Appendix 16. Model Archival Summary for Nitrate plus Nitrite Concentration at U.S. Geological Survey Site 06892350, Kansas River at De Soto, Kansas, during June 2013 through September 2019

This model archival summary summarizes the nitrate plus nitrite (NO_x; U.S. Geological Survey [USGS] parameter code 00631) concentration model developed to compute 15-minute NO_x concentrations from June 2013 onward. This model supersedes all previous models.

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: 06892350
Site name: Kansas River at De Soto, Kansas
Location: Lat 38°59'00", long 94°57'52" referenced to North American Datum of 1927, in NE 1/4 SE 1/4 SE 1/4 sec.28, T.12 S., R.22 E., Leavenworth County, Kans., hydrologic unit 10270104.

Equipment: A YSI 6600 water-quality monitor equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, and turbidity was installed from August 2012 through June 2014. A Xylem YSI EXO2 water-quality monitor equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, turbidity, and chlorophyll and phycocyanin fluorescence was installed during June 2014 through September 2019. A Hach Nitratax plus sc sensor (sNO_x; 5-millimeter path length) that monitors ultraviolet (UV) nitrate concentrations was installed from June 2013 through September 2019. The monitors were housed in side-by-side 4-inch-diameter galvanized steel pipes. Readings from the water-quality and nitrate plus nitrite monitors were recorded every 15 minutes and transmitted by way of satellite, hourly.

Date model was created: March 31, 2020

Model calibration data period: June 17, 2013, through September 24, 2019

Model application date: June 17, 2013, onward

Model-Calibration Dataset

All data were collected using USGS protocols (Wagner and others, 2006; U.S. Geological Survey, variously dated) and are stored in the National Water Information System (U.S. Geological Survey, 2020) database and available to the public. Ordinary least squares analysis was used to develop regression models using R programming language (R Core Team, 2020). Potential explanatory variables that were evaluated individually and in combination included streamflow, water temperature, specific conductance, dissolved oxygen, pH, turbidity, chlorophyll and phycocyanin fluorescence, and sNO_x. The maximum time span between two continuous data points used for interpolation was 2 hours (in order 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 1 hour in the continuous data record resulted in missing interpolated data). Seasonal components (sine and cosine variables) were also evaluated as potential explanatory variables.

The final selected regression model was based on 84 concurrent measurements of NO_x concentration and sensor-measured sNO_x during June 17, 2013, through September 24, 2019. Samples were collected throughout the range of continuously observed hydrologic conditions. Twenty samples had concentrations below laboratory detection limits; therefore, a Tobit regression model was developed to compute estimates of linear regression model parameters using the absolute maximum likelihood estimation approach (Cohen, 1950; Hald, 1949; Helsel and others, 2020; Tobin, 1958). Summary statistics and the complete model-calibration dataset are provided below. Potential outliers were identified using the methods described in Rasmussen and others (2009). Additionally, outlier test criteria, including leverage and Cook's distance (Cook's D; Cook, 1977), were used to estimate potential outlier influence on the final Tobit regression model. One of the NO_x results, from May 9, 2019, was deemed an outlier and removed from the model calibration dataset. This sample was collected during 3 hours of rapidly changing conditions due to a runoff event and may not have been representative of the associated sample time. All other potential outliers were not found to have errors associated with collection, processing, or analysis and were therefore considered valid.

This model is specific to the Kansas River at De Soto, Kans., during this study period and cannot be applied to data collected from other sites on the Kansas River or data collected from other waterbodies.

Nitrate plus Nitrite Sampling Details

Cross-section samples typically were collected either from the downstream side of the bridge or instream within 100 feet of the bridge. The equal-width-increment collection method was used (although multiple vertical, single vertical, and grab samples were occasionally collected), and samples typically were composited for analysis (U.S. Geological Survey, variously dated). During July 2012 through June 2017, cross-section samples were collected every 2 weeks during March through October, once a month during November through February, and during selected reservoir release and runoff events. During July 2017 through September 2019, cross-section samples were collected on a monthly to bimonthly basis, depending on flow conditions. A FISP US DH–81, DH–95, D–95, D–96a, or D–96 depth integrating sampler was used. Additional detail on sample collection is available in Foster and Graham (2016) and Graham and others (2018). Samples were analyzed for NO_x concentration at the USGS National Water Quality Laboratory in Lakewood, Colorado.

