Appendix 22. Model Archival Summary for Chlorophyll *a* Concentration at U.S. Geological Survey Site 06892350, Kansas River at De Soto, Kansas, during June 2014 through September 2019

This model archival summary summarizes the chlorophyll *a* (Chla; U.S. Geological Survey [USGS] parameter code 32209 [June 30, 2014, through May 22, 2017] or 70953 [June 5, 2017, through September 24, 2019]) concentration model developed to compute 15-minute Chla concentrations from June 2014 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 (fCHL) 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: June 30, 2014, through September 24, 2019

Model application date: June 30, 2014, 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, fCHL, 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 78 concurrent measurements of Chla concentration and sensor-measured fCHL during June 30, 2014, through September 24, 2019. Samples were collected throughout the range of continuously observed hydrologic conditions. No samples had concentrations below the laboratory detection limits. Four sample concentrations were qualified as “estimated.” 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, studentized residuals from the final model were inspected for values greater than three or less than negative three. Values outside of that range were considered potential outliers and were investigated. 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.

Chlorophyll *a* 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 Chla concentration at the USGS Kansas Water Science Center in Lawrence, Kans., during June 30, 2014, through May 22, 2017 (USGS parameter code 32209), and at the USGS National Water Quality Laboratory in Lakewood, Colorado during June 5, 2017, through September 24, 2019 (USGS parameter code 70953).

Model Development

Ordinary least squares regression analysis was done using R programming language (R Core Team, 2020) to relate discretely collected Chla concentration to sensor-measured fCHL. 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.

fCHL was selected as a good surrogate for Chla based on residual plots, coefficient of determination (R^2), and model standard percentage 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 regression analysis for Chla concentration at USGS site 06892350:

Chla concentration-based model:

$$\log Chla = 1.08 \times \log fCHL + 0.751$$

where

log = logarithm base 10;

Chla = chlorophyll *a* concentration, in micrograms per liter; and

fCHL = chlorophyll fluorescence, in relative fluorescence units.

fCHL makes physical and statistical sense as an explanatory variable for Chla because chlorophyll *a* pigments fluoresce when irradiated by certain wavelengths of light emitted from the fCHL sensor.

The logarithmically (log) transformed model may be retransformed to the original units so that Chla 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:

$$Chla = 1.08 \times (fCHL^{1.08} \times 10^{0.751})$$

Previous Models

There are no previously published models for hardness as calcium carbonate at this site.

