Appendix 1. Model Archival Summary for Bromide Concentration at U.S. Geological Survey Streamgage 06892350, Kansas River at De Soto, Kansas, during January 2021 through October 2023

This model archival summary summarizes the bromide (Br; U.S. Geological Survey [USGS] parameter code 91000) concentration model developed to compute 15-minute, hourly, or daily Br concentrations from January 2021 onward. This model is specific to the Kansas River at De Soto, Kansas (USGS streamgage 06892350), during this study period and cannot be applied to data collected from other locations on the Kansas River or data collected from other waterbodies.

Streamgage and Model Information

Streamgage number: 06892350

Streamgage 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, Kansas, hydrologic unit 10270104.

Equipment: A Xylem YSI EXO2 water-quality monitor (equipped with sensors for water temperature, specific conductance [SC], dissolved oxygen, pH, turbidity, and chlorophyll and phycocyanin fluorescence) and a Hach Nitratax plus sc monitor (equipped with a 5-millimeter path length nitrate sensor) were deployed during January 2021 through October 2023. Readings from the monitors were recorded every 15 minutes and transmitted by way of satellite, hourly.

Date model was created: March 21, 2024

Model-calibration data period: January 25, 2021, through October 23, 2023

Model-application date: January 25, 2021, onward

Model computations are available at the USGS National Real-Time Water-Quality website (https://nrtwq.usgs.gov/ks/).

Bromide Sampling Details

During January 2021 through October 2023, Br samples were collected on a biweekly to bimonthly basis using the equal-width increment collection method (U.S. Geological Survey, variously dated). All samples were composited for analysis (U.S. Geological Survey, variously dated). A USGS Federal Interagency Sedimentation Project US DH–81, DH–95, or D–96–A1 depth integrating sampler was used (Davis and the Federal Interagency Sedimentation Project, 2005). Samples were analyzed for Br concentration using Environmental Protection Agency Method 300.1 (U.S. Environmental Protection Agency, 1997) by the Water District No. 1 of Johnson County Laboratory in Kansas City, Kansas.

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 USGS National Water Information System (U.S. Geological Survey, 2024) database and available to the public. Ordinary least squares analysis was used to develop regression models using R programming language (R Core Team, 2024). Potential explanatory variables that were evaluated individually and in combination included streamflow, water temperature, SC, dissolved oxygen, pH, turbidity, chlorophyll and phycocyanin fluorescence, and nitrate. These potential explanatory variables were interpolated within the 15-minute continuous record based on sample time. The maximum time span between two continuous data points used for interpolation was 1 hour. Seasonal components (sine and cosine variables) also were evaluated as potential explanatory variables.

The final selected regression model was based on 41 concurrent measurements of Br concentration and sensor-measured SC during January 25, 2021, through October 23, 2023. Samples were collected throughout the range of continuously observed hydrologic conditions. No samples had concentrations less than laboratory minimum reporting limits.

Potential outliers initially were identified using scatterplots of the Br and SC model-calibration data (Rasmussen and others, 2009). Studentized residuals from the model were inspected for values greater than three or less than negative three (Pardoe, 2020). Values outside of that range were considered potential outliers and were investigated. Additionally, computations of leverage, Cook’s distance (Cook’s D), and difference in fits (DFFITS) statistics were used to estimate potential outlier effect on the final selected regression model (Cook, 1977; Helsel and others, 2020). Outliers were investigated for potential removal from the model-calibration dataset by confirming correct database entry, evaluating laboratory analytical performance, and reviewing field notes associated with the sample (Rasmussen and others, 2009). Potential outliers were not determined to have errors associated with sample collection, processing, or analysis and were therefore considered valid.

Model Development

Ordinary least squares regression analysis was done using the *stats* (*v4.3.0*) package in R programming language (R Core Team, 2024) to relate discretely collected Br concentration to sensor-measured SC. The distribution of residuals (the difference between the measured and computed values) was examined for normality, and the plots of residuals were examined for homoscedasticity (departures from zero did not change substantially over the range of computed values).

SC was selected as a good surrogate for Br based on residual plots, coefficient of determination (R^2), and model standard percentage error. Values for all the aforementioned statistics, all relevant sample data, and additional statistical information are included in the “Model Statistics, Data, and Plots” section of this appendix.

Model Summary

The following is a summary of the final regression analysis for Br concentration at USGS streamgage 06892350:

Br concentration-based model:

$$\log Br = 1.224(\log SC) - 1.460$$

where

log = logarithm base 10;

Br = bromide concentration, in micrograms 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 Br because of its positive correlation with charged ionic species (Hem, 1985).

