Appendix 8. Model Archive Summary for Suspended-Sediment Concentration at U.S. Geological Survey Station 07144780, North Fork Ninnescah River above Cheney Reservoir, Kansas, during November 14, 2015, through September 30, 2021

This model archive summary summarizes the suspended-sediment concentration model developed to compute 15-minute, hourly, or daily suspended-sediment concentrations during November 14, 2015, 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).

Site and Model Information

Site number: 07144780

Site name: North Fork Ninnescah River above Cheney Reservoir, Kansas

Location: Lat 37°51'45", long 98°00'49" referenced to North American Datum of 1927, in NE 1/4 SE 1/4 NE 1/4 sec.19, T.25 S., R.6 W., Reno County, Kans., hydrologic unit 11030014, on right bank at upstream side of county highway bridge, 10 miles south of Hutchinson, 18.1 miles upstream from Cheney Dam.

Equipment: A YSI, Inc., EXO water-quality monitor (YSI, Inc., 2017) equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, and turbidity was installed November 14, 2015. The EXO monitor was installed in a 4-inch-diameter metal or polyvinyl chloride (or PVC) pipe suspended from the downstream side of the bridge in the deepest, fastest flowing water. Measurements from the EXO were recorded every 15-minutes to hourly and transmitted hourly via satellite. Real-time stage was measured using a Design Analysis Water Log H–350/355 nonsubmersible pressure transducer.

Date model was created: August 9, 2022

Model calibration data period: April 19, 2016, through August 12, 2021 (dataset consisted of 31 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 National Water Information System (NWIS) 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, seasonality (sine and cosine variables), and streamflow.

The regression model is based on 31 concomitant values of discretely collected suspended sediment and continuously measured turbidity during April 19, 2016, through August 12, 2021. Discrete samples were collected throughout the range of continuously observed hydrologic conditions. No samples had suspended-sediment concentrations that were less than laboratory minimum reporting level. 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. No samples in the model calibration dataset were flagged as outliers.

Suspended Sediment Sampling Details

Discrete water-quality samples were collected over a range of hydrologic conditions primarily using a combination of equal depth- and width-integrated and multiple-vertical sample collection techniques (U.S. Geological Survey, 2006). Equal-width-increment and multiple-vertical sample cross sections included five to 12 sampling points with more than 85 percent of samples including 10 or more sampling points. Samples were collected either instream as a wading sample within 300 feet of the bridge or from the downstream side of the bridge using a Federal Interagency Sedimentation Project depth-integrated sampler with a polytetrafluoroethylene bottle, cap, and nozzle. Discrete samples were collected on a semifixed to event-based schedule two to seven times per year. Samples were analyzed for suspended-sediment concentration and loss-on-ignition in the USGS Sediment Laboratory in Iowa City, Iowa.

Continuous Water-Quality Data

Turbidity was 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). All continuous water-quality data at the North Fork Ninnescah River above Cheney Reservoir 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 07144780.

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.

Turbidity was selected as a good surrogate for suspended-sediment 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 suspended-sediment concentration regression analysis at USGS site 07144780:

Suspended-sediment concentration-based model:

$$\log_{10}(SSC) = (1.07 \times \log_{10}(TBY)) + 0.399,$$

where,

> *SSC* = suspended-sediment concentration, in milligrams per liter (mg/L) (USGS parameter code 80154);
>
> *TBY* = turbidity, monochrome near infra-red light-emitting diode light, 780-900 nanometers detection angle 90 +-2.5 degrees, formazin nephelometric units (FNU) (USGS parameter code 63680); and
>
> $\log_{10}$ = decimal logarithm.

The $\log_{10}$-transformed model may be retransformed to the original units so that suspended-sediment 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 1.10. The retransformed model, accounting for BCF, is as follows:

$$SSC = (TBY^{1.07} \times 10^{0.399}) \times 1.10.$$

Turbidity is an indicator of sediment and other suspended materials in streams and lakes, and therefore is a logical explanatory variable for suspended sediment.

