram per liter as nitrogen. 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)
NO3NO2	Inorganic nitrogen (nitrate plus nitrite), in milligrams per liter as nitrogen (USGS parameter code 00631)
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)
TEMP	Temperature, water, in degrees Celsius (USGS parameter code 00010)
z-score	The estimated coefficient divided by its associated standard error (Helsel and others, 2020)

Model Information

$\log_{10}(NO_3NO_2) = -0.004 \times SPC - 0.043 \times TEMP + 2.936$

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

Variable Summary Statistics

	NO3NO2	SPC	TEMP
Minimum	<0.02	568	0.38
1st Quartile	<0.02	767	15.70
Median	0.08	813	24.18
Mean	0.09	807	19.07
3rd Quartile	0.21	857	25.91
Maximum	0.77	988	28.07
Standard Deviation	0.23	90	8.89

Explanatory Variables

```
Coefficients:
             Estimate Std. Error z-score p-value
```

```
(Intercept)  2.936000  0.6928281   4.238       0
SPC         -0.004151  0.0008079  -5.137       0
TEMP        -0.043129  0.0079945  -5.395       0
```

Basic Model Statistics

```
Estimated residual standard error (Unbiased)    0.4449
Number of observations                          45
Number censored                                 15 (33.3 percent)
Log-likelihood (model)                          -28.54
Log-likelihood (intercept only)                 -46.07
      Chi-square                                35.06
      degrees of freedom                        2
      p-value                                   <0.0001
Pseudo R-squared                                0.5697
Akaike Information Criterion                     65.07
Bayesian Information Criterion                   72.3
Variance inflation factors
      SPC         1.02
      TEMP        1.02
```

Outlier Test Criteria

```
leverage   cooksD
  0.1333   0.8013
```

Flagged Observations

```
    Observations exceeding at least one test criterion
   logNO3NO2  ycen     yhat    resids leverage    cooksD
1    -1.6990  TRUE -1.23384 -0.694762   0.1371 1.497e-01
15   -0.7959 FALSE -0.95038  0.154502   0.1516 8.470e-03
26   -0.3565 FALSE -0.53759  0.181045   0.1844 1.531e-02
27   -0.6576 FALSE -0.65936  0.001783   0.1492 1.104e-06
28   -0.5850 FALSE -0.65240  0.067377   0.1448 1.514e-03
29   -0.4685 FALSE -0.05761 -0.410908   0.1481 5.803e-02
```

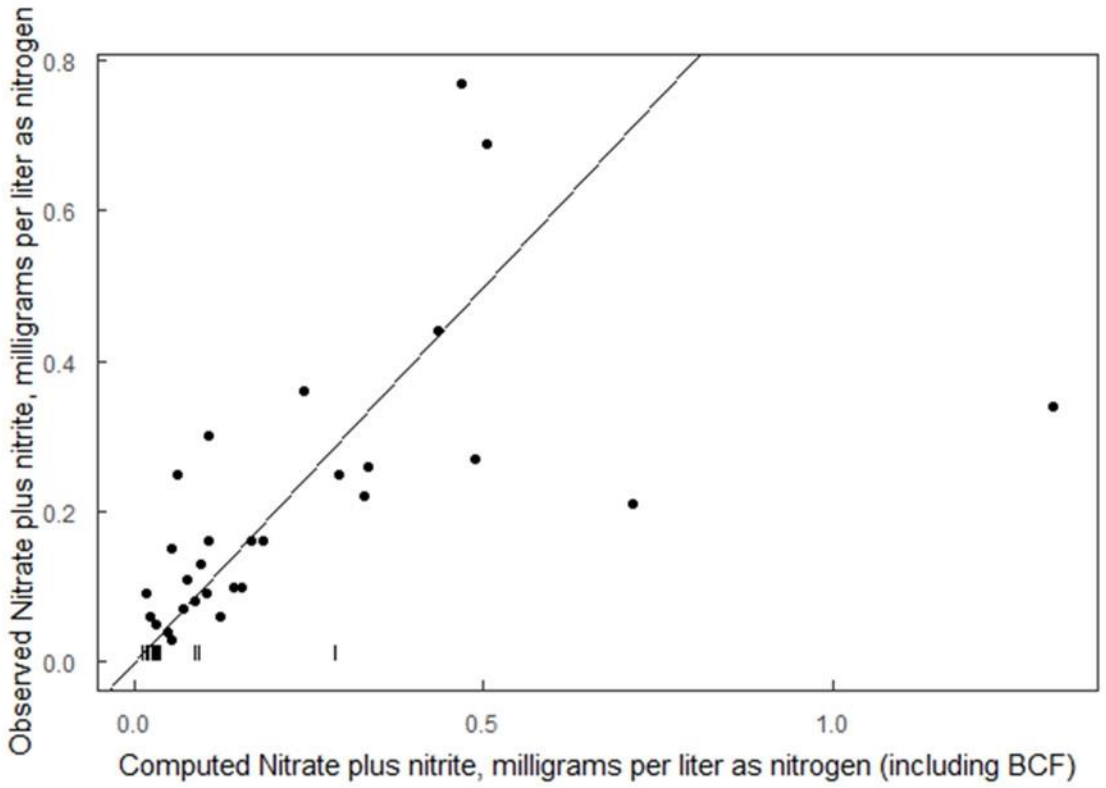
Bias correction factor

```
1.500053
```

