Model Archival Summary for Suspended-Sediment Concentration at U.S. Geological Survey Site 06887000, Big Blue River near Manhattan, Kansas, during July 2018 through August 2023

This model archival summary summarizes the suspended-sediment concentration (SSC; U.S. Geological Survey [USGS] parameter code 80154) model developed to compute 15-minute, hourly, or daily SSC from July 26, 2018, onward. This model is specific to USGS site 06887000, the Big Blue River near Manhattan, Kansas, during this study period and cannot be applied to data collected from other locations on the Big Blue River or data collected from other waterbodies. The methods follow 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: 06887000

Site name: Big Blue River near Manhattan, Kans.

Location: Lat 39°14'14", long 96°34'16" referenced to North American Datum of 1927, in NW 1/4 SW 1/4 NE 1/4 sec.30, T.9 S., R.8 E., Riley County, Kans., hydrologic unit 10270205.

Equipment: A Xylem YSI, Inc., EXO water-quality monitor (YSI, Inc., 2017) equipped with sensors for water temperature, specific conductance, dissolved oxygen (added in December 2018), pH (added in April 2023), and turbidity (TBY), was installed during July 2018 through August 2023. Readings from the water-quality monitor were recorded every 15 minutes and transmitted via satellite, hourly.

Date model was created: November 2, 2023

Model-calibration data period: July 17, 2018, through August 29, 2023

Model-application date: July 26, 2018, onward

Model computations are available at the USGS National Real-Time Water-Quality webpage (https://nrtwq.usgs.gov/explore/dyplot?site_no=06887000).

Suspended-Sediment Sampling Details

During July 2018 through August 2023, suspended-sediment samples were collected over a range of hydrologic conditions (fig. 1–2). Of the 36 samples included in this dataset, 32 suspended-sediment samples were collected using the isokinetic equal-width increment collection method (U.S. Geological Survey, 2006). Three samples, collected during September 2020, were collected using the non-isokinetic single-vertical collection method (U.S. Geological Survey, 2006) due to estimated stream-surface velocities of less than one foot per second and personnel constraints influenced by the COVID-19 pandemic. One sample was collected using a non-isokinetic multiple vertical collection method (U.S. Geological Survey, 2006) on February 21, 2019, due to ice cover. Sample collection was suspended during October 2020 through March 2023 due to project delays and funding constraints.

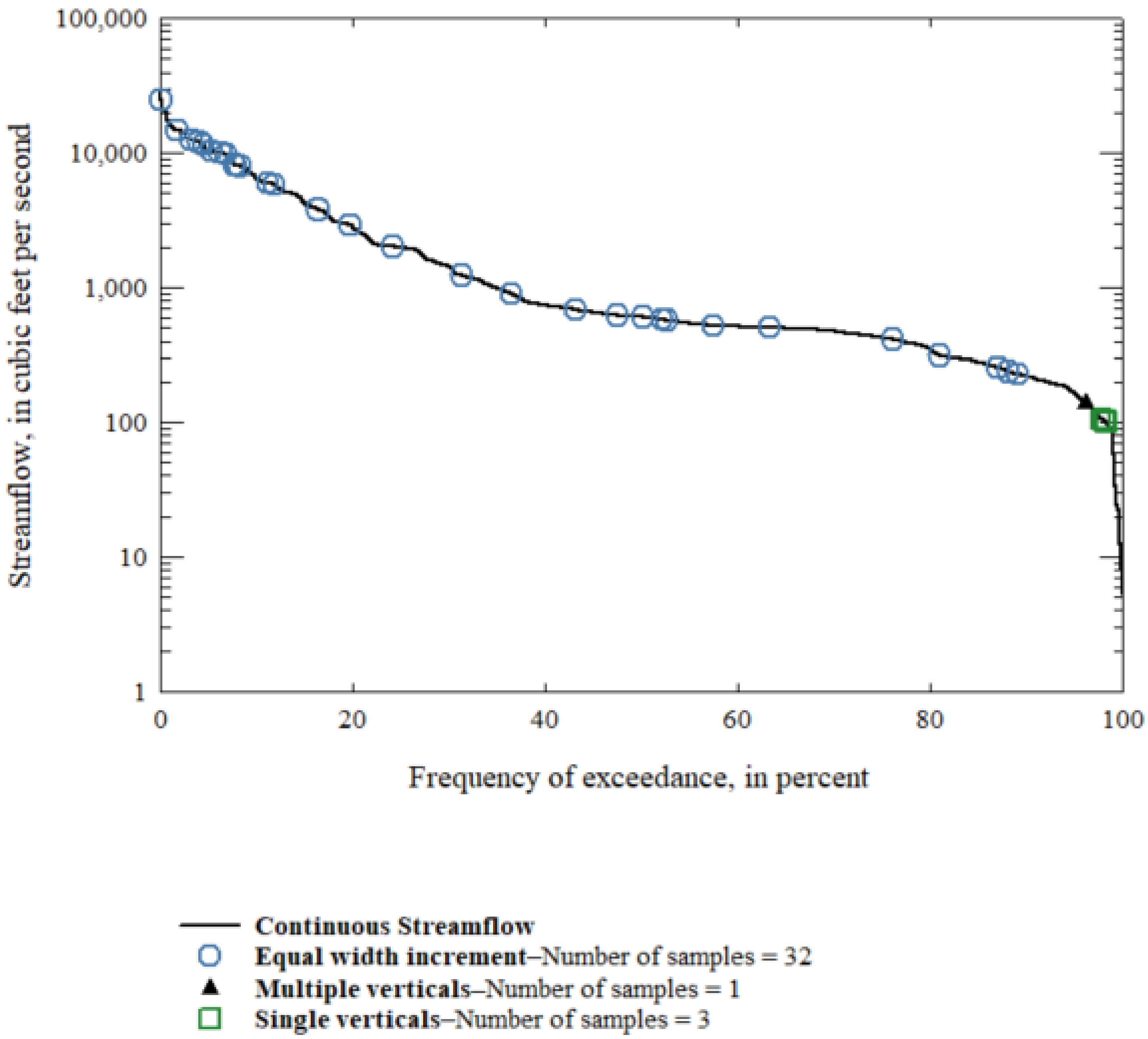

Figure 1. Streamflow duration curve and discrete water-quality samples by sampling method collected at the Big Blue River near Manhattan, Kansas streamgage (U.S. Geological Survey station 06887000) during July 2018 through August 2023.

