# Appendix 2.31. Model Archive Summary for Total Organic Carbon Concentration at U.S. Geological Survey site 07144100; Little Arkansas River near Sedgwick, Kansas, during December 2014 through December 2019

This model archive summary summarizes the total organic carbon model developed to compute hourly or daily total organic carbon. Model development methods follow U.S. Geological Survey (USGS) guidance from Office of Surface Water/Office of Water Quality Technical Memoranda and USGS Techniques and Methods, book 3, chap. C4 (Rasmussen and others, 2009).



## Site and Model Information

Site Number: 07144100

Site Name: Little Arkansas River near Sedgwick, Kansas

Location: Latitude 37°52'59", longitude 97°25'27" referenced to North American Datum of 1927, in NE 1/4 NW 1/4 NW 1/4 sec.15, T.25 S., R.1 W., Sedgwick County, Kansas; hydrologic unit 11030012.

Equipment: A Sutron Satlink II High Data Rate Collection Platform and a Design Analysis Water Log H350/355 nonsubmersible pressure transducer transfers real-time stage and water-quality data via satellite. The primary reference gage is a Type-A wire-weight gage located on the downstream bridge handrail. Check-bar elevation is 33.614 feet. The orifice is enclosed in a well-screen and attached to a concrete pier on the left downstream side of the bridge. Gage height was measured during December 2014 through December 2019. A YSI 6600 water-quality monitor equipped with water temperature, specific conductance, pH, dissolved oxygen, and turbidity (a YSI Model 6026 [September 1998 through December 2006] and YSI Model 6136 [July 2004 through March 2015]) sensors collected data during April 1998 through March 2015. A YSI EXO2 water-quality monitor equipped with water temperature, specific conductance, pH, dissolved oxygen, turbidity, and fluorescent dissolved organic matter sensors collected data during September 2014 through December 2019. A Hach Nitratax monitor collected nitrate data during March 2012 through December 2019.

Date model was developed: June 1, 2020

Model calibration data period: December 9, 2014 through December 11, 2019

## Model Data

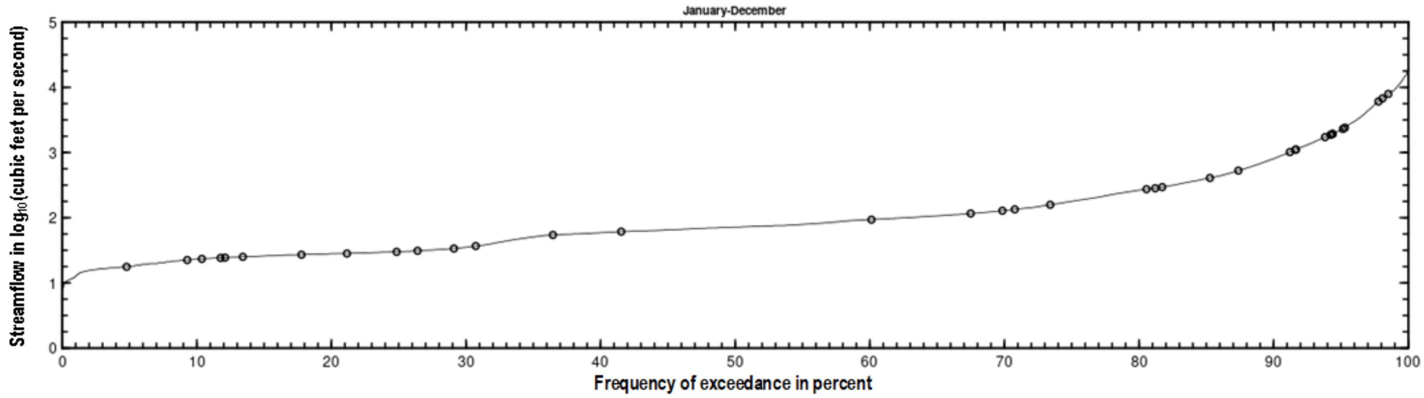
All data were collected using USGS protocols (U.S. Geological Survey, variously dated; Wagner and others, 2006; Sauer and Turnipseed, 2010; Turnipseed and Sauer, 2010) and are stored in the National Water Information System (NWIS) database (U.S. Geological Survey, 2021). Explanatory variables were evaluated individually and in combination. Potential explanatory variables included streamflow, water temperature, specific conductance, pH, dissolved oxygen, YSI EXO2 turbidity, nitrate, and fluorescent dissolved organic matter. Seasonal components (sine and cosine variables) also were evaluated as explanatory variables.