Model Statistics, Data, and Plots

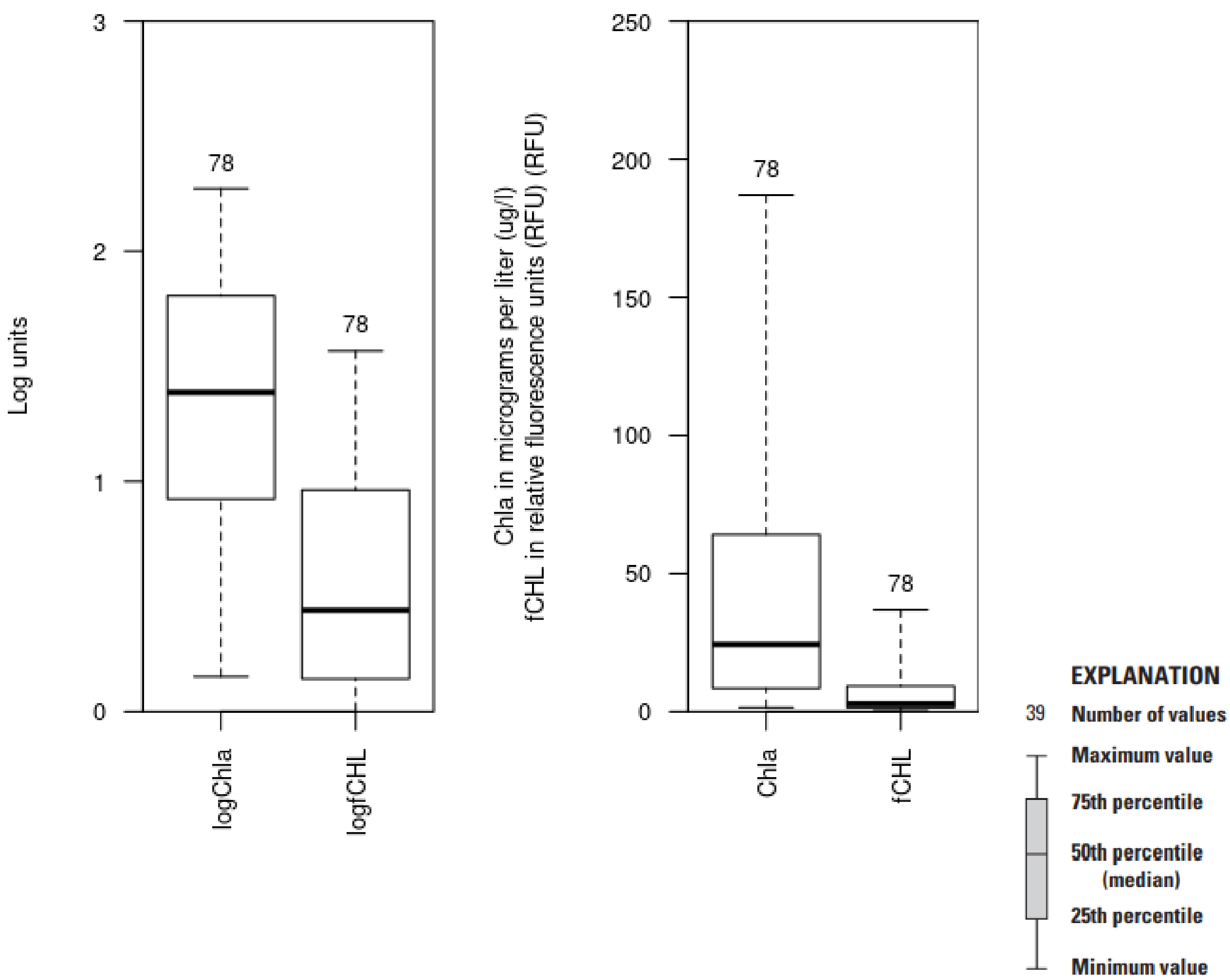
Model

logChla = + 1.08 * logfCHL + 0.751

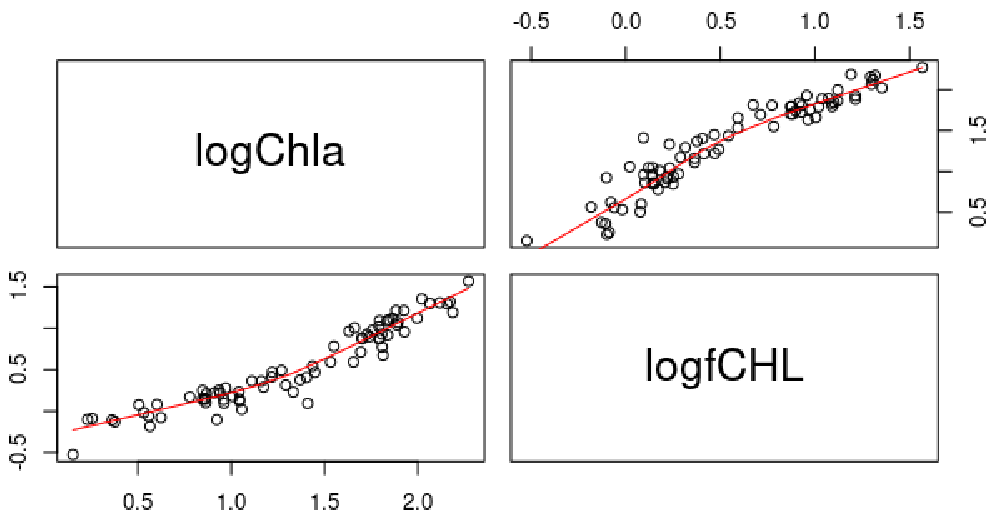
Variable Summary Statistics

	logChla	Chla	logfCHL	fCHL
Minimum	0.152	1.42	-0.523	0.30
1st Quartile	0.923	8.37	0.143	1.39
Median	1.390	24.30	0.440	2.76
Mean	1.340	40.90	0.542	6.21
3rd Quartile	1.810	64.10	0.963	9.19
Maximum	2.270	187.00	1.570	36.90

Box Plots



Exploratory Plots



Red line shows the locally weighted scatterplot smoothing (LOWESS).

