Appendix 5. Model Archival Summary for Magnesium 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 magnesium (Mg; U.S. Geological Survey [USGS] parameter code 00925) concentration model developed to compute 15-minute Mg concentrations from July 2012 onward. This model supersedes all previous models.

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 Mg 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.

Magnesium 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 Mg 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 Mg 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 Mg 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 Mg concentration at USGS site 06887500:

Mg concentration-based model:

$$\log Mg = 0.808 \times \log SC - 1.11$$

where

log = logarithm base 10;

Mg = magnesium 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 Mg 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 Mg 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:

$$Mg = 1.01 \times (SC^{0.808} \times 10^{-1.11})$$

Previous Models

Start Year	End Year	Model Equation	Reference
2012	2019	logMg = 0.768logSC - 0.996	Foster and Graham (2016)
1999	2003	logMg = 0.912logSC - 1.46	Rasmussen and others (2005)

Model Statistics, Data, and Plots

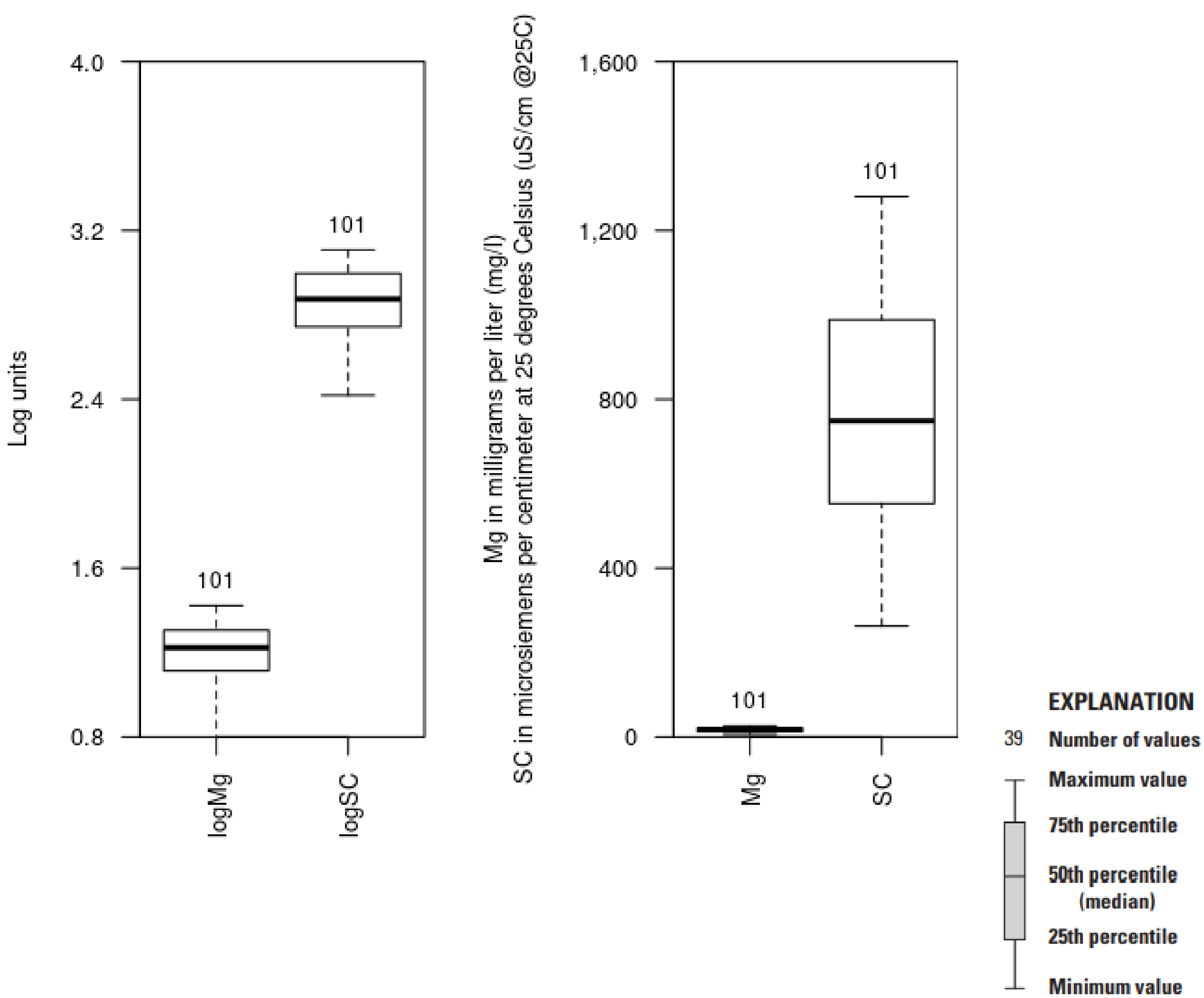
Model

logMg = + 0.808 * logSC - 1.11

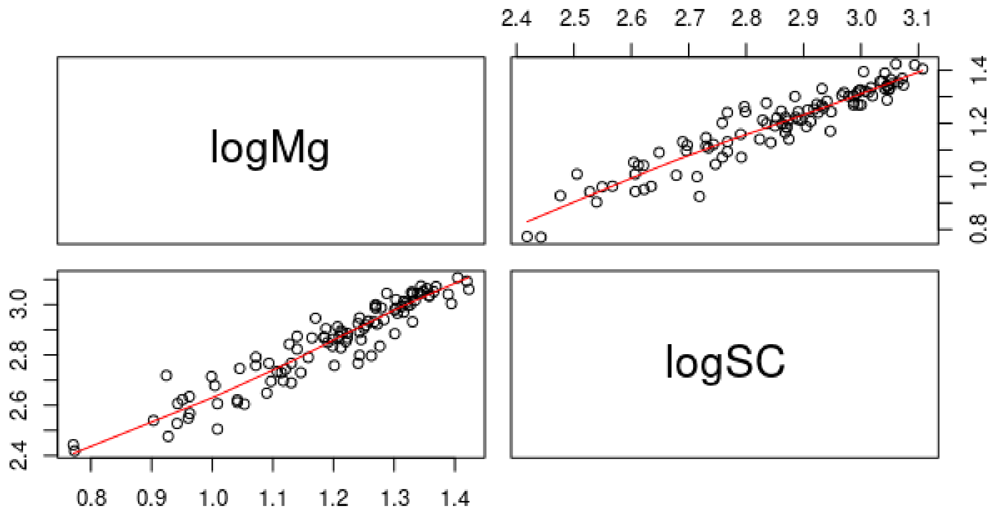
Variable Summary Statistics

	logMg	Mg	logSC	SC
Minimum	0.771	5.9	2.42	262