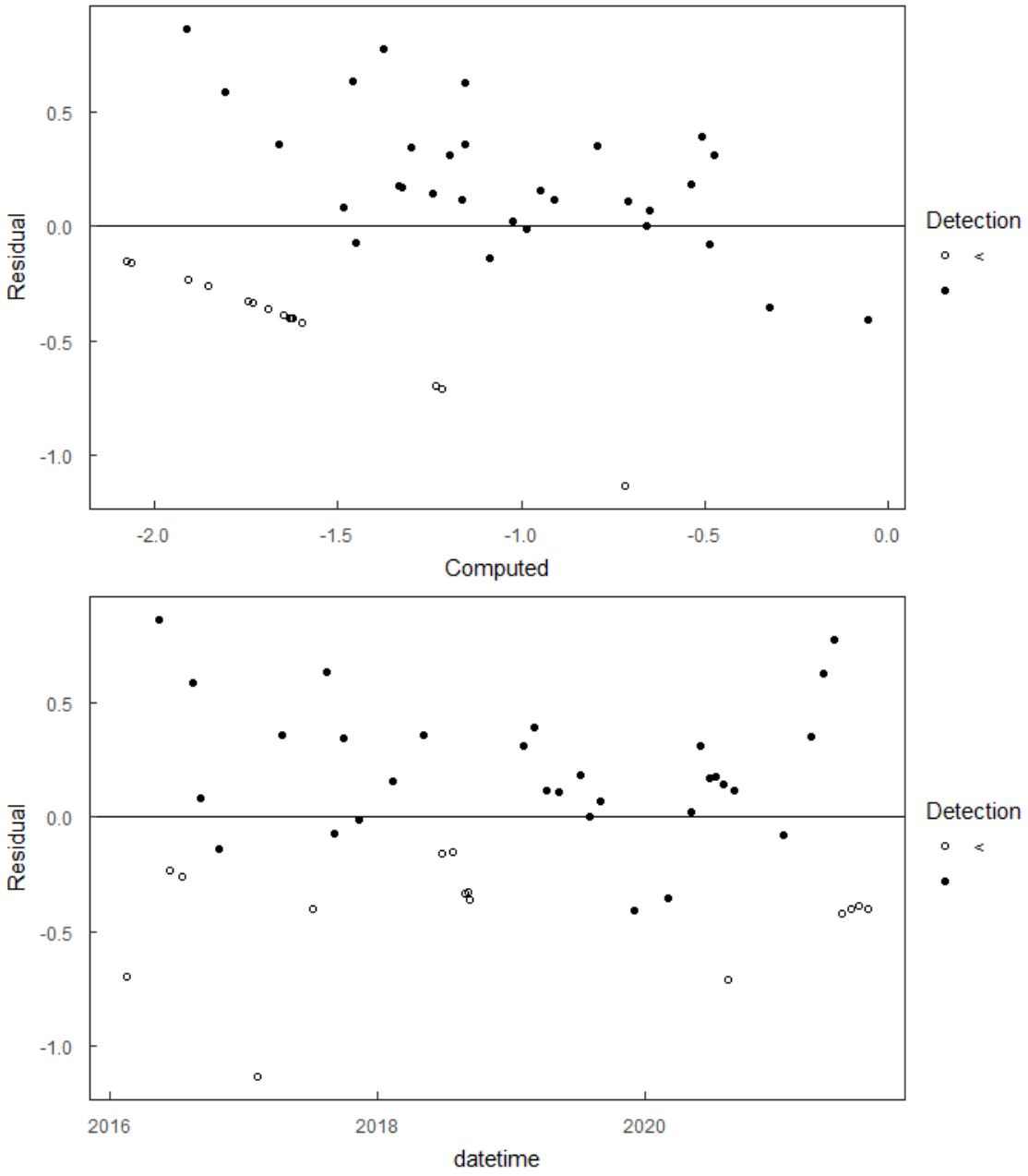
95% Confidence Intervals

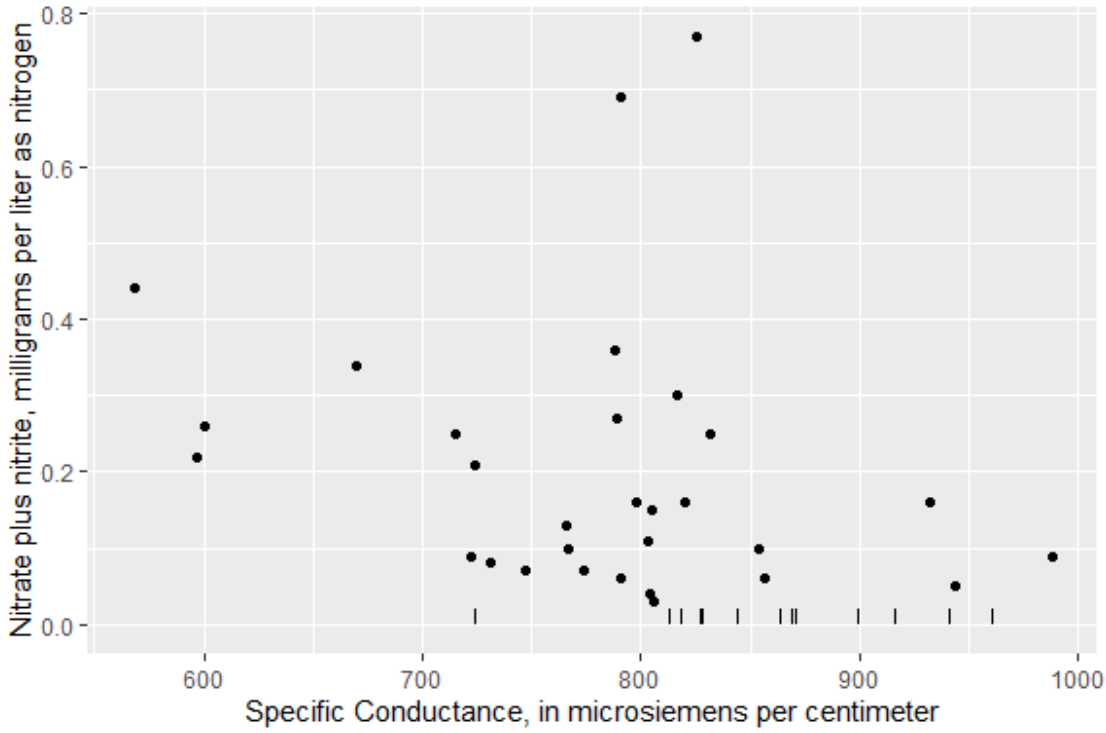
```
                  2.5 %       97.5 %
(Intercept)  1.578081375  4.293917770
SPC         -0.005734115 -0.002567016
TEMP        -0.058797935 -0.027460046
```