Model Development

Discretely collected NO_x was related to sensor-measured sNO_x and other continuous sensor-measured data using stepwise regression analysis in R programming language (R Core Team, 2020). The distribution of residuals was examined for normality, and the plots of residuals (the difference between the measured and computed values) were examined for homoscedasticity (departures from zero did not change substantially over the range of computed values). Previously published explanatory variables were also strongly considered for continuity.

23.8 percent of the model-calibration dataset were censored results (less than the minimum reporting level). Tobit regression models were developed using absolute maximum likelihood estimation methods to relate discretely collected NO_x concentration to sensor-measured sNO_x. Tobit model parameter estimates were calculated using the *smwrQW* (*v0.7.9*) package in R programming language (R Core Team, 2020).

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

Model Summary

The following is a summary of final Tobit regression analysis for NO_x concentration at USGS site 06892350:

NO_x concentration-based model:

$$NO_x = 1.02 \times sNO_x - 0.325$$

where

NO_x = nitrate plus nitrite concentration, in milligrams per liter; and

sNO_x = sensor-measured nitrate plus nitrite, in milligrams per liter.

sNO_x makes physical and statistical sense as an explanatory variable for NO_x.

Previous Models

Start Year	End Year	Model Equation	Reference
1999	2003	NO_x = 0.914logQ - 0.541logWT - 2.04 where Q = streamflow, in cubic feet per second; and, WT = water temperature, in degrees Celsius	Rasmussen and others (2005)

Model Statistics, Data, and Plots

Model

$NO_x = + 1.02 * sNO_x - 0.325$

Computation method: Absolute Maximum Likelihood Estimation (AMLE)

Variable Summary Statistics

	NO_x	sNO_x
Minimum	<0.01	0.068
1st Quartile	0.034	0.373
Median	0.776	1.130
Mean	0.667	1.024
3rd Quartile	1.27	1.519
Maximum	2.20	2.259

Basic Model Statistics

Estimated residual standard error (unbiased)	0.2209	
Number of observations	84	
Number censored	20	(23.8 percent)
Log-likelihood (model)	-3.677	
Log-likelihood (intercept only)	-146	
Chi-square	284.7	
Degrees of freedom	1	
p-value	<0.0001	
Pseudo-R-squared	0.8975	
Akaike Information Criterion	13.35	
Bayesian Information Criterion	20.65	

Explanatory Variables

Coefficients:

	Estimate	Std. Error	z-score	p-value
(Intercept)	-0.3246	0.05456	-5.95	0
sNO_x	1.0160	0.04269	23.80	0