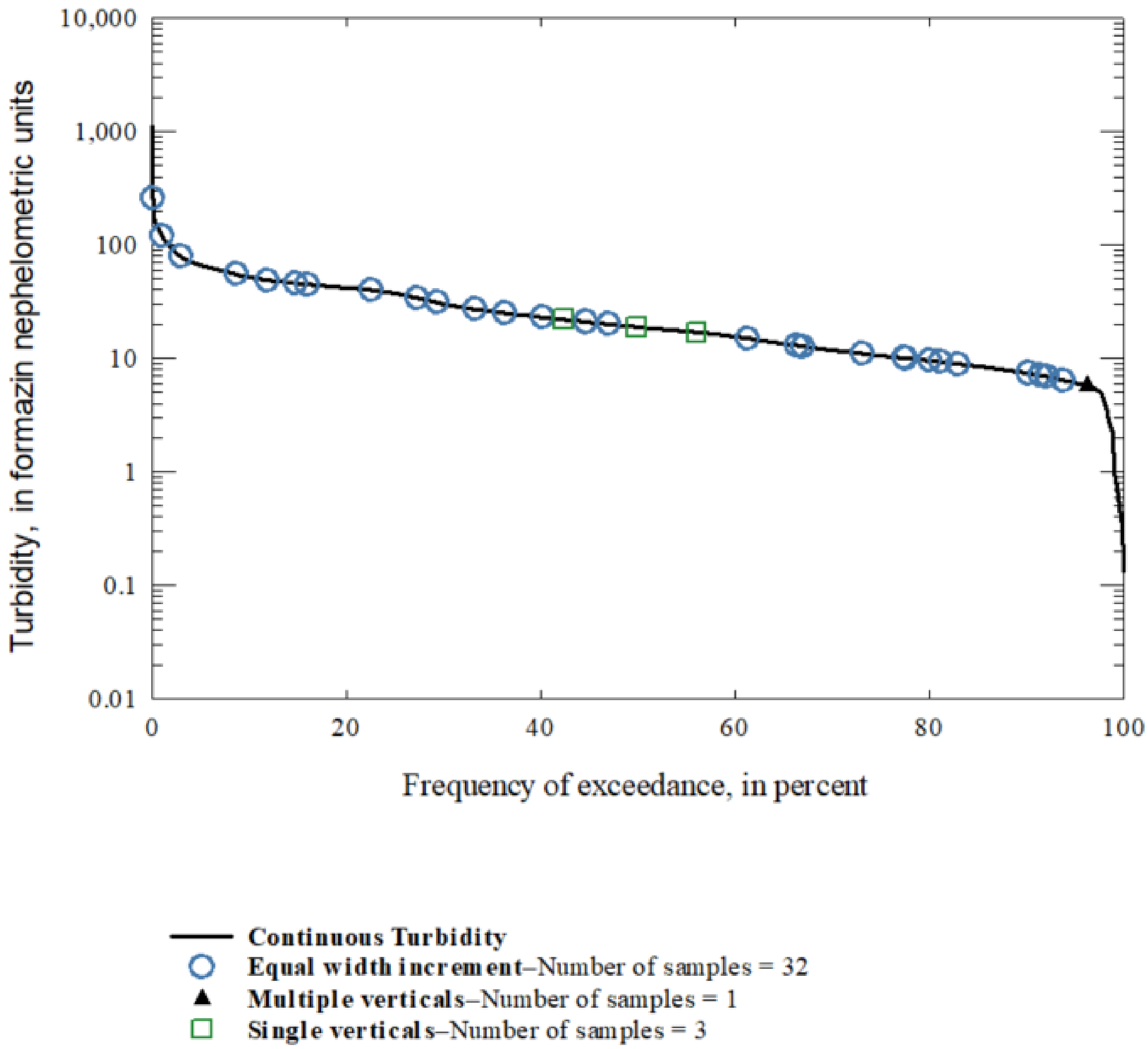

Figure 2. Turbidity duration curve and discrete water-quality samples by sampling method collected at the Big Blue River near Manhattan, Kansas streamgage (U.S. Geological Survey station 06887000) during July 2018 through August 2023.

Models developed using samples collected with different methods have some limitations. Discrete water-quality sampling method differences may affect the applicability of these models to estimate constituent discharge, particularly for suspended constituents, because of uneven particle distribution (U.S. Geological Survey, 2006). Isokinetic, depth-integrated methods (equal-width-increment samples) are designed to produce a discharge-weighted sample, meaning each unit of stream discharge is represented in the sample, whereas non-isokinetic sampling methods (single- and multiple-vertical samples) do not result in a discharge-weighted sample unless the stream is completely mixed laterally and vertically (U.S. Geological Survey, 2006). Due to the few number of non-isokinetic samples collected in this model calibration dataset as well as low turbidity conditions during these samples, there is likely minimal impact on the model and these samples were retained in the dataset.

Samples occasionally were collected during targeted reservoir release and runoff events to get a more representative dataset. A Federal Interagency Sedimentation Project US DH-95, D-95, D-96-A1, or DH-2 depth integrating sampler was used (Davis, 2005). Samples were analyzed for SSC at the USGS Iowa Sediment Laboratory in Iowa City, Iowa, using the methods documented by Guy (1969). All suspended-sediment data are available in the USGS National Water Information System database (https://doi.org/10.5066/F7P55KJN; U.S. Geological Survey, 2023) using site number 06887000.

Continuous Water-Quality Data

Continuous (15-minute) water temperature, specific conductance, and TBY data were measured using a water-quality multiparameter monitor during July 26, 2018, onward. Continuous dissolved oxygen and pH were added to the monitor during December 2018 and April 2023, respectively. Continuous water-quality data collection was suspended during October 2020 through March 2023 due to project delays and funding constraints. The water-quality monitor was operated and maintained according to USGS protocols (Wagner and others, 2006; Bennett and others, 2014) and was deployed by suspension from a bridge about 1 to 3 feet below the water surface at the centroid of flow. All continuous water-quality monitor data for the Big Blue River near Manhattan, Kans. are available in near real time (hourly) in the USGS National Water Information System database (https://doi.org/10.5066/F7P55KJN; U.S. Geological Survey, 2023) using site number 06887000.