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

Basic Model Statistics

Number of Observations	78
Standard error (RMSE)	0.171
Average Model standard percentage error (MSPE)	40.4
Coefficient of determination (R^2)	0.904
Adjusted Coefficient of Determination (Adj. R^2)	0.902
Bias Correction Factor (BCF)	1.08

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.751	0.0292	25.7	3.11e-39
logfCHL	1.080	0.0404	26.7	2.32e-40

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.749
E.vars	-0.749	1.000

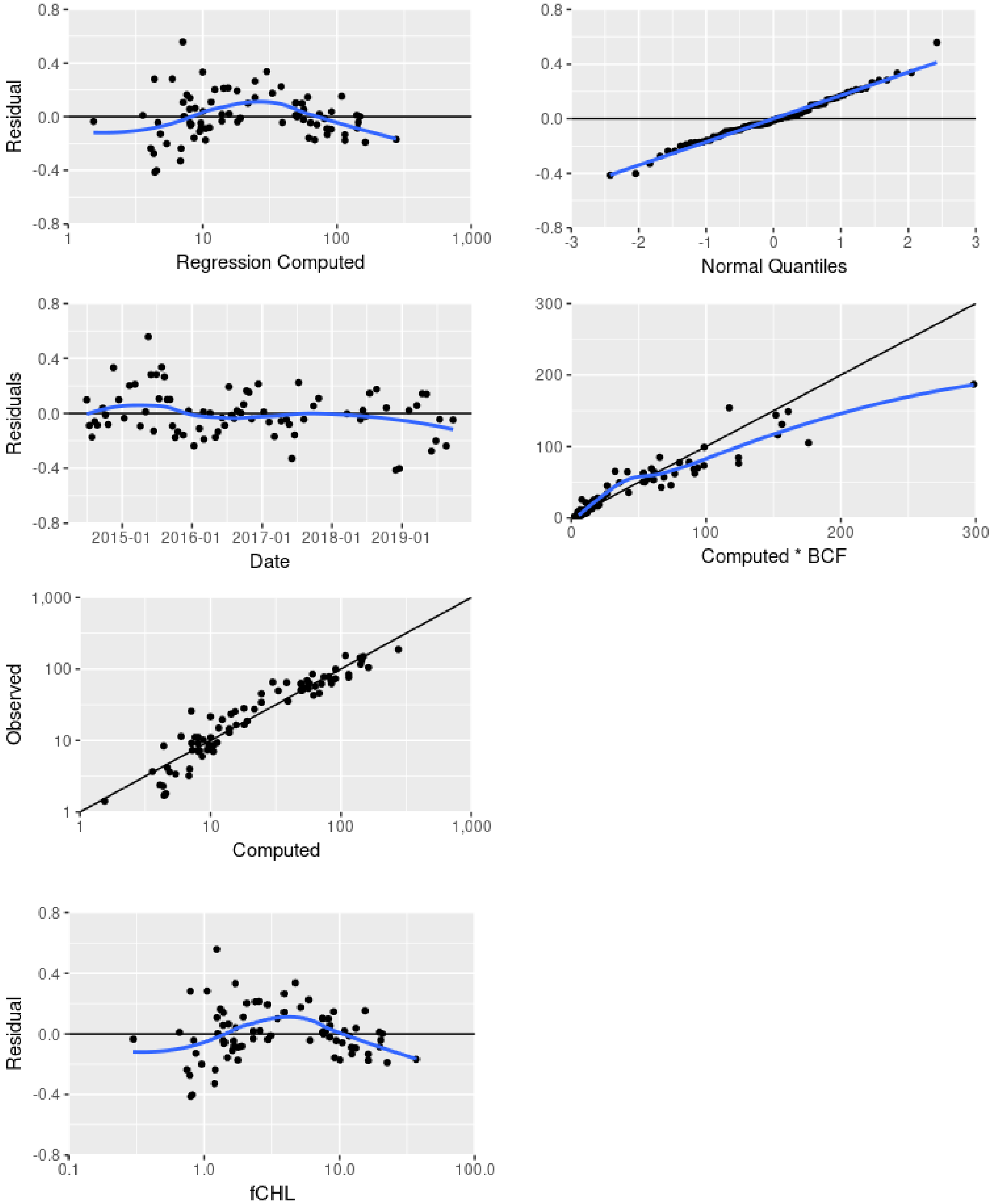
Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.0769	0.1943	0.3203