1st Quartile	1.110	13.0	2.74	553
Median	1.220	16.7	2.87	749
Mean	1.200	16.5	2.85	762
3rd Quartile	1.310	20.2	2.99	988
Maximum	1.420	26.5	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)                                0.0479
Average Model standard percentage error (MSPE)         11.1
Coefficient of determination (R²)                      0.89
Adjusted Coefficient of Determination (Adj. R²)       0.889
Bias Correction Factor (BCF)                           1.01
```

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	-1.110	0.0816	-13.6	2.29e-24
logSC	0.808	0.0285	28.3	3.09e-49

Correlation Matrix

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

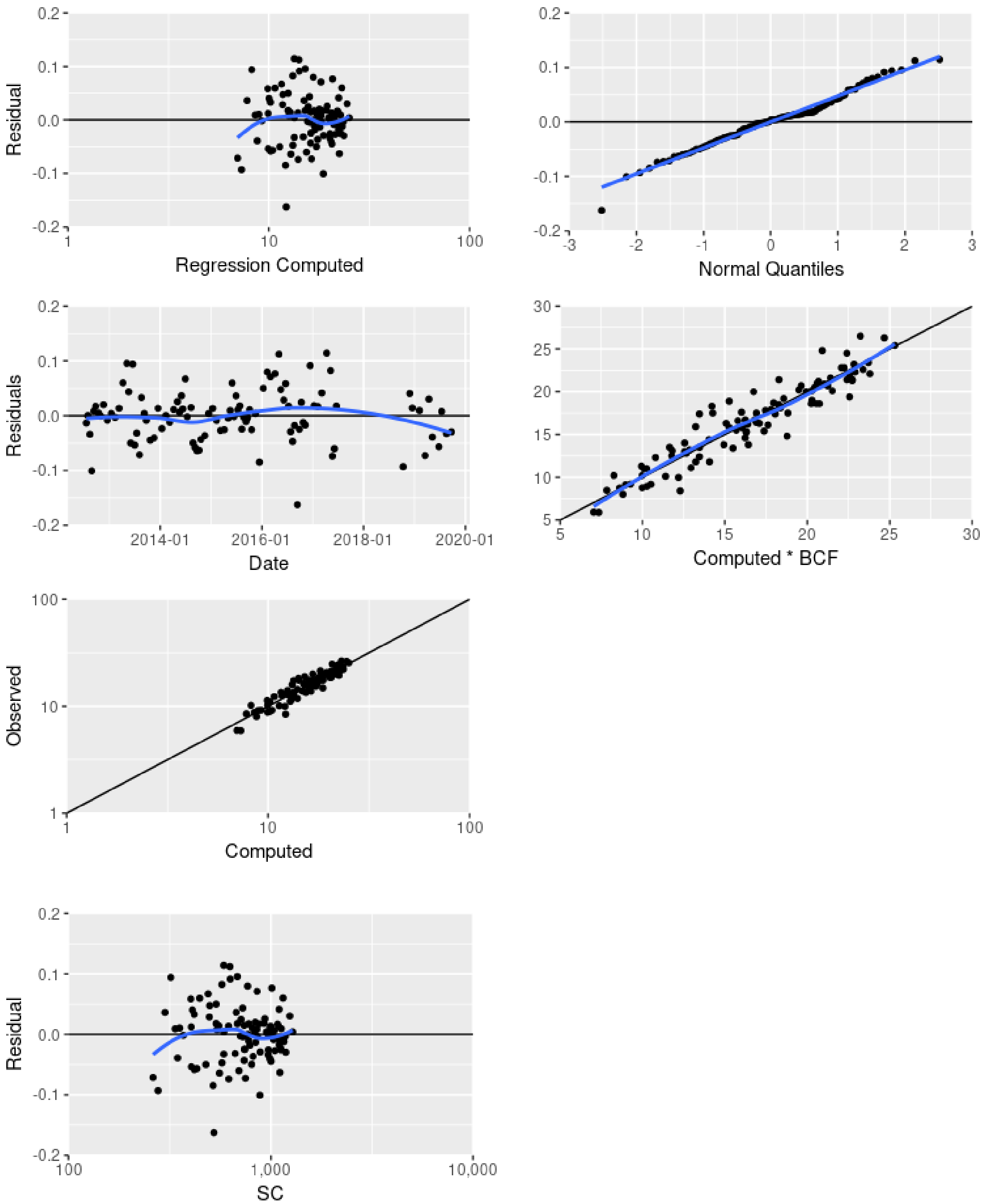
Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.0594	0.1944	0.2814

Flagged Observations

	logMg	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
201306170830	1.010	0.915	0.0940	2.020	2.050	0.0527	0.1130	0.483
201308050730	0.773	0.844	-0.0714	-1.550	-1.560	0.0768	0.0999	-0.450
201506290820	0.927	0.891	0.0362	0.779	0.778	0.0602	0.0195	0.197
201609120850	0.924	1.090	-0.1630	-3.420	-3.620	0.0163	0.0968	-0.466
201810110940	0.771	0.864	-0.0931	-2.010	-2.050	0.0696	0.1520	-0.560

Statistical Plots



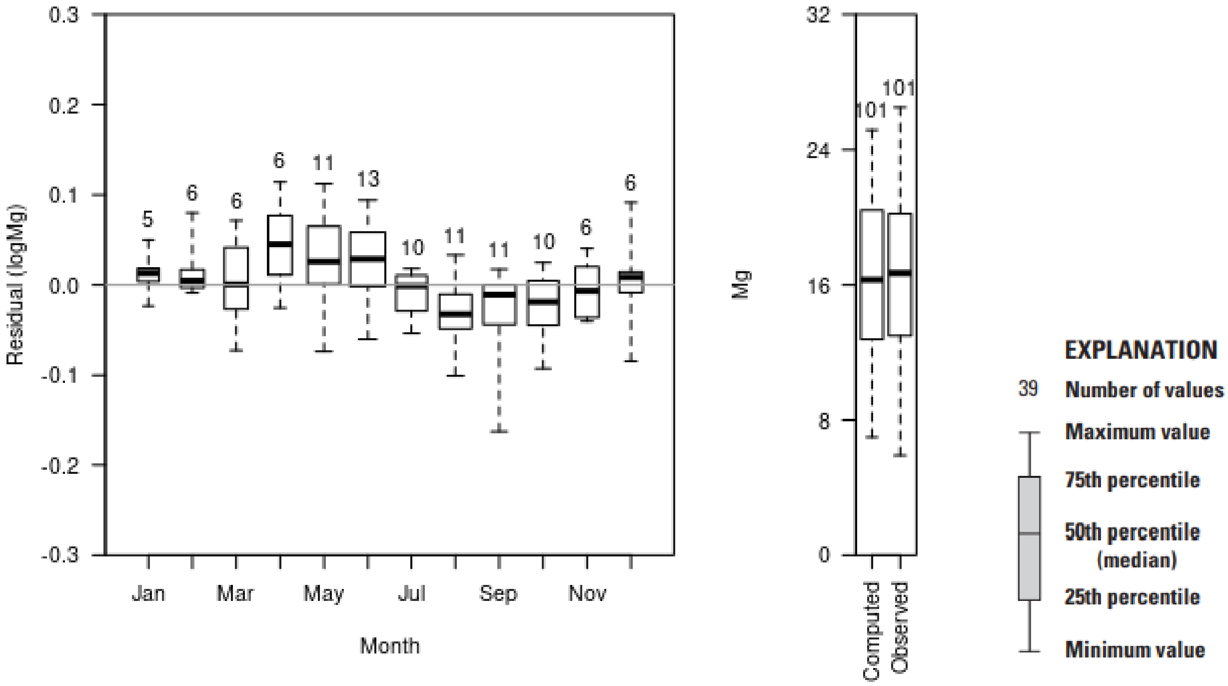
First row (left): Residual Mg related to regression computed Mg with local polynomial regression fitting, or locally estimated scatterplot smoothing (LOESS), indicated by the blue line.