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 suspended-sediment concentration using this model is 608.3 mg/L. 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)
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)
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)
logSSC	Suspended-sediment concentration, in milligrams per liter (mg/L), $\log_{10}$-transformed;
logTBY	Turbidity, monochrome near infra-red LED light, 780-900 nm, formazin nephelometric units (FNU) (USGS parameter code 63680), $\log_{10}$ transformed
MSE	Model standard 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)
RMSE	Root mean square error (Helsel and others, 2020)
SSC	Suspended-sediment concentration, in milligrams per liter (mg/L) (USGS parameter code 80154);
t value	Student's *t* value; the coefficient divided by its associated standard error (Helsel and others, 2020)

Model

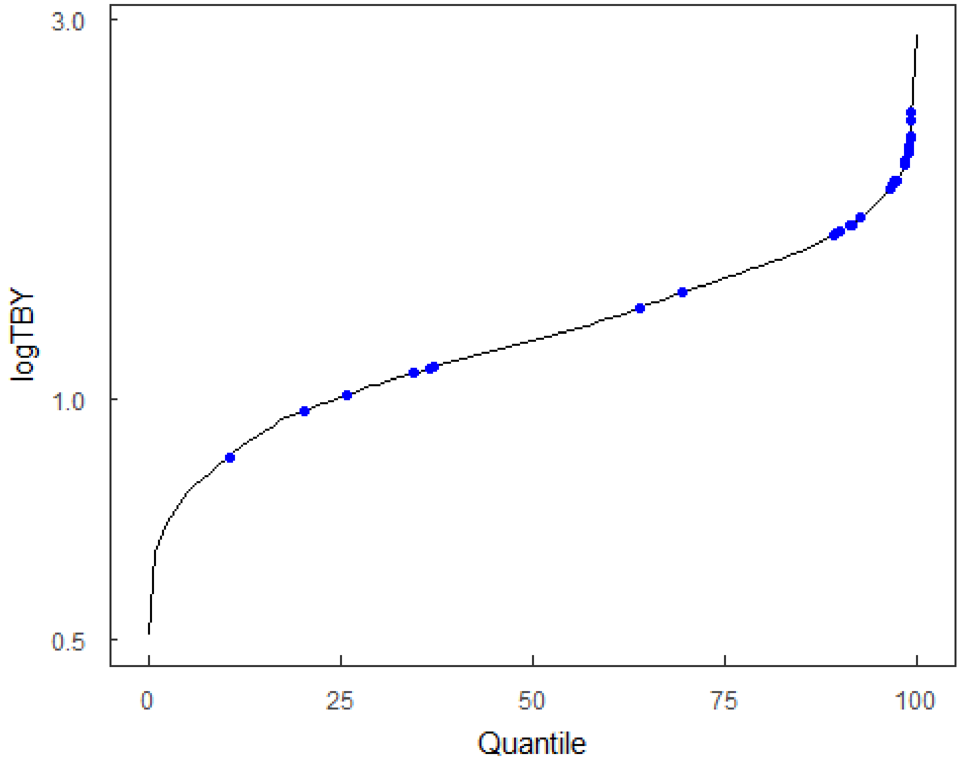
$$\log_{10}(SSC) = (1.07 \times \log_{10}(TBY)) + 0.399$$

