Appendix 17. Model Archive Summary for Chlorophyll *a* 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 chlorophyll *a* concentration model developed to compute 15-minute, hourly, or daily chlorophyll *a* 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 17, 2016, through August 31, 2021 (dataset consisted of 41 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 41 concomitant values of discretely collected chlorophyll *a* concentration and continuously measured reservoir storage, pH, dissolved oxygen, and chlorophyll fluorescence during February 17, 2016, through August 31, 2021. Discrete samples were collected throughout the range of continuously observed hydrologic conditions. No samples had chlorophyll *a* concentrations that were less than laboratory minimum reporting level. One sample had the analyzed chlorophyll *a* concentration flagged as estimated. 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. Three samples in the model calibration dataset were flagged as outliers but all were retained in the dataset after further review.

Chlorophyll *a* 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 five to eight times per year. All samples were collected between 9:15 a.m. and 12:20 p.m. Samples were analyzed for chlorophyll *a* concentration by the Wichita Municipal Water and Wastewater Laboratory in Wichita, Kans., according to standard methods (Eaton and others, 1995).

Continuous Water-Quality Data

Dissolved oxygen, pH, and chlorophyll fluorescence 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). Reservoir storage was computed using a nonsubmersible pressure transducer which was operated and maintained according to standard USGS methods (Turnipseed and Sauer, 2010; Painter and Loving, 2015). 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

Ordinary least squares linear regression was used to develop surrogate regression models that relate continuous water-quality conditions to discretely sampled constituent concentrations. All regressions were computed using the R software environment (R Core Team, 2020). The data and subsequent regression equation must meet the five assumptions necessary to apply ordinary least squares regression: the dependent variable is linearly related to the explanatory variables, data used to fit the model are representative of the data of interest, the variance of the residuals is constant (homoscedastic), the residuals are independent of the explanatory variables, and the residuals are normally distributed (Helsel and others, 2020). Previously published explanatory variables also were considered for continuity.

Reservoir storage, dissolved oxygen, pH, and chlorophyll fluorescence were selected as good surrogates for chlorophyll *a* concentration based on residual plots, coefficient of determination (R^2), and model standard percentage error (MSPE). 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 chlorophyll *a* (Chla) concentration regression analysis at USGS site 07144780:

Chla concentration-based model:

$$\log_{10}(Chla) = (0.89 \times pH) - (9.09 \times 10^{-6} \times RESSTOR) - (0.0928 \times DO) + (0.843 \times \log_{10}(CHL\_RFU)) - 4.29,$$

where,

$Chla$ = chlorophyll *a*, micrograms per liter (μg/L) (USGS parameter code 70953);

pH = pH, standard units (USGS parameter code 00400);

$RESSTOR$ = reservoir storage, acre-feet (ac-ft) (USGS parameter code 00054);

DO = dissolved oxygen, milligrams per liter (mg/L) (USGS parameter code 00300);

CHL_RFU = chlorophyll fluorescence, relative fluorescence units (RFU) (USGS parameter code 32320); and

$\log_{10}$ = decimal logarithm.

The $\log_{10}$-transformed model may be retransformed to the original units so that chlorophyll *a* concentration 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 1.08. The retransformed model, accounting for BCF, is as follows:

$$Chla = (10^{0.89 \times pH} \times 10^{-9.09e^{-6} \times RESSTOR} \times 10^{-0.0928 \times DO} \times CHL\_RFU^{0.843} \times 10^{-4.29}) \times 1.08.$$

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 chlorophyll *a* concentration using this model is 117.7 micrograms per liter. Computed estimates outside that limit are not supported by the current model calibration dataset.

Model statistics, data, and plots

Definitions

Variable	Explanation
BCF	Bias correction factor, used to correct logarithmic bias (Duan 1983)
Chla	Chlorophyll *a*, micrograms per liter (µg/L) (USGS parameter code 70953)
Cook's D	Cook's distance, a measure of influence (Helsel and others, 2020)
DFFITS	Difference in fits, a measure of influence (Helsel and others, 2020)
DO	Dissolved oxygen, milligrams per liter (mg/L) (USGS parameter code 00300)
E.vars	Explanatory variables
Leverage	An outlier's measure in the x direction (Helsel and others, 2020)
LOESS	Local polynomial regression fitting (Helsel and others, 2020)
logChla	Chlorophyll *a*, micrograms per liter (µg/L), $\log_{10}$-transformed
logCHL_RFU	Chlorophyll fluorescence, relative fluorescence units (RFU), $\log_{10}$-transformed
MSE	Model standard error (Helsel and others, 2020)
MSPE	Model standard percentage error (Helsel and others, 2020)
pH	pH, standard units (USGS parameter code 00400)
Pr(>|*t*|)	The probability that the independent variable has no effect on the dependent variable (Helsel and others, 2020)
RESSTOR	Reservoir storage, acre-feet (ac-ft) (USGS parameter code 00054)
RMSE	Root mean square error (Helsel and others, 2020)
t value	Student's *t* value; the coefficient divided by its associated standard error (Helsel and others, 2020)