Flagged Observations

	logChla	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
201505181530	1.410	0.852	0.558	3.30	3.54	0.0240	0.1340	0.556
201506011430	0.923	0.641	0.282	1.68	1.70	0.0359	0.0524	0.328
201811291220	0.230	0.644	-0.414	-2.46	-2.55	0.0357	0.1120	-0.491
201812181040	0.255	0.657	-0.402	-2.39	-2.47	0.0349	0.1030	-0.469
201906031020	0.362	0.636	-0.274	-1.63	-1.65	0.0362	0.0502	-0.320

Statistical Plots



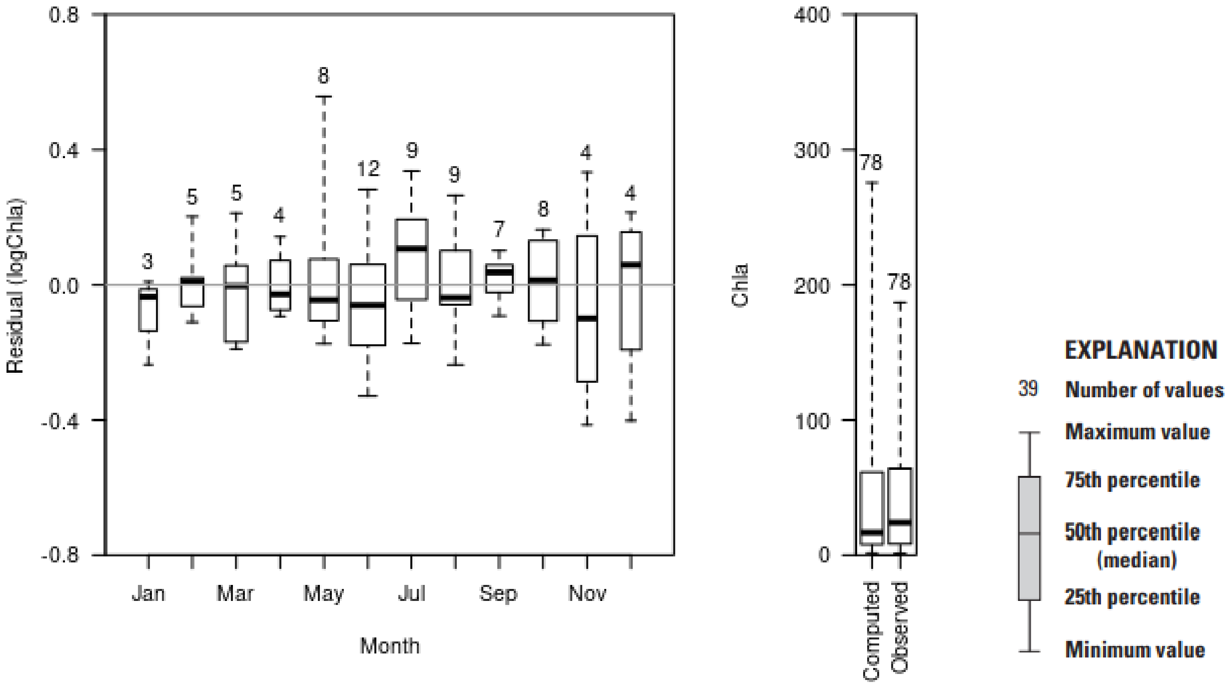
First row (left): Residual Chla related to regression computed Chla with local polynomial regression fitting, or locally estimated scatterplot smoothing (LOESS), indicated by the blue line.