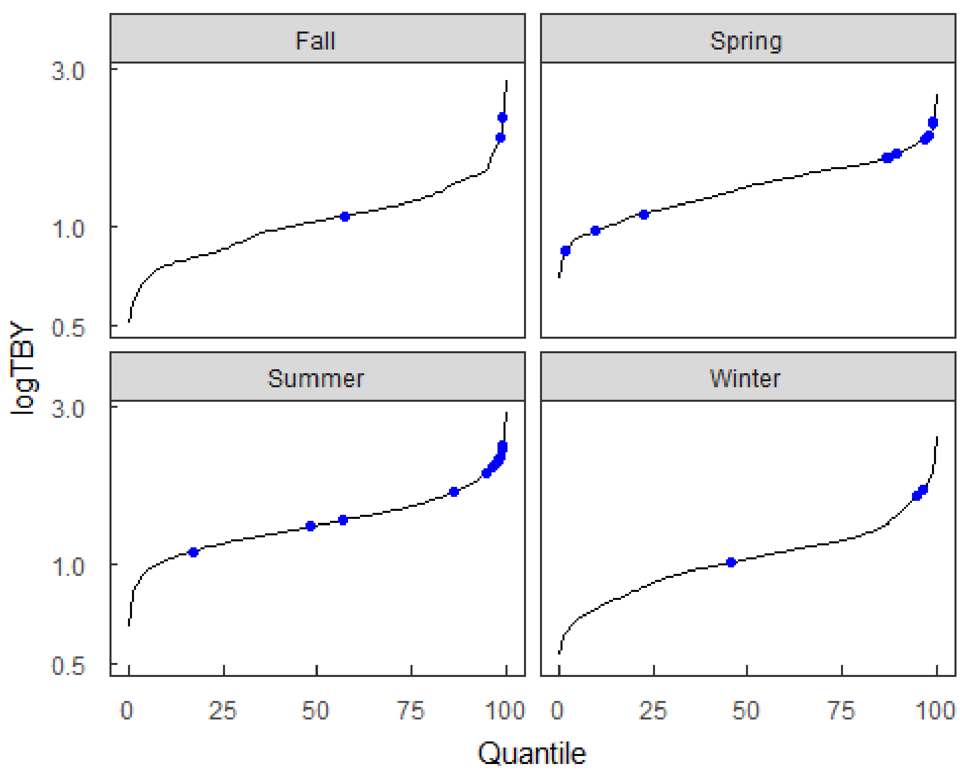
Variable summary statistics

Variable	Minimum	Q1	Median	Mean	Q3	Maximum
logSSC	1.36	1.82	2.44	2.23	2.64	2.74
logTBY	0.847	1.49	1.86	1.72	2.06	2.3
SSC	23	69.5	278	264	432	553