Outlier Test Criteria

Leverage	Cook's D
0.03571	0.69897

Flagged Observations

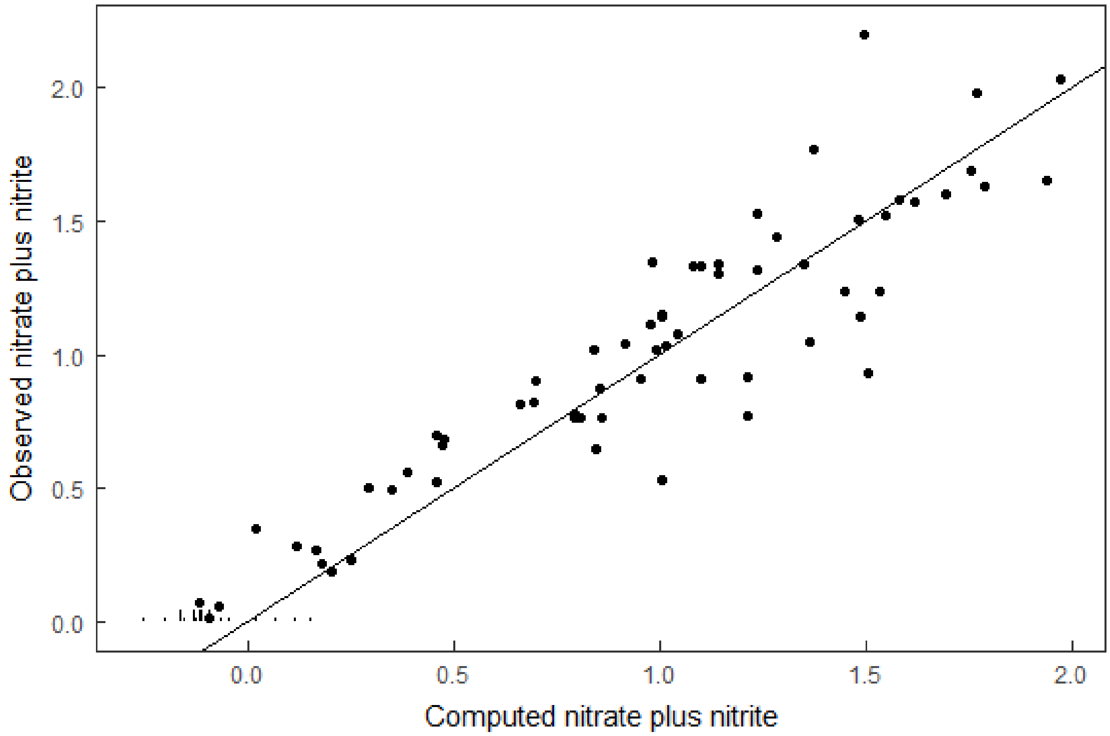
Observations exceeding at least one test criterion

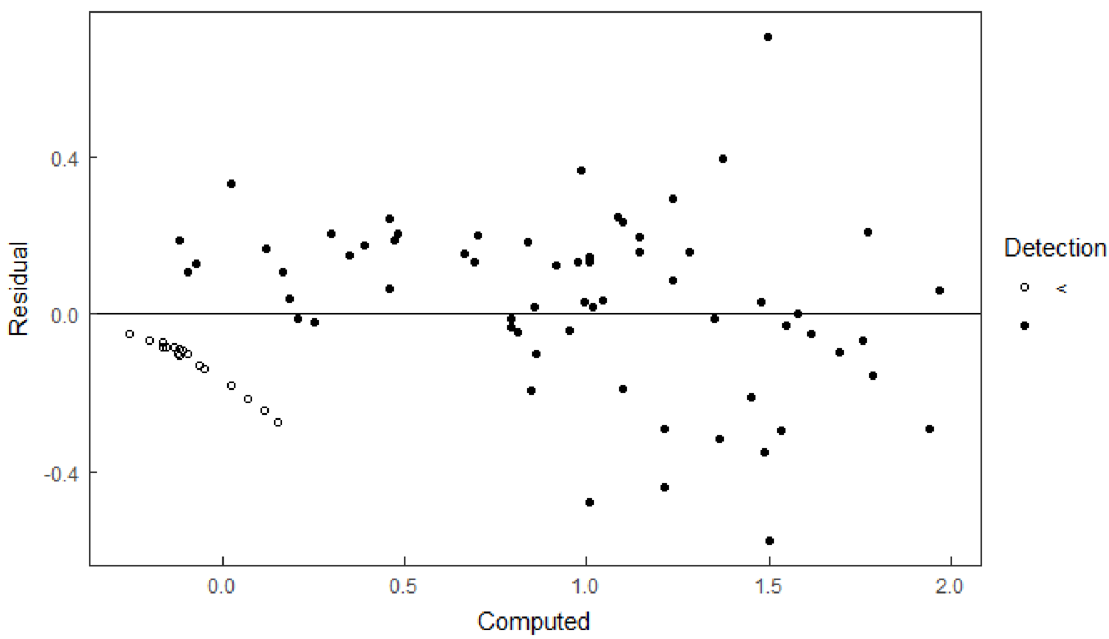
	NOx	ycen	yhat	resids	leverage	cooksD
1	2.03	FALSE	1.9705	0.05953	0.05688	0.002323
2	1.69	FALSE	1.7563	-0.06626	0.04284	0.002104
27	0.01	TRUE	-0.2022	-0.06676	0.03596	0.001767
33	1.98	FALSE	1.7716	0.20844	0.04375	0.021311
35	0.01	TRUE	-0.2558	-0.04822	0.03885	0.001002
50	1.65	FALSE	1.9399	-0.28987	0.05472	0.052742
51	1.60	FALSE	1.6951	-0.09506	0.03931	0.003945
81	1.63	FALSE	1.7869	-0.15686	0.04468	0.012350