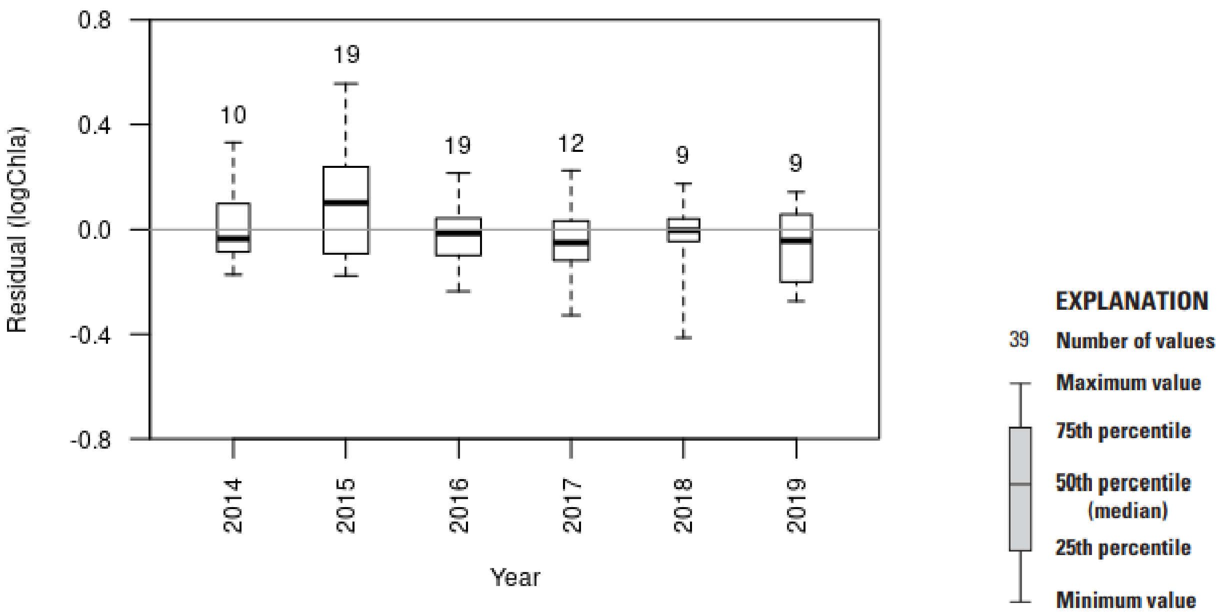
First row (right): Residual Chla related to the corresponding normal quantile of the residual with simple linear regression, indicated by the blue line.

Second row: Residual Chla related to date (left) and regression computed Chla multiplied by the BCF (right) with LOESS, indicated by the blue line.

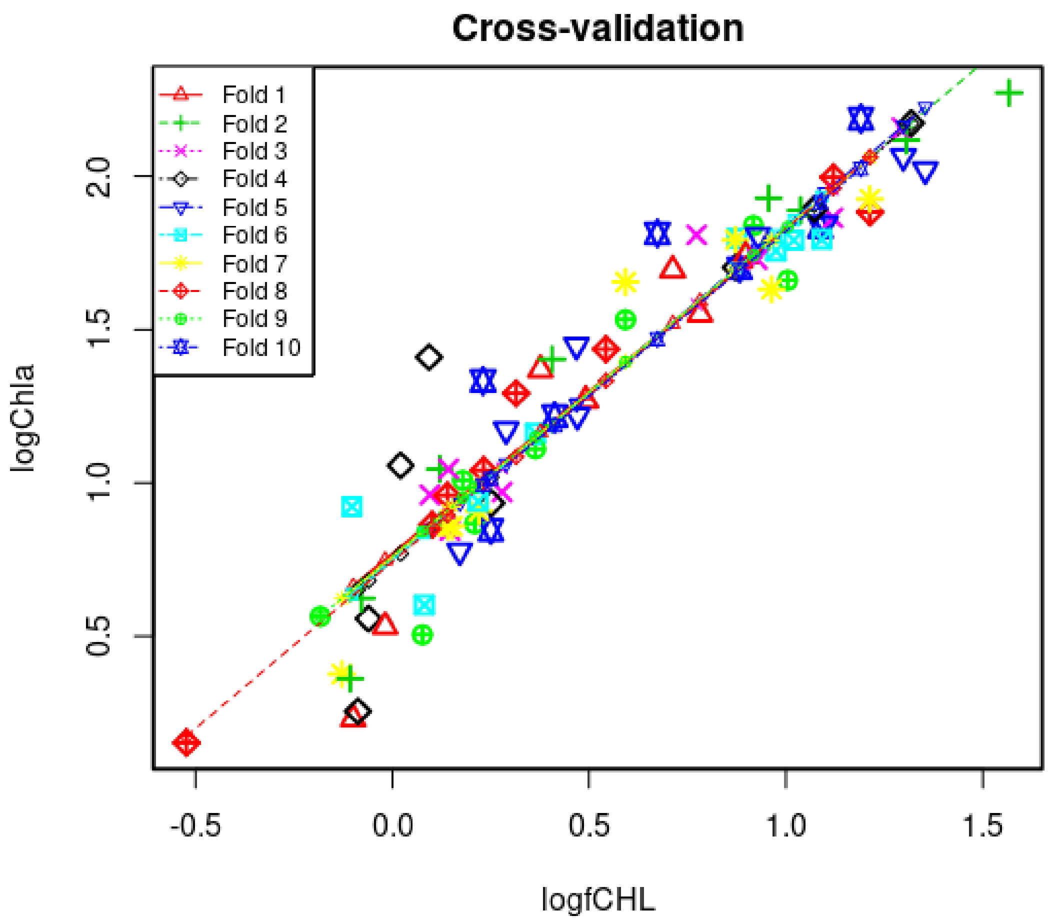
Third row: Observed Chla related to regression computed Chla.

