Appendix 5. Model Archive Summary for Total Organic Carbon 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 total organic carbon concentration model developed to compute 15-minute, hourly, or daily TOC 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).

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: 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 30 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 USGS National Water Information System 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 30 concomitant values of discretely collected total organic carbon concentration and continuously measured turbidity and streamflow during April 19, 2016, through August 12, 2021. Discrete samples were collected throughout the range of continuously observed hydrologic conditions. No samples had total organic carbon 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. Two samples in the model calibration dataset were flagged as outliers but were retained in the dataset after further review.

Total Organic Carbon 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 one to seven times per year. Samples were analyzed for total organic carbon by the Wichita Municipal Water and Wastewater Laboratory in Wichita, Kans., according to standard methods (Eaton and others, 1995).

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). Discharge was computed using a nonsubmersible pressure transducer following standard USGS methods (Turnipseed and Sauer, 2010; Painter and Loving, 2015). 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 and streamflow were selected as a good surrogate for total organic carbon 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 total organic carbon (TOC) concentration regression analysis at USGS site 07144780:

TOC concentration-based model:

$$TOC = 2.43 \times \log_{10}(Q) + 7.01 \times \log_{10}(TBY) - 8.51,$$

where,

> *TOC* = total organic carbon, in milligrams per liter (mg/L) (USGS parameter code 00680);
>
> *Q* = streamflow, instantaneous, cubic feet per second (ft^3/s) (USGS parameter code 00060);
>
> $\log_{10}$ = decimal logarithm; and
>
> *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).

Organic matter is a major component of total suspended solids (Juracek and Rasmussen, 2008; Hem, 1985); therefore, turbidity and streamflow are suitable surrogates for organic carbon.

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 total organic carbon concentration using this model is 21.45 milligrams per liter. Computed estimates outside that limit are not supported by the current model calibration dataset.

Model Statistics, Data, and Plots

Definitions

Variable	Explanation
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)
logQ	Streamflow, instantaneous, cubic feet per second (ft^3/s) (USGS parameter code 00060), 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)
SPC	Specific conductance, in microsiemens per centimeter at 25 degrees Celsius (µS/cm at 25°C) (USGS parameter code 00095)
TOC	Total organic carbon, milligrams per liter (mg/L) (USGS parameter code 00680)
t value	Student's *t* value; the coefficient divided by its associated standard error (Helsel and others, 2020)