The regression model is based on 38 concomitant values of discretely collected total organic carbon and continuously measured turbidity during December 2014 through December 2019. Discrete samples were collected over a range of streamflow and turbidity conditions. No samples had concentrations that were below laboratory detection limits. Summary statistics and the complete model-calibration data are provided below. Outliers and influential points were identified using studentized residuals, DFITS, Cook's D, and leverage. Outliers in previously published versions of this model (Christensen and others, 2003; Rasmussen and others, 2016) were examined and retained in the dataset if there were no clear issues, explanations, or conditions that would cause a result to be invalid for model calibration. All samples were retained in the dataset.

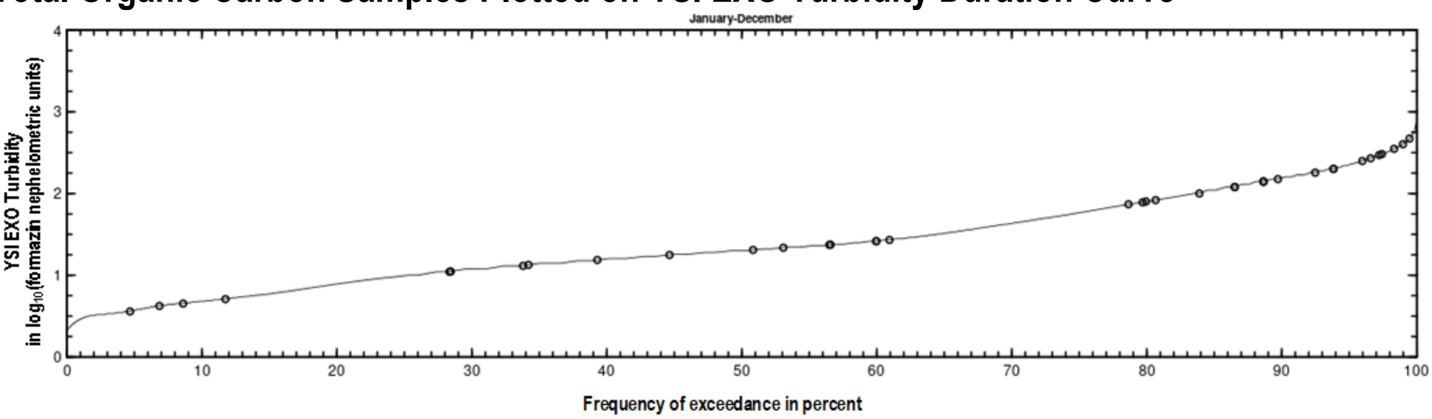
## Total Organic Carbon

Discrete samples were collected from the downstream side of the bridge or instream within 50 feet of the bridge using equal-width-increment, multi-vertical, single vertical or grab-dip methods following U.S. Geological Survey (variously dated) and Rasmussen and others (2014). Discrete samples were collected on a semifixed to event-based schedule ranging from 1 to 9 samples per year with a FISP US DH–95 or D–95 with a Teflon bottle, cap, and nozzle depth-integrating sampler, a DH–81 with a Teflon bottle, cap, and nozzle hand sampler or a grab sample with a Teflon bottle depending on sample location. Samples were analyzed for total organic carbon by the Wichita Municipal Water and Wastewater Laboratory in Wichita, Kansas, or the USGS National Water Quality Laboratory according to standard methods (American Public Health Association and others, 1995).

## Total Organic Carbon Samples Plotted on Streamflow Duration Curve




## Total Organic Carbon Samples Plotted on YSI EXO Turbidity Duration Curve




## Continuous Data

Concomitant turbidity values were time interpolated. If no concomitant continuous data were available within 2 hours of sample collection, the sample was not included in the dataset.

## Model Development

Ordinary least squares regression analysis was done using R (version 4.0.0) programming language (R Core Team, 2020) to relate discretely collected total organic carbon to turbidity and other continuously measured data. The distribution of residuals was examined for normality and plots of residuals (the difference between the measured and model-calculated values) compared to model-computed total organic carbon were examined for homoscedasticity (departures from zero did not change substantially over the range of model-calculated values). Previously published explanatory variables were also strongly considered for continuity; however, the best explanatory variable(s) were ultimately selected.