Fourth row: Residual Chla related to fCHL with LOESS, indicated by the blue line.


0.8
0.4
0.0
-0.4
-0.8
Residual (logChla)
3
5
5
4
8
12
9
9
7
8
4
4
Jan
Mar
May
Jul
Sep
Nov
Month
400
300
200
100
0
Chla
78
78
Computed
Observed
EXPLANATION
39 Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value


0.8
0.4
0.0
-0.4
-0.8
Residual (logChla)
10
19
19
12
9
9
2014
2015
2016
2017
2018
2019
Year
EXPLANATION
39 Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value

Cross-Validation



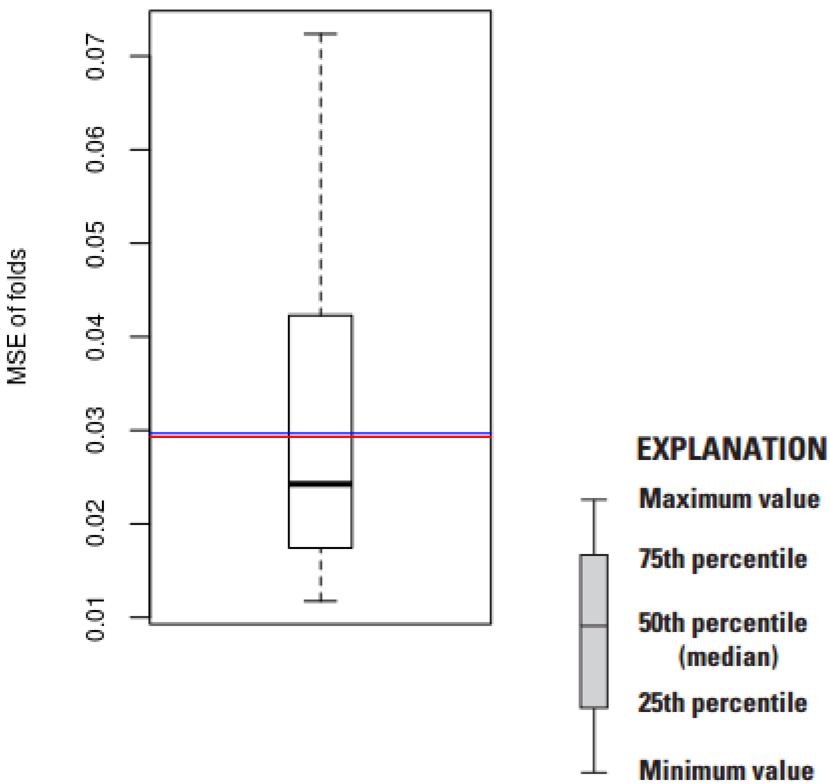

```
Fold - equal partition of the data (10 percent of the data).

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.


              Minimum MSE of folds:  0.0117
                 Mean MSE of folds:  0.0296
               Median MSE of folds:  0.0242
              Maximum MSE of folds:  0.0724
 (Mean MSE of folds) / (Model MSE):  1.0100
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Model-Calibration Dataset

0	Date	logChla	logfCHL	Chla	fCHL	Computed logChla	Computed Chla	Residual	Normal Quantiles	Censored Values
1	2014-06-30	1.84	0.917	69	8.27	1.74	59.5	0.0987	0.556	--
2	2014-07-14	1.83	1.09	68.3	12.2	1.92	90.8	-0.0892	-0.633	--
3	2014-07-28	1.66	1	45.8	10.1	1.83	73.9	-0.173	-1.1	--
4	2014-08-11	1.79	1.02	61.9	10.5	1.85	76.8	-0.059	-0.412	--
5	2014-08-25	2.06	1.3	116	19.9	2.15	153	-0.0862	-0.556	--
6	2014-09-22	2	1.12	99.2	13.2	1.96	98.6	0.0372	0.377	--
7	2014-10-06	1.27	0.491	18.6	3.1	1.28	20.7	-0.0115	0.016	--
8	2014-10-20	0.97	0.279	9.34	1.9	1.05	12.2	-0.0814	-0.519	--
9	2014-11-17	1.33	0.23	21.5	1.7	1	10.8	0.333	1.84	--
10	2014-12-15	1.44	0.543	27.3	3.49	1.34	23.5	0.0999	0.594	--
11	2015-01-12	0.152	-0.523	1.42	0.3	0.188	1.67	-0.0353	-0.112	--
12	2015-02-09	1.29	0.315	19.6	2.06	1.09	13.3	0.202	1.16	--
13	2015-03-09	1.37	0.377	23.4	2.38	1.16	15.5	0.212	1.23	--
14	2015-04-06	1.86	1.12	73.2	13.2	1.96	98.3	-0.0936	-0.713	--
15	2015-05-04	2.16	1.29	144	19.7	2.15	152	0.0116	0.21	--
16	2015-05-18	1.41	0.0934	25.7	1.24	0.852	7.7	0.558	2.42	--
17	2015-06-01	0.923	-0.102	8.37	0.79	0.641	4.74	0.282	1.56	--
18	2015-06-15	0.558	-0.0605	3.61	0.87	0.686	5.25	-0.129	-0.799	--
19	2015-06-29	1.06	0.0212	11.4	1.05	0.774	6.44	0.283	1.68	--
20	2015-07-13	0.961	0.0955	9.15	1.25	0.854	7.74	0.107	0.713	--

21	2015-07-27	1.81	0.673	65.1	4.71	1.48	32.5	0.336	2.04	--
22	2015-08-10	1.66	0.593	45.2	3.91	1.39	26.6	0.265	1.46	--