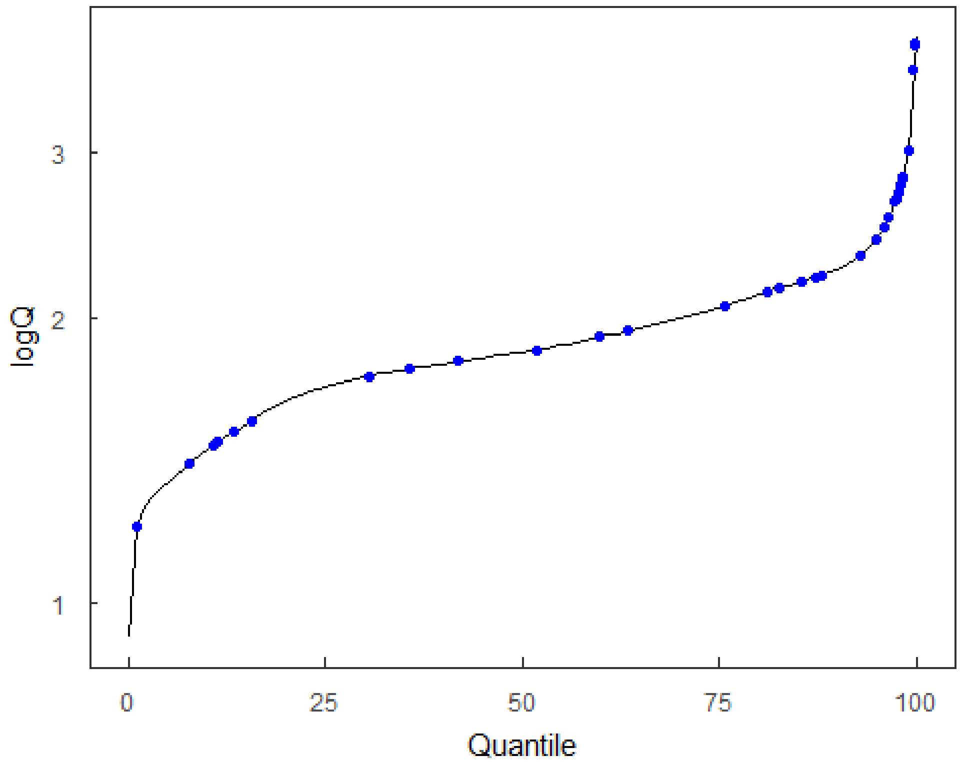
Model

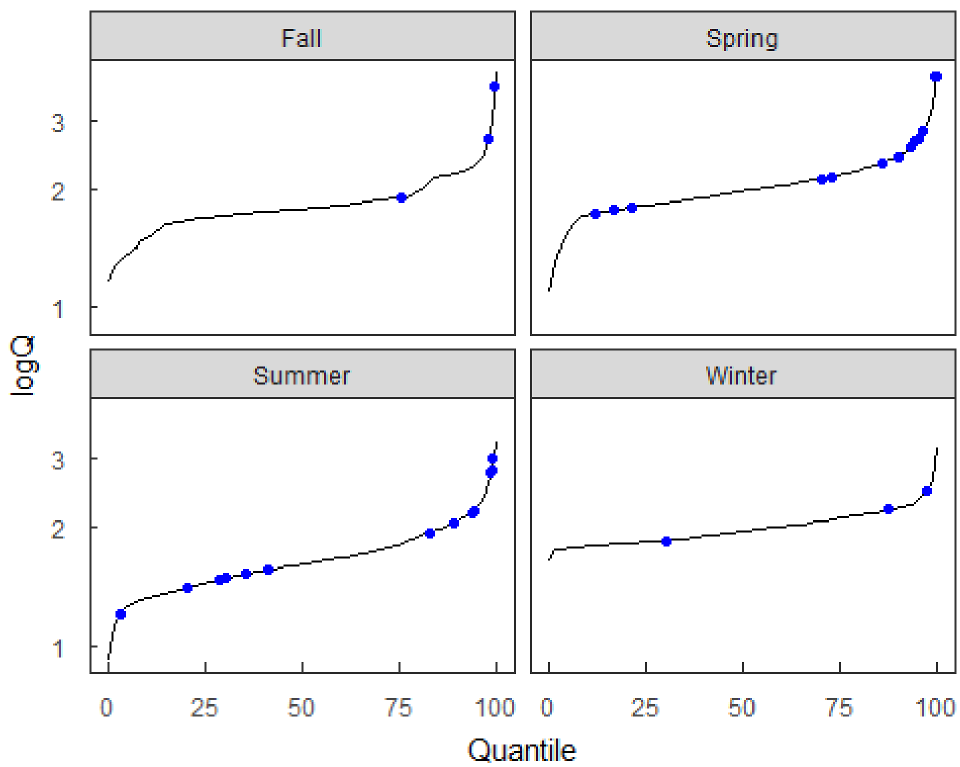
$$TOC = 2.43 \times \log_{10}(Q) + 7.01 \times \log_{10}(TBY) - 8.51$$