Quality-Assurance and Quality-Control

All SSC results collected during July 2018 through August 2023 were reviewed and approved following USGS guidance (Rasmussen and others, 2014). Concurrent replicate quality-control samples were collected for about 10 percent of all SSC samples to characterize variability potentially introduced by sample collection techniques and analytical method (Rasmussen and others, 2014; Mueller and others, 2015). Relative percentage difference (RPD; Zar, 1999) was used to quantify variability among paired concurrent replicate samples. Quality-control objectives were met if the median RPD of SSC concurrent replicate pairs was less than or equal to 10 percent. The median RPD of the SSC concurrent replicate pairs collected from the Big Blue River near Manhattan, Kans. during July 2018 through August 2023 was about 6 percent (n=4; minimum=0 percent; maximum=12 percent).

All continuous water-quality data collected during July 2018 through August 2023 were reviewed and approved quarterly, following USGS guidance (U.S. Geological Survey, 2017). Occasionally, data were corrected or deleted because of sensor calibration drift, fouling, or equipment malfunction (Wagner and others, 2006; Bennett and others, 2014). The continuous water-quality monitor was occasionally removed to prevent damage or loss during sub-freezing temperatures, resulting in missing data. During July 2018 through August 2023, about 3 percent of the water temperature and pH records, 4 percent of the turbidity and dissolved oxygen records, and about 6 percent of the specific conductance record were missing or deleted because of excessive fouling, equipment malfunction, or temporary monitor removal during sub-freezing temperatures.

Model-Calibration Dataset

All data were collected using USGS protocols (Wagner and others, 2006; Bennett and others, 2014; U.S. Geological Survey, 2006), are stored in the USGS National Water Information System (U.S. Geological Survey, 2023) database, and are available to the public. Ordinary least squares analysis was used to develop the SSC regression model using R programming language (R Core Team, 2020). Potential explanatory variables evaluated individually and in combination included streamflow, water temperature, specific conductance, dissolved oxygen, pH, and TBY. These potential explanatory variables were time interpolated between the two nearest continuous water-quality readings. The maximum time span between two continuous data points used for interpolation was 2 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 2 hours in the continuous data record resulted in missing interpolated data; this only occurred once and the field monitor average was used for the sample collected July 17, 2018. Seasonal components (sine and cosine variables) also were evaluated as potential explanatory variables. There have been no previously published SSC models at this location.

The final selected regression model was based on 36 concomitant measurements of SSC and sensor-measured TBY during July 17, 2018, through August 29, 2023. Samples were collected throughout the range of continuously observed hydrologic conditions (fig. 1–2). No sediment samples had concentrations below laboratory minimum reporting limits.

Potential outliers initially were identified using scatterplots of the SSC and TBY 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 in question (Rasmussen and others, 2009). All 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* (*v4.0.2*) package in R programming language (R Core Team, 2020) to relate discretely collected SSC to sensor-measured TBY. The data and resultant regression equation generally conform with the five assumptions for OLS regression: the dependent variable was linearly related to the explanatory variables, data used to fit the model were representative of the data of interest, the variance of the residuals was constant (homoscedastic), the residuals were independent of the explanatory variables, and the residuals were normally distributed (Helsel and others, 2020).

TBY was selected as a good surrogate for SSC 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.

Model Summary

The following is a summary of the final regression analysis for SSC at USGS site 06887000:

SSC-based model:

$$\log SSC = 0.174 + 0.961 * \log TBY$$

where,

log = logarithm base 10.

SSC = suspended-sediment concentration, in milligrams per liter (mg/L; USGS parameter code 80154), and

TBY = turbidity, monochrome near infrared light-emitting diode light, 780-900 nanometers detection angle 90 +-2.5 degrees, formazin nephelometric units (FNU; USGS parameter code 63680)

TBY is a logical explanatory variable, physically and statistically, for SSC because it is an indicator of sediment and other suspended materials in streams and lakes.

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

$$SSC = 1.08 \times (TBY^{0.961} \times 10^{0.174})$$

This model was developed using continuous and discrete water-quality data collected during July 2018 through September 2023. These data were collected throughout the observed range of streamflow conditions during this time. Extrapolation, defined as computation beyond the range of the model calibration dataset, should 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 430.1 mg/L. Computed estimates exceeding that limit are not supported by the current model calibration dataset.

Previous Models

There are no previously published models at this site.

Model statistics, data, and plots

Definitions

Variable	Explanation
BCF	Duan's bias correction factor (Duan, 1983)
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 (Rasmussen and others, 2009).
$Pr(>\|t\|)$	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 (U.S. Geological Survey parameter code 80154; U.S. Geological Survey method code SED16).
t value	Student's t value; the coefficient divided by its associated standard error (Helsel and others, 2020)
TBY	Turbidity, water, unfiltered, monochrome near infrared light-emitting diode, 780−900 nm, detection angle 90±2.5 degrees, in formazin nephelometric units (U.S. Geological Survey parameter code 63680; U.S. Geological Survey method code TS213).

Model

$$\log SSC = 0.174 + 0.961 * \log TBY$$

Variable summary statistics

Variable	Minimum	Q1	Median	Mean	Q3	Maximum
logSSC	0.845	1.18	1.38	1.42	1.57	2.59
logTBY	0.463	0.994	1.25	1.3	1.56	2.4
SSC	7	15	24	48.5	37.5	391
TBY	2.91	9.86	17.8	31.4	36.4	251