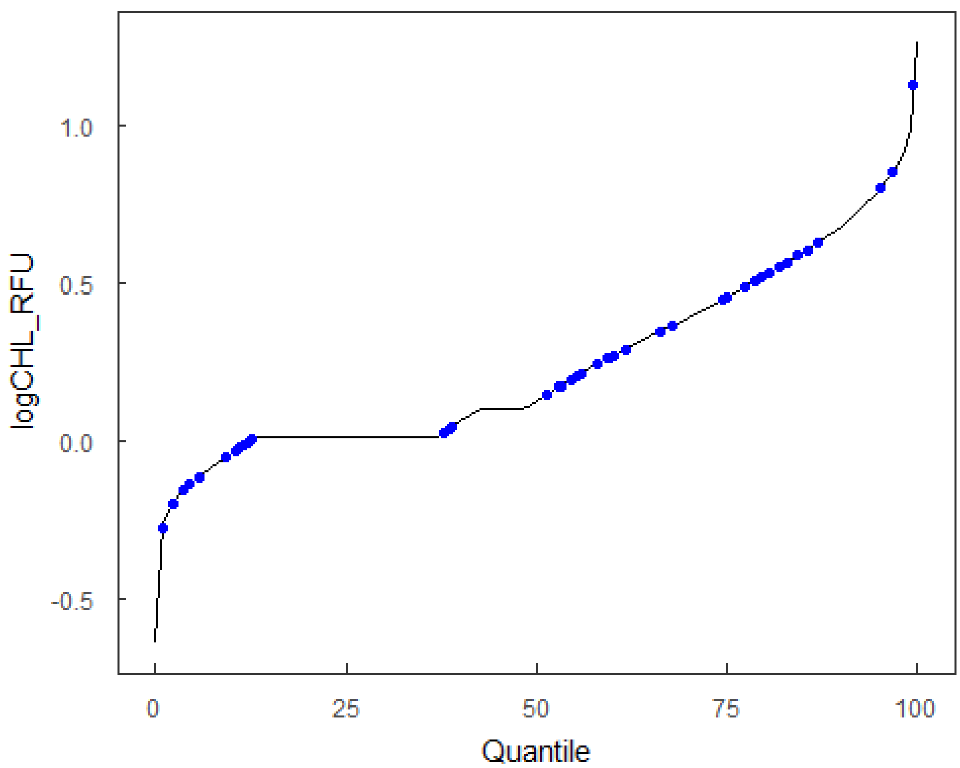
Model

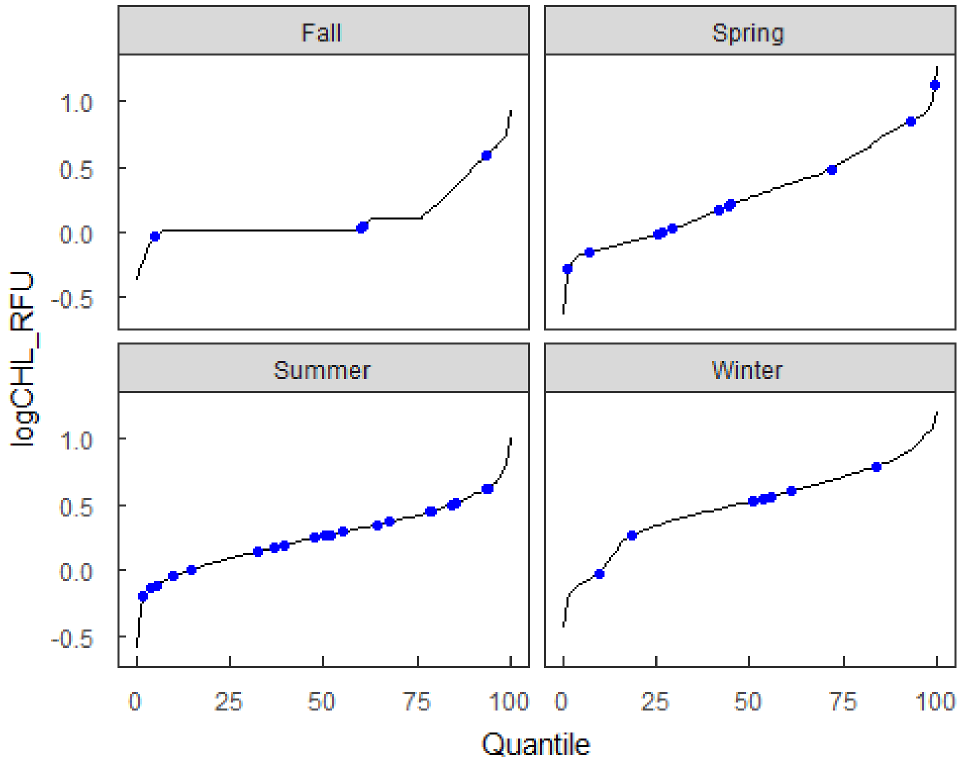
$$\log_{10}(Chla) = (0.89 \times pH) - (9.09 \times 10^{-6} \times RESSTOR) - (0.0928 \times DO) + (0.843 \times \log_{10}(CHL\_RFU)) - 4.29$$

