Appendix 24. Model Archival Summary for Total Suspended Solids Concentration at U.S. Geological Survey Site 06892350, Kansas River at De Soto, 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: 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 (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. A Hach Nitratax plus sc sensor (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: April 2, 2020

Model calibration data period: September 23, 2013, through September 24, 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, chlorophyll and phycocyanin fluorescence, and UV nitrate sensor data. 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 79 concurrent measurements of TSS concentration and sensor-measured TBY during September 23, 2013, through September 24, 2019. Samples were collected throughout the range of continuously observed hydrologic conditions. Four samples had concentrations below the 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 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.

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.

5.1 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, 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 TSS concentration at USGS site 06892350:

TSS concentration-based model:

$$\log TSS = 1.03 \times \log TBY + 0.226$$

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 1.03. The retransformed model, accounting for BCF is as follows:

$$TSS = 1.03 \times (TBY^{1.03} \times 10^{0.226})$$

Previous Models

Start Year	End Year	Model Equation	Reference
1999	2003	logTSS = 0.863logTBY + 0.204	Rasmussen and others (2005)

Model Statistics, Data, and Plots

Model

logTSS = + 1.03 * logTBY + 0.226

Computation method: Absolute Maximum Likelihood Estimation (AMLE)

Variable Summary Statistics

	TSS	TBY
Minimum	<15.0	6.1
1st Quartile	54.0	26.0
Median	96.0	58.0
Mean	339.9	165.8
3rd Quartile	356.0	160.0
Maximum	2920.0	1530.0

Basic Model Statistics

Estimated residual standard error (unbiased)	0.1103	
Number of observations	79	
Number censored	4	(5.1 percent)
Log-likelihood (model)	57.2	
Log-likelihood (intercept only)	-98.31	
Chi-square	311	
Degrees of freedom	1	
p-value	<0.0001	
Pseudo-R-squared	0.9657	
Akaike Information Criterion	-108.4	
Bayesian Information Criterion	-101.3	
Bias Correction Factor	1.03162	

Explanatory Variables

Coefficients:

	Estimate	Std. Error	z-score	p-value
(Intercept)	0.2261	0.04421	5.115	0
logTBY	1.0349	0.02292	45.158	0

Outlier Test Criteria

Leverage	Cook's D
0.03797	0.69934

Flagged Observations

Observations exceeding at least one test criterion:

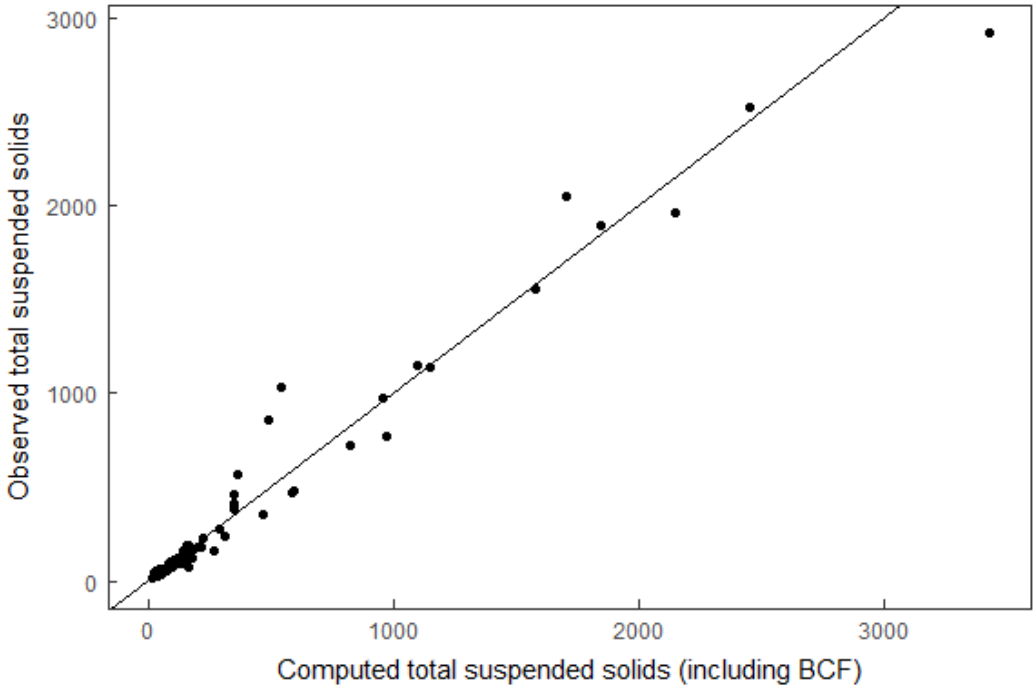
	logTSS	ycen	yhat	resids	leverage	cooksD
6	1.176	TRUE	1.198	-0.15593	0.04434	0.0186828
7	1.322	TRUE	1.159	-0.05655	0.04752	0.0026515
8	1.531	FALSE	1.256	0.27539	0.03988	0.0519235
24	1.176	TRUE	1.186	-0.14837	0.04526	0.0173000
25	1.255	FALSE	1.216	0.03886	0.04288	0.0011186
26	1.176	TRUE	1.185	-0.14753	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