Duration plots

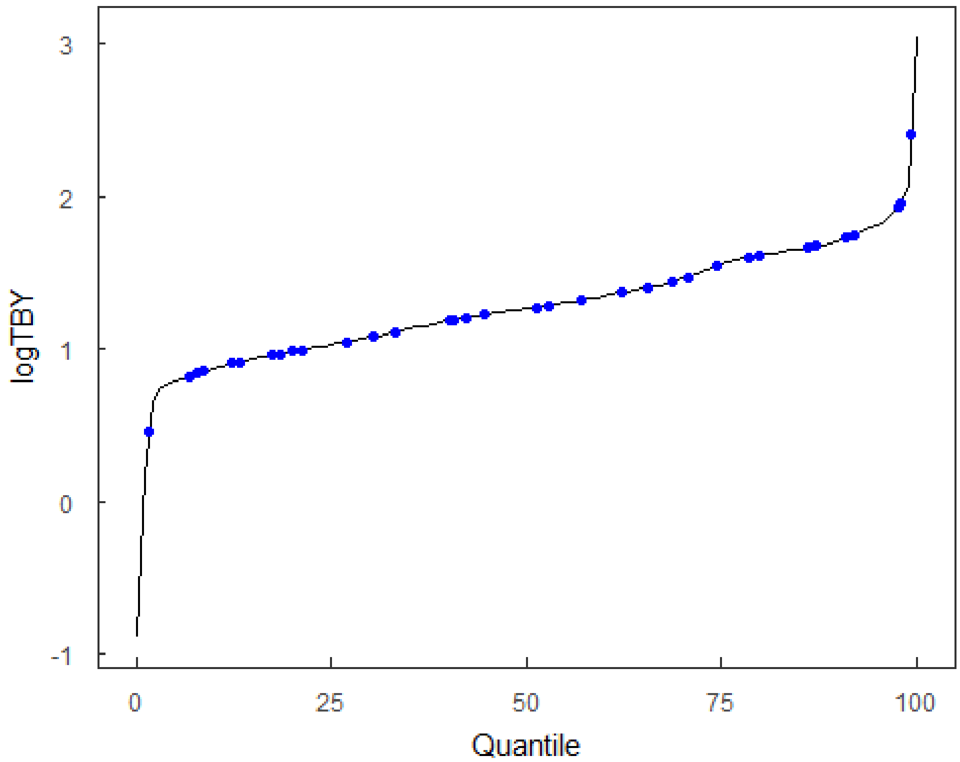

Figure 3. Duration plot of continuous log-scale turbidity (black line) and turbidity observations during discrete sample collection (blue dots) by quantile.

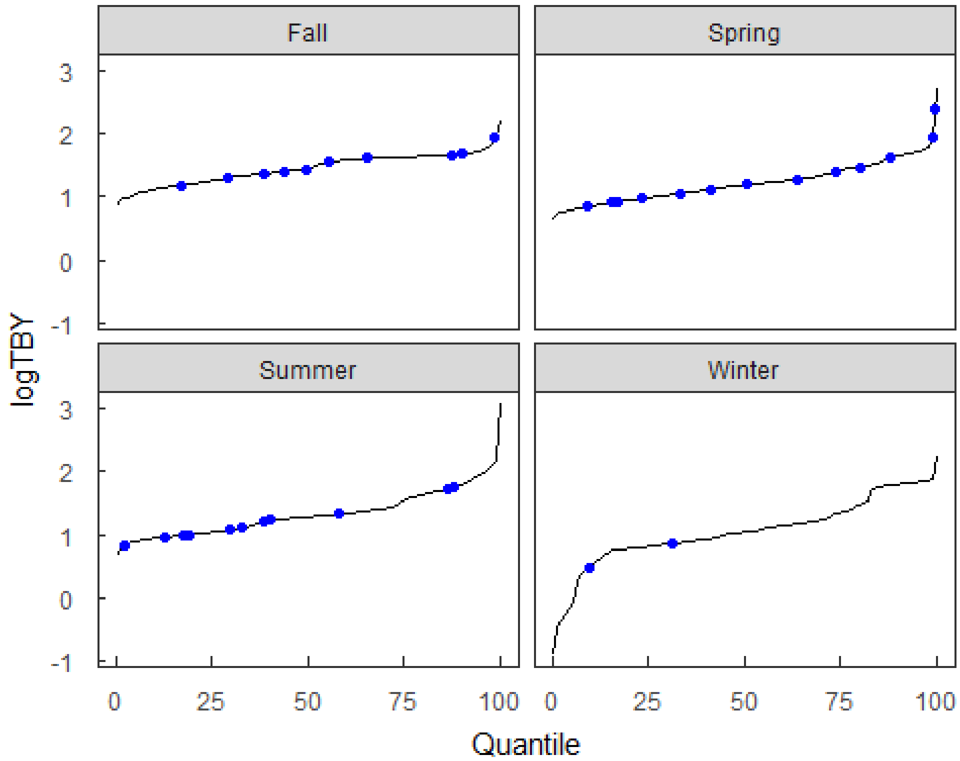

Figure 4. Seasonal duration plots of continuous log-scale turbidity (black line) and observed turbidity during discrete sample collection (blue dots) by quantile.

Box plots

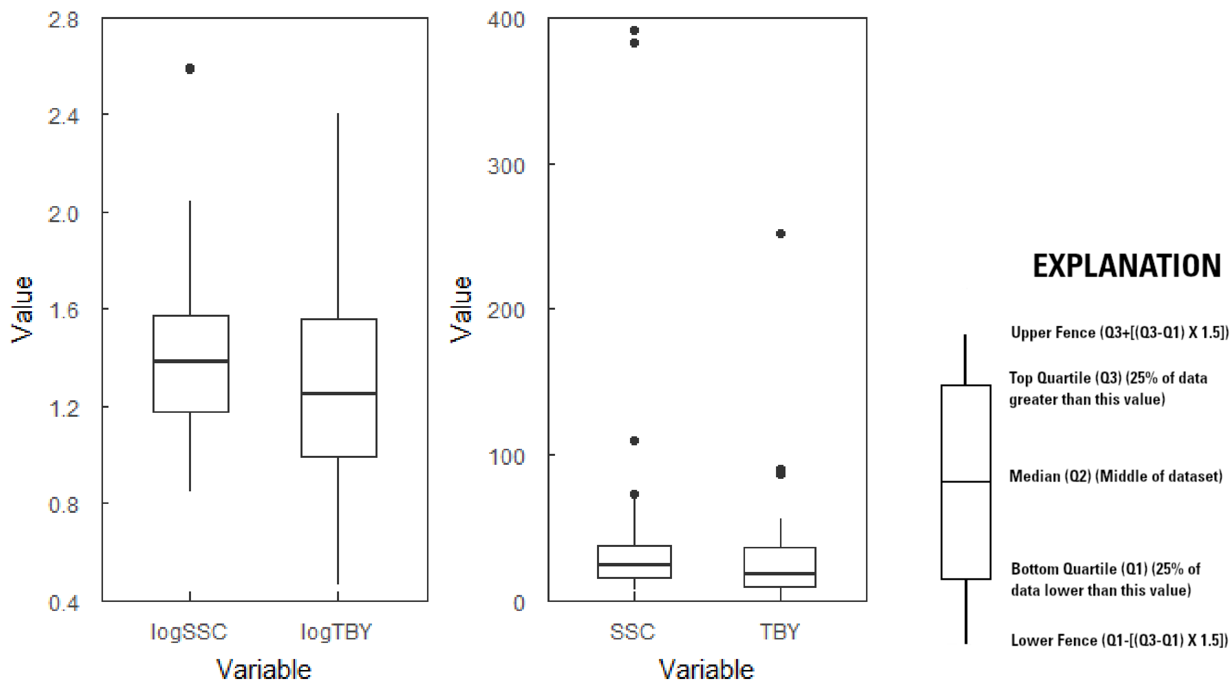

Figure 5. Box plots of log-transformed (left) and linear (right) SSC and turbidity values used in the calibration dataset.