Turbidity was selected as the best predictor of total organic carbon based on residual plots, high coefficient of determination ($R^2$), and low model standard percentage error (MSPE). Turbidity was positively correlated with total organic carbon because turbidity measures light scattered by particulates in water.

## Model Summary

Summary of final total organic carbon regression analysis at USGS site number 07143672:

Total organic carbon-based model:

$$\log_{10}(TOC) = 0.445 \times \log_{10}(TBY) + 0.192$$

where,

$\log_{10}$ = logarithm base 10;
*TOC* = total organic carbon, in milligrams per liter (mg/L); and
*TBY* = turbidity, in formazin nephelometric units (FNU)

The log-transformed model may be retransformed to original units so that TOC 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 1.02. The retransformed model, accounting for BCF is:

$$TOC = 1.587 \times TBY^{0.445}$$

## Model Statistics, Data, and Plots

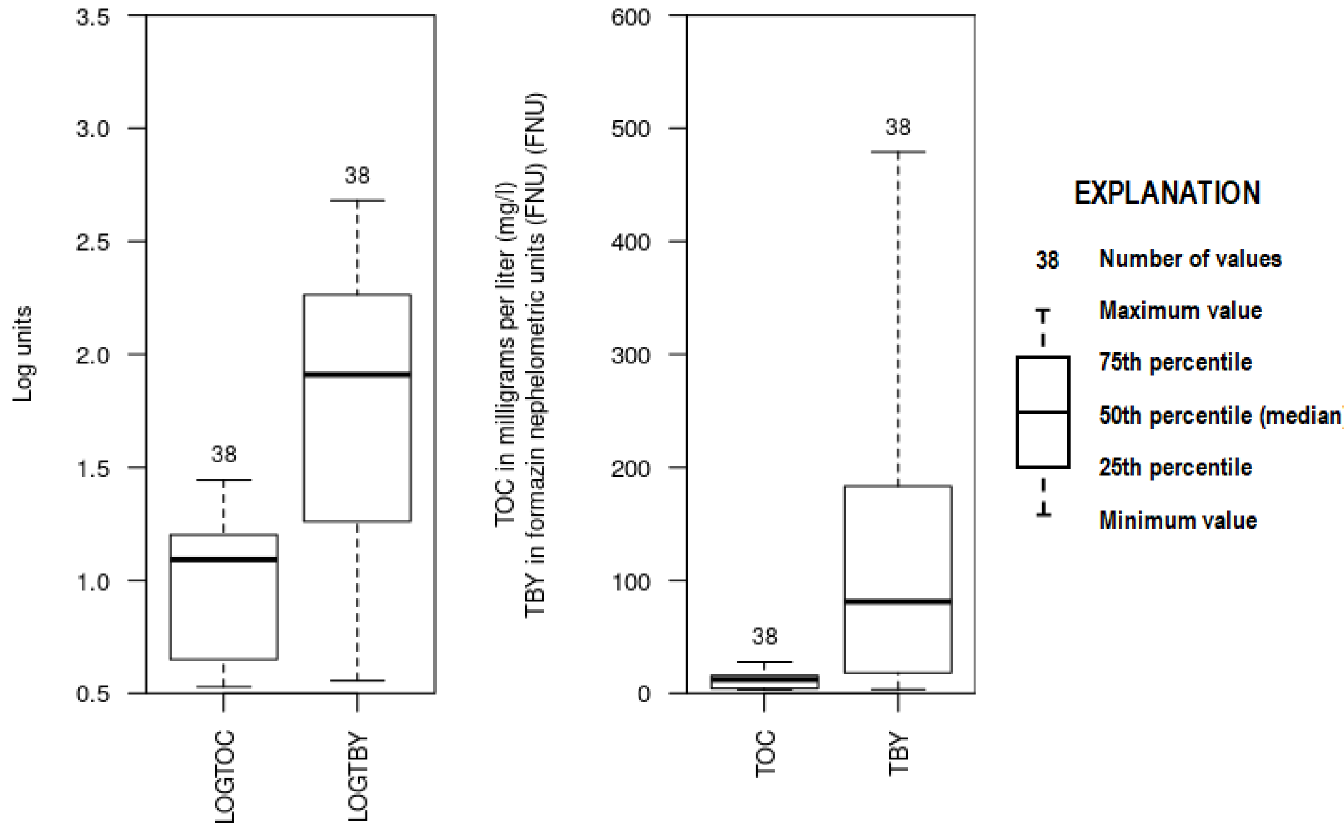
### Model

LOGTOC = + 0.445 * LOGTBY + 0.192

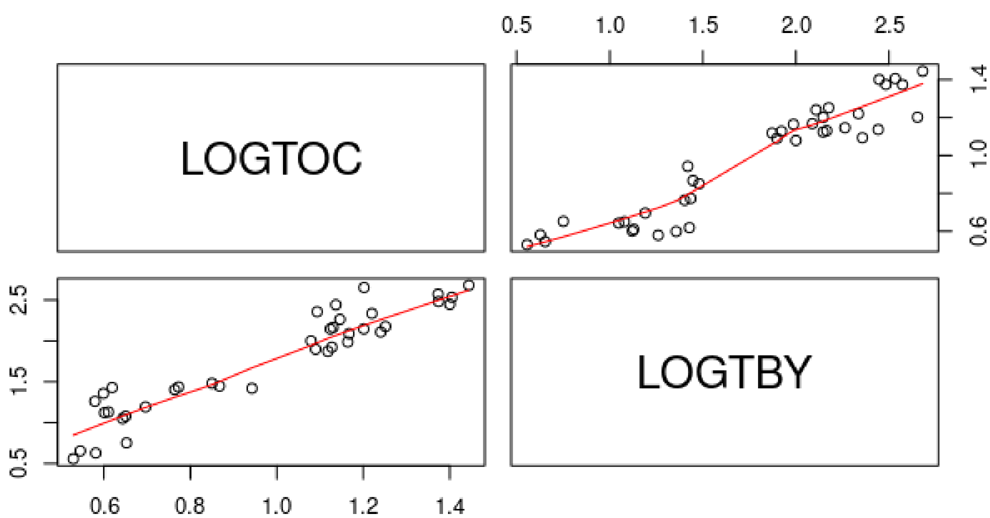
### Variable Summary Statistics

| | LOGTOC | TOC | LOGTBY | TBY |