The logarithmically (log) transformed model may be retransformed to the original units so that Br 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 1.012. The retransformed model, accounting for BCF is as follows:

$$Br = 1.012 \times (SC^{1.224} \times 10^{-1.460})$$

This model was developed using continuous and discrete water-quality data collected during January 2021 through October 2023. These data were collected throughout the observed range of streamflow conditions during this time. However, a limitation in model accuracy during conditions outside of those observed during January 2021 through October 2023 warrants consideration when interpreting model computations beyond October 2023. Extrapolation, defined as computation beyond the range of the model calibration dataset, should be used no more than 10 percent beyond the range of the calibration data used to fit the model and is therefore limited. The extrapolation limit for Br concentration using this model is 276 micrograms per liter. Computed estimates exceeding that limit are not supported by the current model calibration dataset.

Previous Models

No Br models at this streamgage have been published previously. However, similar models for other constituents have been published at this streamgage and other Kansas River streamgages, as documented by Rasmussen and others (2005), Foster and Graham (2016), and Williams (2021, 2023).

Model Statistics, Data, and Plots

Definitions

Variable	Explanation
BCF	Duan's bias correction factor (Duan, 1983).
Br	Bromide concentration, in micrograms per liter (USGS parameter code 91000; USGS method code IC041).
Cook's D	Cook's distance (Cook, 1977; Helsel and others, 2020).
DFFITS	Difference in fits statistic (Helsel and others, 2020).
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).
log	Common logarithm with base 10.
MSE	Mean square error (Helsel and others, 2020).
MSPE	Model standard percentage error (Helsel and others, 2020).
Pr(>|t|)	The probability that the independent variable has no effect on the dependent variable (Helsel and others, 2020).
Q1	The value at which 25 percent of the data fall under when data are arranged in ascending order (25th percentile).
Q2	The value at which 50 percent of the data fall under when data are arranged in ascending order (median).
Q3	The value at which 75 percent of the data fall under when data are arranged in ascending order (75th percentile).
R^2	Coefficient of determination.
RMSE	Root mean square error (Helsel and others, 2020).
SC	Specific conductance, in microsiemens per centimeter at 25 degrees Celsius (USGS parameter code 00095; USGS method code SC001).
t value	Student's t value; the coefficient divided by its associated standard error (Helsel and others, 2020).

Model

$\log Br = 1.224(\log SC) - 1.460$

Variable Summary Statistics

Variable	Minimum	Q1	Median	Mean	Q3	Maximum
Br	31.9	94.3	136	131	166	251
SC	359	656	830	823	975	1240
logBr	1.5	1.97	2.13	2.08	2.22	2.4
logSC	2.56	2.82	2.92	2.9	2.99	3.09

Duration Plots

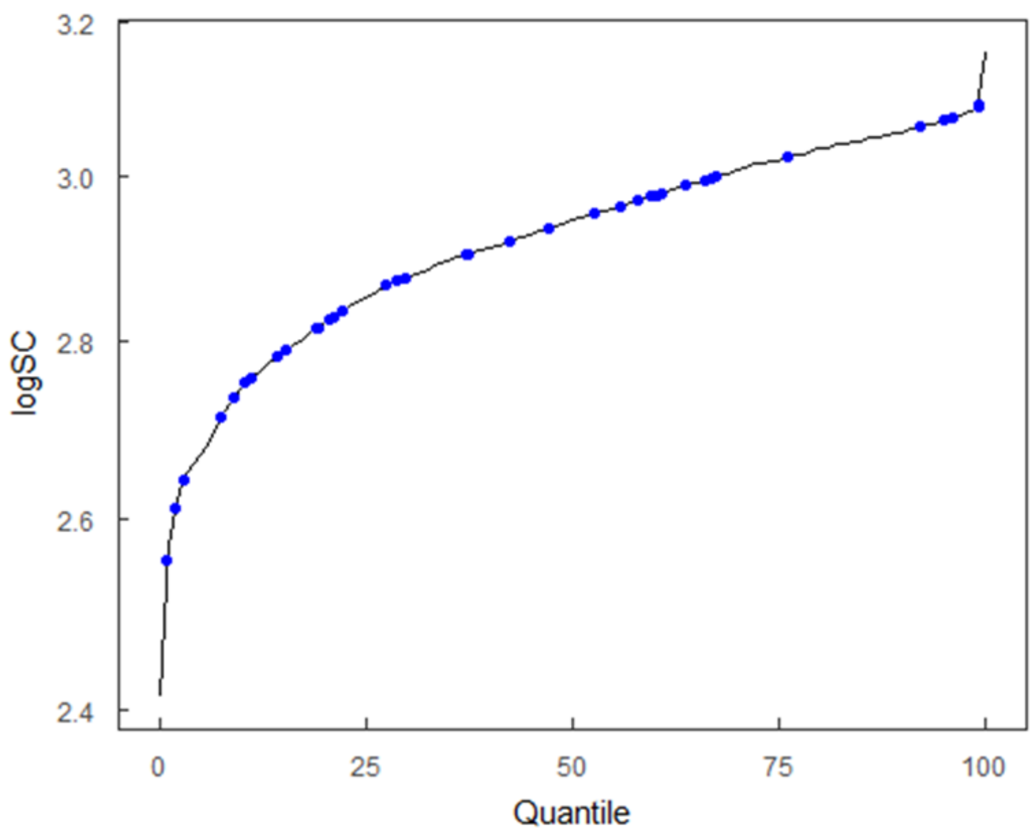

Figure 1. Duration plot of continuous log-transformed specific conductance (SC; black line) and measured specific conductance during discrete sample collection (blue dots) by quantile.