Variable summary statistics

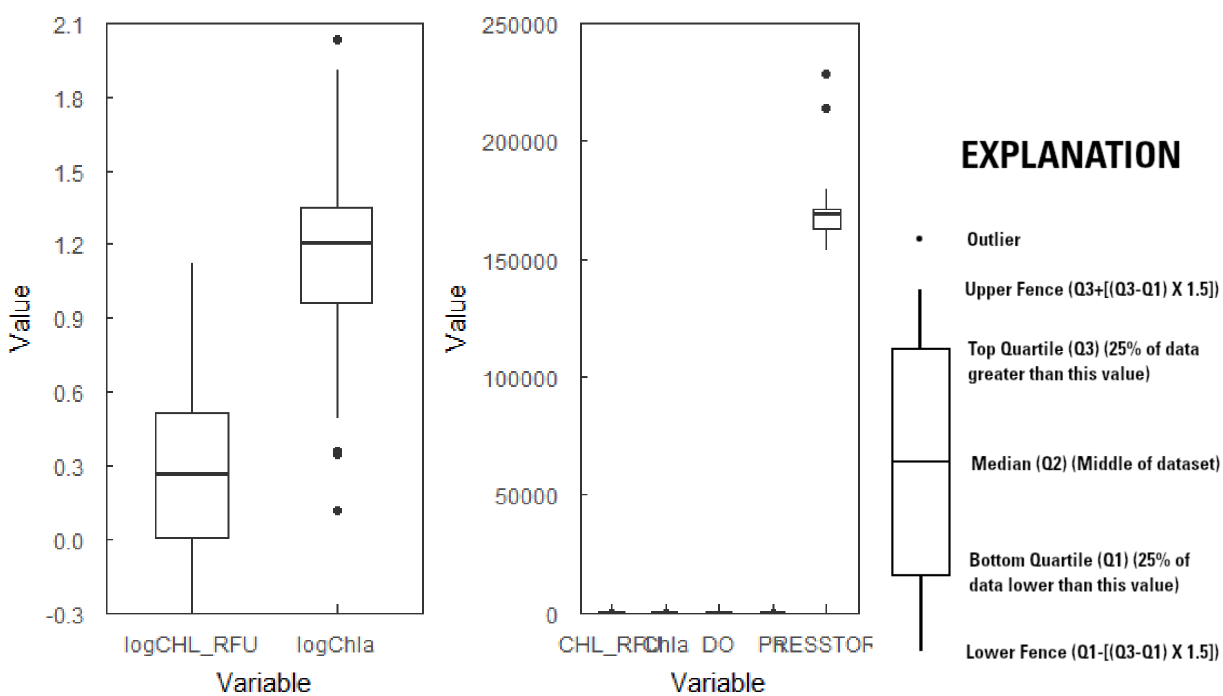
Variable	Minimum	Q1	Median	Mean	Q3	Maximum
CHL_RFU	0.53	1.01	1.82	2.47	3.28	13.2
Chla	1.3	9.1	15.9	20.1	22.3	107
DO	4.48	6.4	7.65	8.68	10.6	14.2
logCHL_RFU	-0.276	0.00432	0.26	0.272	0.516	1.12
logChla	0.114	0.959	1.2	1.13	1.35	2.03
pH	7.99	8.34	8.45	8.46	8.55	9.00
RESSTOR	153,000	163,000	169,000	169,000	171,000	229,000