Plots

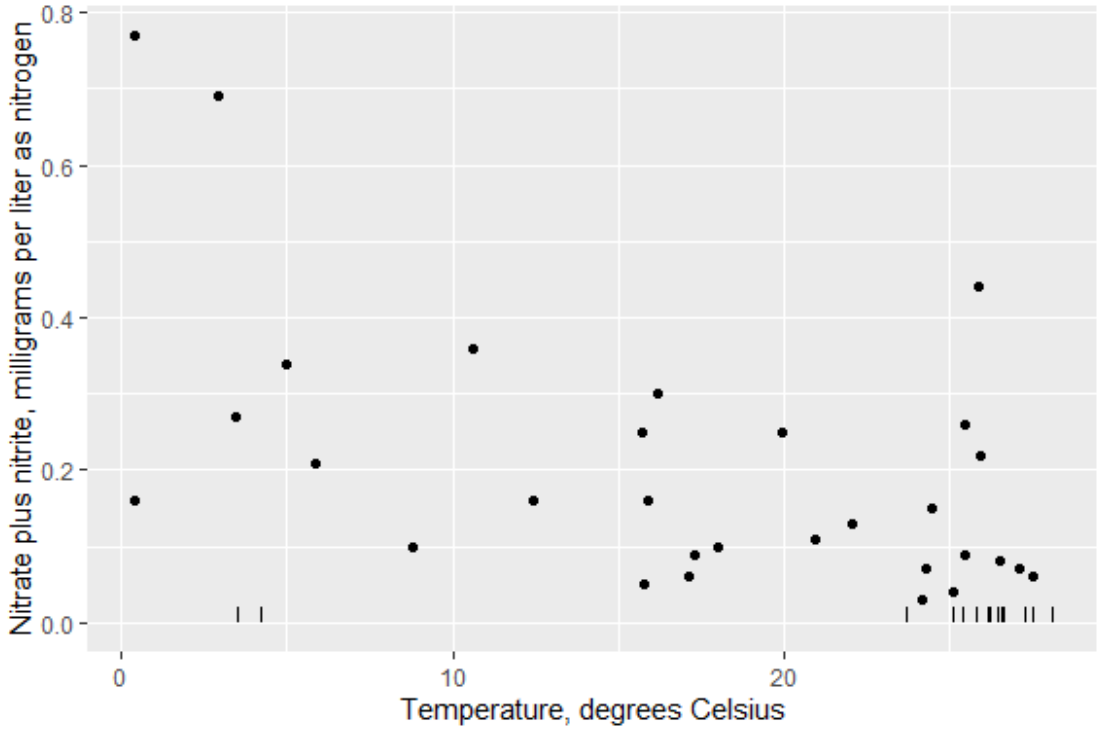


The black vertical lines correspond to the censored results in the model calibration dataset as they are distributed in the model computations. The black dots represent observations. The trend line represents the 1:1 line.


