

Appendix 9. Model Archival Summary for Sulfate Concentration at U.S. Geological Survey Site 06887500, Kansas River at Wamego, Kansas, during July 2012 through September 2019

This model archival summary summarizes the sulfate (SO₄; U.S. Geological Survey [USGS] parameter code 00945) concentration model developed to compute 15-minute SO₄ concentrations from July 2012 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 (SC), 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, SC, dissolved oxygen, pH, turbidity, 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: March 30, 2020

Model calibration data period: July 19, 2012, through September 23, 2019

Model application date: July 19, 2012, 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, SC, dissolved oxygen, pH, turbidity, and chlorophyll and phycocyanin fluorescence. 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 101 concurrent measurements of SO₄ concentration and sensor-measured SC during July 19, 2012, through September 23, 2019. Samples were collected throughout the range of continuously observed hydrologic conditions. No samples had concentrations below laboratory detection limits. 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. One of the sensor-measured SC samples, from July 28, 2014, was deemed an outlier and removed from the model calibration dataset. The removed sensor-measured SC value was significantly lower than the field monitor and laboratory result during this sample. 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 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.

Sulfate 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 SO₄ concentration at the USGS National Water Quality Laboratory in Lakewood, Colorado.

Model Development

Ordinary least squares regression analysis was done using R programming language (R Core Team, 2020) to relate discretely collected SO₄ concentration to sensor-measured SC. 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.

SC was selected as a good surrogate for SO₄ 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 SO₄ concentration at USGS site 06887500:

SO₄ concentration-based model:

$$\log SO_4 = 1.10 \times \log SC - 1.18$$

where

log = logarithm base 10;

SO₄ = sulfate concentration, in milligrams per liter; and

SC = specific conductance, in microsiemens per centimeter at 25 degrees Celsius.

SC makes physical and statistical sense as an explanatory variable for SO₄ because of its positive correlation with charged ionic species (Hem, 1992).

The logarithmically (log) transformed model may be retransformed to the original units so that SO₄ 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 1.01. The retransformed model, accounting for BCF is as follows:

$$SO_4 = 1.01 \times (SC^{1.1} \times 10^{-1.18})$$

Previous Models

Start Year	End Year	Model Equation	Reference
2012	2019	$\log SO_4 = 1.05 \log SC - 1.06$	Foster and Graham (2016)
1999	2003	$\log SO_4 = 1.05 \log SC - 1.05$	Rasmussen and others (2005)