Variable summary statistics

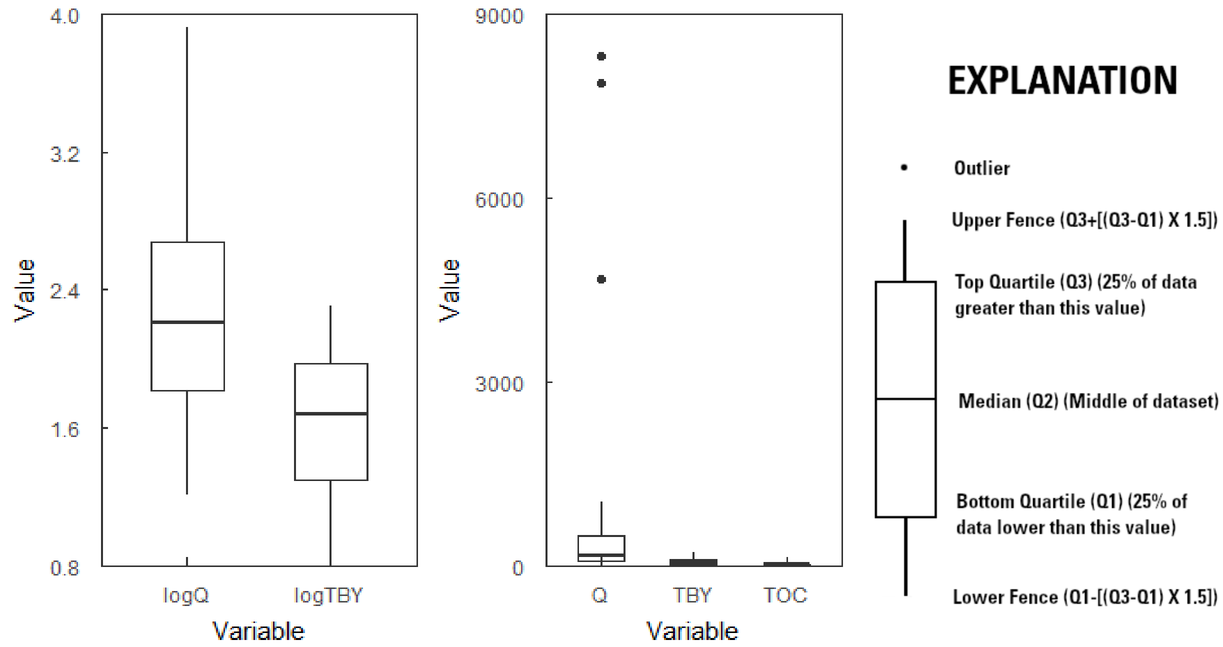
Variable	Minimum	Q1	Median	Mean	Q3	Maximum
logQ	1.21	1.82	2.2	2.31	2.68	3.92
logTBY	0.847	1.29	1.68	1.64	1.97	2.3
Q	16	65.5	160	911	479	8,290
TBY	7.03	19.6	47.4	65.4	93.7	201
TOC	1.72	3.53	9.32	8.58	12.8	19.5