|---|---|---|---|---|
| Minimum | 0.529 | 3.38 | 0.556 | 3.6 |
| 1st Quartile | 0.650 | 4.47 | 1.260 | 18.2 |
| Median | 1.090 | 12.30 | 1.910 | 81.3 |
| Mean | 0.974 | 11.60 | 1.750 | 123.0 |
| 3rd Quartile | 1.200 | 15.90 | 2.260 | 183.0 |
| Maximum | 1.440 | 27.80 | 2.680 | 479.0 |

## Box Plots




## Exploratory Plots



## Basic Model Statistics

| | |
|---|---|
| Number of Observations | 38 |
| Standard error (RMSE) | 0.0985 |
| Average Model standard percentage error (MSPE) | 22.9 |
| Coefficient of determination ($R^2$) | 0.892 |
| Adjusted Coefficient of Determination (Adj. $R^2$) | 0.889 |
| Bias Correction Factor (BCF) | 1.02 |

## Explanatory Variables

| | Coefficients | Standard Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.192 | 0.0482 | 3.99 | 3.12e-04 |
| LOGTBY | 0.445 | 0.0259 | 17.20 | 6.01e-19 |

## Correlation Matrix

| | Intercept | E.vars |
|---|---|---|
| Intercept | 1.000 | -0.943 |
| E.vars | -0.943 | 1.000 |

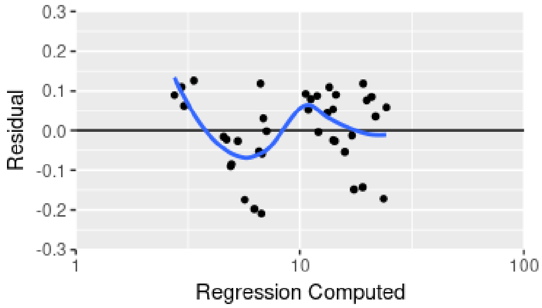
## Outlier Test Criteria

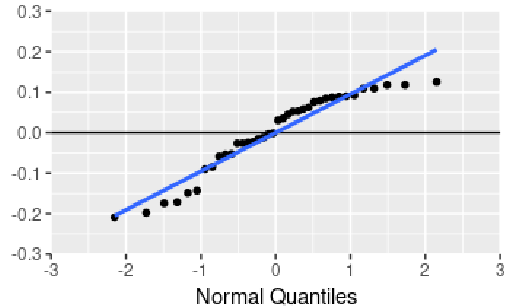
| Leverage | Cook's D | DFFITS |
|---|---|---|
| 0.158 | 0.194 | 0.459 |