Duration plots

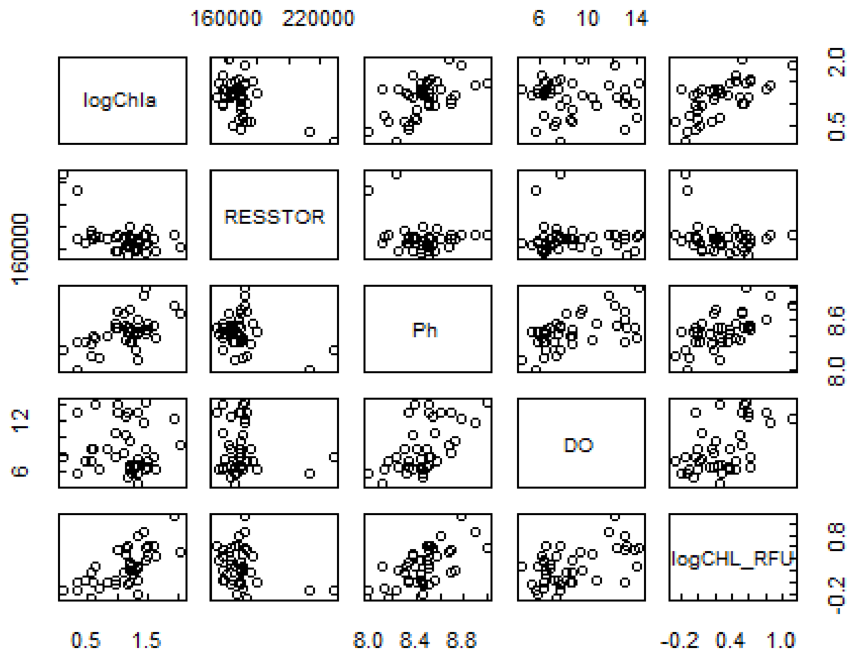



Fall
Spring
Summer
Winter
logCHL_RFU
1.0
0.5
0.0
-0.5
0
25
50
75
100
Quantile

Box plots


2.1
1.8
1.5
1.2
0.9
0.6
0.3
0.0
-0.3
Value
logCHL_RFU
logChla
Variable
250000
200000
150000
100000
50000
0
Value
DO
Variable
EXPLANATION
Outlier
Upper Fence (Q3+[(Q3-Q1) X 1.5])
Top Quartile (Q3) (25% of data greater than this value)
Median (Q2) (Middle of dataset)
Bottom Quartile (Q1) (25% of data lower than this value)
Lower Fence (Q1-[(Q3-Q1) X 1.5])

Scatter plots



The x- and y-axis labels for a given bivariate plot are defined by the intersecting row and column labels.

Basic model statistics

Statistic	Value
Observations	41
R^2	0.832
Adjusted R^2	0.813
RMSE	0.181
Upper MSPE (90%)	51.8
Lower MSPE (90%)	34.1
BCF	1.08

Model coefficients

	Estimate	Standard error	t value	Pr(>|t|)
(Intercept)	-4.2887306	1.7221061	-2.490398	0.0175116
RESSTOR	-0.0000091	0.0000023	-4.027598	0.0002781
Ph	0.8903590	0.1971783	4.515502	0.0000653
DO	-0.0927987	0.0127710	-7.266341	0.0000000
logCHL_RFU	0.8428707	0.1231086	6.846561	0.0000001

Correlation matrix

	logChla	RESSTOR	Ph	DO	logCHL_RFU
logChla	1.0000000	-0.5333856	0.5605849	-0.0240985	0.6641501
RESSTOR	-0.5333856	1.0000000	-0.2881052	-0.1118365	-0.2954378
Ph	0.5605849	-0.2881052	1.0000000	0.5591649	0.6265197
DO	-0.0240985	-0.1118365	0.5591649	1.0000000	0.5179341
logCHL_RFU	0.6641501	-0.2954378	0.6265197	0.5179341	1.0000000