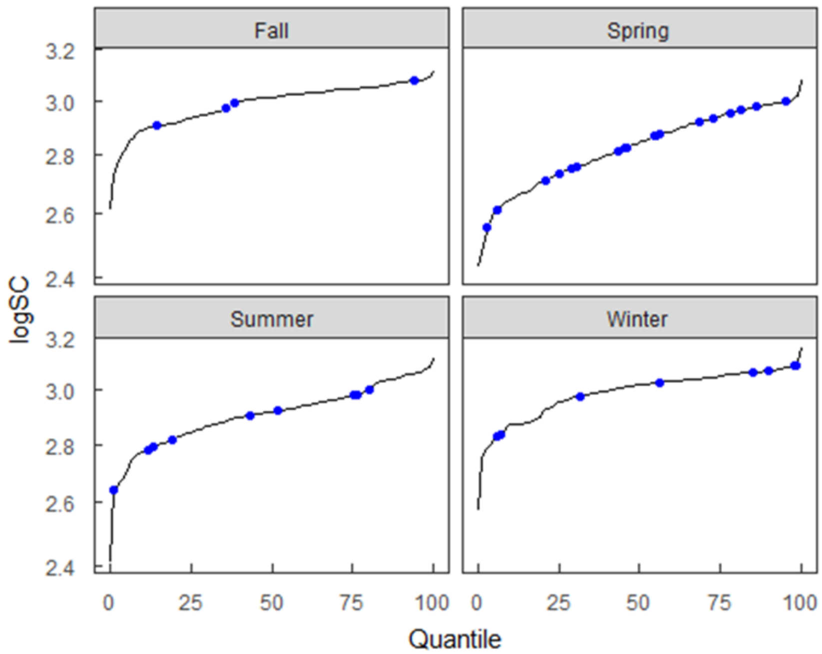

Figure 2. Seasonal duration plots of continuous log-transformed specific conductance (SC; black line) and measured specific conductance during discrete sample collection (blue dots) by quantile.

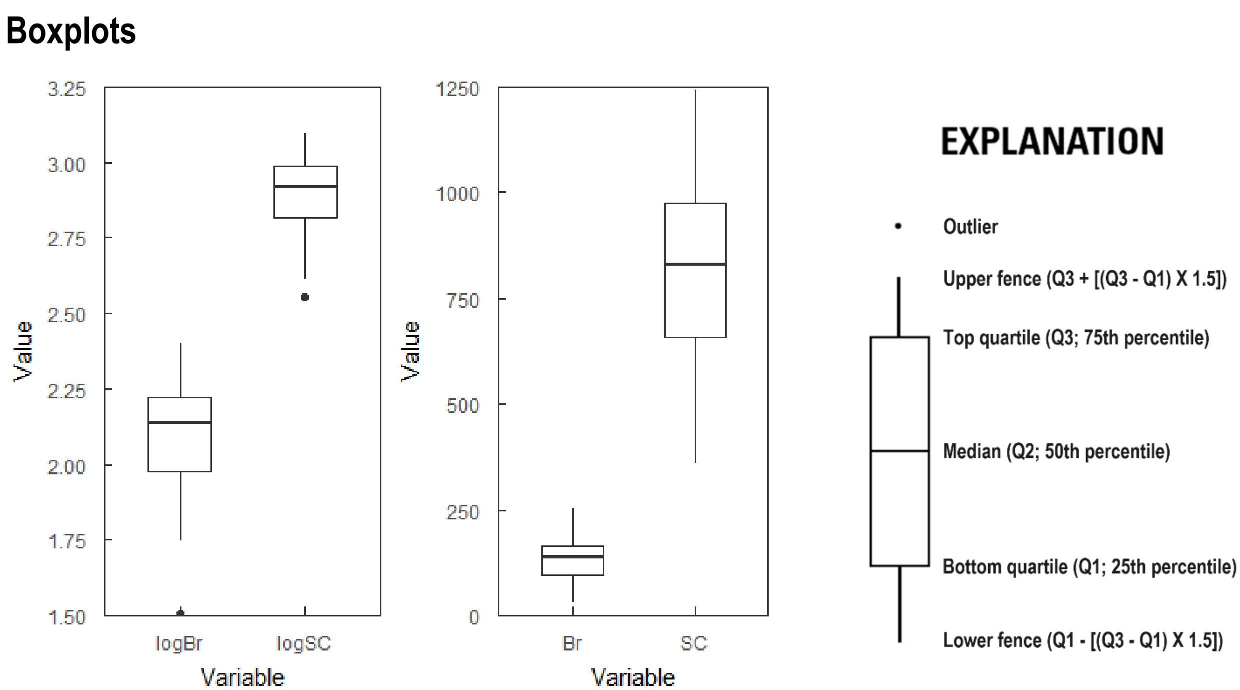

Figure 3. Boxplots of log-transformed (left) and untransformed (right) bromide concentration (Br) and specific conductance (SC) sample results used in the model-calibration dataset.

