

Appendix 23. Model Archival Summary for Total Suspended Solids Concentration at U.S. Geological Survey Site 06887500, Kansas River at Wamego, Kansas, during September 2013 through September 2019

This model archival summary summarizes the total suspended solids (TSS; U.S. Geological Survey [USGS] parameter code 00530) concentration model developed to compute 15-minute TSS concentrations from September 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: 06887500

Site name: Kansas River at Wamego, Kansas

Location: Lat 39°11'54", long 96°18'19" referenced to North American Datum of 1927, in SW 1/4 NW 1/4 SE 1/4 sec.9, T.10 S., R.10 E., Pottawatomie County, Kans., hydrologic unit 10270102.

Equipment: A YSI 6600 water-quality monitor equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, and turbidity (TBY) 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, TBY, and chlorophyll and phycocyanin fluorescence was installed during June 2014 through September 2019. The monitor was housed in a 4-inch-diameter galvanized steel pipe. Readings from the water-quality monitor were recorded every 15 minutes and transmitted by way of satellite, hourly.

Date model was created: May 5, 2020

Model calibration data period: September 23, 2013, through September 23, 2019

Model application date: September 23, 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, TBY, and chlorophyll and phycocyanin fluorescence. These potential explanatory variables were interpolated within the 15-minute continuous record based on sample time. 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 78 concurrent measurements of TSS concentration and sensor-measured TBY during September 23, 2013, through September 23, 2019. Samples were collected throughout the range of continuously observed hydrologic conditions. Five 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. All 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 Wamego, 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.

Total Suspended Solids 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 TSS concentration at the USGS National Water Quality Laboratory in Lakewood, Colorado.

Model Development

Discretely collected TSS was related to sensor-measured TBY 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.

6.4 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 TSS concentration to sensor-measured TBY. Tobit model parameter estimates were calculated using the *smwrQW* (v0.7.9) package in R programming language (R Core Team, 2020).

TBY was selected as a good surrogate for TSS based on residual plots, pseudocoeficient 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 TSS concentration at USGS site 06887500:

TSS concentration-based model:

$$\log TSS = 1.05 \times \log TBY + 0.167$$

where

\log = logarithm base 10;

TSS = total suspended solids concentration, in milligrams per liter; and

TBY = turbidity, in formazin nephelometric units.

TBY makes physical and statistical sense as an explanatory variable for TSS because of its positive correlation with suspended material.

The logarithmically (\log) transformed model may be retransformed to the original units so that TSS can be calculated directly. The retransformation introduces a bias in the calculated constituent. This bias may be corrected using Duan's bias correction factor (BCF; Duan, 1983). For this model, the calculated BCF is 1.08. The retransformed model, accounting for BCF is as follows:

$$TSS = 1.08 \times (TBY^{1.05} \times 10^{0.167})$$

Previous Models

Start Year	End Year	Model Equation	Reference
1999	2003	$\log TSS = 0.923 \log TBY + 0.087$	Rasmussen and others (2005)

Model Statistics and Data

Model

$\log\text{TSS} = + 1.05 * \log\text{TBY} + 0.167$

Computation method: Absolute Maximum Likelihood Estimation (AMLE)

Variable Summary Statistics

	TSS	TBY
Minimum	<15.0	7.42
1st Quartile	38.0	23.32
Median	84.0	50.15
Mean	199.6	101.42
3rd Quartile	210.0	115.60
Maximum	1830.0	556.70

Basic Model Statistics

Estimated residual standard error (unbiased)	0.1777
Number of observations	78
Number censored	5 (6.4 percent)
Log-likelihood (model)	20.12
Log-likelihood (intercept only)	-91.97
Chi-square	224.2
Degrees of freedom	1
p-value	<0.0001
Pseudo-R-squared	0.8926
Akaike Information Criterion	-34.24
Bayesian Information Criterion	-27.17
Bias Correction Factor	1.076639

Explanatory Variables

Coefficients:

	Estimate	Std. Error	z-score	p-value
(Intercept)	0.1674	0.07897	2.12	0.0377
logTBY	1.0490	0.04333	24.21	0.0000