Model Statistics, Data, and Plots

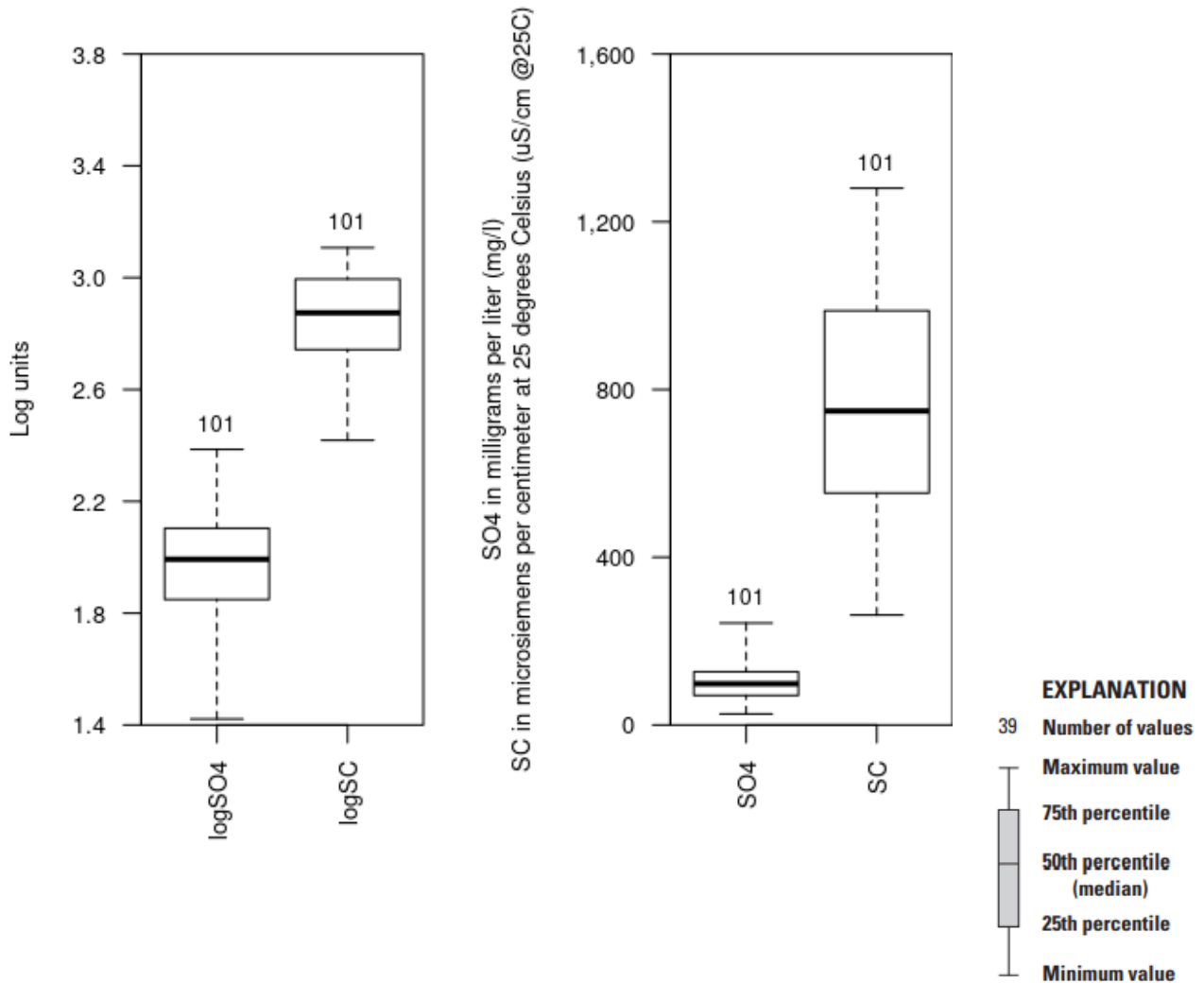
Model

$$\log\text{SO4} = + 1.1 * \log\text{SC} - 1.18$$

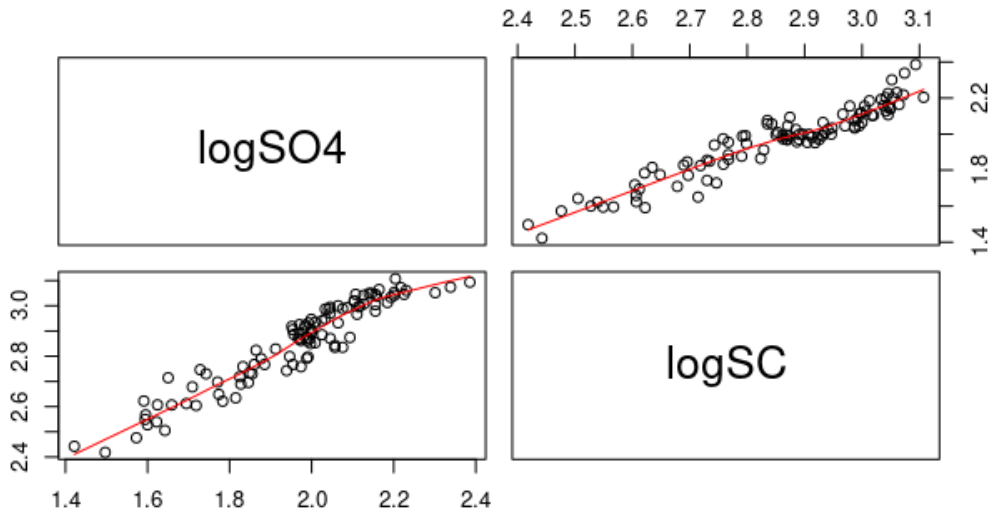
Variable Summary Statistics

	logSO4	SO4	logSC	SC
Minimum	1.42	26.4	2.42	262
1st Quartile	1.85	70.6	2.74	553
Median	1.99	98.3	2.87	749
Mean	1.96	101.0	2.85	762
3rd Quartile	2.10	127.0	2.99	988
Maximum	2.39	243.0	3.11	1280

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	101
Standard error (RMSE)	0.0614
Average Model standard percentage error (MSPE)	14.2
Coefficient of determination (R^2)	0.902
Adjusted Coefficient of Determination (Adj. R^2)	0.901
Bias Correction Factor (BCF)	1.01