Scatterplots

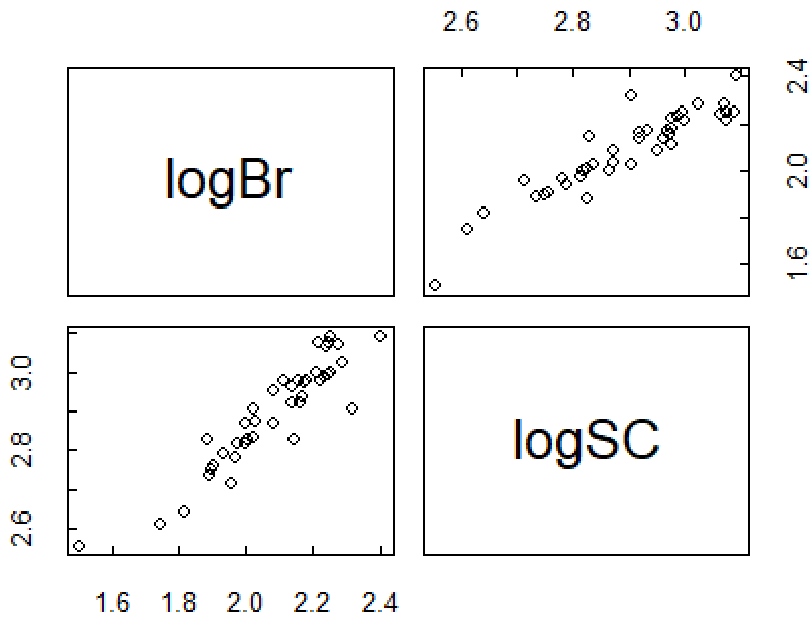

Figure 4. Bivariate plots of log-transformed bromide concentration (Br) and log-transformed specific conductance (SC). The x- and y-axis labels for a given bivariate plot are defined by the intersecting row and column labels.

Basic Model Statistics

Statistic	Value
Observations	41
R^2	0.858
Adjusted R^2	0.854
RMSE	0.0676
Upper MSPE (90 percent)	16.9
Lower MSPE (90 percent)	14.4
BCF	1.012

Model Coefficients

	Estimate	Standard Error	t value	Pr(>|t|)
(Intercept)	-1.459881	0.2314467	-6.307636	2e-07
logSC	1.223701	0.0798250	15.329803	0e+00

Correlation Matrix

	logBr	logSC
logBr	1.0000000	0.9261024
logSC	0.9261024	1.0000000

Outlier Test Criteria

Leverage	DFFITS	Cook’s D
0.1463	0.4417	0.1938

Flagged Observations

datetime	logBr	Cook’s D	DFFITS	Leverage	Studentized Residual	Flag*
2021-05-18 11:40:00	1.5	0.82	-1.4	0.187	-2.92	CL
2023-10-23 11:10:00	2.32	0.137	0.607	0.0245	3.83	DS

*C: Cook’s distance; L: Leverage; D: Difference in fits statistic; S: Studentized residual