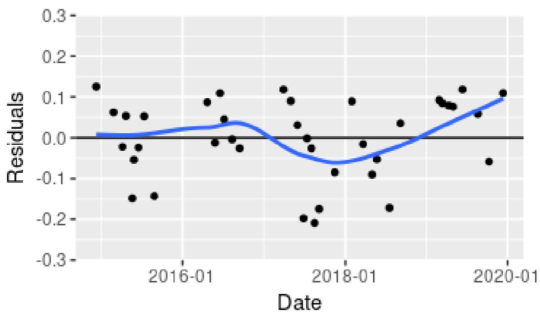
## Flagged Observations

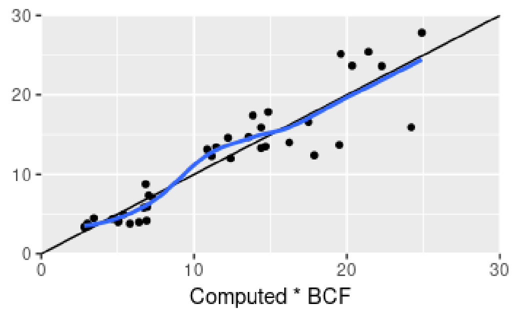
| | LOGTOC | Estimate | Residual | Standard Residual | Studentized Residual | Leverage | Cook's D | DFFITS |
|---|---|---|---|---|---|---|---|---|
| 7/19/2018 11:30 | 1.2 | 1.37 | -0.172 | -1.82 | -1.89 | 0.0822 | 0.149 | -0.564 |