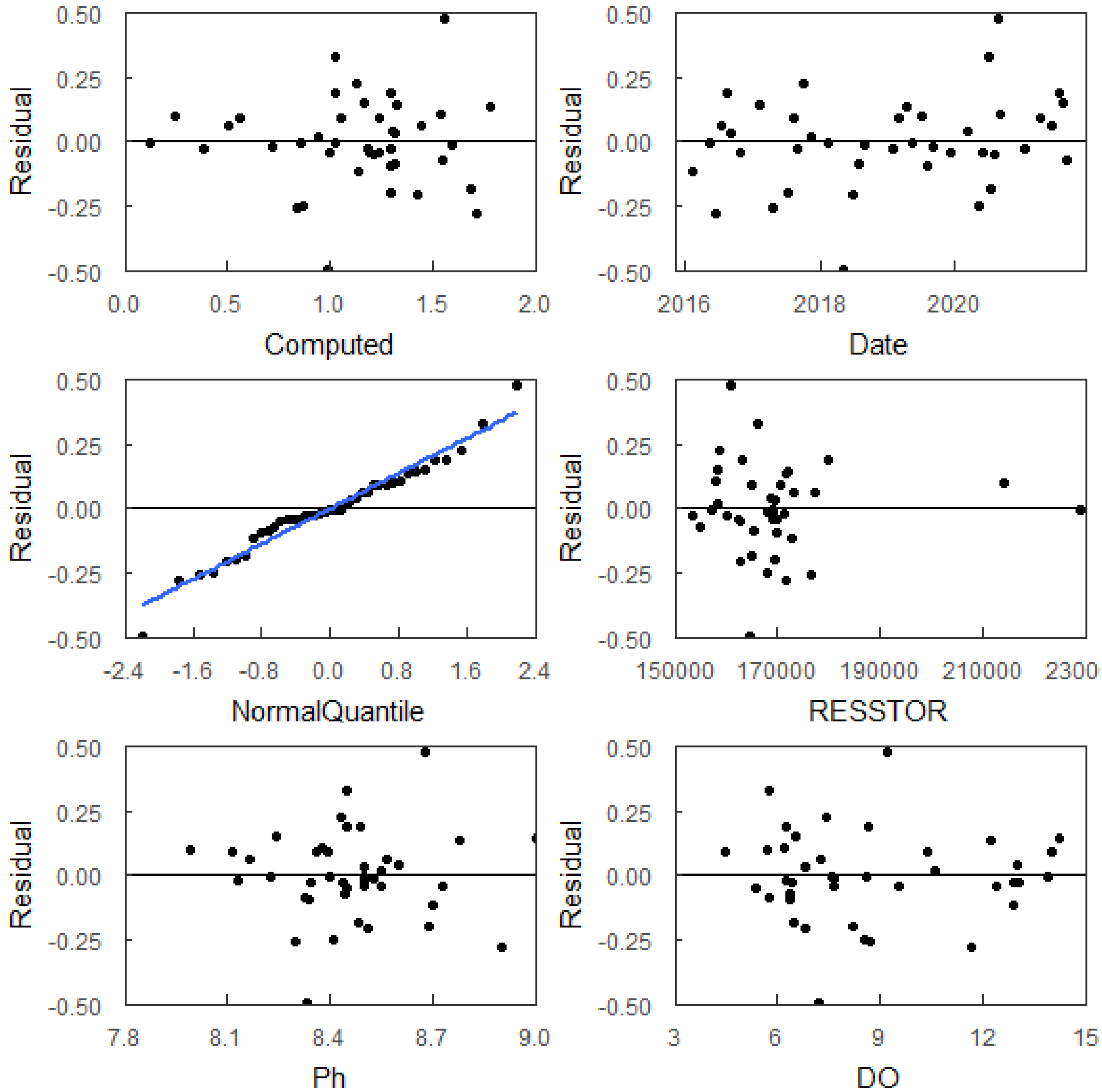
Outlier test criteria

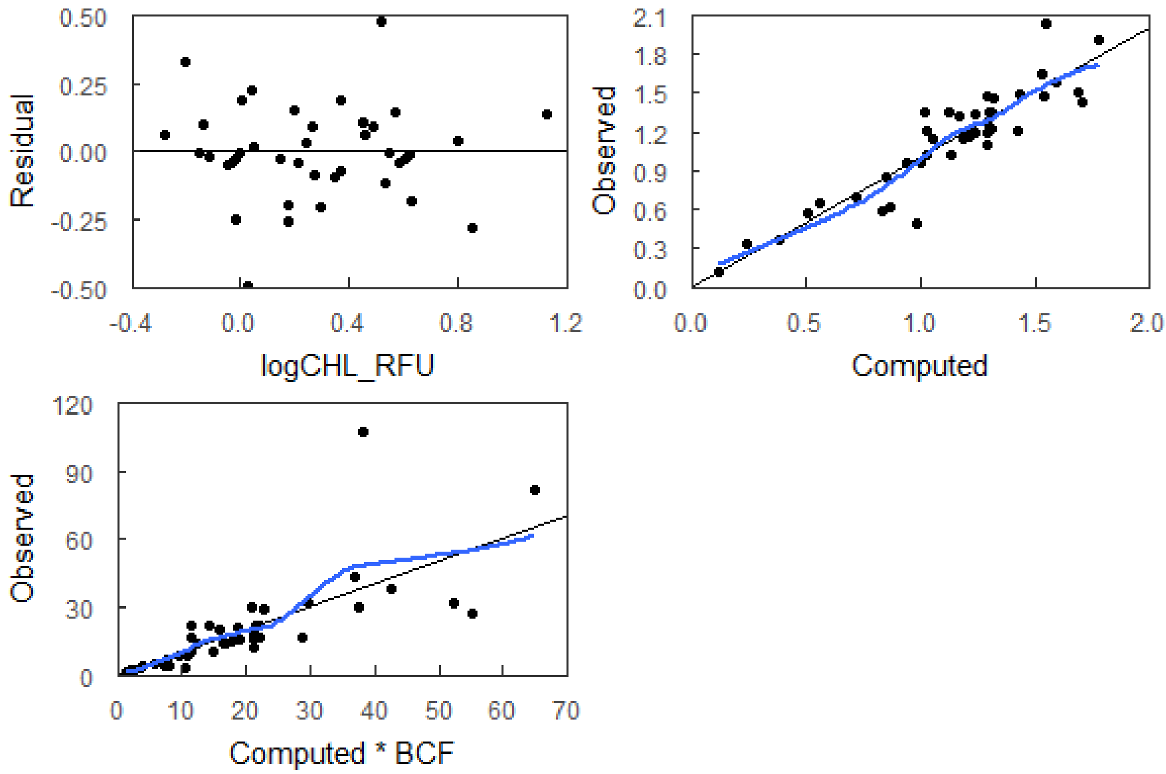
Leverage	DFFITS	CooksD
0.3659	0.6984	0.3587