23	2015-08-24	1.8	0.875	62.6	7.51	1.7	53.6	0.102	0.672	--
24	2015-09-08	1.79	0.872	62	7.44	1.69	53.1	0.102	0.633	--
25	2015-09-21	1.85	1.1	70	12.6	1.94	93.7	-0.0924	-0.672	--
26	2015-10-05	1.88	1.21	76.4	16.4	2.06	124	-0.176	-1.23	--
27	2015-10-19	1.93	1.21	84.2	16.3	2.06	124	-0.134	-0.89	--
28	2015-11-16	1.63	0.963	42.8	9.19	1.79	66.7	-0.158	-0.99	--
29	2015-12-14	1.16	0.364	14.5	2.31	1.14	15.1	0.0181	0.243	--
30	2016-01-11	0.377	-0.128	2.38	0.744	0.613	4.44	-0.237	-1.46	--
31	2016-02-08	0.867	0.21	7.36	1.62	0.977	10.3	-0.11	-0.755	--
32	2016-02-29	1.7	0.872	50.4	7.45	1.69	53.2	0.011	0.177	--
33	2016-03-03	2.02	1.35	105	22.6	2.21	176	-0.19	-1.3	--
34	2016-04-04	2.17	1.32	149	20.8	2.17	161	0.00102	0.0802	--
35	2016-05-02	0.847	0.25	7.03	1.78	1.02	11.4	-0.174	-1.16	--
36	2016-05-16	1.79	1.09	62.3	12.3	1.93	91.6	-0.133	-0.843	--
37	2016-06-06	1.11	0.364	12.9	2.31	1.14	15.1	-0.0327	-0.0802	--
38	2016-06-20	0.934	0.252	8.6	1.79	1.02	11.4	-0.0885	-0.594	--
39	2016-07-11	1.45	0.468	28.1	2.94	1.26	19.5	0.193	1.1	--
40	2016-07-25	1.89	1.07	78.1	11.8	1.91	87.3	-0.0141	-0.016	--
41	2016-08-08	1.22	0.47	16.6	2.95	1.26	19.6	-0.0382	-0.145	--
42	2016-08-22	1.89	1.04	77.4	10.9	1.87	80.2	0.0191	0.276	--
43	2016-09-12	0.863	0.102	7.3	1.26	0.861	7.86	0.00259	0.112	--
44	2016-09-26	1.01	0.179	10.2	1.51	0.944	9.52	0.0644	0.519	--
45	2016-10-11	1.05	0.121	11.1	1.32	0.881	8.24	0.164	0.99	--
46	2016-10-24	2.19	1.19	154	15.5	2.03	117	0.153	0.939	--
47	2016-11-07	2.12	1.31	131	20.2	2.16	156	-0.042	-0.177	--
48	2016-12-12	1.4	0.406	25.3	2.55	1.19	16.7	0.214	1.3	--
49	2017-01-09	0.565	-0.183	3.67	0.657	0.554	3.88	0.0103	0.145	--
50	2017-02-06	0.845	0.146	7	1.4	0.909	8.78	-0.0637	-0.447	--
51	2017-03-06	2.27	1.57	187	36.9	2.44	299	-0.169	-1.04	--
52	2017-04-10	0.859	0.152	7.23	1.42	0.915	8.91	-0.0563	-0.377	--
53	2017-05-08	0.859	0.141	7.22	1.38	0.903	8.66	-0.0447	-0.276	--
54	2017-05-22	0.908	0.22	8.1	1.66	0.989	10.5	-0.0801	-0.483	--
55	2017-06-05	0.505	0.0768	3.2	1.19	0.834	7.39	-0.329	-1.84	--
56	2017-06-19	0.778	0.171	6	1.48	0.936	9.34	-0.158	-0.939	--
57	2017-07-10	1.81	0.773	64.4	5.93	1.58	41.6	0.225	1.38	--
58	2017-08-07	1.55	0.782	35.5	6.05	1.59	42.5	-0.0438	-0.243	--
59	2017-09-26	1.81	0.93	64.1	8.5	1.75	61.4	0.0535	0.447	--
60	2017-10-23	1.17	0.289	14.9	1.95	1.06	12.5	0.11	0.755	--
61	2018-03-20	1.7	0.882	50.1	7.63	1.7	54.6	-0.00261	0.0481	--