Scatter plots

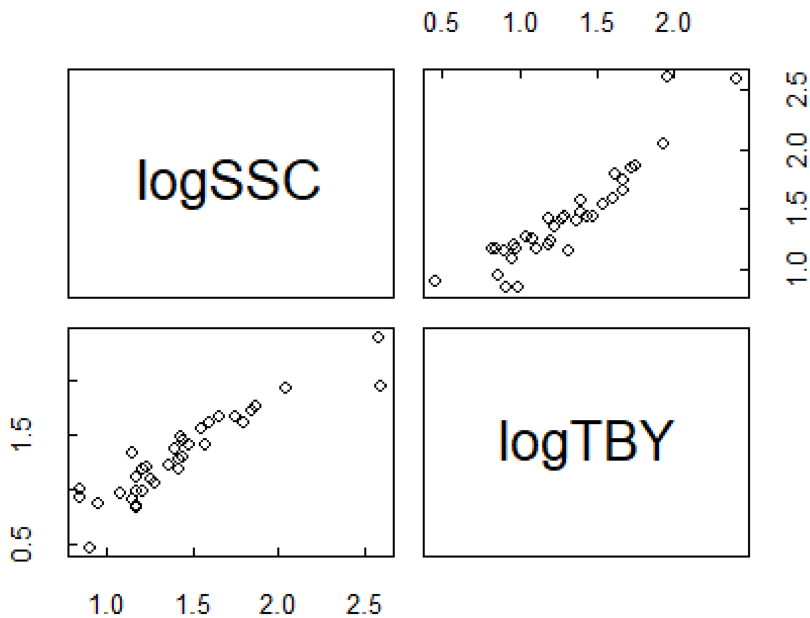

Figure 6. Scatter plot of log-transformed SSC and log-transformed turbidity.

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	36
R2	0.85
adjusted R2	0.846
RMSE	0.159
Upper MSPE (90%)	44.3
Lower MSPE (90%)	30.7
BCF	1.08

Model coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.1738191	0.0936411	1.856226	0.0721051
logTBY	0.9614444	0.0691714	13.899453	0.0000000

Correlation matrix

	logSSC	logTBY
logSSC	1.0000000	0.9221435
logTBY	0.9221435	1.0000000

Outlier test criteria

Leverage	DFFITS	CooksD
0.1667	0.4714	0.1936

Flagged observations

datetime	logSSC	CooksD	DFFITS	Leverage	StudResidual	Flag*
2019-02-21 10:10:00	0.903	0.358	0.885	0.159	2.03	CD
2019-05-08 14:30:00	2.58	0.0952	0.433	0.257	0.737	L
2019-05-29 16:10:00	2.59	0.789	1.57	0.109	4.5	CDS