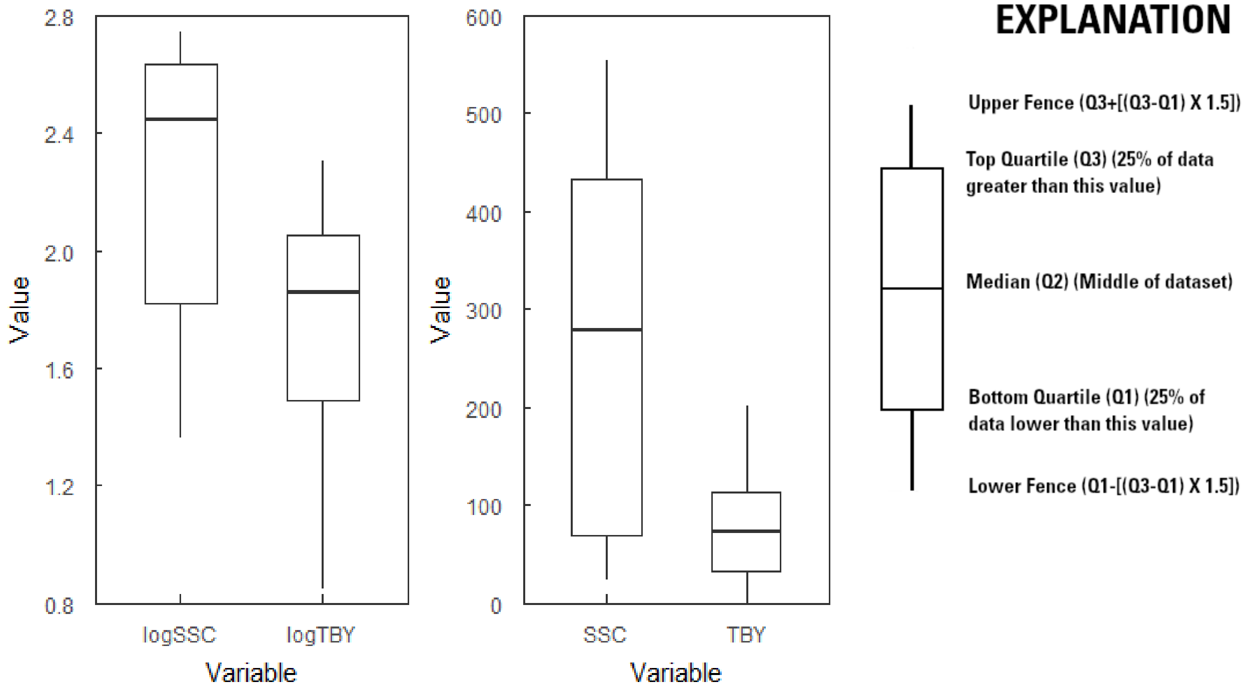
TBY	7.03	32	72.3	74.9	114	201

Duration plots

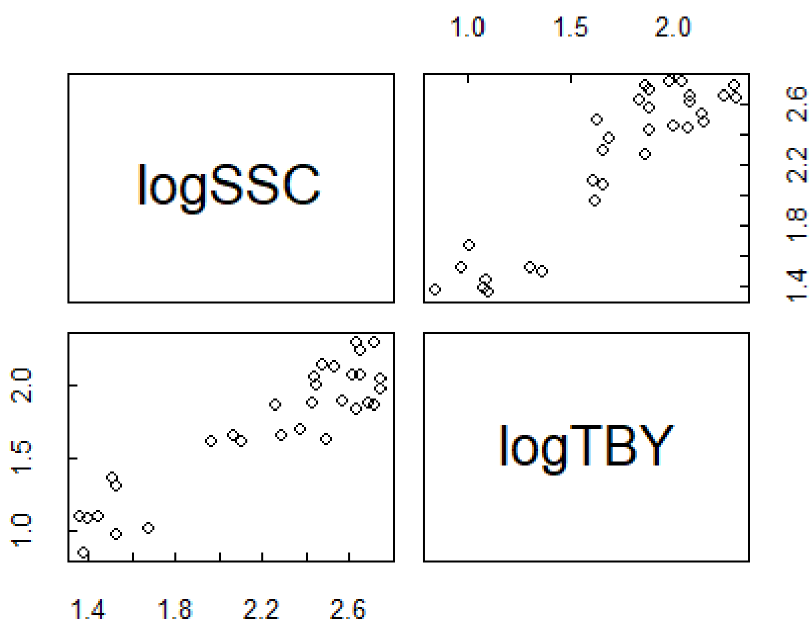




Box plots



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	31
R^2	0.848
Adjusted R^2	0.843
RMSE	0.195
Upper MSPE (90%)	56.7
Lower MSPE (90%)	36.2
BCF	1.1

Model coefficients

	Estimate	Standard error	t value	Pr(>|t|)
(Intercept)	0.3985774	0.1480923	2.691413	0.0116874
logTBY	1.0668335	0.0838652	12.720805	0.0000000