62	2018-05-29	1.76	0.975	57.1	9.44	1.8	68.6	-0.0455	-0.309	--
63	2018-06-11	1.74	0.897	55	7.9	1.72	56.7	0.0216	0.343	--
64	2018-06-25	1.73	0.927	53.5	8.46	1.75	61	-0.0225	-0.0481	--
65	2018-07-16	1.93	0.956	84.8	9.04	1.78	65.5	0.146	0.89	--
66	2018-08-21	1.69	0.712	49.5	5.16	1.52	35.8	0.175	1.04	--
67	2018-10-11	1.04	0.232	11	1.71	1	10.9	0.0398	0.412	--
68	2018-11-29	0.23	-0.0994	1.7	0.795	0.644	4.77	-0.414	-2.42	--
69	2018-12-18	0.255	-0.0873	1.8	0.818	0.657	4.92	-0.402	-2.04	--
70	2019-02-06	1.22	0.413	16.5	2.59	1.2	17	0.0212	0.309	--
71	2019-03-19	0.959	0.14	9.1	1.38	0.902	8.64	0.057	0.483	--
72	2019-04-16	1.53	0.592	34.1	3.91	1.39	26.6	0.143	0.843	--
73	2019-05-09	1.05	0.143	11.1	1.39	0.905	8.7	0.14	0.799	--
74	2019-06-03	0.362	-0.107	2.3	0.782	0.636	4.68	-0.274	-1.68	--
75	2019-06-26	0.531	-0.018	3.4	0.959	0.732	5.84	-0.2	-1.38	--

76	2019-07-16	0.623	-0.0782	4.2	0.835	0.667	5.03	-0.0437	-0.21	--
77	2019-08-20	0.602	0.0816	4	1.21	0.839	7.48	-0.237	-1.56	--
78	2019-09-24	0.94	0.218	8.7	1.65	0.987	10.5	-0.0472	-0.343	--

Definitions

Chla: Chlorophyll *a*, in micrograms per liter (32209 during June 30, 2014, through May 22, 2017; 70953 during June 5, 2017, through September 24, 2019).

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

DFFITS: Difference in fits statistic (Helsel and others, 2020).

E.vars: Explanatory variables.

fCHL: Chlorophyll fluorescence, in relative fluorescence units (32320).

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

LOESS: Local polynomial regression fitting, or locally estimated scatterplot smoothing (Helsel and others, 2020).

LOWESS: Locally weighted scatterplot smoothing (Cleveland, 1979; Helsel and others, 2020).

MSE: Model standard error (Helsel and others, 2020).

MSPE: Model standard percentage error (Helsel and others, 2020).

Probability(>|t|): The probability that the independent variable has no effect on the dependent variable (Helsel and others, 2020).

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).

References Cited

Cleveland, W.S., 1979, Robust locally weighted regression and smoothing scatterplots: Journal of the American Statistical Association, v. 74, no. 368, p. 829-836.

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.]

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.]

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.]

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.]