Flagged observations

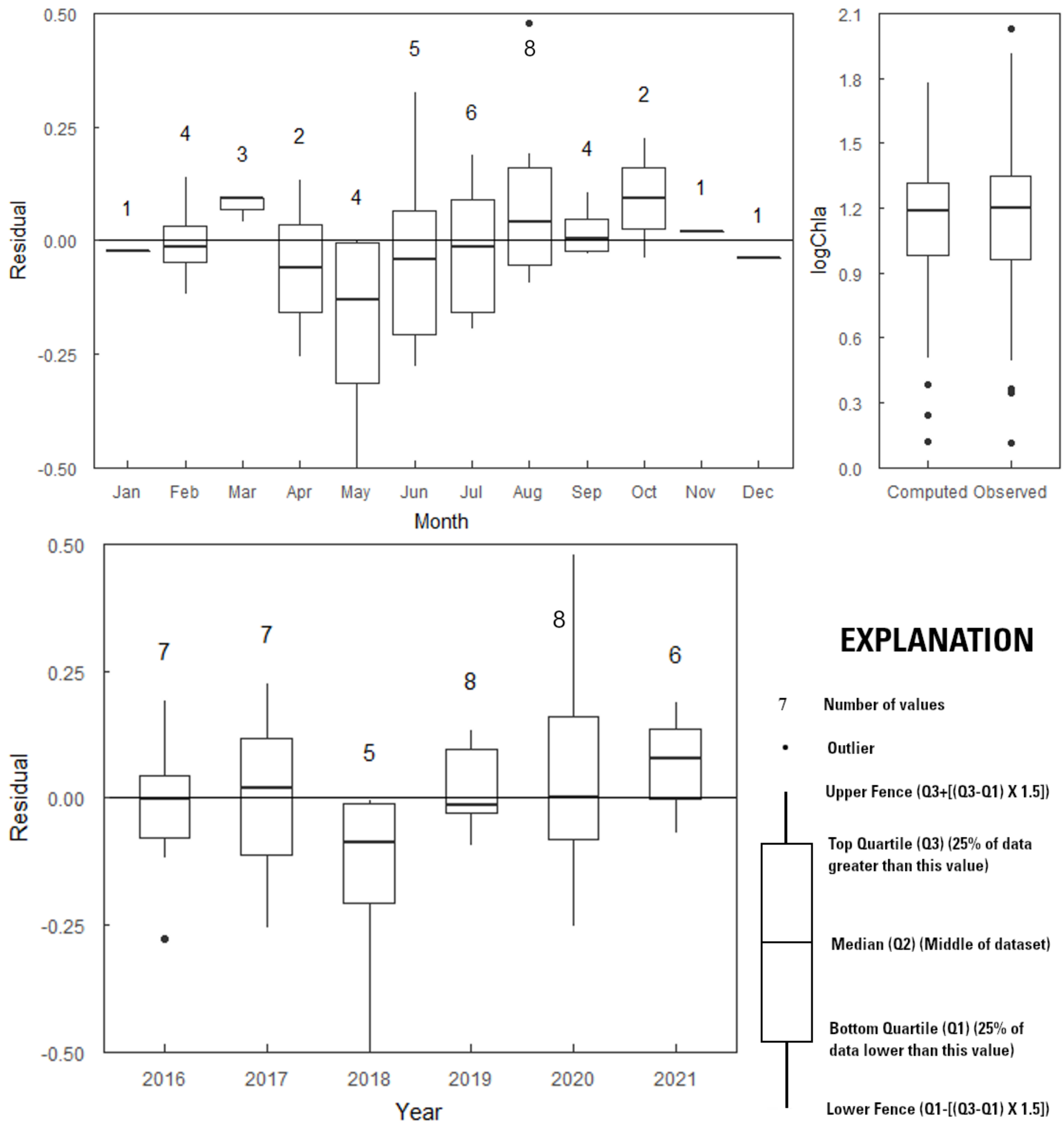
datetime	logChla	CooksD	DFFITS	Leverage	Studentized Residual
2019-05-14 11:10:00	0.114	0.000589	-0.0535	0.512	-0.0522
2020-06-25 11:30:00	1.35	0.116	0.793	0.134	2.02
2020-08-18 11:40:00	2.03	0.0981	0.776	0.0617	3.02