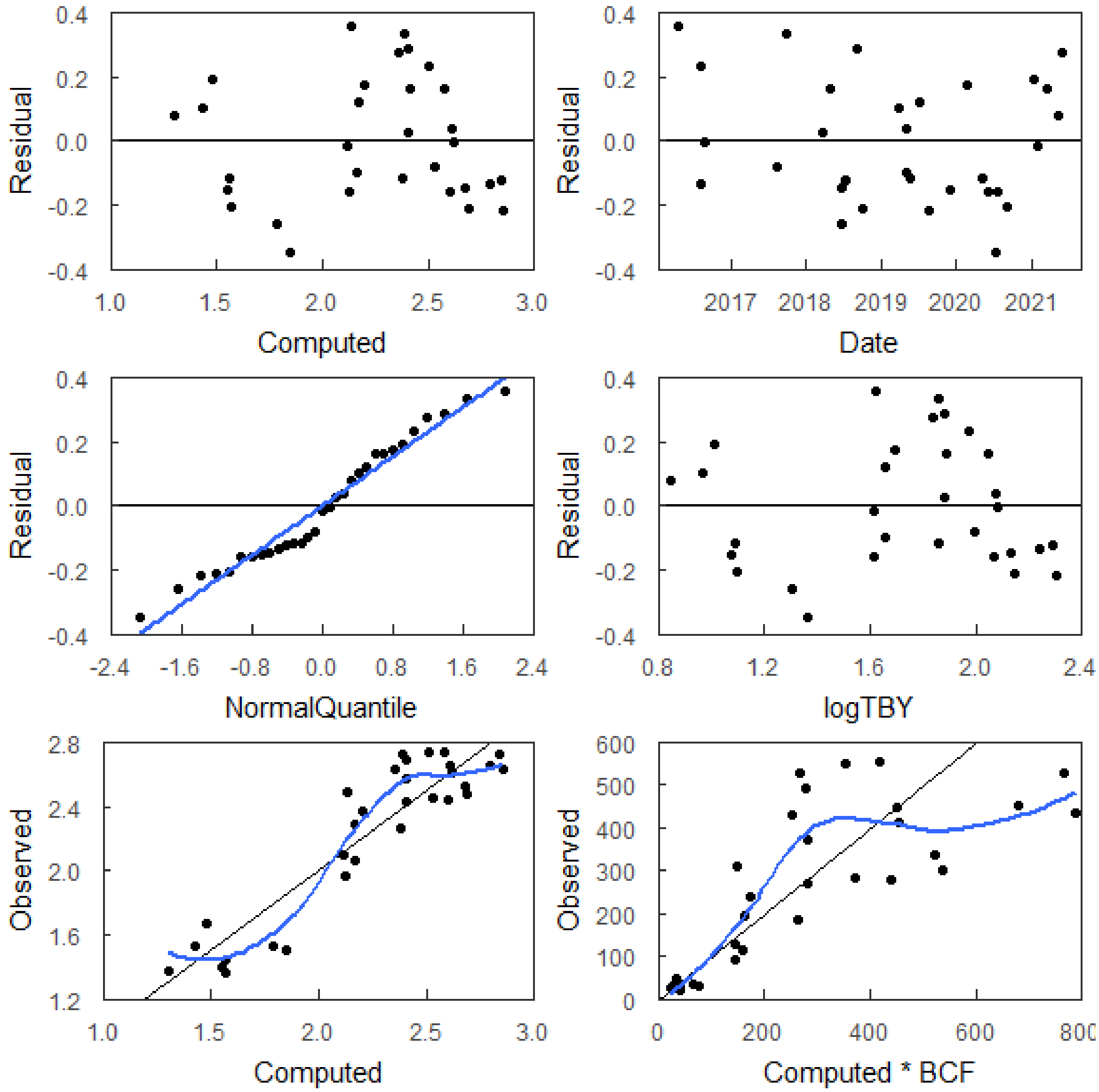
Correlation matrix

	logSSC	logTBY
logSSC	1.0000000	0.9208819
logTBY	0.9208819	1.0000000