Duration plots

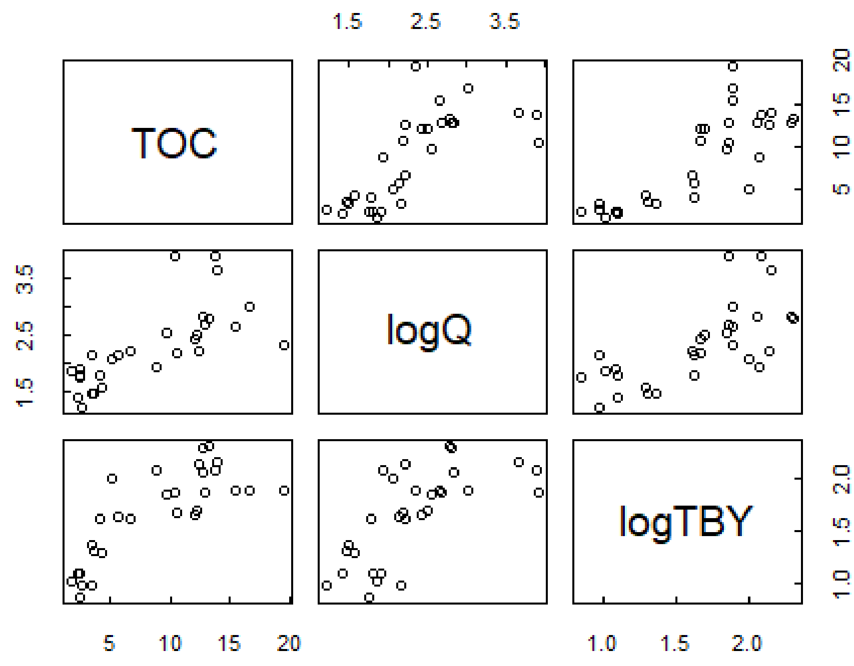




Box plots



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	30
R^2	0.707
Adjusted R^2	0.685
RMSE	2.93
Upper MSPE (90%)	34.1
Lower MSPE (90%)	-34.1

Model coefficients

	Estimate	Standard error	t value	Pr(>|t|)
(Intercept)	-8.505353	2.185008	-3.892597	0.0005876
logQ	2.429733	1.077309	2.255373	0.0324212
logTBY	7.008205	1.725619	4.061271	0.0003763

Correlation matrix

	TOC	logQ	logTBY
TOC	1.0000000	0.7265167	0.8072572
logQ	0.7265167	1.0000000	0.6890046
logTBY	0.8072572	0.6890046	1.0000000