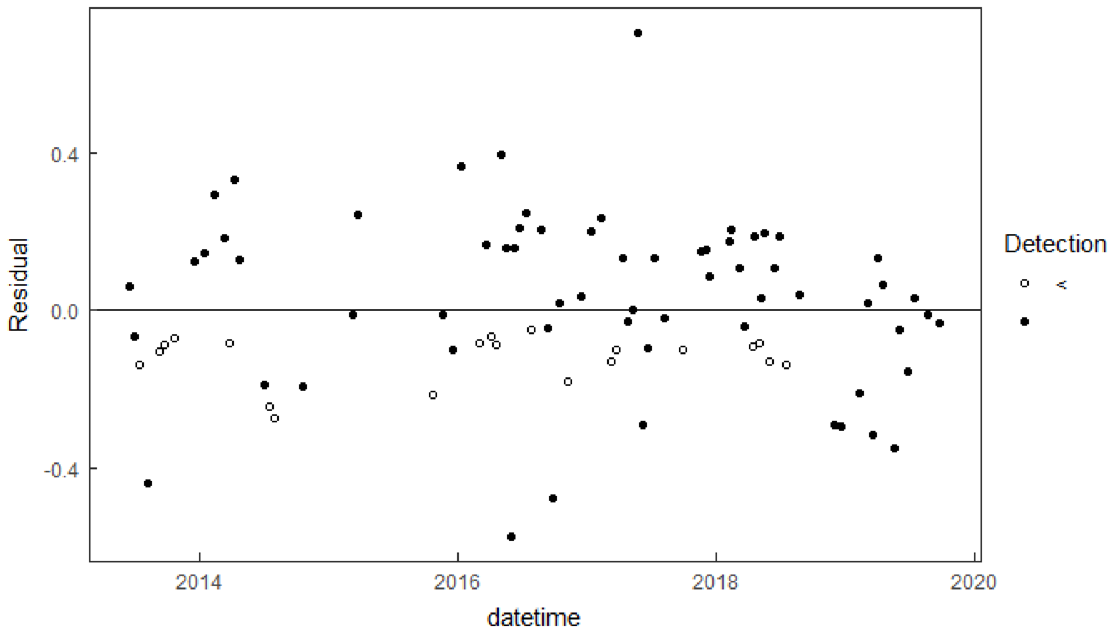
95 Percent Confidence Interval

	2.5 %	97.5 %
(Intercept)	-0.4315937	-0.2177029
sNO_x	0.9323271	1.0996520

Plots







Model-Calibration Dataset

	datetime	NO_x	sNO_x	Computed_NOx
1	2013-06-17 12:30:00	2.03	2.2590	1.9705
2	2013-07-01 13:20:00	1.69	2.0482	1.7563
3	2013-07-15 10:40:00	<0.01	0.2711	-0.0492
4	2013-08-05 11:30:00	0.772	1.5135	1.2131
5	2013-09-09 10:30:00	<0.01	0.2033	-0.1180
6	2013-09-23 09:40:00	<0.04	0.2033	-0.1180
7	2013-10-21 10:30:00	<0.04	0.1581	-0.1639
8	2013-12-16 08:00:00	1.04	1.2199	0.9148
9	2014-01-13 08:00:00	1.15	1.3102	1.0066
10	2014-02-10 07:40:00	1.53	1.5361	1.2361
11	2014-03-10 13:40:00	1.02	1.1446	0.8383
12	2014-03-24 12:50:00	<0.04	0.1883	-0.1333
13	2014-04-07 11:40:00	0.35	0.3388	0.0197
14	2014-04-21 12:00:00	0.057	0.2485	-0.0721
15	2014-06-30 12:20:00	0.91	1.4006	1.0984
16	2014-07-14 12:50:00	<0.01	0.4292	0.1115
17	2014-07-28 13:50:00	<0.01	0.4669	0.1497
18	2014-10-20 14:00:00	0.65	1.1521	0.8459
19	2015-03-09 12:20:00	0.78	1.0994	0.7924
20	2015-03-23 10:50:00	0.696	0.7681	0.4557
21	2015-10-19 09:30:00	<0.01	0.3840	0.0656
22	2015-11-16 09:10:00	0.19	0.5196	0.2033
23	2015-12-14 15:10:00	0.76	1.1671	0.8612
24	2016-01-11 10:10:00	1.35	1.2876	0.9836
25	2016-03-03 09:00:00	<0.01	0.1581	-0.1639
26	2016-03-21 10:50:00	0.285	0.4367	0.1191
27	2016-04-04 09:20:00	<0.01	0.1205	-0.2022
28	2016-04-18 09:30:00	<0.04	0.2033	-0.1180
29	2016-05-02 09:30:00	1.77	1.6717	1.3738
30	2016-05-16 09:10:00	1.44	1.5813	1.2820
31	2016-05-28 13:20:00	0.93	1.7997	1.5039
32	2016-06-06 11:30:00	1.3	1.4458	1.1443
33	2016-06-20 08:20:00	1.98	2.0632	1.7716
34	2016-07-11 08:40:00	1.33	1.3855	1.0831

```
35 2016-07-25 08:30:00  <0.01 0.0678      -0.2557
36 2016-08-22 09:00:00    0.5 0.6099       0.2951
37 2016-09-12 08:50:00   0.76 1.1144       0.8077
38 2016-09-26 10:30:00   0.53 1.3102       1.0066
39 2016-10-11 09:20:00   0.87 1.1596       0.8536
40 2016-11-07 08:40:00  <0.01 0.3388       0.0197
41 2016-12-12 09:50:00   1.08 1.3479       1.0448
42 2017-01-09 09:50:00    0.9 1.0090       0.7006