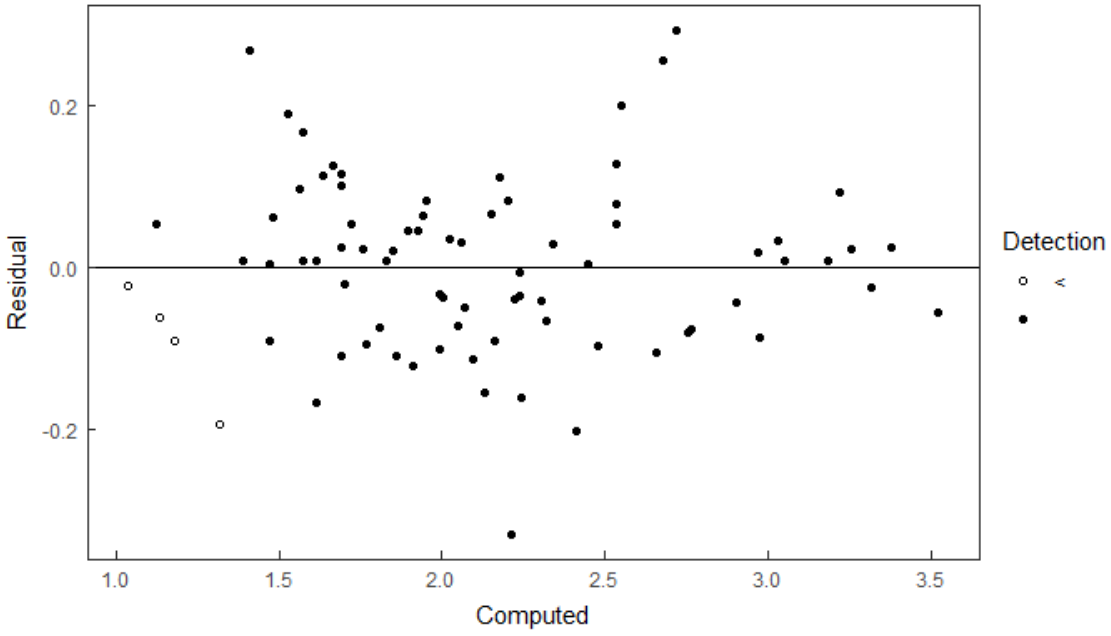
```
47  2.827 FALSE 2.795  0.03203  0.04521 0.0008050
55  2.776 FALSE 2.800 -0.02409  0.04559 0.0004596
61  1.176  TRUE 1.225 -0.17465  0.04218 0.0221979
66  3.013 FALSE 2.919  0.09361  0.05595 0.0087065
72  3.061 FALSE 3.048  0.01310  0.06871 0.0002151
74  3.127 FALSE 3.045  0.08164  0.06849 0.0083250
75  2.732 FALSE 2.857 -0.12489  0.05039 0.0137924
```