Statistical plots



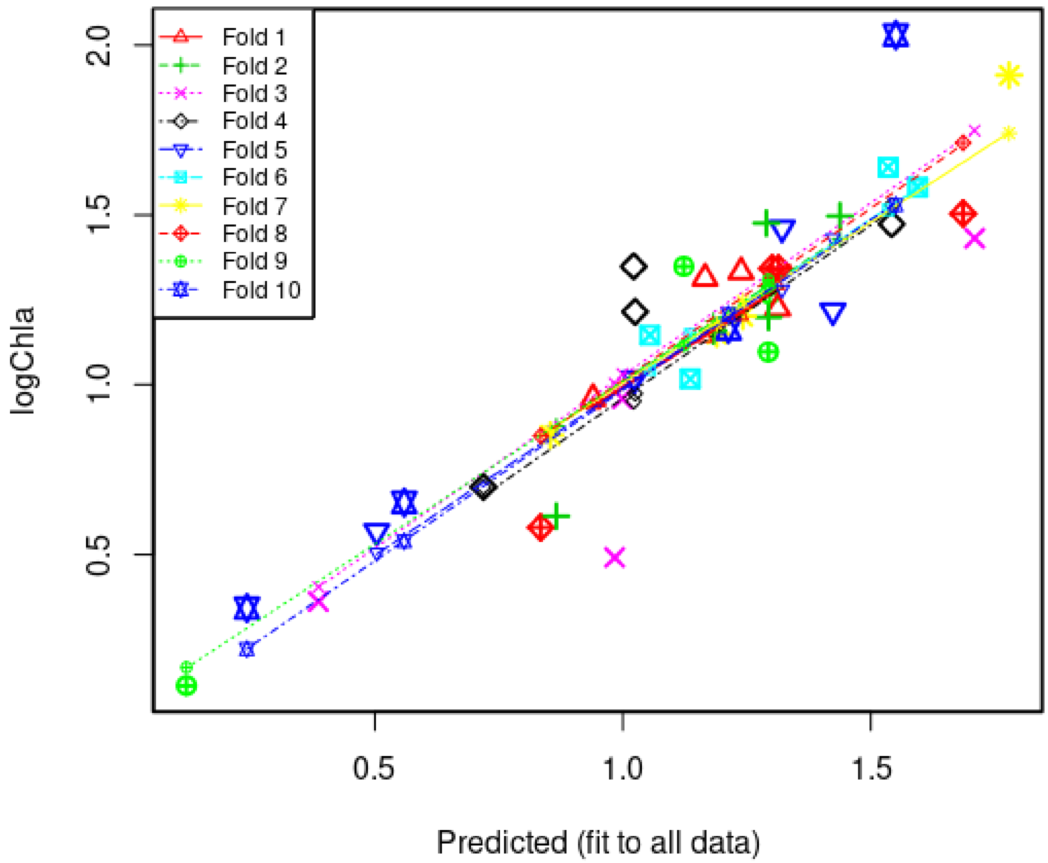


The blue line shows the locally estimated scatterplot smoothing (LOESS). The black dots correspond to observed values. The black line represents the 1:1 line.



0.50
0.25
0.00
-0.25
-0.50
Residual
1
4
3
2
4
5
6
8
4
2
1
1
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
2.1
1.8
1.5
1.2
0.9
0.6
0.3
0.0
logChla
Computed
Observed
7
7
5
8
8
6
2016
2017
2018
2019
2020
2021
Year
EXPLANATION
7 Number of values
Outlier
Upper Fence (Q3+[(Q3-Q1) X 1.5])
Top Quartile (Q3) (25% of data greater than this value)
Median (Q2) (Middle of dataset)
Bottom Quartile (Q1) (25% of data lower than this value)
Lower Fence (Q1-[(Q3-Q1) X 1.5])

Cross Validation



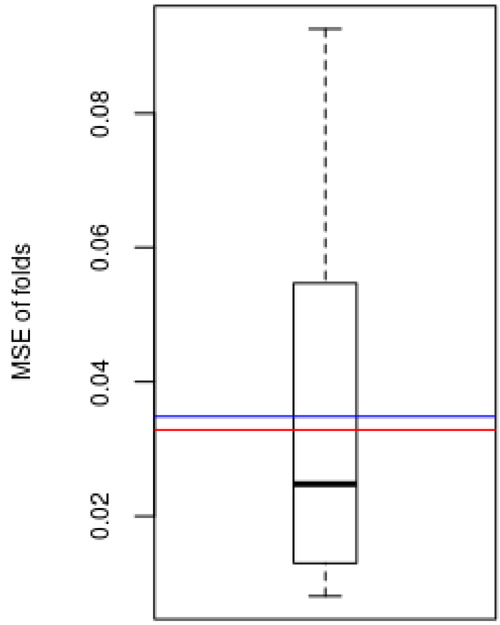
Fold - equal partition of the data (10 percent of the data).

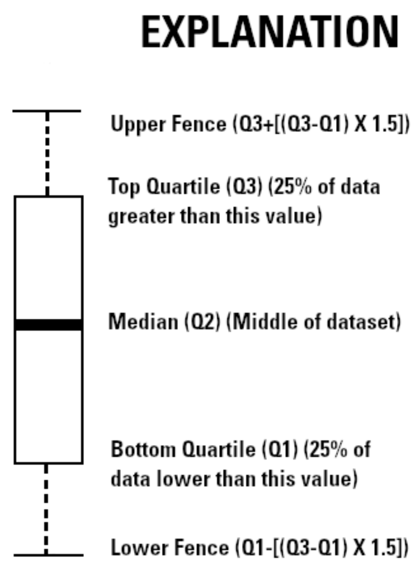
Large symbols – observed value of a data point removed in a fold.

Small symbols – recomputed value of a data point removed in a fold.

Recomputed regression lines – adjusted regression line with one fold removed.