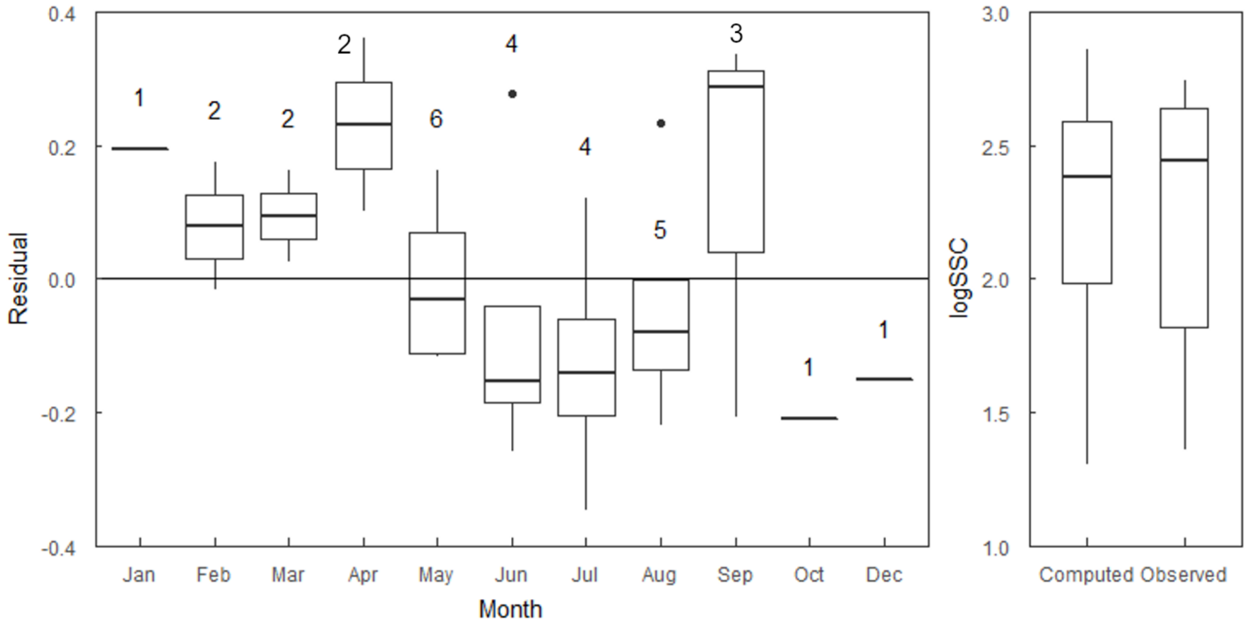
Outlier test criteria

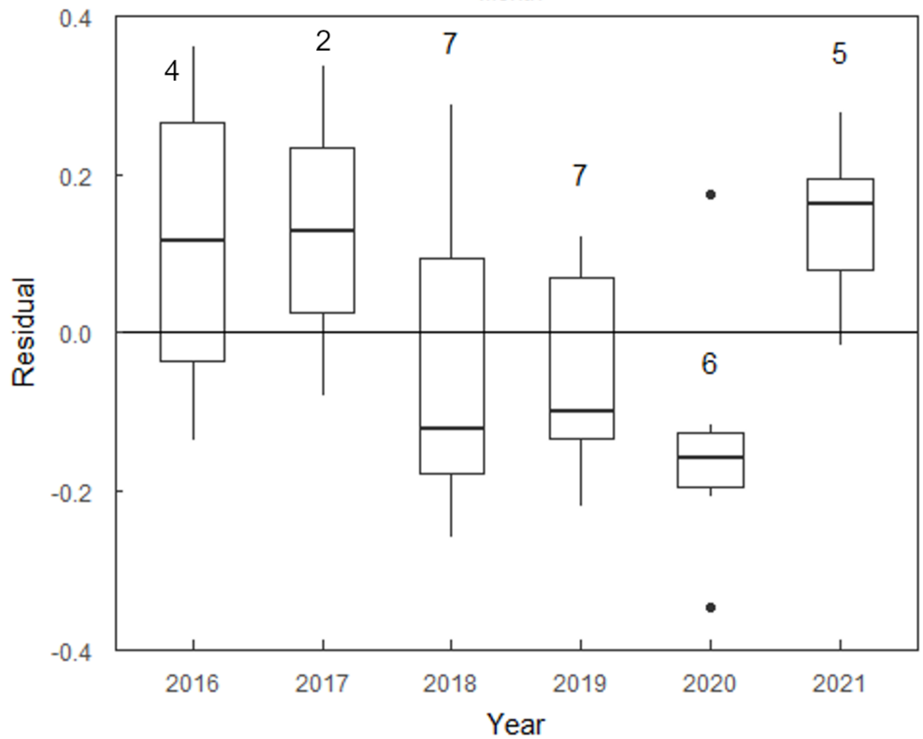
Leverage	DFFITS	CooksD
0.1935	0.508	0.1935