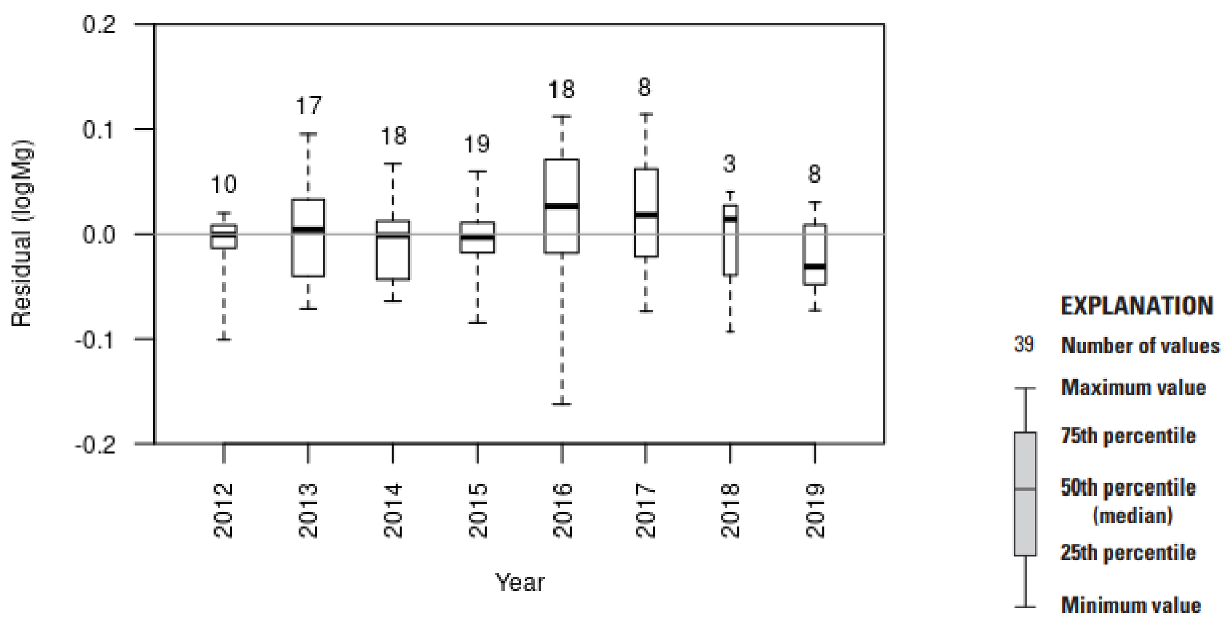
First row (right): Residual Mg related to the corresponding normal quantile of the residual with simple linear regression, indicated by the blue line.

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

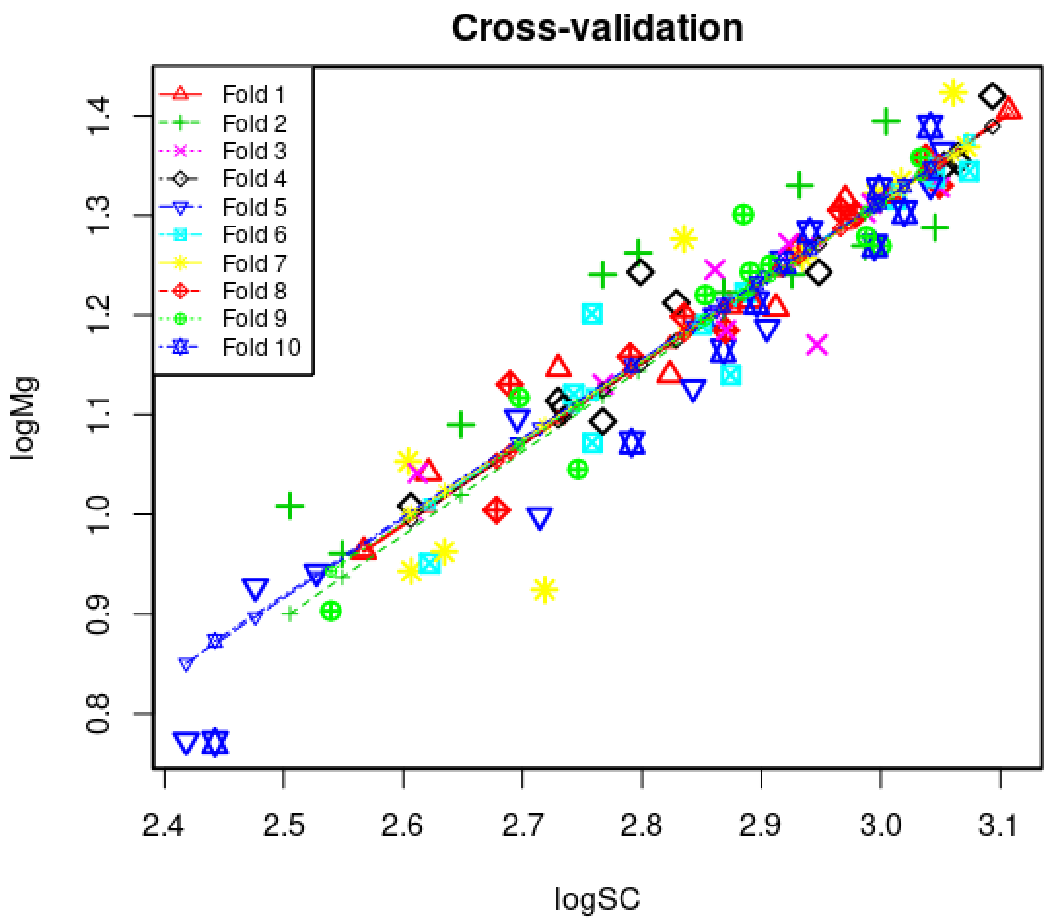
Third row: Observed Mg related to regression computed Mg.

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


Residual (logMg)
Month
Jan
Mar
May
Jul
Sep
Nov
Mg
Computed
Observed
101
EXPLANATION
39 Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value


Residual (logMg)
Year
2012
2013
2014
2015
2016
2017
2018
2019
EXPLANATION
39 Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value

Cross-Validation



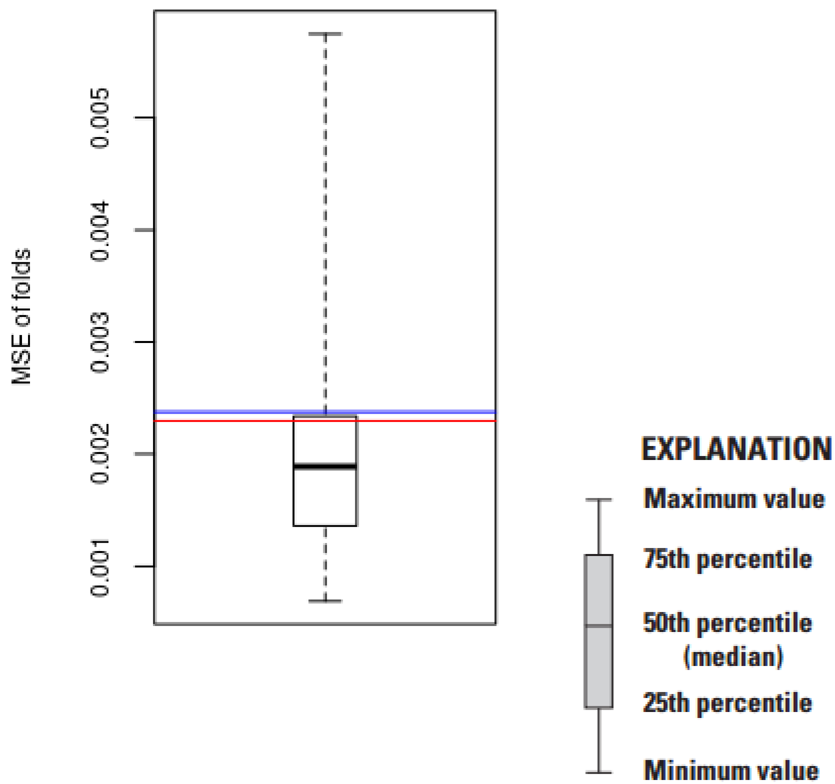