Outlier Test Criteria

Leverage	Cook's D
0.03846	0.69942

Flagged Observations

Observations exceeding at least one test criterion:

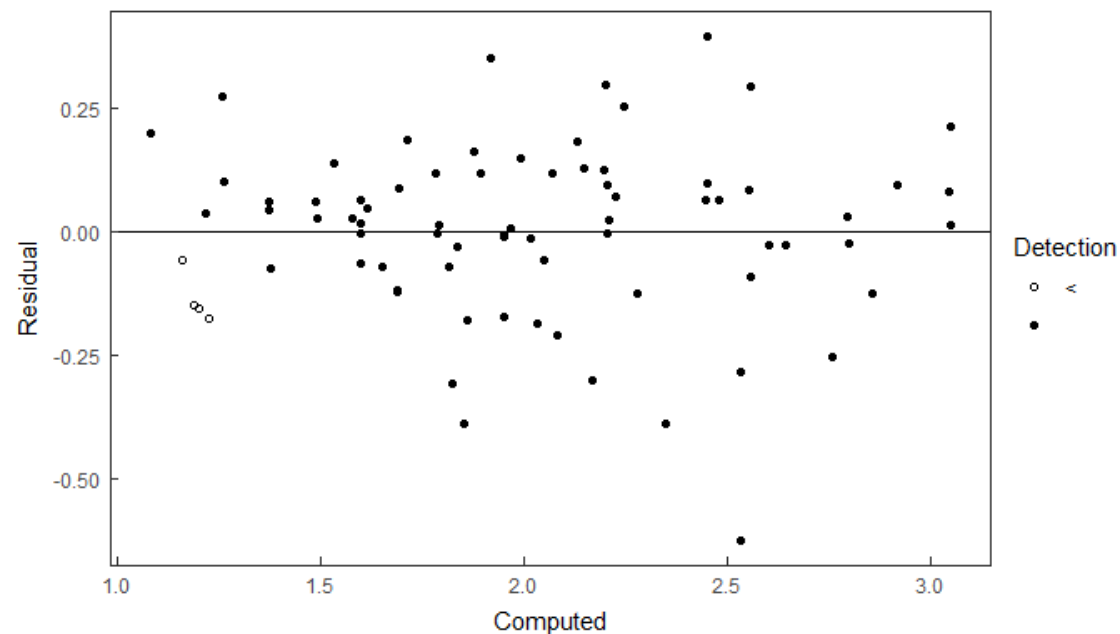
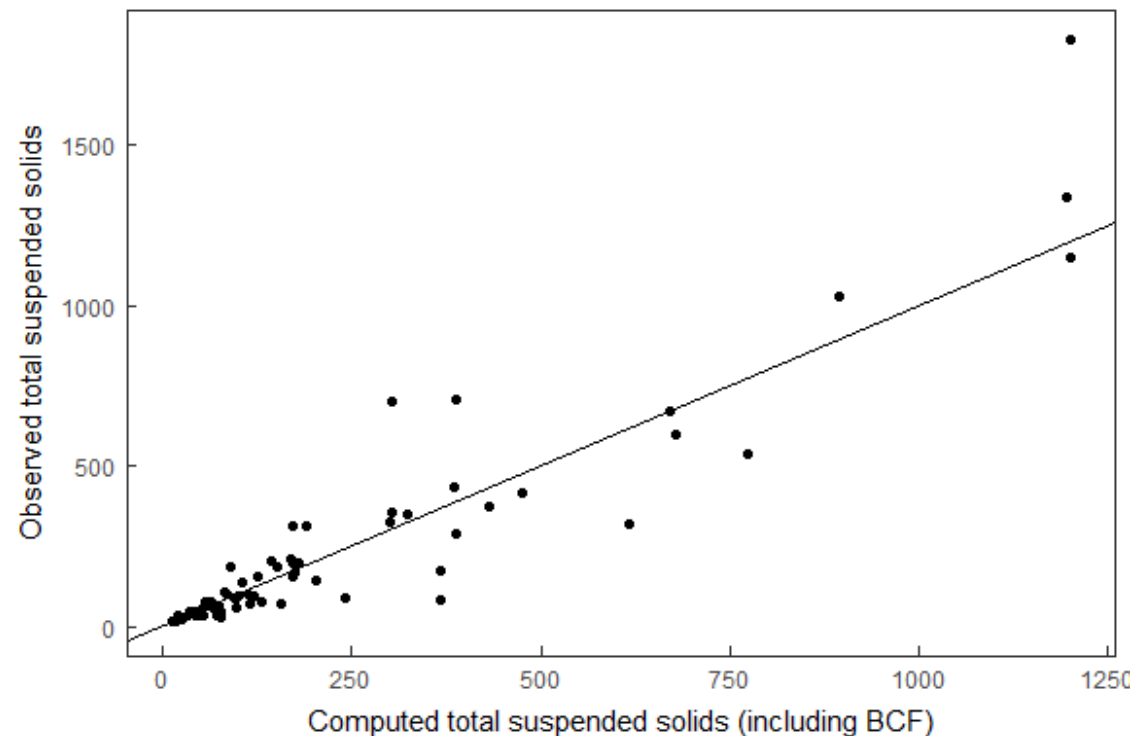
	logTSS	ycen	yhat	resids	leverage	cooksD
6	1.176	TRUE	1.198	-0.15593	0.04434	0.0186828
7	1.322	TRUE	1.159	-0.05655	0.04752	0.0026515
8	1.531	FALSE	1.256	0.27539	0.03988	0.0519235
24	1.176	TRUE	1.186	-0.14837	0.04526	0.0173000
25	1.255	FALSE	1.216	0.03886	0.04288	0.0011186
26	1.176	TRUE	1.185	-0.14753	0.04537	0.0171474
27	1.279	FALSE	1.080	0.19829	0.05435	0.0378226
29	2.505	FALSE	2.758	-0.25313	0.04229	0.0467586
42	3.262	FALSE	3.048	0.21488	0.06871	0.0578909
46	1.362	FALSE	1.260	0.10189	0.03960	0.0070552
47	2.827	FALSE	2.795	0.03203	0.04521	0.0008050
55	2.776	FALSE	2.800	-0.02409	0.04559	0.0004596
61	1.176	TRUE	1.225	-0.17465	0.04218	0.0221979