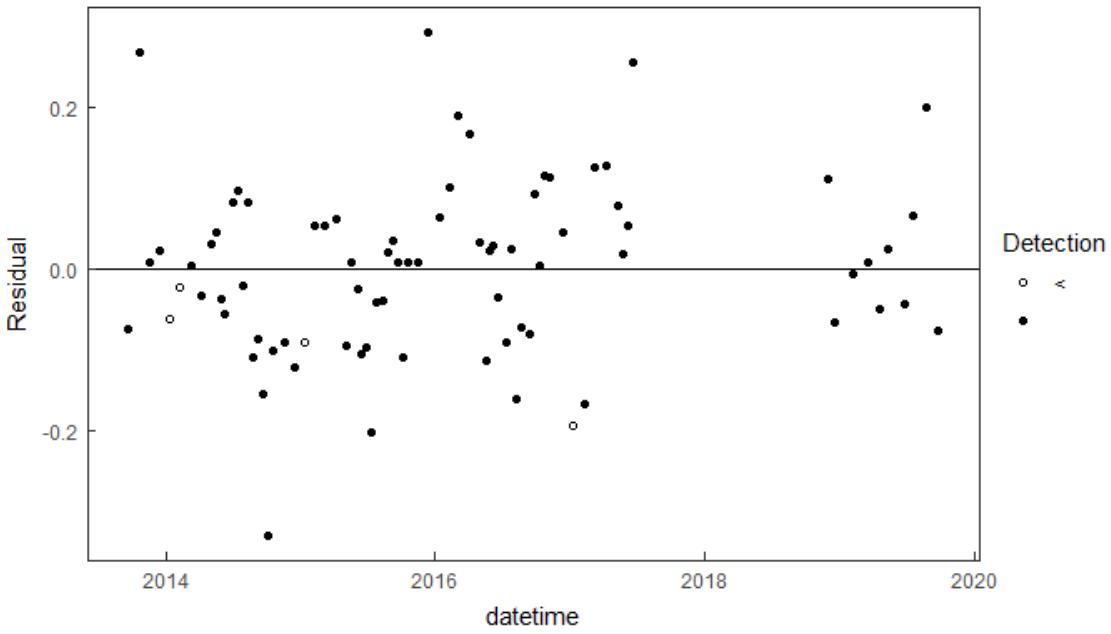
95 Percent Confidence Interval

```
                2.5 %    97.5 %
(Intercept) 0.1394503 0.3127347
logTBY      0.9899910 1.0798258
```

Plots







Model-Calibration Dataset

	datetime	logTSS	logTBY	TSS	TBY	Computed_logTSS	Computed_TSS
1	2013-09-23 09:40:00	1.73	1.528	54	33.70	1.81	66.2
2	2013-10-21 10:30:00	1.68	1.146	48	14.00	1.41	26.7
3	2013-11-18 14:00:00	1.62	1.342	42	22.00	1.62	42.5
4	2013-12-16 08:00:00	1.78	1.477	60	30.00	1.75	58.7
5	2014-01-13 08:00:00	<1.18	0.875	<15	7.50	1.13	14.0
6	2014-02-10 07:40:00	<1.18	0.785	<15	6.10	1.04	11.3
7	2014-03-10 13:40:00	1.48	1.204	30	16.00	1.47	30.6
8	2014-04-07 11:40:00	1.96	1.708	91	51.00	1.99	101.6
9	2014-05-05 14:00:00	2.09	1.771	123	59.00	2.06	118.1
10	2014-05-19 12:20:00	1.94	1.613	87	41.00	1.90	81.0
11	2014-06-02 13:00:00	1.96	1.716	92	52.00	2.00	103.6
12	2014-06-11 16:00:00	3.47	3.185	2920	1530.00	3.52	3431.3
13	2014-06-30 12:20:00	2.28	1.908	192	81.00	2.20	163.9
14	2014-07-14 12:50:00	1.66	1.294	46	19.67	1.56	37.9
15	2014-07-28 13:50:00	1.68	1.426	48	26.67	1.70	51.9
16	2014-08-11 15:10:00	2.03	1.666	108	46.33	1.95	92.0
17	2014-08-25 14:20:00	1.75	1.576	56	37.67	1.86	74.2
18	2014-09-08 12:40:00	2.89	2.655	772	452.00	2.97	971.5
19	2014-09-22 10:00:00	1.97	1.839	94	69.00	2.13	138.9
20	2014-10-06 13:30:00	1.88	1.918	76	82.80	2.21	167.7
21	2014-10-20 14:00:00	1.89	1.708	78	51.00	1.99	101.6
22	2014-11-17 12:00:00	1.38	1.204	24	16.00	1.47	30.6
23	2014-12-15 10:00:00	1.79	1.625	61	42.20	1.91	83.5
24	2015-01-12 11:30:00	<1.18	0.923	<15	8.37	1.18	15.7
25	2015-02-09 13:50:00	1.78	1.447	60	28.00	1.72	54.6
26	2015-03-09 12:20:00	1.18	0.865	15	7.33	1.12	13.6
27	2015-04-06 14:50:00	1.54	1.213	35	16.33	1.48	31.3
28	2015-05-04 13:30:00	1.67	1.490	47	30.88	1.77	60.4
29	2015-05-18 15:30:00	3.19	2.860	1560	724.10	3.19	1582.1
30	2015-06-06 19:50:00	3.29	2.988	1960	972.22	3.32	2146.1
31	2015-06-15 14:50:00	2.55	2.348	356	222.96	2.66	467.5
32	2015-06-29 12:30:00	2.38	2.176	240	150.00	2.48	310.2
33	2015-07-13 14:30:00	2.21	2.114	163	130.00	2.41	267.5
34	2015-07-27 14:20:00	2.26	2.009	183	102.00	2.30	208.1