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(> t)
(Intercept)	-1.18	0.1040	-11.3	1.83e-19
logSC	1.10	0.0366	30.1	1.20e-51

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.998
E.vars	-0.998	1.000

Outlier Test Criteria

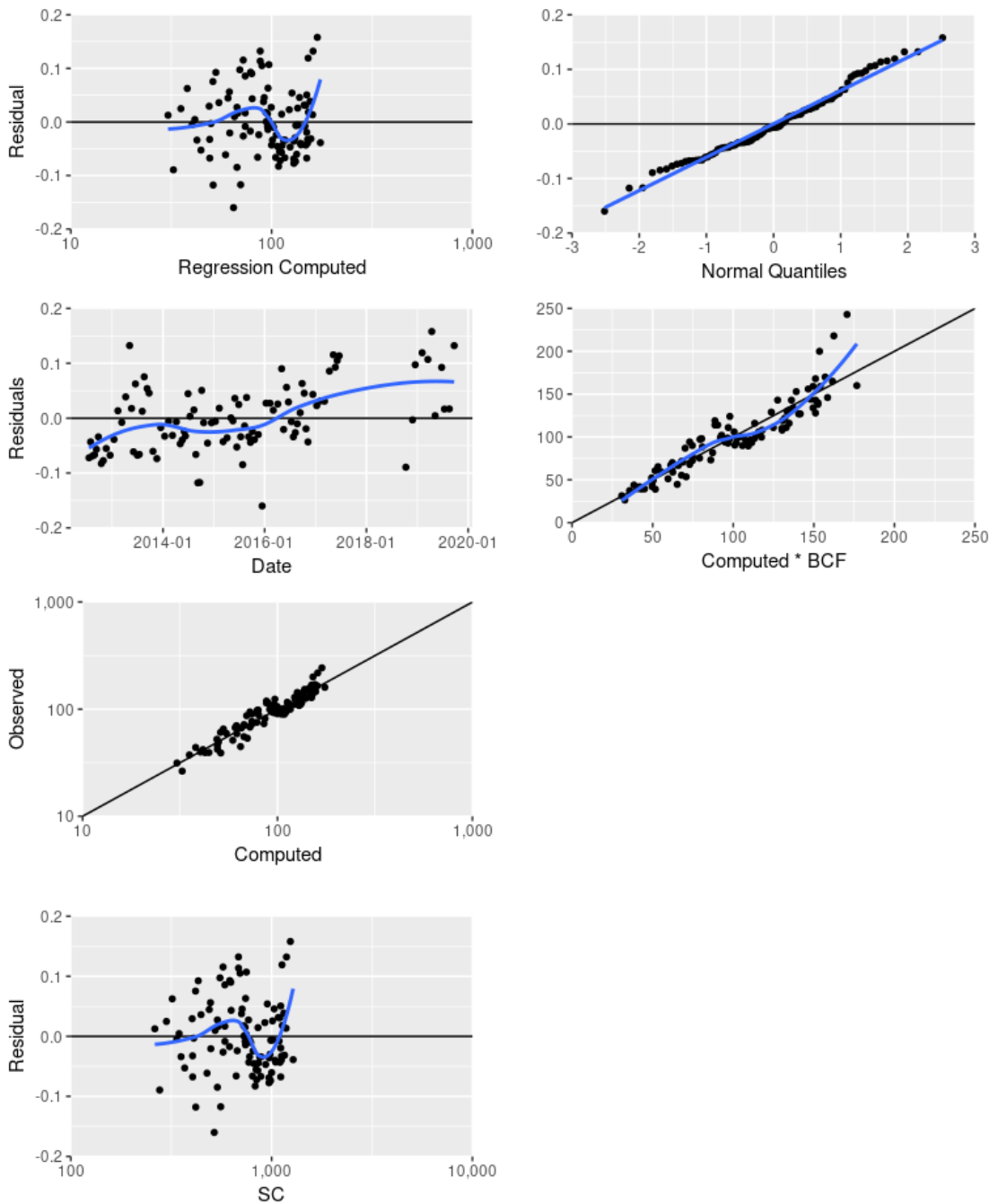
Leverage	Cook's D	DFFITS
0.0594	0.1944	0.2814