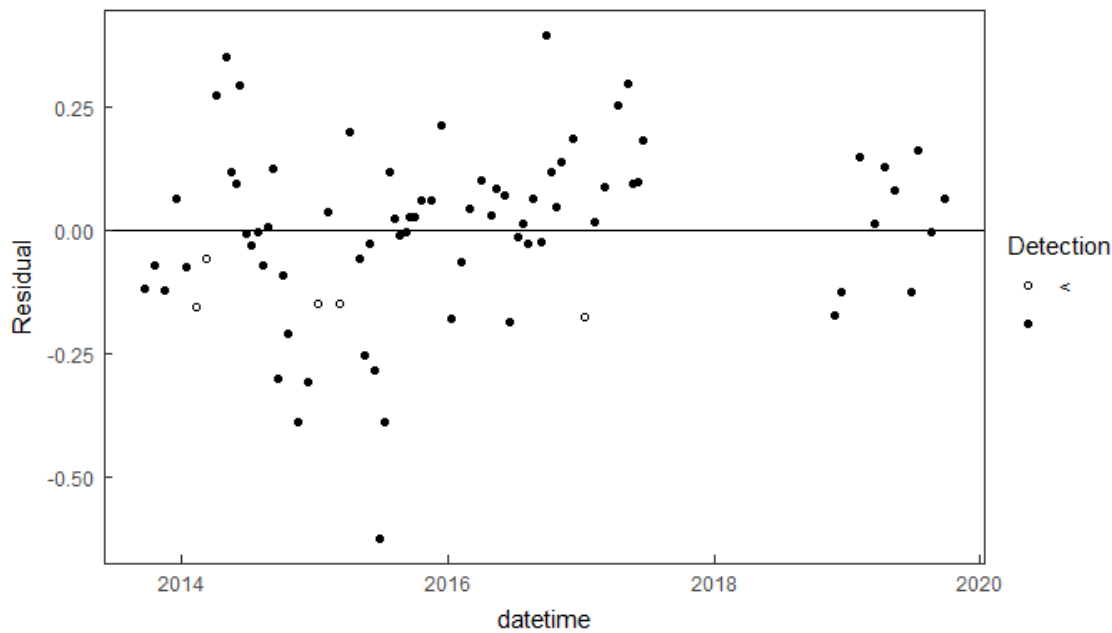
66	3.013	FALSE	2.919	0.09361	0.05595	0.0087065
72	3.061	FALSE	3.048	0.01310	0.06871	0.0002151
74	3.127	FALSE	3.045	0.08164	0.06849	0.0083250
75	2.732	FALSE	2.857	-0.12489	0.05039	0.0137924