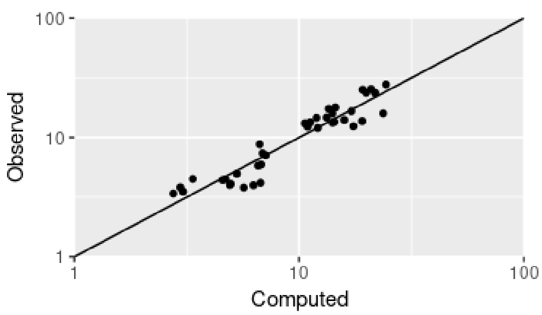
# Statistical Plots

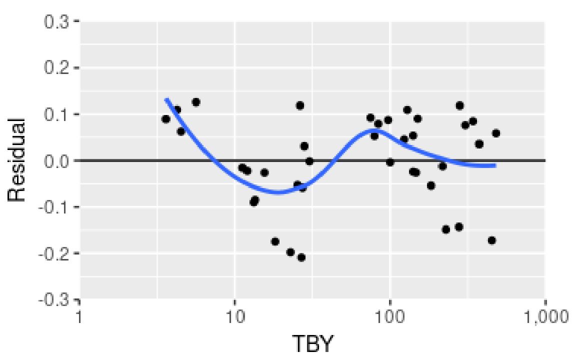


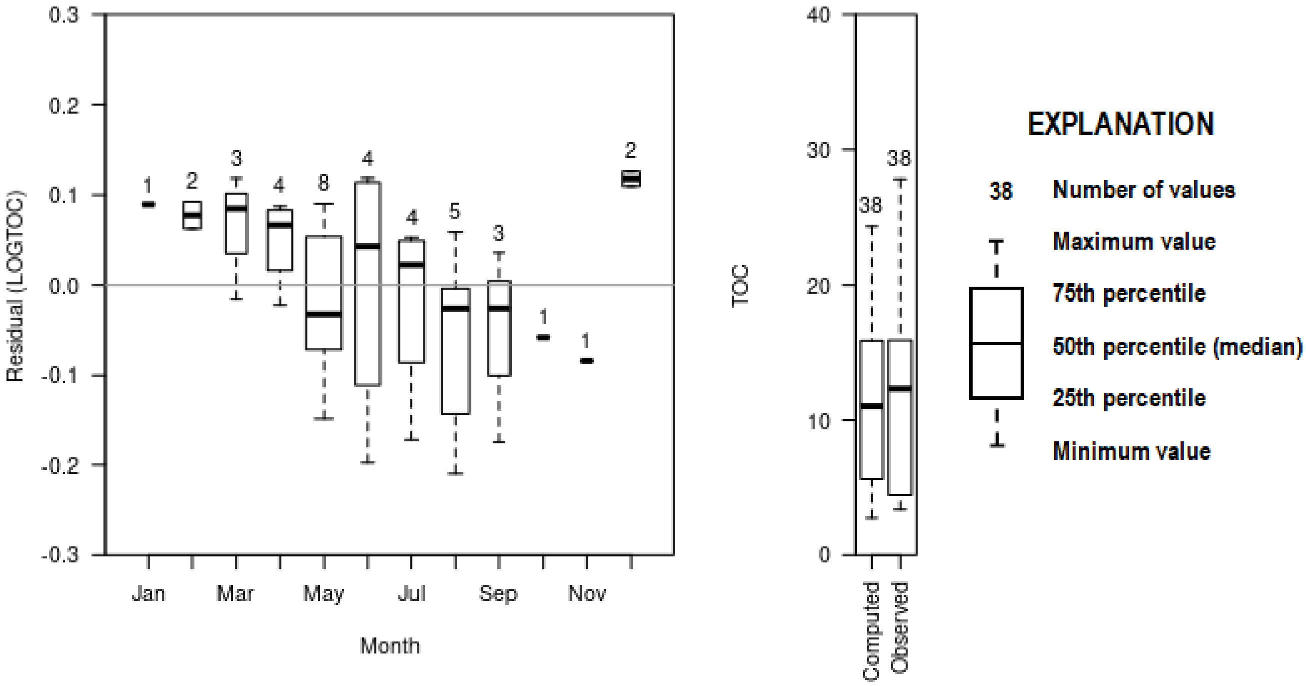



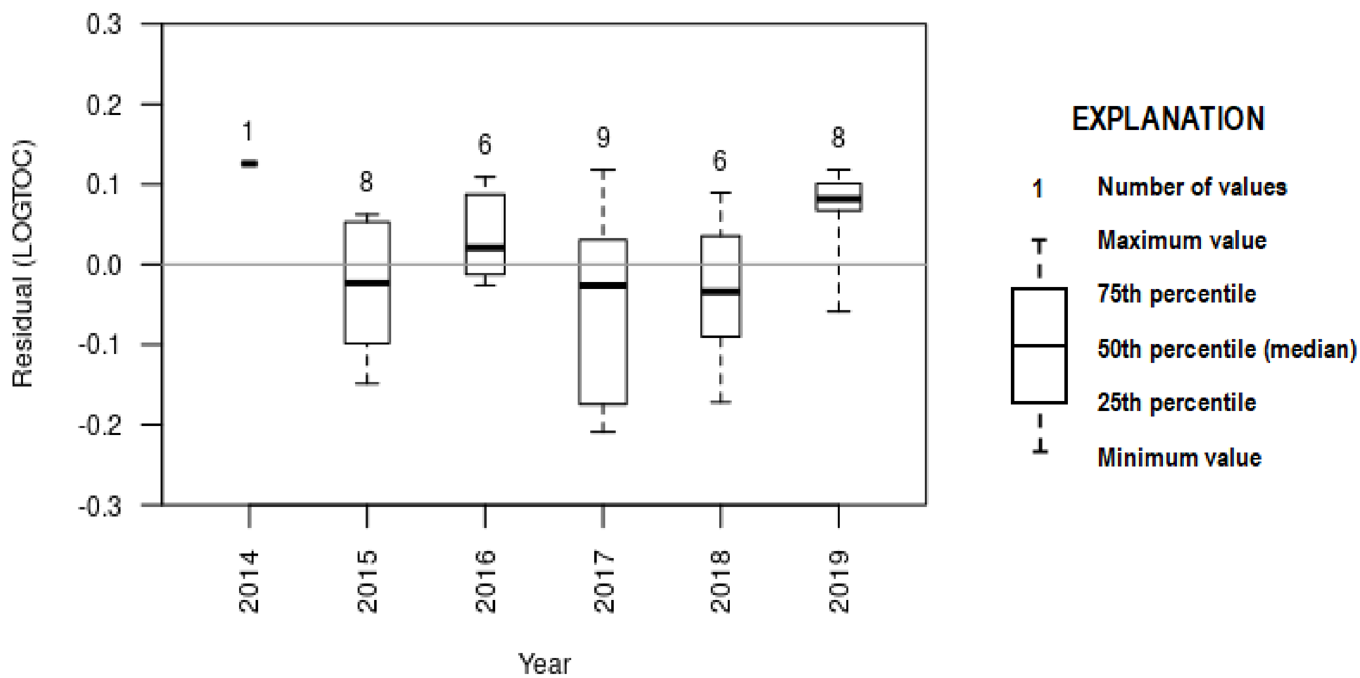



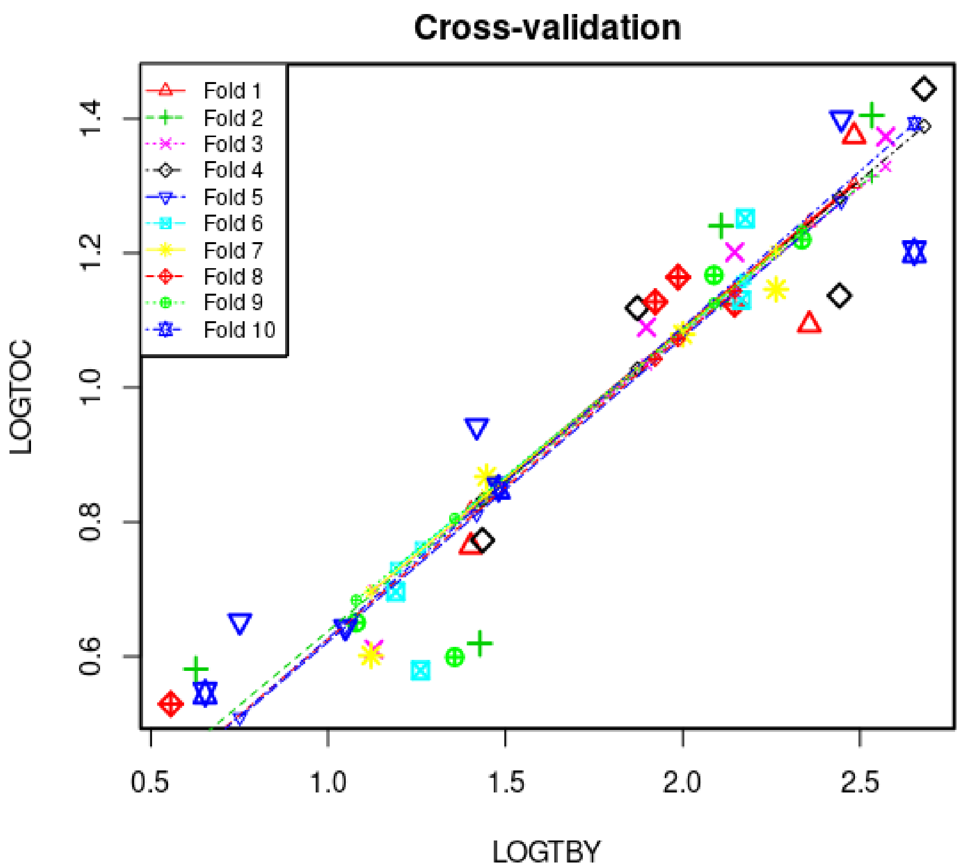


0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
Residual (LOGTOC)
1
2
3
4
8
4
4
5
3
1
1
2
Jan
Mar
May
Jul
Sep
Nov