43 2017-02-06 10:00:00   1.33 1.4006       1.0984
44 2017-03-06 10:40:00  <0.01 0.2560      -0.0645
45 2017-03-20 10:30:00  <0.04 0.2259      -0.0951
46 2017-04-10 09:20:00   1.11 1.2801       0.9760
47 2017-04-25 10:10:00   1.52 1.8449       1.5498
48 2017-05-08 09:20:00   1.58 1.8750       1.5804
49 2017-05-22 09:20:00    2.2 1.7921       1.4962
50 2017-06-05 08:50:00   1.65 2.2289       1.9399
51 2017-06-19 09:10:00    1.6 1.9879       1.6951
52 2017-07-10 09:10:00  0.825 1.0015       0.6929
53 2017-08-07 09:30:00   0.23 0.5647       0.2492
54 2017-09-26 10:20:00  <0.01 0.1958      -0.1257
55 2017-11-17 14:40:00  0.496 0.6626       0.3486
56 2017-12-01 09:00:00  0.814 0.9714       0.6623
57 2017-12-11 10:30:00   1.32 1.5361       1.2361
58 2018-02-05 10:00:00   0.56 0.7003       0.3869
59 2018-02-09 09:30:00  0.684 0.7906       0.4787
60 2018-03-05 10:10:00   0.27 0.4819       0.1650
61 2018-03-20 08:20:00   0.91 1.2575       0.9530
62 2018-04-11 13:00:00  <0.04 0.2090      -0.1123
63 2018-04-16 09:30:00   0.07 0.2033      -0.1180
64 2018-04-30 09:20:00  <0.01 0.1657      -0.1563
65 2018-05-07 10:40:00   1.02 1.2952       0.9913
66 2018-05-15 09:30:00   1.34 1.4458       1.1443
67 2018-05-29 08:40:00  <0.01 0.2560      -0.0645
68 2018-06-11 10:00:00   0.01 0.2259      -0.0951
69 2018-06-25 07:20:00   0.66 0.7831       0.4710
70 2018-07-16 11:00:00  <0.01 0.2711      -0.0492
71 2018-08-21 10:40:00   0.22 0.4970       0.1803
72 2018-11-29 12:20:00   0.92 1.5135       1.2131
73 2018-12-18 10:40:00   1.24 1.8298       1.5345
74 2019-02-06 10:20:00   1.24 1.7470       1.4503
75 2019-03-06 07:50:00   1.03 1.3177       1.0142
76 2019-03-19 09:20:00   1.05 1.6641       1.3662
77 2019-04-01 12:00:00   1.14 1.3102       1.0066
78 2019-04-16 09:30:00   0.52 0.7681       0.4557
79 2019-05-20 12:30:00   1.14 1.7846       1.4886
80 2019-06-03 10:20:00   1.57 1.9126       1.6186
81 2019-06-26 12:30:00   1.63 2.0783       1.7869
82 2019-07-16 10:10:00   1.51 1.7771       1.4809
83 2019-08-20 09:20:00   1.34 1.6491       1.3509
84 2019-09-24 09:50:00   0.76 1.0994       0.7924
```