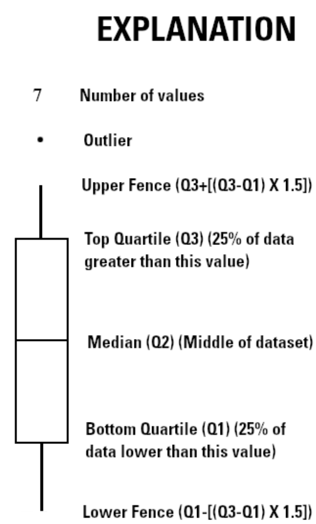
Statistical plots



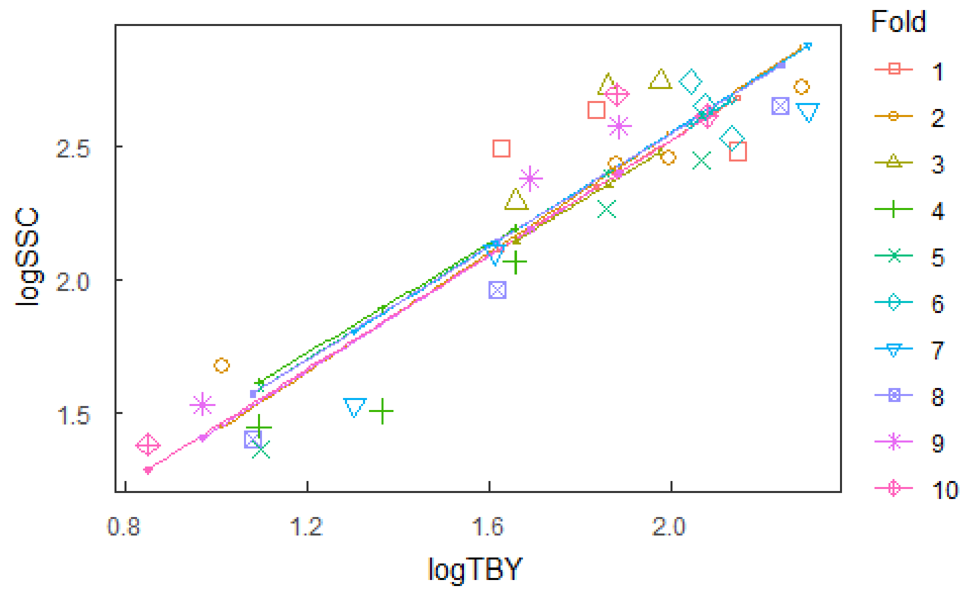
The blue line shows the locally estimated scatterplot smoothing (LOESS). The black dots correspond to observed values. The black line represents the 1:1 line.







Cross Validation



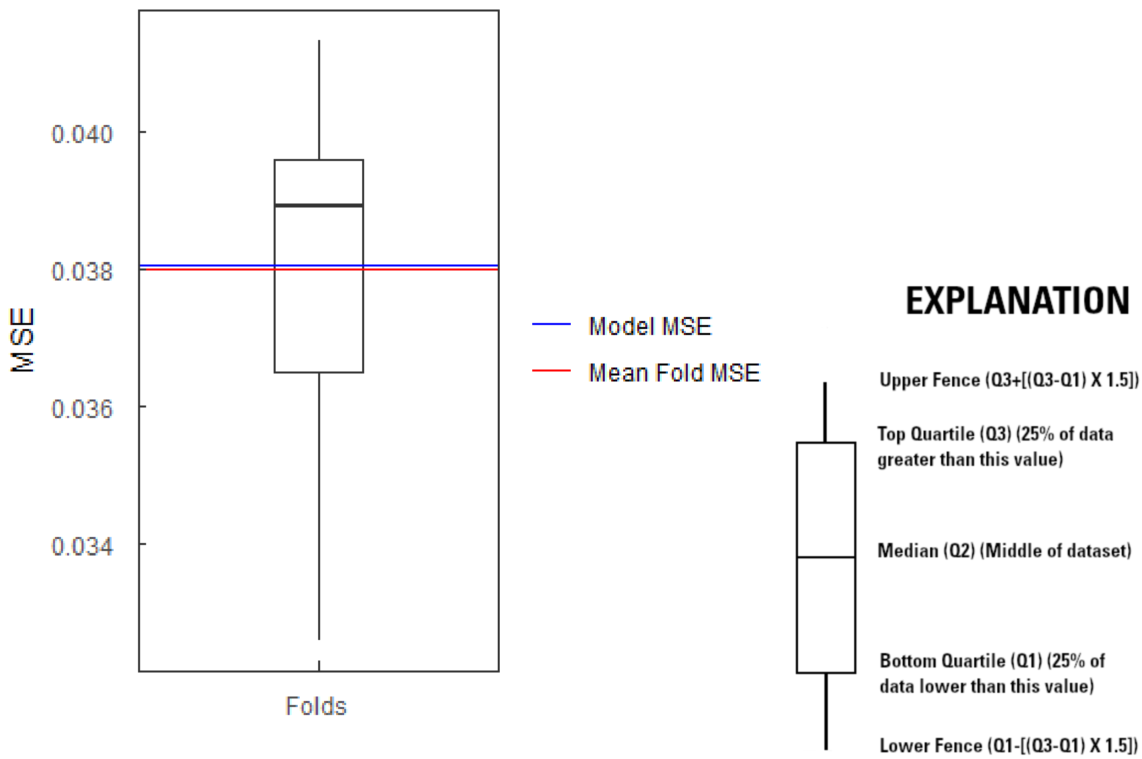
Fold - equal partition of the data (10 percent of the data).