Statistical Plots

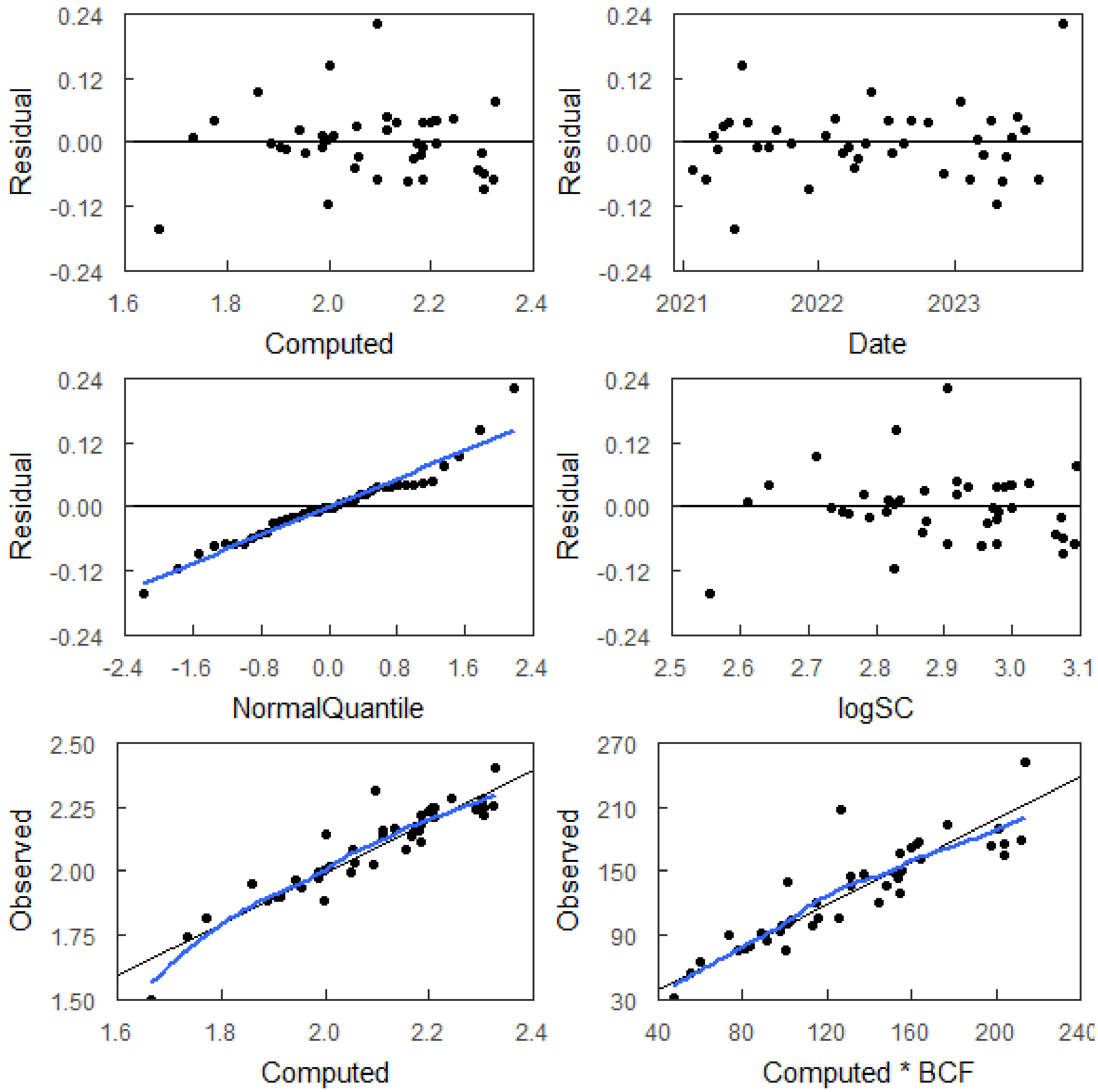

Figure 5. Statistical plots of model residuals relative to regression-computed bromide concentration, date, normal quantiles, log-transformed specific conductance (SC); and observed bromide concentration relative to regression-computed bromide concentration. Blue line shows the locally estimated scatterplot smoothing (LOESS). BCF=Duan's bias correction factor.

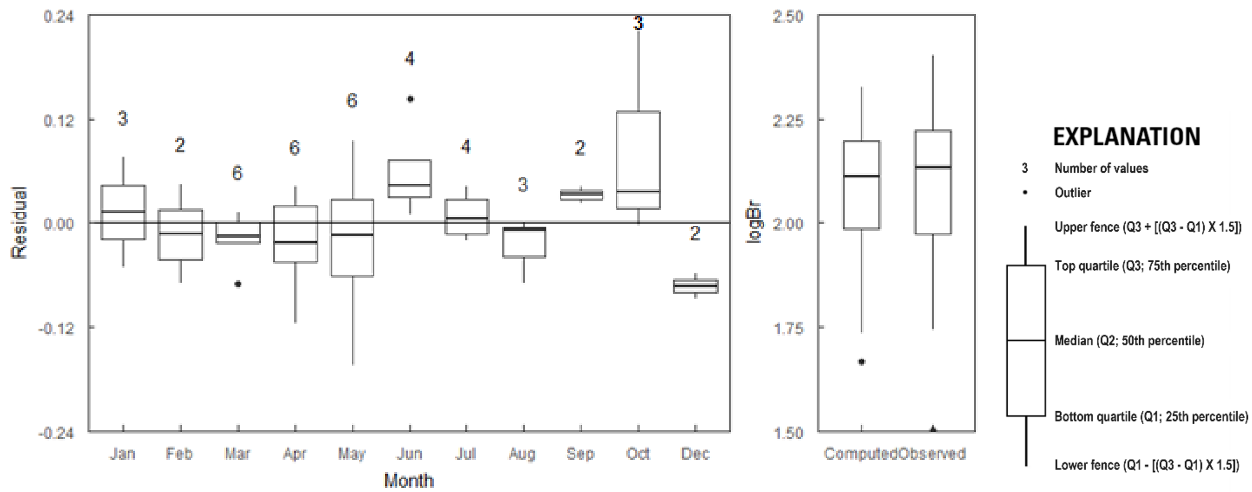

Figure 6. Boxplots of model residuals by month (left) and log-transformed computed and observed bromide concentrations (Br; right).