Flagged Observations

	logSO4	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
201308050730	1.50	1.48	0.0126	0.213	0.212	0.0768	0.00189	0.0611
201409080800	1.59	1.71	-0.1180	-1.950	-1.980	0.0287	0.05610	-0.3400
201506290820	1.57	1.55	0.0248	0.417	0.415	0.0602	0.00556	0.1050
201512140910	1.65	1.81	-0.1600	-2.630	-2.710	0.0167	0.05860	-0.3530
201810110940	1.42	1.51	-0.0894	-1.510	-1.520	0.0696	0.08520	-0.4160
201902050950	2.30	2.18	0.1190	1.970	1.990	0.0240	0.04740	0.3120

201904150950	2.39	2.23	0.1580	2.610	2.690	0.0304	0.10700	0.4770
201909231020	2.34	2.21	0.1320	2.180	2.230	0.0273	0.06680	0.3730

Statistical Plots



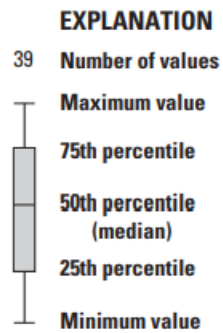
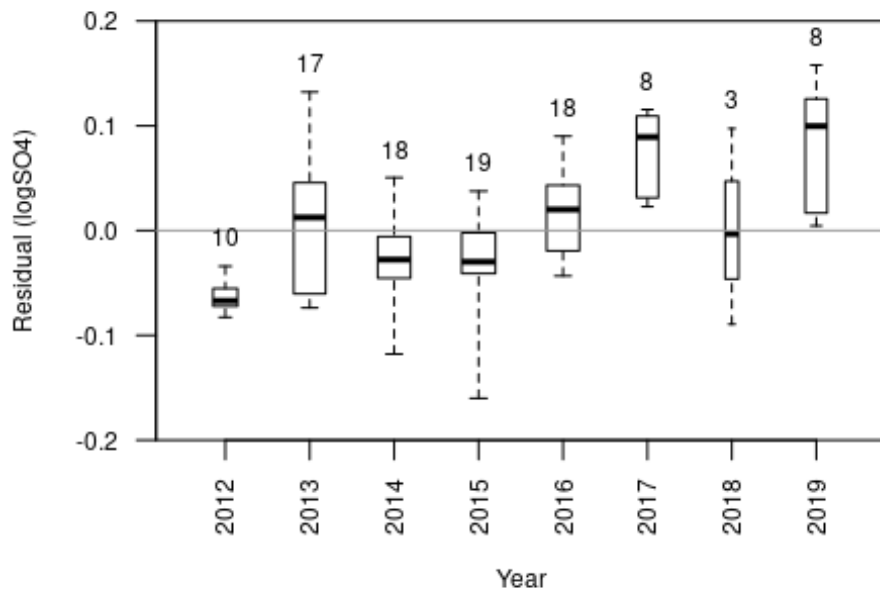
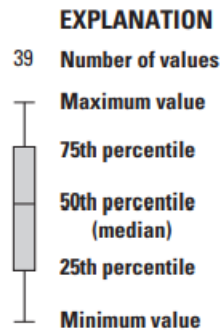
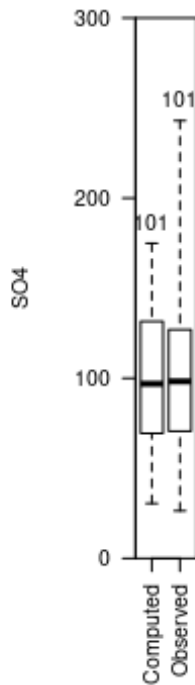
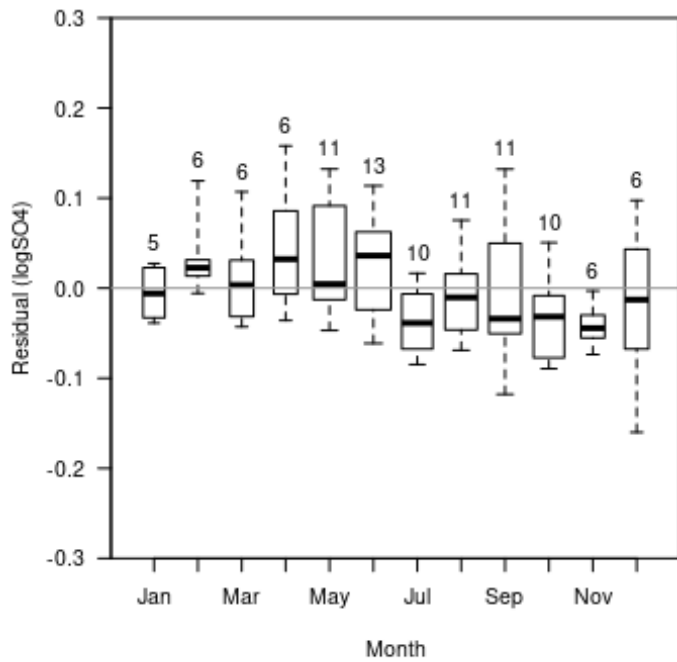
First row (left): Residual S04 related to regression computed S04 with local polynomial regression fitting, or locally estimated scatterplot smoothing (LOESS), indicated by the blue line.