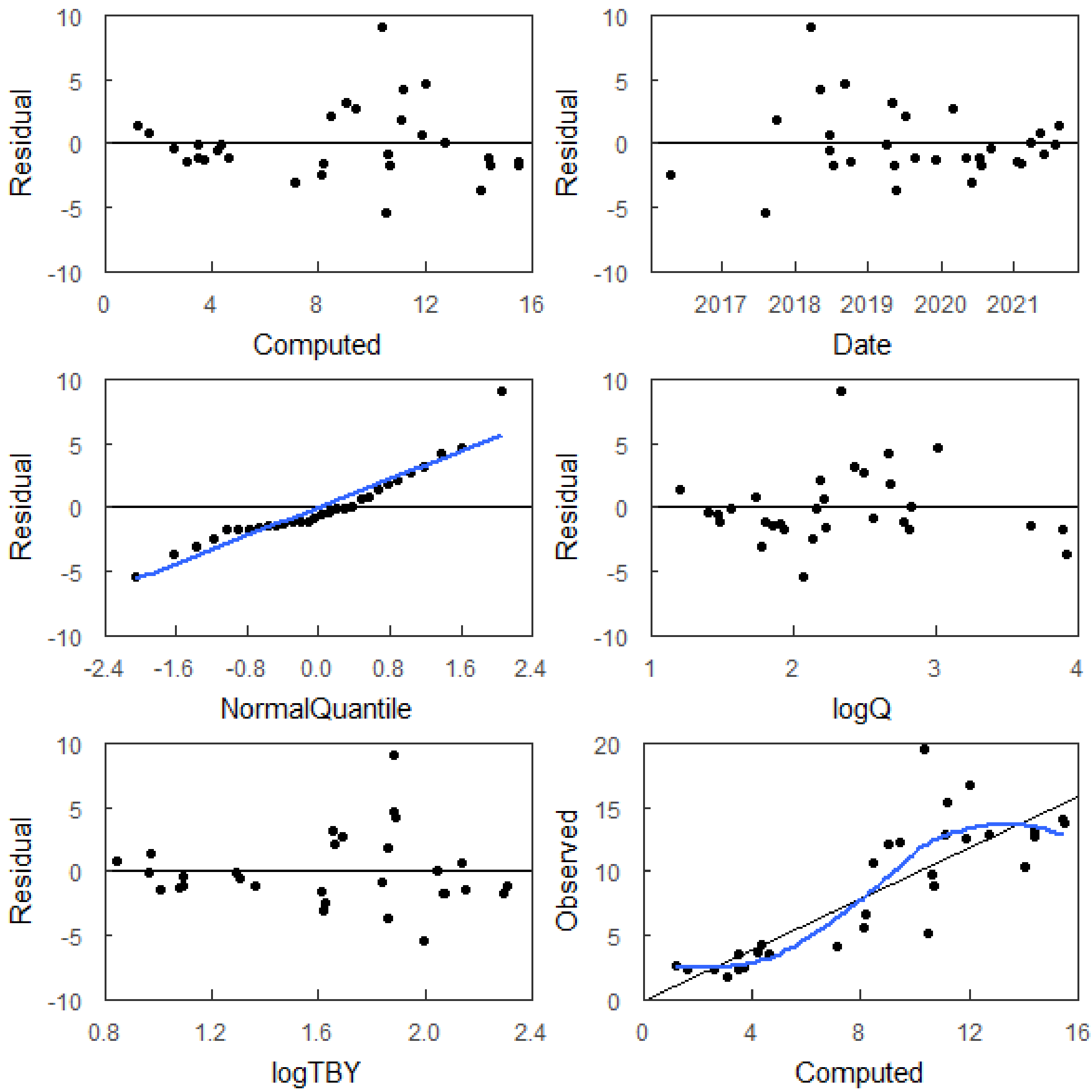
Outlier test criteria

Leverage	DFFITS	CooksD
0.3	0.6325	0.2615

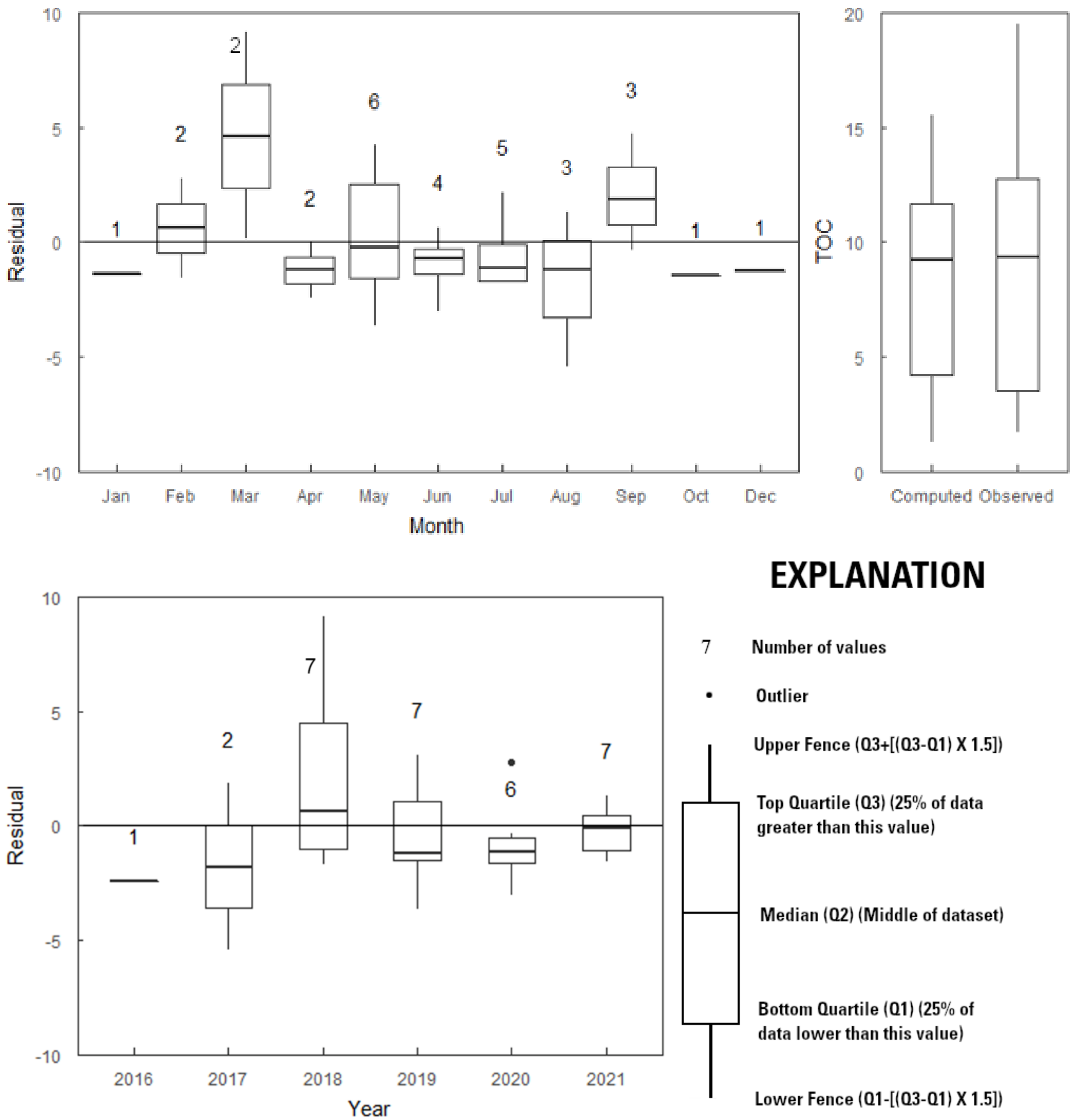
Flagged observations

datetime	TOC	CooksD	DFFITS	Leverage	Studentized Residual
2018-03-20 10:30:00	19.5	0.187	0.935	0.0518	4
2019-05-21 12:30:00	10.4	0.306	-0.981	0.295	-1.52