Month
40
30
20
10
0
TOC
38
38
Computed
Observed
EXPLANATION
38
Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value

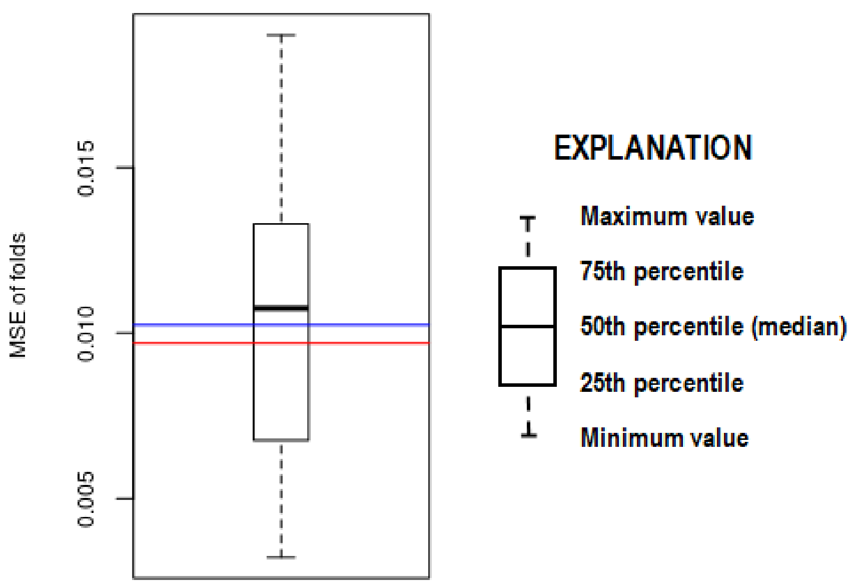



0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
Residual (LOGTOC)
1
8
6
9
6
8
2014
2015
2016
2017
2018
2019
Year
EXPLANATION
1
Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value