First row (right): Residual S04 related to the corresponding normal quantile of the residual with simple linear regression, indicated by the blue line.

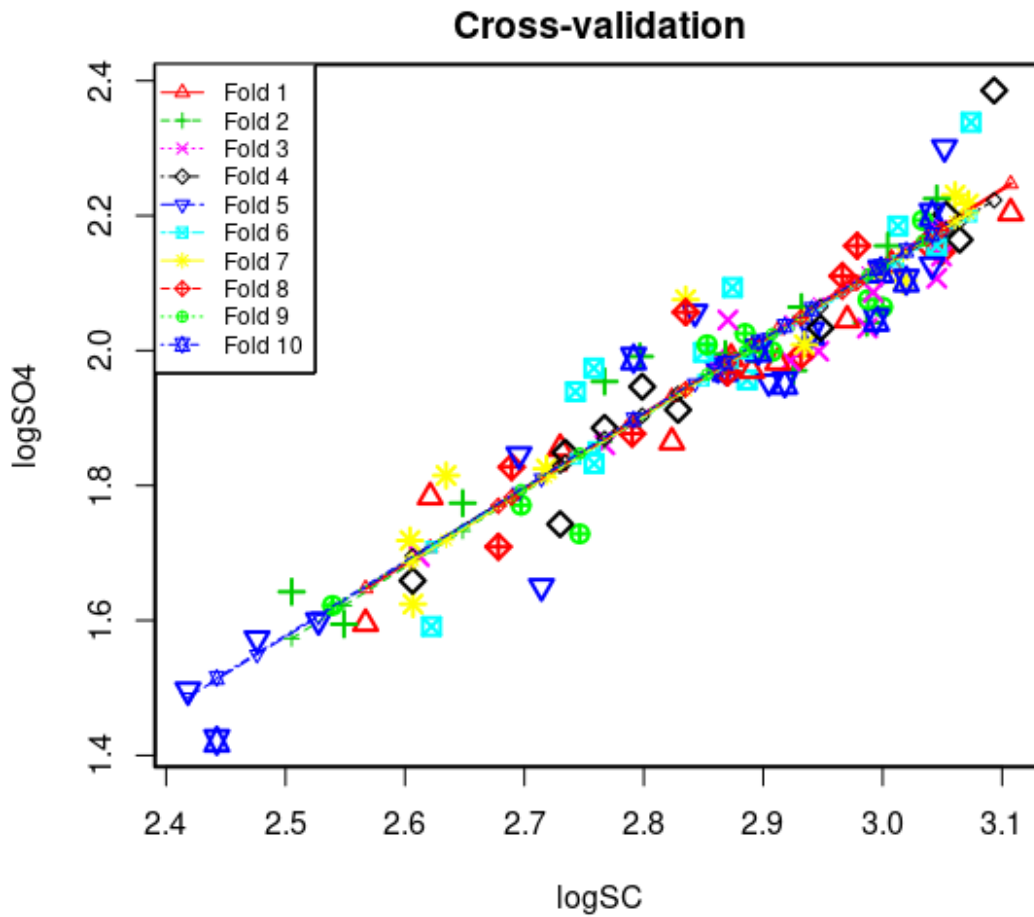
Second row: Residual S04 related to date (left) and regression computed S04 multiplied by the BCF (right) with LOESS, indicated by the blue line.

Third row: Observed S04 related to regression computed S04.

Fourth row: Residual S04 related to SC with LOESS, indicated by the blue line.