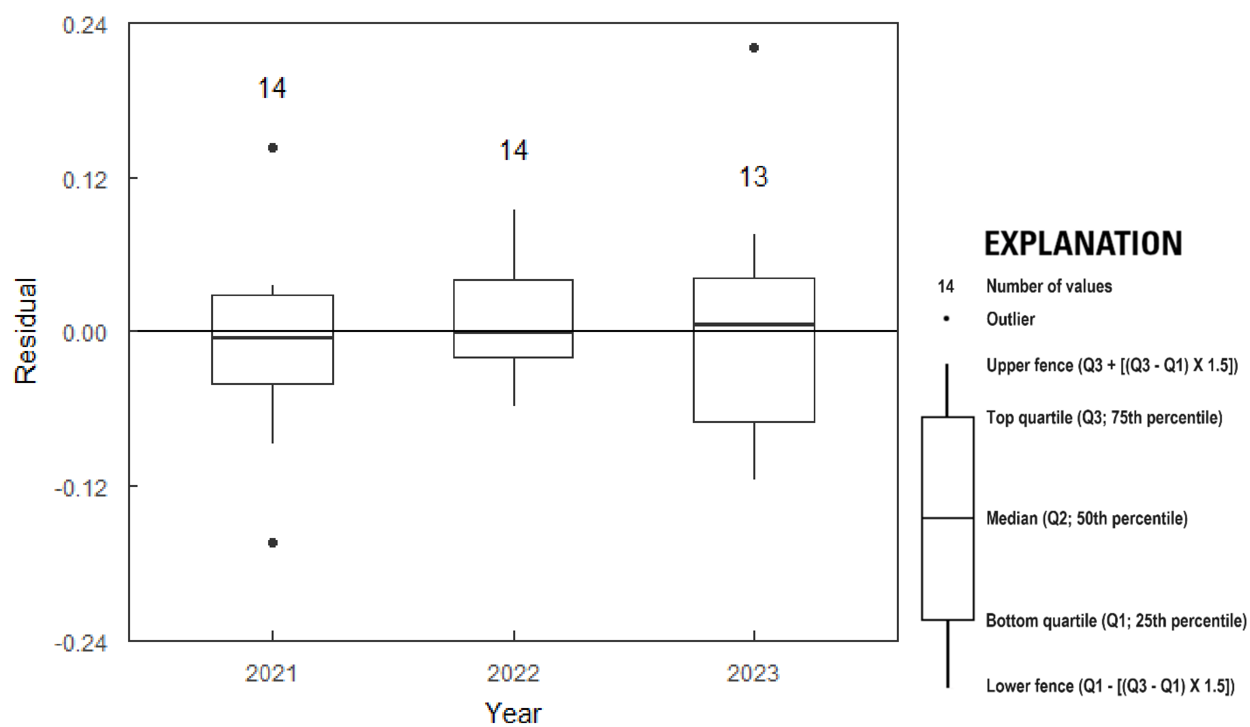

Figure 7. Boxplots of model residuals by year.

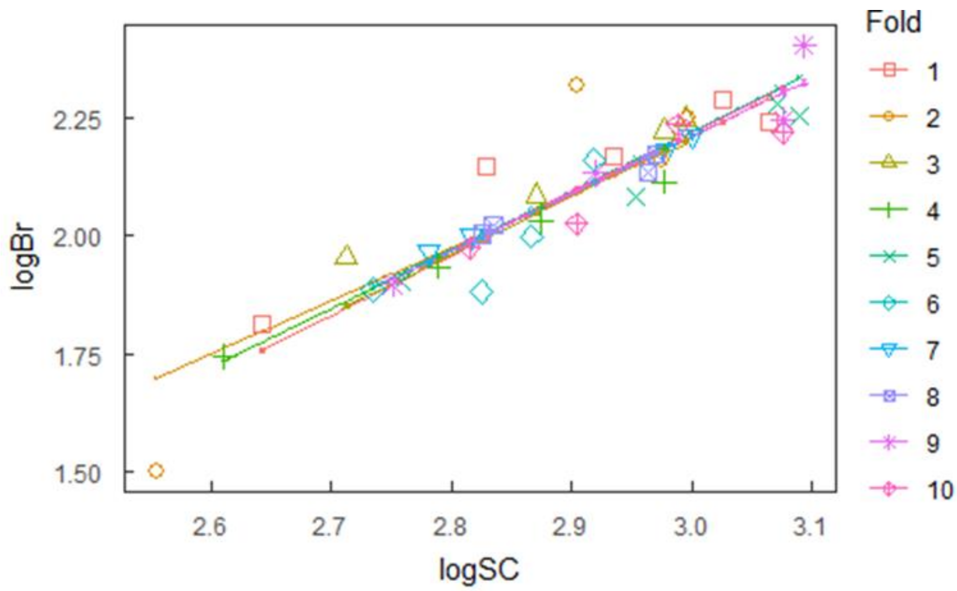

Figure 8. A 10-fold cross-validation plot (*fold*: equal partition of the data [10 percent of the data]; *large symbols*: observed value of a data point removed in a fold; *small symbols*: recomputed value of a data point removed in a fold; *recomputed regression lines*: adjusted regression line with one fold removed). Br=bromide concentration; SC=specific conductance.