95 Percent Confidence Interval

	2.5%	97.5%
(Intercept)	0.01262439	0.3221704
logTBY	0.96409304	1.1339410

Plots





Model-Calibration Dataset

	datetime	logTSS	logTBY	TSS	TBY	Computed_logTSS	Computed_TSS
1	2013-09-23 10:15:00	1.57	1.447	37	28.00	1.69	52.2
2	2013-10-21 10:15:00	1.58	1.415	38	26.00	1.65	48.3
3	2013-11-18 09:30:00	1.57	1.450	37	28.20	1.69	52.6
4	2013-12-16 12:30:00	1.66	1.364	46	23.10	1.60	42.6
5	2014-01-13 11:00:00	1.3	1.152	20	14.20	1.38	25.6
6	2014-02-10 10:30:00	<1.18	0.982	<15	9.60	1.20	17.0
7	2014-03-10 08:20:00	<1.32	0.945	<21	8.81	1.16	15.5
8	2014-04-07 10:00:00	1.53	1.038	34	10.91	1.26	19.4
9	2014-05-05 09:00:00	2.27	1.668	186	46.60	1.92	89.0
10	2014-05-19 08:00:00	1.9	1.538	79	34.50	1.78	65.0
11	2014-06-02 08:30:00	2.3	1.940	198	87.10	2.20	171.6
12	2014-06-11 09:00:00	2.85	2.279	710	190.00	2.56	388.9
13	2014-06-30 08:30:00	1.94	1.701	88	50.20	1.95	96.3
14	2014-07-14 08:20:00	1.81	1.591	64	39.00	1.84	73.9
15	2014-07-28 10:00:00	1.59	1.362	39	23.00	1.60	42.5
16	2014-08-11 09:20:00	1.74	1.568	55	37.00	1.81	69.9
17	2014-08-25 08:40:00	1.97	1.716	94	52.00	1.97	99.9
18	2014-09-08 08:00:00	2.32	1.934	210	86.00	2.20	169.3
19	2014-09-22 09:20:00	1.86	1.905	73	80.33	2.17	157.7
20	2014-10-06 09:00:00	2.47	2.279	292	190.00	2.56	388.9
21	2014-10-20 09:00:00	1.88	1.826	75	67.00	2.08	130.3
22	2014-11-17 12:30:00	1.46	1.606	29	40.40	1.85	76.7
23	2014-12-15 14:50:00	1.52	1.580	33	38.00	1.82	71.9
24	2015-01-12 08:40:00	<1.18	0.971	<15	9.36	1.19	16.5
25	2015-02-09 10:20:00	1.26	1.000	18	10.00	1.22	17.7
26	2015-03-09 09:20:00	<1.18	0.970	<15	9.33	1.18	16.5
27	2015-04-06 08:50:00	1.28	0.870	19	7.42	1.08	13.0
28	2015-05-04 07:50:00	1.99	1.792	98	62.00	2.05	120.1
29	2015-05-18 10:30:00	2.51	2.470	320	295.00	2.76	617.0
30	2015-06-01 08:00:00	2.62	2.362	414	230.00	2.64	475.3
31	2015-06-15 07:50:00	2.25	2.254	177	179.52	2.53	366.5
32	2015-06-29 08:20:00	1.91	2.255	81	180.00	2.53	367.5
33	2015-07-13 08:00:00	1.96	2.079	91	120.00	2.35	240.2
34	2015-07-27 09:10:00	2.01	1.643	102	44.00	1.89	83.8