Cross-Validation



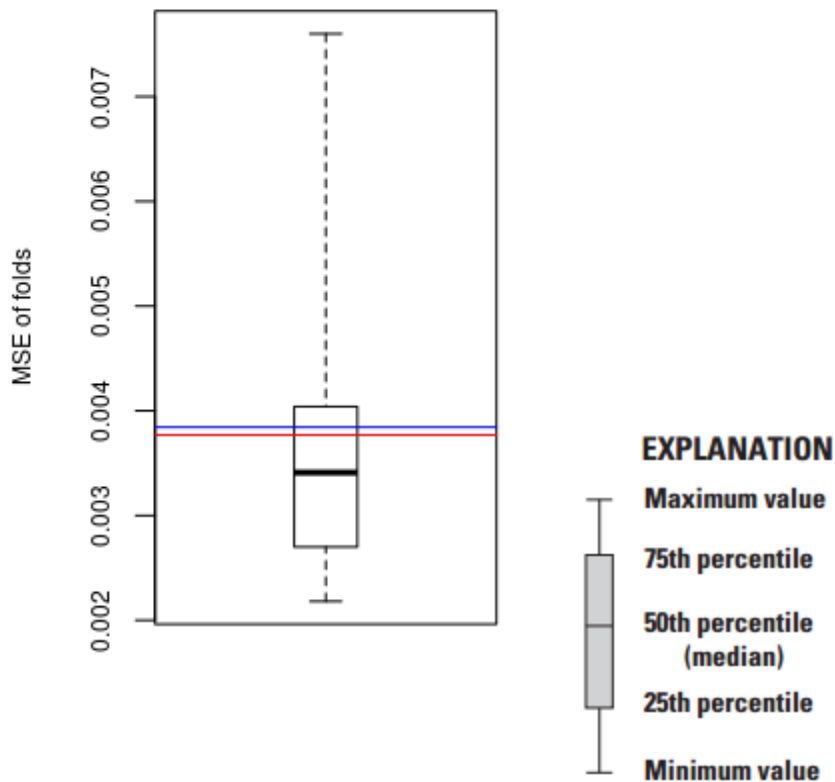
Fold - equal partition of the data (10 percent of the data).