Statistic	Value
Minimum MSE of folds	0.00271
25th Percentile	0.00459
Median MSE of folds	0.00472
Mean MSE of folds	0.00456
75th percentile	0.00489
Maximum MSE of folds	0.00507
Model MSE	0.00457

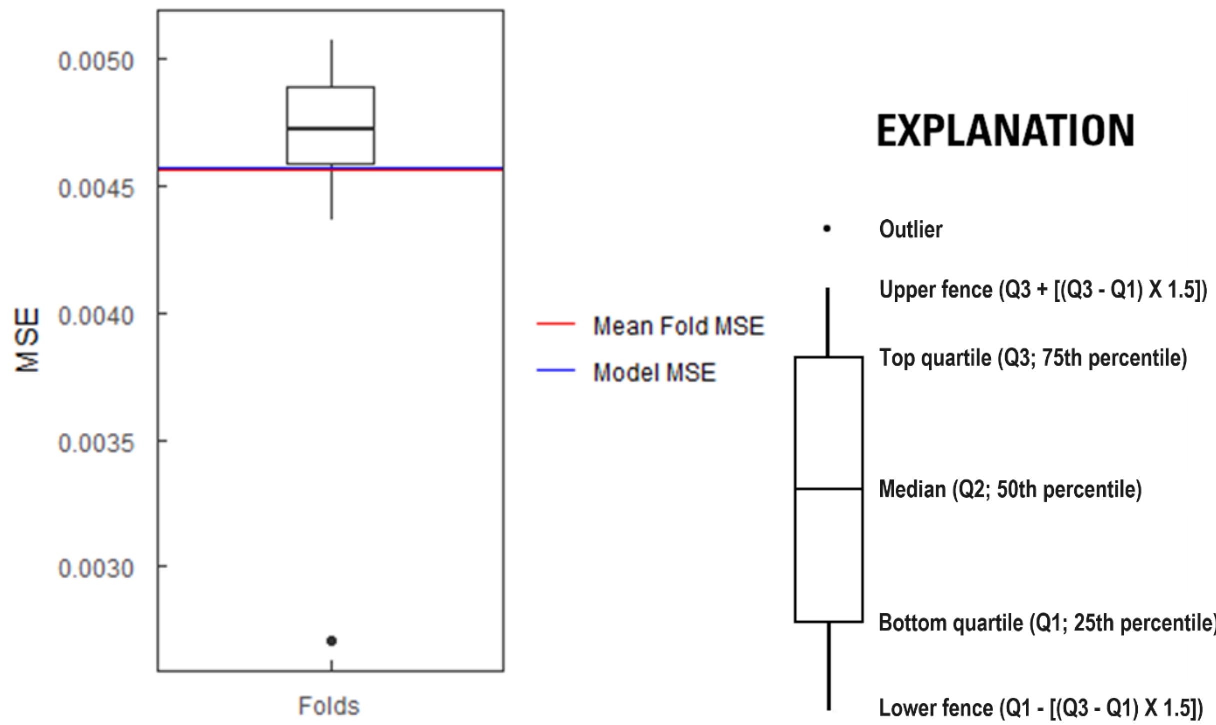

Figure 9. Boxplot of mean square error (MSE) of folds from cross validation.

Model-Calibration Dataset

Date/Time[1]	logBr	logSC	Br	SC	Computed logBr	Retransformed Br
2021-01-25 10:10:00	2.24	3.06	174	1160	2.29	197
2021-03-03 09:50:00	2.25	3.09	179	1230	2.32	212
2021-03-23 09:40:00	2.00	2.82	99.7	656	1.99	98.2
2021-04-05 10:20:00	1.90	2.76	79.8	573	1.92	83.3
2021-04-20 09:20:00	2.08	2.87	121	743	2.05	114
2021-05-03 10:00:00	2.17	2.94	147	862	2.13	137
2021-05-18 11:40:00	1.50	2.56	31.9	359	1.67	47.0
2021-06-07 09:50:00	2.14	2.83	139	674	2.00	102
2021-06-22 10:10:00	2.22	2.98	166	949	2.18	154
2021-07-19 10:20:00	1.97	2.82	94.3	654	1.99	97.8
2021-08-18 09:40:00	2.18	2.98	151	952	2.19	155
2021-09-08 09:10:00	1.97	2.78	92.6	604	1.94	88.9
2021-10-19 10:50:00	2.17	2.97	149	933	2.17	151
2021-12-06 11:10:00	2.22	3.08	165	1190	2.30	204
2022-01-18 10:10:00	2.02	2.84	105	684	2.01	103