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

Statistical plots

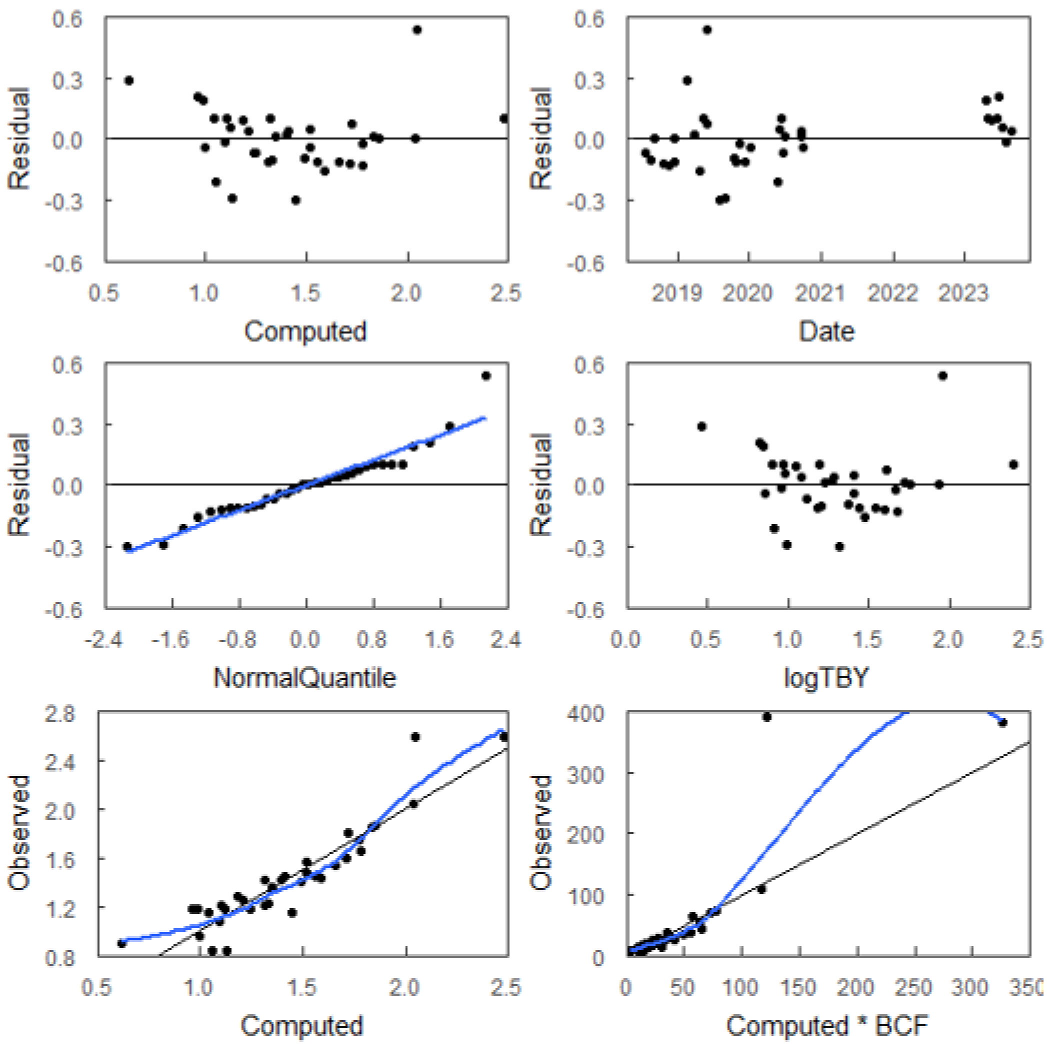

Figure 7. Statistical plots of model residuals and observed and computed SSC. Blue line shows the locally estimated scatterplot smoothing (LOESS).

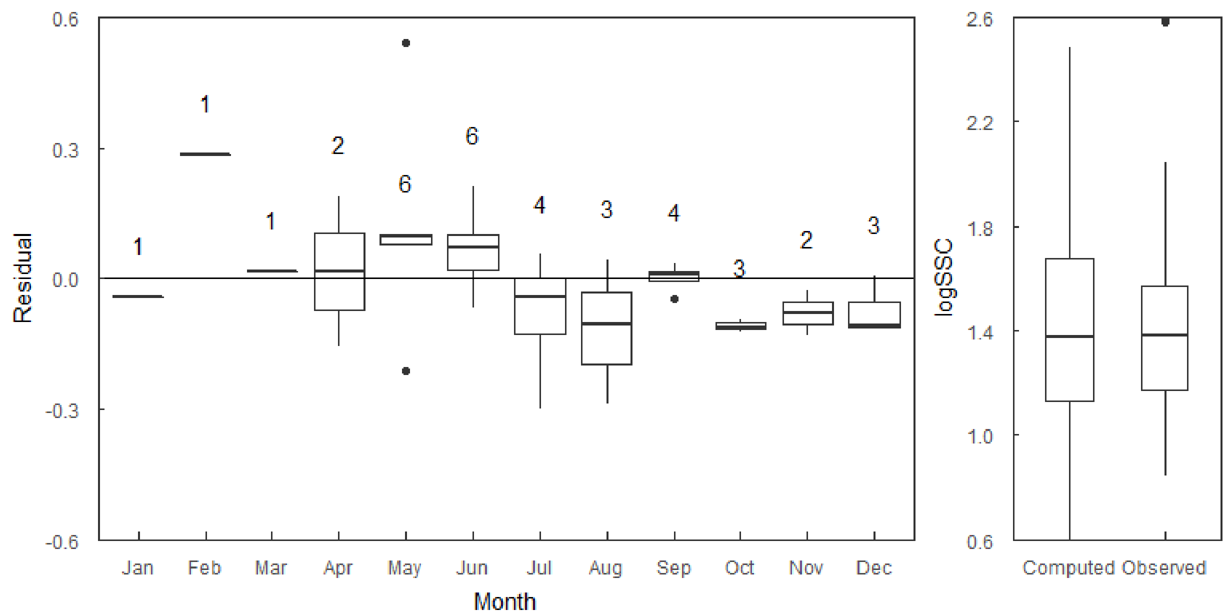

Figure 8. Box plots of SSC model residuals by month.

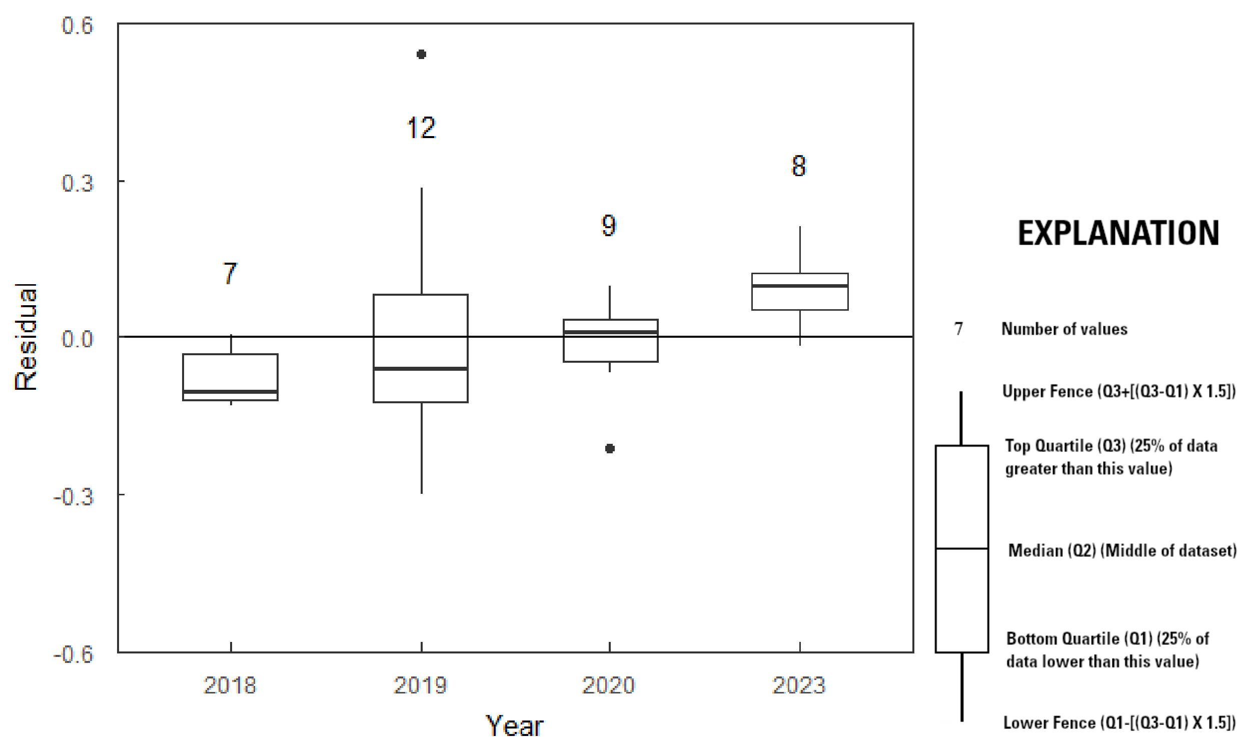

Figure 9. Box plots of SSC model residuals by year.