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: 0.00218
Mean MSE of folds: 0.00384
Median MSE of folds: 0.00341
Maximum MSE of folds: 0.00760
(Mean MSE of folds) / (Model MSE): 1.02000
```



Red line - Model MSE

Blue line - Mean MSE of folds

Model-Calibration Dataset

	Date	logS04	logSC	S04	SC	Computed logS04	Computed S04	Residual	Normal Quantiles	Censored Values
0										
1	2012-07-19	1.97	2.93	93.4	842	2.04	111	-0.0722	-1.37	--
2	2012-07-30	1.96	2.89	90.5	770	2	101	-0.0432	-0.696	--
3	2012-08-13	2.04	2.99	110	970	2.11	130	-0.0689	-1.31	--
4	2012-08-27	2	2.95	99.8	884	2.07	118	-0.0667	-1.15	--
5	2012-09-10	1.99	2.93	98.3	854	2.05	113	-0.0568	-0.939	--
6	2012-09-24	2.03	2.94	106	872	2.06	116	-0.034	-0.489	--
7	2012-10-15	1.95	2.92	89.5	828	2.03	109	-0.0828	-1.6	--
8	2012-10-29	2.03	2.99	108	971	2.11	130	-0.0773	-1.51	--
9	2012-11-19	1.98	2.92	96.5	837	2.04	111	-0.0552	-0.901	--
10	2012-12-17	2.11	3.05	128	1110	2.17	151	-0.0675	-1.2	--
11	2013-01-14	2.2	3.11	160	1280	2.24	177	-0.0387	-0.575	--
12	2013-02-11	2.22	3.07	165	1180	2.2	162	0.0135	0.25	--
13	2013-03-11	2.16	3.04	144	1090	2.17	148	-0.00767	-0.0496	--
14	2013-04-08	2.23	3.06	170	1150	2.19	157	0.0388	0.728	--
15	2013-05-06	2.08	2.84	119	684	1.94	88.6	0.132	2.15	--
16	2013-05-20	1.99	2.86	97.6	726	1.97	94.6	0.0177	0.381	--
17	2013-06-03	1.71	2.68	51.2	477	1.77	59.5	-0.0613	-1.02	--
18	2013-06-17	1.64	2.51	43.9	320	1.58	38.4	0.0625	1.02	--
19	2013-07-01	1.62	2.61	42.1	404	1.69	49.7	-0.0676	-1.25	--
20	2013-07-15	1.86	2.82	73.2	666	1.93	86.1	-0.066	-1.06	--

21	2013-08-05	1.5	2.42	31.4	262	1.48	30.8	0.0126	0.225	--
22	2013-08-19	1.78	2.62	60.7	418	1.71	51.5	0.0755	1.11	--
23	2013-09-09	2.16	2.98	143	952	2.1	128	0.054	0.939	--
24	2013-09-23	2.18	3.01	153	1030	2.14	139	0.0457	0.864	--
25	2013-10-21	2.06	3	116	1000	2.12	135	-0.0604	-0.978	--
26	2013-11-18	2.05	2.99	111	988	2.12	133	-0.0737	-1.44	--
27	2013-12-16	1.88	2.79	75.3	617	1.89	79.1	-0.0171	-0.174	--
28	2014-01-13	2.15	3.05	140	1120	2.18	153	-0.0329	-0.407	--
29	2014-02-10	2.13	3.01	134	1020	2.13	137	-0.00578	0.0496	--
30	2014-03-10	2.16	3.06	146	1160	2.2	159	-0.0314	-0.354	--
31	2014-04-07	2.11	3	130	991	2.12	133	-0.00656	-0.0248	--
32	2014-05-05	2.05	2.97	111	934	2.09	125	-0.0469	-0.829	--
33	2014-05-19	2.1	3.02	127	1040	2.14	141	-0.0398	-0.635	--
34	2014-06-02	1.91	2.83	81.7	674	1.94	87.2	-0.024	-0.25	--
35	2014-06-11	1.66	2.61	45.6	404	1.69	49.6	-0.0325	-0.381	--
36	2014-06-30	1.83	2.69	67.2	489	1.78	61.2	0.0446	0.795	--
37	2014-07-14	1.99	2.87	97.4	747	1.99	97.6	0.00342	0.149	--
38	2014-08-11	1.99	2.87	98.8	738	1.98	96.4	0.015	0.302	--
39	2014-08-25	1.95	2.9	89.9	803	2.02	106	-0.0662	-1.11	--
40	2014-09-08	1.59	2.62	39	419	1.71	51.7	-0.118	-2.15	--
41	2014-09-22	1.73	2.75	53.5	558	1.85	70.8	-0.117	-1.95	--
42	2014-10-06	2.23	3.05	168	1110	2.17	151	0.0506	0.901	--
43	2014-10-20	1.97	2.87	93.6	738	1.98	96.4	-0.00828	-0.0744	--
44	2014-11-17	1.98	2.91	96.1	817	2.03	108	-0.0455	-0.795	--
45	2014-12-15	1.86	2.77	72.5	585	1.87	74.7	-0.00839	-0.0992	--
46	2015-01-12	2.12	3	131	997	2.12	134	-0.00597	0.0248	--
47	2015-02-09	2.2	3.05	159	1130	2.18	154	0.0181	0.407	--
48	2015-03-09	2.1	3.02	127	1050	2.15	142	-0.0428	-0.665	--
49	2015-04-06	2.08	2.99	119	972	2.11	131	-0.0358	-0.546	--
50	2015-05-04	1.97	2.86	94.1	728	1.97	94.9	0.000561	0.124	--
51	2015-05-18	1.6	2.53	39.8	337	1.6	40.6	-0.00486	0.0744	--
52	2015-06-01	1.77	2.65	59.4	445	1.74	55.2	0.0361	0.665	--
53	2015-06-15	1.6	2.57	39.4	369	1.65	44.9	-0.0526	-0.864	--