Residual
0.5
0.0
-0.5
-1.0
-2.0
-1.5
-1.0
-0.5
0.0
Computed
Detection
<
Residual
0.5
0.0
-0.5
-1.0
2016
2018
2020
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	logNO3NO2	NO3NO2	SPC	Computed_logNO3NO2	Computed_NO3NO2	TEMP
1	2016-02-17 10:45:00	<-1.7	<0.02	961	-1.2343	0.0875	4.200
2	2016-05-17 10:20:00	-1.05	0.09	988	-1.9113	0.0184	17.300
3	2016-06-15 09:15:00	<-1.7	<0.02	899	-1.9096	0.0185	25.825
4	2016-07-18 10:40:00	<-1.7	<0.02	871	-1.8555	0.0209	27.267
5	2016-08-15 10:30:00	-1.22	0.06	857	-1.8075	0.0234	27.500
6	2016-09-06 10:40:00	-1.4	0.04	804	-1.4839	0.0492	25.100
7	2016-10-25 10:15:00	-1.22	0.06	791	-1.0849	0.1234	17.100
8	2017-02-09 10:40:00	<-1.7	<0.02	844	-0.7198	0.2860	3.500
9	2017-04-17 10:30:00	-0.796	0.16	820	-1.1536	0.1053	15.900
10	2017-07-10 11:40:00	<-1.7	<0.02	813	-1.6271	0.0354	27.520
11	2017-08-15 10:00:00	-0.824	0.15	805	-1.4592	0.0521	24.430
12	2017-09-07 10:00:00	-1.52	0.03	806	-1.4532	0.0528	24.180
13	2017-10-03 10:20:00	-0.959	0.11	803	-1.3016	0.0749	20.930
14	2017-11-13 12:00:00	-1	0.1	854	-0.9876	0.1543	8.780
15	2018-02-13 10:40:00	-0.796	0.16	932	-0.9508	0.1680	0.377
16	2018-05-08 10:30:00	-1.3	0.05	944	-1.6631	0.0326	15.780
17	2018-06-25 12:00:00	<-1.7	<0.02	941	-2.0643	0.0129	25.370
18	2018-07-26 11:40:00	<-1.7	<0.02	916	-2.0763	0.0126	28.070
19	2018-08-29 11:00:00	<-1.7	<0.02	864	-1.7339	0.0277	25.120
20	2018-09-05 12:20:00	<-1.7	<0.02	864	-1.7445	0.0270	25.367
21	2018-09-11 09:40:00	<-1.7	<0.02	869	-1.6921	0.0305	23.670
22	2019-02-05 11:20:00	-0.161	0.69	791	-0.4735	0.5042	2.923
23	2019-03-07 10:50:00	-0.114	0.77	825	-0.5073	0.4664	0.418
24	2019-04-09 10:30:00	-0.796	0.16	798	-0.9124	0.1835	12.425
25	2019-05-14 11:10:00	-0.602	0.25	715	-0.7105	0.2922	15.700
26	2019-07-09 10:15:00	-0.357	0.44	568	-0.5378	0.4348	25.878
27	2019-08-06 11:00:00	-0.658	0.22	597	-0.6596	0.3285	25.910
28	2019-09-03 10:40:00	-0.585	0.26	600	-0.6527	0.3338	25.460
29	2019-12-04 10:50:00	-0.469	0.34	669	-0.0579	1.3128	4.997
30	2020-03-04 11:00:00	-0.678	0.21	724	-0.3225	0.7139	5.870
31	2020-05-06 10:30:00	-1	0.1	767	-1.0246	0.1418	18.010
32	2020-06-03 10:20:00	-0.886	0.13	766	-1.1949	0.0958	22.057
33	2020-06-25 11:30:00	-1.15	0.07	774	-1.3249	0.0710	24.300
34	2020-07-15 10:00:00	-1.15	0.07	747	-1.3336	0.0696	27.100
35	2020-08-04 11:30:00	-1.1	0.08	731	-1.2413	0.0861	26.500
36	2020-08-18 11:40:00	<-1.7	<0.02	724	-1.2150	0.0914	26.563
37	2020-09-01 10:50:00	-1.05	0.09	723	-1.1617	0.1034	25.457
38	2021-01-13 10:30:00	-0.569	0.27	789	-0.4876	0.4881	3.450
39	2021-03-31 10:30:00	-0.444	0.36	788	-0.7934	0.2414	10.630
40	2021-05-04 10:20:00	-0.523	0.3	817	-1.1526	0.1056	16.167
41	2021-06-03 10:30:00	-0.602	0.25	832	-1.3772	0.0629	19.930
42	2021-06-21 11:00:00	<-1.7	<0.02	818	-1.6003	0.0377	26.450
43	2021-07-20 10:40:00	<-1.7	<0.02	828	-1.6313	0.0351	26.207
44	2021-08-10 10:00:00	<-1.7	<0.02	828	-1.6496	0.0336	26.630
45	2021-08-31 11:40:00	<-1.7	<0.02	827	-1.6251	0.0356	26.160

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.]

Lane, S.L., Flanagan, S., and Wilde, F.D., 2003, Selection of equipment for water sampling (ver. 2.0): U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A2, accessed September 2022 at http://pubs.water.usgs.gov/twri9A2/

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.