Date/Time[1]	logBr	logSC	Br	SC	Computed logBr	Retransformed Br
2022-02-14 10:40:00	2.29	3.03	193	1060	2.24	177
2022-03-08 11:00:00	2.28	3.07	190	1180	2.30	202
2022-03-22 10:50:00	1.90	2.75	78.8	564	1.91	81.6
2022-04-04 10:30:00	2.00	2.87	99.7	736	2.05	113
2022-04-19 10:50:00	2.13	2.96	136	918	2.17	148
2022-05-09 11:10:00	1.89	2.74	77.0	543	1.89	78.0
2022-05-24 10:00:00	1.95	2.71	90.0	516	1.86	73.2
2022-07-06 10:20:00	1.81	2.64	65.3	439	1.77	60.0
2022-07-19 10:00:00	1.93	2.79	85.9	616	1.95	91.0
2022-08-16 10:40:00	2.21	3.00	162	1000	2.21	165
2022-09-07 10:00:00	2.25	3.00	178	995	2.21	164
2022-10-25 10:50:00	2.23	2.99	171	975	2.20	160
2022-12-05 11:10:00	2.25	3.08	176	1190	2.30	204
2023-01-17 10:40:00	2.40	3.09	251	1240	2.33	214
2023-02-13 10:30:00	2.11	2.98	130	948	2.18	154
2023-03-07 10:50:00	2.00	2.83	101	670	2.00	101
2023-03-21 11:00:00	2.16	2.98	144	945	2.18	154
2023-04-11 11:00:00	2.25	3.00	176	989	2.21	162
2023-04-24 10:10:00	1.88	2.82	76.1	668	2.00	101
2023-05-08 10:40:00	2.08	2.95	121	900	2.16	145
2023-05-23 10:30:00	2.03	2.87	107	749	2.06	116
2023-06-05 10:10:00	1.75	2.61	55.6	409	1.74	55.1
2023-06-20 11:10:00	2.16	2.92	145	830	2.11	131
2023-07-10 10:10:00	2.13	2.92	136	831	2.11	131
2023-08-15 09:40:00	2.02	2.90	106	802	2.09	126
2023-10-23 11:10:00	2.32	2.91	207	804	2.10	126

[1]Dates are formatted as “year-month-day” and times are formatted as “hours:minutes:seconds.”

References Cited

Cook, R.D., 1977, Detection of influential observations in linear regression: Technometrics, v. 19, no. 1, p. 15–18, accessed Mach 2024 at https://doi.org/10.2307/1268249.

Davis, B.E., and the Federal Interagency Sedimentation Project, 2005, A guide to the proper selection and use of federally approved sediment and water-quality samplers: U.S. Geological Survey, Open File Report 2005–1087, accessed July 2024 at https://doi.org/10.3133/ofr20051087.

Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: Journal of the American Statistical Association, v. 78, no. 383, p. 605–610, accessed March 2024 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 February 2022 at https://doi.org/10.3133/ofr20161040.

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., accessed March 2024 at https://doi.org/10.3133/tm4a3. [Supersedes USGS Techniques of Water-Resources Investigations, book 4, chap. A3, ver. 1.1.]

Hem, J.D., 1985, Study and interpretation of the chemical characteristics of natural water (3d ed.): U.S. Geological Survey Water-Supply Paper 2254, 263 p., 4 pls., accessed March 2024 at https://doi.org/10.3133/wsp2254.

Pardoe, I., 2020, Applied regression modeling (3d ed.): New York, Wiley, 336 p. [Also available at https://onlinelibrary.wiley.com/doi/book/10.1002/9781119615941.]

R Core Team, 2024, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing: accessed March 2024 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., accessed March 2024 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., accessed March 2024 at https://doi.org/10.3133/sir20055165.

U.S. Environmental Protection Agency, 1997, Method 300.1—Determination of inorganic anions in drinking water by ion chromatography: Cincinnati, Ohio, National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, 40 p., accessed March 2024 at https://www.epa.gov/sites/production/files/2015-06/documents/epa-300.1.pdf.

U.S. Geological Survey, 2024, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed March 2024 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 March 2024 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., accessed March 2024 at https://doi.org/10.3133/tm1D3.] [Supersedes USGS Water-Resources Investigations Report 2000–4252.]

Williams, T.J., 2021, Linear regression model documentation and updates for computing water-quality constituent concentrations or densities using continuous real-time water-quality data for the Kansas River, Kansas, July 2012 through September 2019: U.S. Geological Survey Open-File Report 2021–1018, 18 p., accessed March 2024 at https://doi.org/10.3133/ofr20211018.

Williams, T.J., 2023, Linear regression model documentation for computing water-quality constituent concentrations or densities using continuous real-time water-quality data for the Kansas River above Topeka Weir at Topeka, Kansas, November 2018 through June 2021: U.S. Geological Survey Scientific Investigations Report 2022–5130, 14 p., accessed March 2024 at https://doi.org/10.3133/sir20225130.