54	2015-06-29	1.57	2.48	37.4	299	1.55	35.7	0.0248	0.462	--
55	2015-07-13	1.59	2.55	39.3	354	1.63	42.9	-0.0339	-0.462	--
56	2015-07-27	1.74	2.73	55.3	537	1.83	67.9	-0.0848	-1.69	--
57	2015-08-10	1.97	2.87	92.8	741	1.98	96.8	-0.0142	-0.149	--
58	2015-08-24	2	2.85	99.4	708	1.96	92.1	0.0377	0.696	--
59	2015-09-08	1.97	2.89	93.4	777	2	102	-0.0338	-0.434	--
60	2015-09-21	2.01	2.93	102	860	2.05	114	-0.0441	-0.761	--
61	2015-10-05	2	2.91	99.7	808	2.02	106	-0.0242	-0.276	--
62	2015-10-19	2.14	3.05	138	1120	2.18	153	-0.0391	-0.605	--
63	2015-11-16	2.09	2.99	122	982	2.12	132	-0.0298	-0.328	--
64	2015-12-14	1.65	2.71	44.7	518	1.81	65.3	-0.16	-2.52	--
65	2016-01-11	1.85	2.73	71.6	537	1.83	67.9	0.0274	0.546	--
66	2016-02-08	2.03	2.88	106	767	2	101	0.0273	0.518	--
67	2016-03-03	2.06	2.93	116	855	2.05	113	0.0145	0.276	--
68	2016-04-04	2.16	3	143	1010	2.13	136	0.0258	0.489	--
69	2016-05-02	1.99	2.8	98	626	1.9	80.4	0.0901	1.2	--
70	2016-05-16	1.77	2.7	59	498	1.79	62.5	-0.0206	-0.225	--
71	2016-06-06	1.85	2.7	70.1	496	1.79	62.2	0.0562	0.978	--
72	2016-06-20	1.72	2.6	52.3	402	1.69	49.4	0.0294	0.575	--
73	2016-07-11	2	2.89	99.5	777	2	102	-0.00635	0	--
74	2016-07-25	2.03	2.95	108	887	2.07	118	-0.0341	-0.518	--
75	2016-08-08	1.83	2.76	68	573	1.86	73	-0.0263	-0.302	--

76	2016-08-22	2	2.9	100	787	2.01	103	-0.0103	-0.124	--
77	2016-09-12	1.82	2.72	66.8	523	1.81	65.9	0.00989	0.199	--
78	2016-09-26	2.05	2.87	111	742	1.98	96.9	0.0632	1.06	--
79	2016-10-11	2.01	2.85	102	713	1.96	92.8	0.0455	0.829	--
80	2016-10-24	2.16	3.05	143	1110	2.17	151	-0.0194	-0.199	--
81	2016-11-07	2.13	3.04	134	1100	2.17	150	-0.0433	-0.728	--
82	2016-12-12	1.95	2.8	88.4	629	1.9	80.8	0.0433	0.761	--
83	2017-01-09	2.11	2.97	129	925	2.09	124	0.0229	0.434	--
84	2017-02-06	2.19	3.03	156	1080	2.16	147	0.0315	0.635	--
85	2017-03-06	2.2	3.04	159	1100	2.17	150	0.031	0.605	--
86	2017-04-10	1.95	2.77	90	585	1.87	74.6	0.0858	1.15	--
87	2017-05-08	1.97	2.76	94.2	573	1.86	72.9	0.116	1.69	--
88	2017-05-22	1.99	2.79	97.3	619	1.9	79.3	0.0929	1.31	--
89	2017-06-05	2.06	2.84	114	696	1.95	90.4	0.105	1.44	--
90	2017-06-19	2.06	2.84	114	684	1.94	88.6	0.114	1.6	--
91	2018-10-11	1.42	2.44	26.4	277	1.51	32.8	-0.0894	-1.8	--
92	2018-11-28	1.69	2.61	49.5	409	1.7	50.4	-0.0031	0.0992	--
93	2018-12-17	1.94	2.74	86.9	553	1.84	70.1	0.0974	1.37	--
94	2019-02-05	2.3	3.05	200	1130	2.18	154	0.119	1.8	--
95	2019-03-18	2.09	2.87	124	749	1.99	97.9	0.107	1.51	--
96	2019-04-15	2.39	3.09	243	1240	2.23	171	0.158	2.52	--
97	2019-05-09	1.62	2.54	41.9	346	1.62	41.9	0.00456	0.174	--
98	2019-06-25	1.81	2.63	65.3	431	1.72	53.3	0.0925	1.25	--
99	2019-07-15	1.85	2.73	70.6	542	1.83	68.6	0.0166	0.328	--
100	2019-08-19	1.89	2.77	76.8	585	1.87	74.6	0.0169	0.354	--
101	2019-09-23	2.34	3.07	218	1190	2.21	162	0.132	1.95	--

Definitions

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

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

E.vars: Explanatory variables.

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).

SC: Specific conductance, in microsiemens per centimeter at 25 degrees Celsius (00095).

S04: Sulfate, in milligrams per liter (00945).

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.]
- Hem, J.D., 1992, Study and interpretation of the chemical characteristics of natural water (3d. ed): U.S. Geological Survey Water-Supply Paper 2254, 264 p. [Also available at <https://doi.org/10.3133/wsp2254>.]
- 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.]