Statistic	Value
Minimum MSE of folds	0.00803
Median MSE of folds	0.02480
Mean MSE of folds	0.03480
Maximum MSE of folds	0.09260
(Mean MSE of folds) / (Model MSE)	1.06000





Red line - Model MSE

Blue line - Mean MSE of folds

Model calibration dataset

datetime	logChla	RESSTOR	pH	DO	logCHL_RFU	Chla	Computed	Retransformed
2016-02-17 10:45:00	1.02	173,000	8.70	12.9	0.531	10.4	1.14	14.8
2016-05-17 10:20:00	0.851	169,000	8.40	8.60	-0.00436	7.1	0.854	7.71
2016-06-15 09:15:00	1.43	172,000	8.90	11.6	0.848	27	1.71	55.2
2016-07-18 10:40:00	1.5	177,000	8.57	7.27	0.454	31.4	1.44	29.6
2016-08-15 10:30:00	1.21	180,000	8.45	6.25	0.00432	16.4	1.02	11.4
2016-09-06 10:40:00	1.34	169,000	8.50	6.80	0.244	22	1.31	22.2
2016-10-25 10:15:00	0.959	170,000	8.50	7.66	-0.0298	9.1	0.999	10.8
2017-02-09 10:40:00	1.46	172,000	9.00	14.2	0.565	28.9	1.32	22.6
2017-04-17 10:30:00	0.58	176,000	8.30	8.70	0.171	3.8	0.835	7.38
2017-07-10 11:40:00	1.1	170,000	8.69	8.18	0.173	12.5	1.29	21.2
2017-08-15 10:00:00	1.33	165,000	8.11	4.48	0.26	21.5	1.24	18.7

2017-09-07 10:00:00	1.27	160,000	8.44	6.40	0.145	18.5	1.3	21.3
2017-10-03 10:20:00	1.35	159,000	8.43	7.40	0.0383	22.3	1.12	14.3
2017-11-13 12:00:00	0.959	158,000	8.55	10.6	0.0453	9.1	0.94	9.41
2018-02-13 10:40:00	1.02	157,000	8.50	13.9	0.547	10.4	1.02	11.4
2018-05-08 10:30:00	0.491	165,000	8.33	7.20	0.0253	3.1	0.984	10.4
2018-06-25 12:00:00	1.22	163,000	8.51	6.81	0.29	16.5	1.42	28.6
2018-07-26 11:40:00	1.23	165,000	8.32	5.76	0.27	16.8	1.31	22.2
2018-08-29 11:00:00	1.58	168,000	8.53	7.65	0.623	38.2	1.59	42.4
2019-02-05 11:20:00	0.362	169,000	8.34	12.9	-0.0208	2.3	0.387	2.63
2019-03-07 10:50:00	0.653	170,000	8.39	14.0	0.264	4.5	0.56	3.92
2019-04-09 10:30:00	1.91	172,000	8.78	12.2	1.12	81.5	1.78	64.9
2019-05-14 11:10:00	0.114	229,000	8.22	7.60	-0.153	1.3	0.121	1.43
2019-07-09 10:15:00	0.342	214,000	7.99	5.68	-0.137	2.2	0.243	1.89
2019-08-06 11:00:00	1.2	170,000	8.34	6.36	0.344	15.8	1.29	21.2
2019-09-03 10:40:00	0.699	171,000	8.13	6.24	-0.114	5	0.719	5.66
2019-12-04 10:50:00	1.15	163,000	8.55	12.4	0.585	14.1	1.19	16.7
2020-03-04 11:00:00	1.34	169,000	8.60	13.0	0.796	22	1.3	21.6
2020-05-06 10:30:00	0.613	168,000	8.41	8.54	-0.0153	4.1	0.866	7.94
2020-06-03 10:20:00	1.2	169,000	8.73	9.52	0.213	15.9	1.24	18.8
2020-06-25 11:30:00	1.35	166,000	8.45	5.74	-0.201	22.3	1.02	11.4
2020-07-15 10:00:00	1.5	165,000	8.48	6.48	0.624	31.9	1.69	52.4
2020-08-04 11:30:00	1.16	163,000	8.45	5.37	-0.0503	14.6	1.21	17.6
2020-08-18 11:40:00	2.03	161,000	8.67	9.22	0.516	107	1.55	38.3
2020-09-01 10:50:00	1.64	158,000	8.37	6.17	0.447	43.7	1.53	37
2021-01-13 10:30:00	1.16	153,000	8.50	13.0	0.602	14.4	1.18	16.4
2021-03-31 10:30:00	1.15	170,000	8.36	10.4	0.486	14	1.05	12.2
2021-06-03 10:30:00	0.568	173,000	8.16	7.23	-0.276	3.7	0.504	3.45
2021-07-20 10:40:00	1.48	163,000	8.49	8.65	0.366	29.9	1.29	21
2021-08-10 10:00:00	1.31	158,000	8.24	6.53	0.193	20.6	1.17	15.8
2021-08-31 11:40:00	1.47	155,000	8.44	6.34	0.366	29.7	1.54	37.6

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.

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.

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/

Painter, C.C., and Loving, B.L., 2015, U.S. Geological Survey quality-assurance plan for surface-water activities in Kansas, 2015: U.S. Geological Survey Open-File Report 2015-1074, 33 p., https://doi.org/10.3133/ofr20151074.

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/.

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.]

Turnipseed, D.P., and Sauer, V.B., 2010, Discharge measurements at gaging stations: U.S. Geological Survey Techniques and Methods book 3, chap. A8, 87 p., accessed July 13, 2022, at https://doi.org/10.3133/tm3A8.

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.