```
Fold - equal partition of the data (10 percent of the data).

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:  0.000686
                 Mean MSE of folds:  0.002380
               Median MSE of folds:  0.001890
              Maximum MSE of folds:  0.005750
 (Mean MSE of folds) / (Model MSE):  1.030000
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Model-Calibration Dataset

0	Date	logMg	logSC	Mg	SC	Computed logMg	Computed Mg	Residual	Normal Quantiles	Censored Values
1	2012-07-19	1.24	2.93	17.4	842	1.25	18.1	-0.0133	-0.407	--
2	2012-07-30	1.22	2.89	16.7	770	1.22	16.8	0.000174	-0.0248	--
3	2012-08-13	1.27	2.99	18.6	970	1.3	20.2	-0.034	-0.795	--
4	2012-08-27	1.17	2.95	14.8	884	1.27	18.8	-0.101	-2.15	--
5	2012-09-10	1.27	2.93	18.5	854	1.26	18.3	0.00831	0.174	--
6	2012-09-24	1.28	2.94	19.2	872	1.27	18.6	0.0171	0.518	--
7	2012-10-15	1.25	2.92	17.9	828	1.25	17.8	0.00484	0.0992	--
8	2012-10-29	1.3	2.99	20.1	971	1.3	20.3	-0.000696	-0.0744	--
9	2012-11-19	1.27	2.92	18.7	837	1.25	18	0.02	0.665	--
10	2012-12-17	1.34	3.05	22	1110	1.35	22.6	-0.00839	-0.25	--
11	2013-01-14	1.4	3.11	25.4	1280	1.4	25.3	0.00404	0.0496	--
12	2013-02-11	1.37	3.07	23.4	1180	1.37	23.7	-0.00305	-0.174	--
13	2013-03-11	1.36	3.04	22.8	1090	1.34	22.2	0.0135	0.381	--
14	2013-04-08	1.42	3.06	26.5	1150	1.36	23.2	0.06	1.25	--
15	2013-05-06	1.28	2.84	18.9	684	1.18	15.3	0.0955	1.95	--
16	2013-05-20	1.25	2.86	17.6	726	1.2	16	0.0436	1.02	--
17	2013-06-03	1	2.68	10.1	477	1.05	11.4	-0.05	-1.11	--
18	2013-06-17	1.01	2.51	10.2	320	0.915	8.26	0.094	1.8	--
19	2013-07-01	0.943	2.61	8.77	404	0.997	9.98	-0.0536	-1.15	--
20	2013-07-15	1.14	2.82	13.8	666	1.17	14.9	-0.0318	-0.728	--