35	2015-08-10 13:50:00	2.18	1.931	153	85.33	2.22	173.0
36	2015-08-24 09:50:00	1.87	1.568	74	37.00	1.85	72.9
37	2015-09-08 09:50:00	2.06	1.735	114	54.33	2.02	108.5
38	2015-09-21 09:30:00	1.84	1.550	69	35.50	1.83	69.8
39	2015-10-05 09:20:00	1.58	1.415	38	26.00	1.69	50.6
40	2015-10-19 09:30:00	1.58	1.301	38	20.00	1.57	38.6
41	2015-11-16 09:10:00	1.4	1.125	25	13.33	1.39	25.3
42	2015-12-14 15:10:00	3.01	2.409	1030	256.67	2.72	540.8
43	2016-01-11 10:10:00	2	1.656	101	45.33	1.94	89.9
44	2016-02-08 08:30:00	1.79	1.415	62	26.00	1.69	50.6
45	2016-03-03 09:00:00	1.72	1.255	52	18.00	1.53	34.6
46	2016-04-04 09:20:00	1.74	1.301	55	20.00	1.57	38.6
47	2016-05-02 09:30:00	3.06	2.708	1150	510.00	3.03	1100.7
48	2016-05-16 09:10:00	1.98	1.806	96	64.00	2.10	128.5
49	2016-05-28 13:20:00	3.28	2.925	1890	841.67	3.25	1848.6
50	2016-06-06 11:30:00	2.37	2.041	233	110.00	2.34	225.0
51	2016-06-20 08:20:00	2.2	1.944	160	88.00	2.24	178.6
52	2016-07-11 08:40:00	2.07	1.867	117	73.67	2.16	148.6
53	2016-07-25 08:30:00	1.72	1.415	52	26.00	1.69	50.6
54	2016-08-08 10:10:00	2.08	1.949	121	89.00	2.24	180.7
55	2016-08-22 09:00:00	1.98	1.763	95	58.00	2.05	116.0
56	2016-09-12 08:50:00	2.67	2.442	470	276.67	2.75	584.5
57	2016-09-26 10:30:00	3.31	2.892	2050	780.00	3.22	1708.6
58	2016-10-11 09:20:00	2.45	2.146	282	140.00	2.45	288.8
59	2016-10-24 09:30:00	1.81	1.415	64	26.00	1.69	50.6
60	2016-11-07 08:40:00	1.75	1.362	56	23.00	1.64	44.6
61	2016-12-12 09:50:00	1.97	1.640	93	43.67	1.92	86.5
62	2017-01-09 09:50:00	<1.18	1.054	<15	11.33	1.32	21.4
63	2017-02-06 10:00:00	1.45	1.342	28	22.00	1.62	42.5
64	2017-03-06 10:40:00	1.79	1.392	62	24.67	1.67	47.9
65	2017-04-10 09:20:00	2.66	2.230	459	170.00	2.53	353.1
66	2017-05-08 09:20:00	2.61	2.230	410	170.00	2.53	353.1
67	2017-05-22 09:20:00	2.99	2.650	970	446.67	2.97	959.6
68	2017-06-05 08:50:00	2.59	2.230	388	170.00	2.53	353.1
69	2017-06-19 09:10:00	2.93	2.368	856	233.33	2.68	490.0
70	2018-11-29 12:20:00	2.29	1.883	193	76.37	2.17	154.3
71	2018-12-18 10:40:00	2.26	2.024	180	105.77	2.32	216.1
72	2019-02-06 10:20:00	2.23	1.946	171	88.33	2.24	179.3
73	2019-03-19 09:20:00	3.06	2.727	1140	532.93	3.05	1152.0
74	2019-04-16 09:30:00	2.02	1.782	105	60.60	2.07	121.4
75	2019-05-09 11:10:00	3.4	3.044	2520	1105.62	3.38	2451.6
76	2019-06-26 12:30:00	2.86	2.586	722	385.50	2.90	823.9
77	2019-07-16 10:10:00	2.22	1.861	165	72.60	2.15	146.4
78	2019-08-20 09:20:00	2.75	2.247	564	176.60	2.55	367.3
79	2019-09-24 09:50:00	2.68	2.451	484	282.23	2.76	596.7

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.]