35	2015-08-10 07:50:00	2.23	1.944	170	88.00	2.21	173.5
36	2015-08-24 10:20:00	1.94	1.699	87	50.00	1.95	95.9
37	2015-09-08 09:10:00	1.79	1.544	61	35.00	1.79	65.9
38	2015-09-21 09:30:00	1.6	1.342	40	22.00	1.58	40.5
39	2015-10-05 09:50:00	1.52	1.262	33	18.30	1.49	33.4
40	2015-10-19 09:50:00	1.54	1.255	35	18.00	1.48	32.8
41	2015-11-16 09:40:00	1.43	1.146	27	14.00	1.37	25.2
42	2015-12-14 09:10:00	3.26	2.746	1830	556.67	3.05	1201.2
43	2016-01-11 09:30:00	1.68	1.613	48	41.00	1.86	77.9
44	2016-02-08 09:40:00	1.53	1.362	34	23.00	1.60	42.5
45	2016-03-03 09:30:00	1.41	1.146	26	14.00	1.37	25.2
46	2016-04-04 09:00:00	1.36	1.041	23	11.00	1.26	19.6
47	2016-05-02 09:20:00	2.83	2.505	672	320.00	2.80	672.0
48	2016-05-16 08:50:00	2.64	2.275	434	188.33	2.55	385.4
49	2016-06-06 08:20:00	2.3	1.961	198	91.33	2.22	180.4
50	2016-06-20 12:10:00	1.85	1.776	70	59.67	2.03	115.4
51	2016-07-11 09:00:00	2	1.763	101	58.00	2.02	112.0
52	2016-07-25 08:40:00	1.81	1.548	64	35.33	1.79	66.6
53	2016-08-08 08:10:00	2.58	2.322	377	210.00	2.60	432.0
54	2016-08-22 09:00:00	2.55	2.204	351	160.00	2.48	324.8
55	2016-09-12 08:50:00	2.78	2.510	597	323.33	2.80	679.3
56	2016-09-26 09:00:00	2.85	2.176	702	150.00	2.45	303.5
57	2016-10-11 09:00:00	2.19	1.813	154	65.00	2.07	126.2
58	2016-10-24 09:50:00	1.66	1.380	46	24.00	1.62	44.4
59	2016-11-07 09:00:00	1.67	1.301	47	20.00	1.53	36.7
60	2016-12-12 09:20:00	1.9	1.472	79	29.67	1.71	55.4
61	2017-01-09 10:10:00	<1.18	1.009	<15	10.20	1.23	18.1
62	2017-02-06 09:40:00	1.61	1.362	41	23.00	1.60	42.5
63	2017-03-06 09:50:00	1.78	1.452	60	28.33	1.69	52.8
64	2017-04-10 09:40:00	2.5	1.982	316	96.00	2.25	190.1
65	2017-05-08 09:20:00	2.5	1.938	316	86.67	2.20	170.7
66	2017-05-22 09:20:00	3.01	2.623	1030	420.00	2.92	893.8
67	2017-06-05 07:50:00	2.55	2.176	354	150.00	2.45	303.5
68	2017-06-19 08:00:00	2.31	1.869	205	74.00	2.13	144.6
69	2018-11-28 11:20:00	1.78	1.700	60	50.10	1.95	96.1
70	2018-12-17 09:10:00	2.15	2.010	142	102.41	2.28	203.4
71	2019-02-05 09:50:00	2.14	1.739	138	54.83	1.99	105.6
72	2019-03-18 09:40:00	3.06	2.746	1150	556.70	3.05	1201.2
73	2019-04-15 09:50:00	2.27	1.886	188	76.90	2.15	150.6
74	2019-05-09 15:20:00	3.13	2.744	1340	554.10	3.05	1195.3
75	2019-06-25 09:00:00	2.73	2.564	540	366.60	2.86	775.0
76	2019-07-15 09:10:00	2.04	1.630	110	42.67	1.88	81.2
77	2019-08-19 09:00:00	2.2	1.941	158	87.30	2.20	172.0
78	2019-09-23 10:20:00	2.51	2.173	324	149.07	2.45	301.5

Definitions

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

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

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).

TBY: Turbidity, in formazin nephelometric units (63680).

TSS: Total suspended solids, in milligrams per liter (00530).

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.]