21	2013-08-05	0.773	2.42	5.93	262	0.844	7.03	-0.0714	-1.51	--
22	2013-08-19	1.04	2.62	11	418	1.01	10.3	0.0331	0.829	--
23	2013-09-09	1.3	2.98	20	952	1.3	19.9	0.00407	0.0744	--
24	2013-09-23	1.32	3.01	20.7	1030	1.32	21.2	-0.00861	-0.276	--
25	2013-10-21	1.27	3	18.6	1000	1.31	20.7	-0.0447	-0.978	--
26	2013-11-18	1.27	2.99	18.6	988	1.31	20.5	-0.0405	-0.901	--
27	2013-12-16	1.16	2.79	14.4	617	1.14	14	0.0135	0.407	--
28	2014-01-13	1.33	3.05	21.4	1120	1.35	22.7	-0.0235	-0.489	--
29	2014-02-10	1.32	3.01	20.9	1020	1.32	21	1.97e-05	-0.0496	--
30	2014-03-10	1.35	3.06	22.6	1160	1.37	23.4	-0.0122	-0.354	--
31	2014-04-07	1.32	3	21	991	1.31	20.6	0.0112	0.302	--
32	2014-05-05	1.32	2.97	20.7	934	1.29	19.6	0.0257	0.728	--
33	2014-05-19	1.33	3.02	21.6	1040	1.33	21.4	0.00648	0.124	--
34	2014-06-02	1.21	2.83	16.3	674	1.18	15.1	0.0363	0.901	--
35	2014-06-11	1.01	2.61	10.2	404	0.996	9.97	0.0123	0.328	--
36	2014-06-30	1.13	2.69	13.5	489	1.06	11.6	0.067	1.31	--
37	2014-07-14	1.21	2.87	16.2	747	1.21	16.4	-0.00223	-0.124	--
38	2014-08-11	1.22	2.87	16.7	738	1.21	16.2	0.0149	0.462	--
39	2014-08-25	1.19	2.9	15.4	803	1.24	17.4	-0.0497	-1.06	--
40	2014-09-08	0.951	2.62	8.93	419	1.01	10.3	-0.0583	-1.25	--
41	2014-09-22	1.05	2.75	11.1	558	1.11	12.9	-0.0641	-1.44	--
42	2014-10-06	1.29	3.05	19.4	1110	1.35	22.6	-0.063	-1.37	--
43	2014-10-20	1.16	2.87	14.6	738	1.21	16.2	-0.0433	-0.939	--
44	2014-11-17	1.21	2.91	16.1	817	1.24	17.6	-0.0365	-0.829	--
45	2014-12-15	1.13	2.77	13.5	585	1.13	13.5	0.00397	0.0248	--
46	2015-01-12	1.33	3	21.2	997	1.31	20.7	0.0133	0.354	--
47	2015-02-09	1.35	3.05	22.3	1130	1.36	22.9	-0.00878	-0.302	--
48	2015-03-09	1.3	3.02	20.1	1050	1.33	21.5	-0.027	-0.635	--
49	2015-04-06	1.28	2.99	19	972	1.3	20.3	-0.0256	-0.605	--
50	2015-05-04	1.2	2.86	15.8	728	1.2	16	-0.00421	-0.225	--
51	2015-05-18	0.942	2.53	8.75	337	0.933	8.62	0.00928	0.199	--
52	2015-06-01	1.09	2.65	12.3	445	1.03	10.8	0.0597	1.2	--
53	2015-06-15	0.963	2.57	9.18	369	0.965	9.27	-0.0017	-0.0992	--
54	2015-06-29	0.927	2.48	8.46	299	0.891	7.83	0.0362	0.864	--
55	2015-07-13	0.96	2.55	9.13	354	0.95	8.97	0.0105	0.25	--
56	2015-07-27	1.11	2.73	13	537	1.1	12.6	0.0178	0.605	--
57	2015-08-10	1.18	2.87	15.3	741	1.21	16.3	-0.0245	-0.518	--
58	2015-08-24	1.19	2.85	15.5	708	1.19	15.7	-0.00277	-0.149	--
59	2015-09-08	1.21	2.89	16.4	777	1.23	16.9	-0.0109	-0.328	--
60	2015-09-21	1.26	2.93	18.1	860	1.26	18.4	-0.00364	-0.199	--
61	2015-10-05	1.25	2.91	17.8	808	1.24	17.5	0.011	0.276	--
62	2015-10-19	1.33	3.05	21.3	1120	1.35	22.7	-0.0256	-0.575	--
63	2015-11-16	1.31	2.99	20.5	982	1.31	20.4	0.00391	0	--
64	2015-12-14	0.999	2.71	9.97	518	1.08	12.2	-0.0848	-1.8	--
65	2016-01-11	1.15	2.73	14	537	1.1	12.6	0.05	1.11	--
66	2016-02-08	1.3	2.88	20	767	1.22	16.7	0.0799	1.51	--
67	2016-03-03	1.33	2.93	21.4	855	1.26	18.3	0.0711	1.37	--
68	2016-04-04	1.39	3	24.8	1010	1.32	20.9	0.0767	1.44	--
69	2016-05-02	1.26	2.8	18.3	626	1.15	14.2	0.112	2.15	--
70	2016-05-16	1.12	2.7	13.1	498	1.07	11.8	0.0476	1.06	--
71	2016-06-06	1.1	2.7	12.5	496	1.07	11.8	0.0286	0.761	--
72	2016-06-20	1.05	2.6	11.3	402	0.995	9.93	0.0585	1.15	--
73	2016-07-11	1.24	2.89	17.5	777	1.23	16.9	0.0173	0.546	--
74	2016-07-25	1.24	2.95	17.5	887	1.27	18.8	-0.0291	-0.665	--
75	2016-08-08	1.07	2.76	11.8	573	1.12	13.2	-0.0472	-1.02	--