Definitions

Cook's D: Cook's distance (Helsel and others, 2020).

Leverage: An outlier's measure in the x direction (Helsel and others, 2020).

NO_x: Inorganic nitrogen (nitrate and nitrite), in milligrams per liter as nitrogen (00631).

p-value: The probability that the independent variable has no effect on the dependent variable (Helsel and others, 2020).

Pseudo-R-squared: Pseudocoefficient of determination. An estimation of the proportion of variance in the response variable explained by the model (McKelvey and Zavoina, 1975).

sNO_x: Sensor measured inorganic nitrogen (nitrate and nitrite), in milligrams per liter as nitrogen (99133).

z-score: The estimated coefficient divided by its associated standard error (Helsel and others, 2020).

References Cited

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.

Cook, R.D., 1977, Detection of influential observations in linear regression: Technometrics, v. 19, no. 1, p. 15–18. [Also available at https://www.jstor.org/stable/1268249.]

Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: Journal of the American Statistical Association, v. 78, no. 383, p. 605–610. [Also available at https://doi.org/10.1080/01621459.1983.10478017.]

Foster, G.M., and Graham, J.L., 2016, Logistic and linear regression model documentation for statistical relations between continuous real-time and discrete water-quality constituents in the Kansas River, Kansas, July 2012 through June 2015: U.S. Geological Survey Open-File Report 2016–1040, 27 p., accessed July 2020 at https://doi.org/10.3133/ofr20161040.

Graham, J.L., Foster, G.M., Williams, T.J., Mahoney, M.D., May, M.R., and Loftin, K.A., 2018, Water-quality conditions with an emphasis on cyanobacteria and associated toxins and taste-and-odor compounds in the Kansas River, Kansas, July 2012 through September 2016: U.S. Geological Survey Scientific Investigations Report 2018–5089, 55 p. [Also available at https://doi.org/10.3133/sir20185089.]

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, ver. 1.1.]

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.]

R Core Team, 2020, R—A language and environment for statistical computing, version 4.0.3: Vienna, Austria, R Foundation for Statistical Computing, accessed December 2020 at https://www.R-project.org/.

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, 53 p. [Also available at https://doi.org/10.3133/tm3C4.]

Rasmussen, T.J., Ziegler, A.C., and Rasmussen, P.P., 2005, Estimation of constituent concentrations, densities, loads, and yields in lower Kansas River, northeast Kansas, using regression models and continuous water-quality monitoring, January 2000 through December 2003: U.S. Geological Survey Scientific Investigations Report 2005–5165, 117 p. [Also available at https://doi.org/10.3133/sir20055165.]

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, 2020, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 2020 at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey, variously dated, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1–A9 [variously paged], accessed July 2020 at https://water.usgs.gov/owq/FieldManual/.

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.] [Supersedes USGS Water-Resources Investigations Report 2000–4252.]