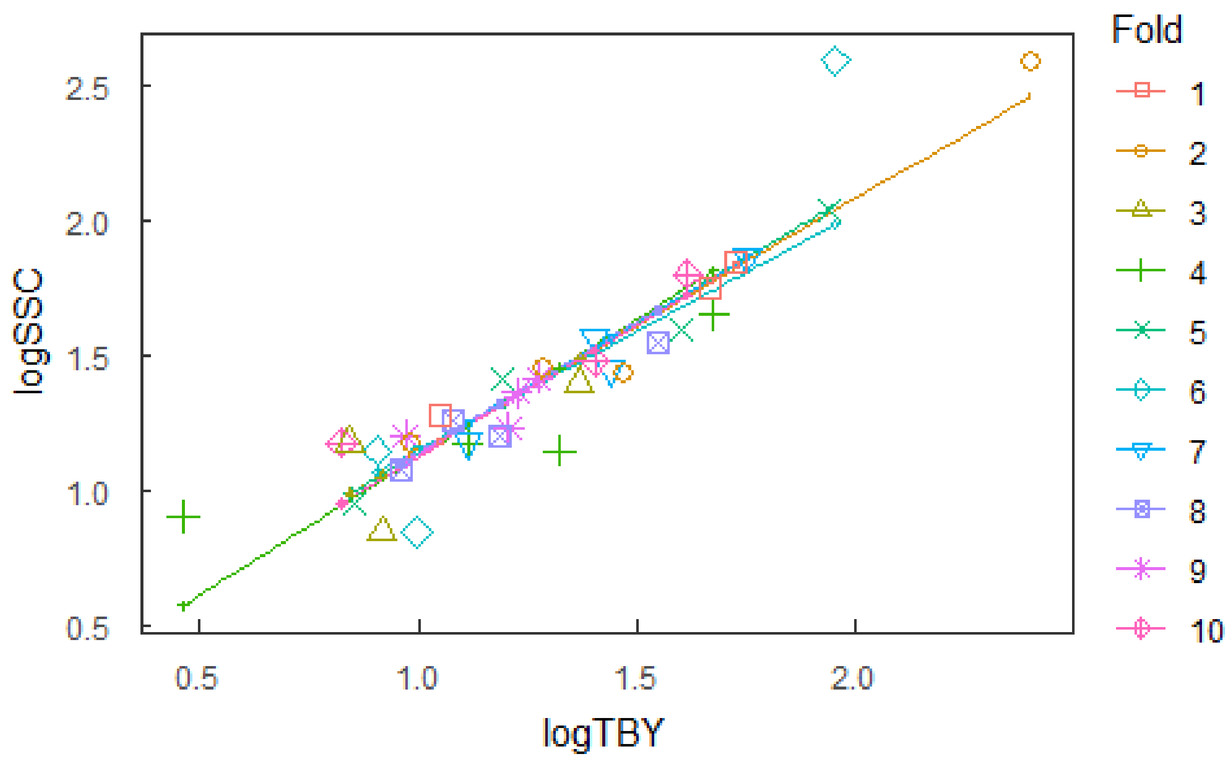

Figure 10. 10-fold cross validation plot

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.

Statistic	Value
Minimum MSE of folds	0.0141
25th Percentile	0.0252
Median MSE of folds	0.0274
Mean MSE of folds	0.0253
75th percentile	0.0278
Maximum MSE of folds	0.0281
Model MSE	0.0253

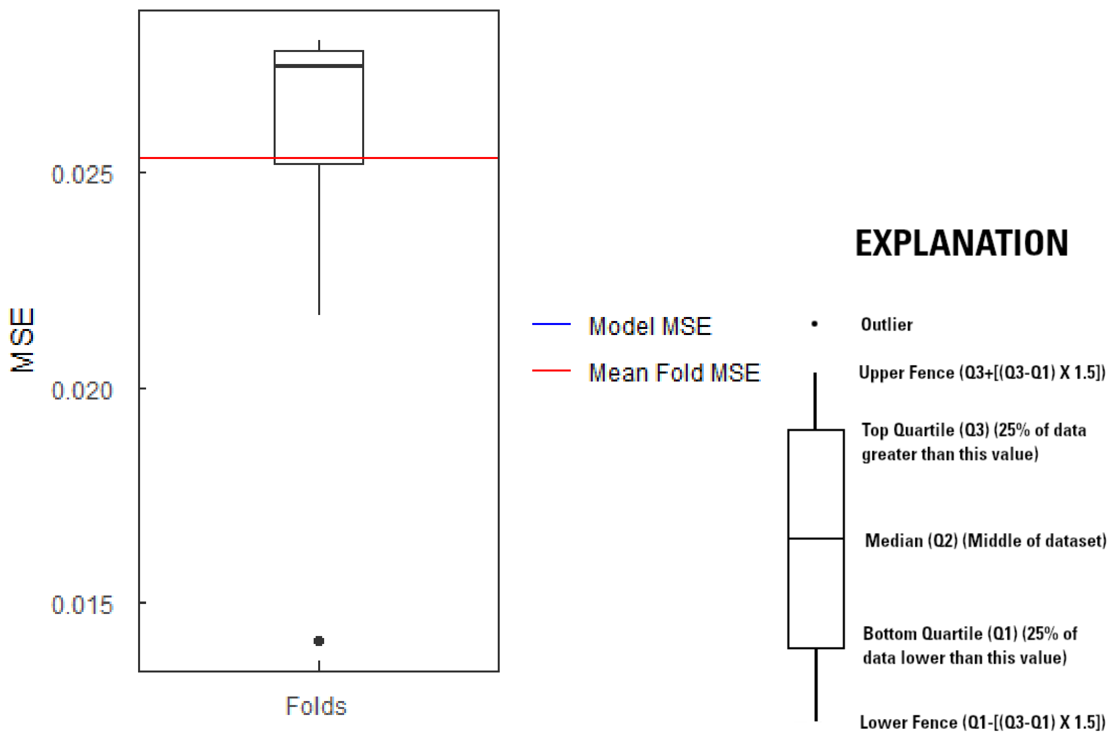

Figure 11. Mean square error of folds from cross validation.