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	0.0326
25th Percentile	0.0365
Median MSE of folds	0.0389
Mean MSE of folds	0.038
75th percentile	0.0396
Maximum MSE of folds	0.0413
Model MSE	0.0381


0.040
0.038
0.036
0.034
MSE
Folds
Model MSE
Mean Fold MSE
EXPLANATION
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])

Model calibration dataset

datetime	logSSC	logTBY	SSC	Computed	Retransformed
2016-04-19 10:25:00	2.49	1.62	310	2.13	149
2016-08-07 12:30:00	2.65	2.24	451	2.79	681
2016-08-13 11:30:00	2.74	1.98	552	2.51	355
2016-08-27 09:20:00	2.61	2.08	411	2.62	456
2017-08-11 11:00:00	2.45	2	282	2.53	372
2017-09-28 10:30:00	2.72	1.86	527	2.39	268
2018-03-20 10:30:00	2.43	1.88	270	2.41	280
2018-05-04 10:00:00	2.57	1.89	374	2.41	283
2018-06-21 10:10:00	1.53	1.3	34	1.79	67.8
2018-06-26 13:20:00	2.53	2.13	339	2.68	522
2018-07-14 12:00:00	2.72	2.29	526	2.84	766
2018-09-05 09:55:00	2.69	1.88	493	2.41	280
2018-10-09 10:10:00	2.48	2.15	302	2.69	538
2019-04-02 10:50:00	1.53	0.968	34	1.43	29.7
2019-05-02 11:20:00	2.06	1.66	116	2.17	161
2019-05-08 12:00:00	2.65	2.07	447	2.61	451
2019-05-21 12:30:00	2.26	1.86	184	2.38	265
2019-07-08 11:30:00	2.29	1.66	195	2.17	163
2019-08-26 11:30:00	2.64	2.3	433	2.86	789
2019-12-03 10:20:00	1.4	1.08	25	1.55	39.1
2020-02-26 10:30:00	2.38	1.69	238	2.2	175
2020-05-07 10:30:00	1.45	1.09	28	1.56	40.4
2020-06-04 10:20:00	1.96	1.62	92	2.12	146
2020-07-08 11:00:00	1.51	1.36	32	1.85	78.4
2020-07-21 10:10:00	2.44	2.07	278	2.6	441
2020-09-03 10:20:00	1.36	1.1	23	1.57	40.8
2021-01-12 10:10:00	1.67	1.01	47	1.48	33.1
2021-02-01 11:00:00	2.1	1.61	126	2.12	144
2021-03-23 11:40:00	2.74	2.04	553	2.58	418
2021-05-10 10:50:00	1.38	0.847	24	1.3	22.1
2021-06-01 10:40:00	2.63	1.84	431	2.36	251

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.

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.]

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.]

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.