76	2016-08-22	1.21	2.9	16.3	787	1.23	17.1	-0.018	-0.462	--
77	2016-09-12	0.924	2.72	8.4	523	1.09	12.3	-0.163	-2.52	--
78	2016-09-26	1.18	2.87	15.3	742	1.21	16.3	-0.0249	-0.546	--
79	2016-10-11	1.22	2.85	16.6	713	1.2	15.8	0.0245	0.696	--
80	2016-10-24	1.34	3.05	21.8	1110	1.35	22.6	-0.0124	-0.381	--
81	2016-11-07	1.33	3.04	21.4	1100	1.35	22.4	-0.0172	-0.434	--
82	2016-12-12	1.24	2.8	17.5	629	1.15	14.3	0.0914	1.69	--
83	2017-01-09	1.31	2.97	20.2	925	1.29	19.5	0.0184	0.635	--
84	2017-02-06	1.36	3.03	22.8	1080	1.34	22.1	0.0167	0.489	--
85	2017-03-06	1.39	3.04	24.5	1100	1.35	22.4	0.0415	0.978	--
86	2017-04-10	1.24	2.77	17.4	585	1.13	13.5	0.114	2.52	--
87	2017-05-08	1.2	2.76	15.9	573	1.12	13.2	0.0825	1.6	--
88	2017-05-22	1.07	2.79	11.8	619	1.15	14.1	-0.0739	-1.69	--
89	2017-06-05	1.13	2.84	13.4	696	1.19	15.5	-0.0602	-1.31	--
90	2017-06-19	1.2	2.84	15.8	684	1.18	15.3	0.0177	0.575	--
91	2018-10-11	0.771	2.44	5.9	277	0.864	7.35	-0.0931	-1.95	--
92	2018-11-28	1.04	2.61	11	409	1	10.1	0.0405	0.939	--
93	2018-12-17	1.12	2.74	13.2	553	1.11	12.9	0.0141	0.434	--
94	2019-02-05	1.37	3.05	23.2	1130	1.36	22.8	0.00944	0.225	--
95	2019-03-18	1.14	2.87	13.8	749	1.21	16.4	-0.0728	-1.6	--
96	2019-04-15	1.42	3.09	26.3	1240	1.39	24.7	0.0303	0.795	--
97	2019-05-09	0.903	2.54	8	346	0.942	8.81	-0.0391	-0.864	--
98	2019-06-25	0.962	2.63	9.17	431	1.02	10.5	-0.0566	-1.2	--
99	2019-07-15	1.11	2.73	12.8	542	1.1	12.7	0.00762	0.149	--
100	2019-08-19	1.09	2.77	12.4	585	1.13	13.5	-0.0327	-0.761	--
101	2019-09-23	1.34	3.07	22.1	1190	1.37	23.8	-0.0296	-0.696	--

Definitions

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

DFFITS: 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).

Mg: Magnesium, in milligrams per liter (00925).

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).

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.]