Model calibration dataset

datetime	logSSC	logTBY	SSC	TBY	Computed	Retransformed	Computed SSC, lower 90% prediction interval	Computed SSC, upper 90% prediction interval
2018-07-17 10:40:00	1.18	1.11	15	13.00	1.24	18.9	10.1	35.5
2018-08-14 09:10:00	1.23	1.21	17	16.11	1.33	23.2	12.4	43.5
2018-09-05 15:00:00	1.86	1.75	73	56.40	1.86	77.5	40.8	146.9
2018-10-24 10:10:00	1.59	1.6	39	40.13	1.72	55.8	29.6	105.2
2018-11-19 10:50:00	1.65	1.67	45	47.27	1.78	65.4	34.6	123.5
2018-12-10 11:40:00	1.54	1.55	35	35.18	1.66	49.2	26.2	92.6
2018-12-14 12:20:00	2.04	1.94	110	86.42	2.04	117	60.9	224.0
2019-02-21 10:10:00	0.903	0.463	8	2.91	0.619	4.48	2.3	8.7
2019-03-27 10:30:00	1.41	1.27	26	18.78	1.4	26.9	14.4	50.5
2019-04-24 10:50:00	1.43	1.47	27	29.68	1.59	41.8	22.3	78.5
2019-05-08 14:30:00	2.58	2.4	383	251.25	2.48	326	162.4	651.9
2019-05-29 16:10:00	2.59	1.95	391	89.83	2.05	121	63.1	232.7

2019-05-30 12:50:00	1.8	1.61	63	41.14	1.73	57.2	30.3	107.8
2019-07-31 10:50:00	1.15	1.32	14	21.06	1.45	30	16.0	56.2
2019-08-27 12:00:00	0.845	0.997	7	9.93	1.13	14.6	7.7	27.4
2019-10-16 12:50:00	1.4	1.37	25	23.42	1.49	33.3	17.8	62.4
2019-10-23 10:40:00	1.45	1.44	28	27.56	1.56	38.9	20.8	73.1
2019-11-13 14:10:00	1.75	1.67	56	46.30	1.78	64.1	33.9	121.1
2019-12-08 15:20:00	1.2	1.19	16	15.35	1.31	22.2	11.8	41.5
2020-01-08 13:10:00	0.954	0.857	9	7.19	0.998	10.7	5.6	20.3
2020-05-20 12:30:00	0.845	0.918	7	8.27	1.06	12.2	6.5	23.2
2020-06-04 10:40:00	1.57	1.4	37	25.30	1.52	35.8	19.1	67.2
2020-06-09 12:00:00	1.41	1.19	26	15.52	1.32	22.4	11.9	42.0
2020-06-17 13:10:00	1.18	1.11	15	13.03	1.25	18.9	10.1	35.5
2020-06-29 11:20:00	1.85	1.73	70	53.20	1.83	73.2	38.7	138.8
2020-09-18 12:40:00	1.36	1.22	23	16.79	1.35	24.2	12.9	45.3
2020-09-23 11:50:00	1.45	1.29	28	19.42	1.41	27.8	14.8	52.0

2020-09-28 11:50:00	1.48	1.4	30	25.30	1.52	35.8	19.1	67.2
2023-04-19 10:40:00	1.18	0.846	15	7.01	0.987	10.4	5.5	19.8
2023-05-01 10:20:00	1.2	0.97	16	9.34	1.11	13.7	7.3	25.9
2023-05-15 10:40:00	1.28	1.05	19	11.22	1.18	16.4	8.7	30.8
2023-06-12 10:20:00	1.15	0.905	14	8.04	1.04	11.9	6.3	22.5
2023-06-26 10:00:00	1.18	0.824	15	6.67	0.966	9.95	5.3	19.0
2023-07-17 10:20:00	1.18	0.985	15	9.67	1.12	14.2	7.5	26.8
2023-07-31 11:30:00	1.08	0.959	12	9.09	1.1	13.4	7.1	25.3
2023-08-29 11:20:00	1.26	1.08	18	12.04	1.21	17.6	9.4	33.1

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.

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

Davis, B.E., 2005, A Guide to the Proper Selection and Use of Federally Approved Sediment and Water-Quality Samplers: Vicksburg, MS, U.S. Geological Survey, Open File Report 2005-1087, p. 20, accessed November 2023 at https://pubs.usgs.gov/of/2005/1087/pdf/OFR_2005-1087.pdf.

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.]

Guy, H.P., 1969, Laboratory theory and methods for sediment analysis: U.S. Geological Survey Techniques of Water-Resources Investigations, book 5, chap. C1, 58 p. [Also available at https://doi.org/10.3133/twri05C1.]

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.]

Mueller, D.K., Schertz, T.L., Martin, J.D., and Sandstrom, M.W., 2015, Design, analysis, and interpretation of field quality-control data for water-sampling projects: U.S. Geological Survey Techniques and Methods, book 4, chap. C4, 54 p., accessed November 2023 at https://doi.org/10.3133/tm4C4.

Pardoe, I., 2020, Applied regression modeling: United Kingdom, John Wiley & Sons, 336 p.

R Core Team, 2020, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, accessed November 2, 2023, 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.]

Rasmussen, T.J., Bennett, T.J., Stone, M.L., Foster, G.M., Graham, J.L., and Putnam, J.E., 2014, Quality-assurance and data-management plan for water-quality activities in the Kansas Water Science Center, 2014: U.S. Geological Survey Open-File Report 2014–1233, 41 p., accessed November 2023 at https://dx.doi.org/10.3133/ofr20141233.

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, 166 p., accessed June 17, 2021, 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/admin/memo/SW/sw.2016.07+wq.2016.10.pdf.

U.S. Geological Survey, 2017, Procedures for processing, approving, publishing, and auditing time-series records for water data: U.S. Geological Survey Office of Water Quality Technical Memorandum no. 2017.07, accessed November 2023 at https://water.usgs.gov/admin/memo/QW/qw2017.07.pdf.

U.S. Geological Survey, 2023, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed December 2023 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.

Zar, J.H., 1999, Biostatistical analysis (4th ed.): New Jersey, Prentice-Hall Inc., 663 p.