Statistical plots

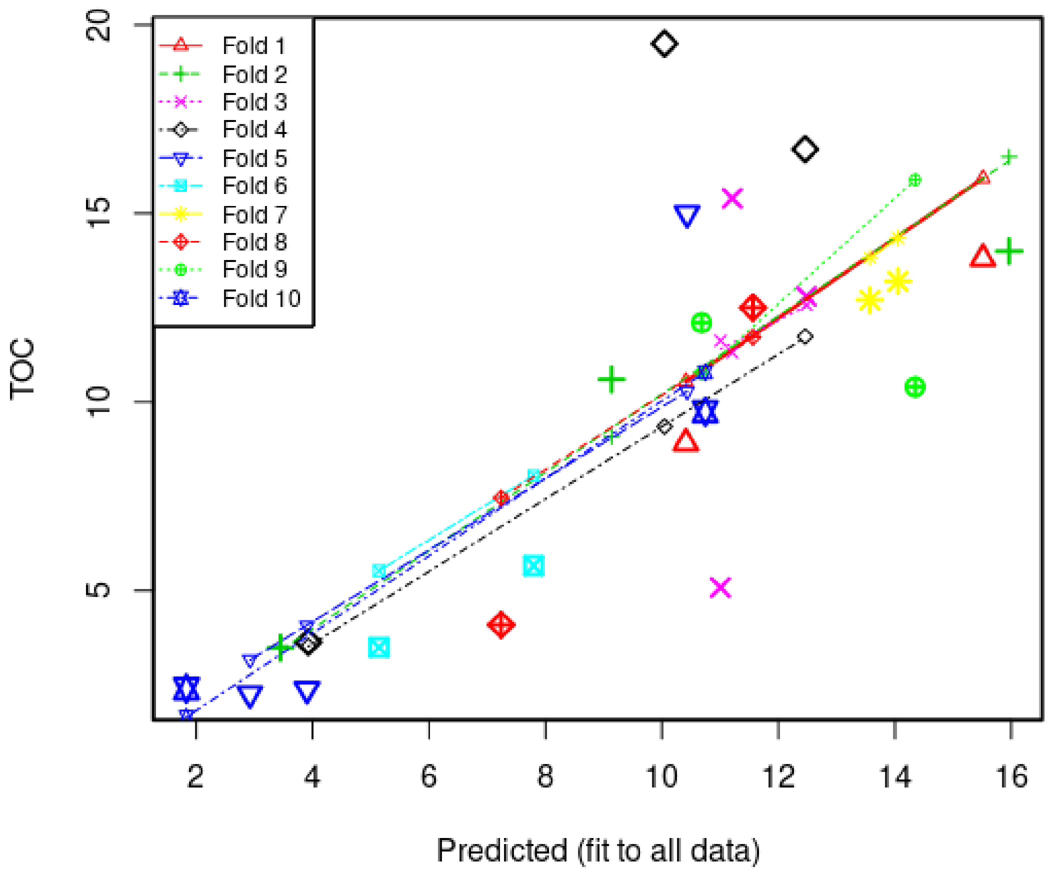


The blue line shows the locally estimated scatterplot smoothing (LOESS). The black dots correspond to observed values. The black line represents the 1:1 line.



Residual
10
5
0
-5
-10
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Dec
Month
TOC
20
15
10
5
0
Computed
Observed
Residual
2016
2017
2018
2019
2020
2021
Year
EXPLANATION
7 Number of values
Outlier
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])

Cross Validation



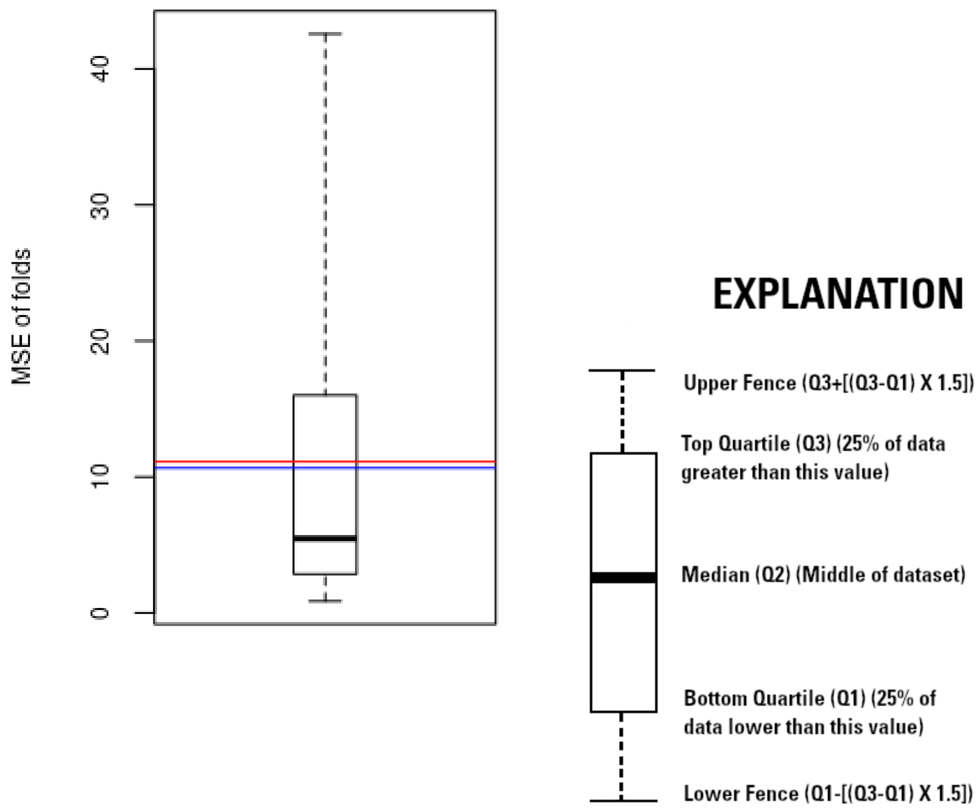
Fold - equal partition of the data (10 percent of the data).

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	0.835
Median MSE of folds	5.460
Mean MSE of folds	10.700
Maximum MSE of folds	42.600
(Mean MSE of folds) / (Model MSE)	0.958




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Model calibration dataset

datetime	TOC	logQ	logTBY	Computed
2016-04-19 10:25:00	5.66	2.14	1.62	8.08
2017-08-11 11:00:00	5.08	2.07	2	10.5
2017-09-28 10:30:00	12.9	2.69	1.86	11.1
2018-03-20 10:30:00	19.5	2.34	1.88	10.4
2018-05-04 10:00:00	15.4	2.66	1.89	11.2
2018-06-21 10:10:00	3.63	1.47	1.3	4.2
2018-06-26 13:20:00	12.5	2.22	2.13	11.8
2018-07-14 12:00:00	12.7	2.82	2.29	14.4
2018-09-05 09:55:00	16.7	3.01	1.88	12
2018-10-09 10:10:00	14.0	3.67	2.15	15.5
2019-04-02 10:50:00	3.48	2.16	0.968	3.53

datetime	TOC	logQ	logTBY	Computed
2019-05-02 11:20:00	12.1	2.43	1.66	9
2019-05-08 12:00:00	13.8	3.9	2.07	15.5
2019-05-21 12:30:00	10.4	3.92	1.86	14
2019-07-08 11:30:00	10.6	2.19	1.66	8.45
2019-08-26 11:30:00	13.2	2.78	2.3	14.4
2019-12-03 10:20:00	2.42	1.91	1.08	3.71
2020-02-26 10:30:00	12.2	2.5	1.69	9.42
2020-05-07 10:30:00	2.38	1.8	1.09	3.53
2020-06-04 10:20:00	4.09	1.77	1.62	7.13
2020-07-08 11:00:00	3.49	1.48	1.36	4.65
2020-07-21 10:10:00	8.91	1.94	2.07	10.7
2020-09-03 10:20:00	2.26	1.4	1.1	2.59
2021-01-12 10:10:00	1.72	1.86	1.01	3.09
2021-02-01 11:00:00	6.65	2.23	1.61	8.2
2021-03-23 11:40:00	12.8	2.83	2.04	12.7
2021-05-10 10:50:00	2.40	1.74	0.847	1.66
2021-06-01 10:40:00	9.74	2.56	1.84	10.6
2021-07-22 10:40:00	4.24	1.56	1.29	4.32
2021-08-12 11:00:00	2.57	1.21	0.973	1.25

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.

Eaton, A.D., Clesceri, L.S., and Greenberg, A.E., eds., 1995, Standard methods for the examination of water and wastewater (19th ed.): New York, American Public Health Association, 905 p.

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.]

Hem, J.D., 1985, 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.]

Juracek, K.E., and Rasmussen, P.P., 2008, Sediment quality and comparison to historical water quality, Little Arkansas River Basin, south-central Kansas, 2007: U.S. Geological Survey Scientific Investigations Report 2008–5187, 47 p., accessed July 13, 2022, at https://doi.org/10.3133/sir20085187.

Painter, C.C., and Loving, B.L., 2015, U.S. Geological Survey quality-assurance plan for surface-water activities in Kansas, 2015: U.S. Geological Survey Open-File Report 2015-1074, 33 p., http://dx.doi.org/10.3133/ofr20151074.

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.]

Turnipseed, D.P., and Sauer, V.B., 2010, Discharge measurements at gaging stations: U.S. Geological Survey Techniques and Methods book 3, chap. A8, 87 p., accessed July 13, 2022, at https://doi.org/10.3133/tm3A8.

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.