## Cross Validation




```
              Minimum MSE of folds:  0.00323
                 Mean MSE of folds:  0.01020
               Median MSE of folds:  0.01070
              Maximum MSE of folds:  0.01900
 (Mean MSE of folds) / (Model MSE):  1.06000
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

## Model-Calibration Dataset

| | Date | LOGTOC | LOGTBY | TOC | TBY | Computed LOGTOC | Computed TOC | Residual | Normal Quantiles |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 12/9/2014 | 0.652 | 0.751 | 4.49 | 5.64 | 0.527 | 3.44 | 0.126 | 2.15 |
| 2 | 2/25/2015 | 0.545 | 0.653 | 3.51 | 4.5 | 0.483 | 3.11 | 0.0623 | 0.44 |
| 3 | 4/6/2015 | 0.65 | 1.08 | 4.47 | 12 | 0.673 | 4.82 | -0.0225 | -0.3 |
| 4 | 4/22/2015 | 1.2 | 2.15 | 15.9 | 140 | 1.15 | 14.4 | 0.0533 | 0.3 |
| 5 | 5/20/2015 | 1.09 | 2.36 | 12.4 | 228 | 1.24 | 17.9 | -0.149 | -1.17 |
| 6 | 5/27/2015 | 1.15 | 2.26 | 14 | 183 | 1.2 | 16.2 | -0.0541 | -0.67 |
| 7 | 6/17/2015 | 1.12 | 2.15 | 13.3 | 140 | 1.15 | 14.4 | -0.0242 | -0.369 |
| 8 | 7/13/2015 | 1.09 | 1.9 | 12.3 | 79 | 1.04 | 11.2 | 0.0525 | 0.232 |
| 9 | 8/27/2015 | 1.14 | 2.44 | 13.7 | 277 | 1.28 | 19.5 | -0.143 | -1.05 |
| 10 | 4/21/2016 | 1.16 | 1.99 | 14.6 | 97 | 1.08 | 12.2 | 0.0873 | 0.755 |
| 11 | 5/26/2016 | 1.22 | 2.34 | 16.6 | 217 | 1.23 | 17.5 | -0.0124 | -0.165 |
| 12 | 6/17/2016 | 1.24 | 2.11 | 17.4 | 128 | 1.13 | 13.8 | 0.109 | 1.17 |
| 13 | 7/6/2016 | 1.17 | 2.09 | 14.7 | 123 | 1.12 | 13.6 | 0.0451 | 0.165 |
| 14 | 8/11/2016 | 1.08 | 2 | 12 | 100 | 1.08 | 12.4 | -0.00379 | -0.0986 |
| 15 | 9/13/2016 | 1.13 | 2.17 | 13.5 | 146 | 1.16 | 14.7 | -0.0262 | -0.44 |
| 16 | 3/30/2017 | 1.4 | 2.45 | 25.1 | 280 | 1.28 | 19.6 | 0.118 | 1.49 |
| 17 | 5/1/2017 | 1.25 | 2.18 | 17.8 | 150 | 1.16 | 14.8 | 0.0902 | 0.943 |
| 18 | 5/31/2017 | 0.867 | 1.45 | 7.37 | 28 | 0.837 | 7.03 | 0.0308 | 0.0328 |
| 19 | 6/28/2017 | 0.599 | 1.36 | 3.97 | 22.8 | 0.797 | 6.41 | -0.198 | -1.73 |
| 20 | 7/13/2017 | 0.85 | 1.48 | 7.08 | 30.2 | 0.852 | 7.27 | -0.00158 | -0.0328 |
| 21 | 8/2/2017 | 0.696 | 1.19 | 4.97 | 15.5 | 0.723 | 5.4 | -0.0263 | -0.514 |
| 22 | 8/16/2017 | 0.619 | 1.43 | 4.16 | 26.8 | 0.828 | 6.89 | -0.209 | -2.15 |
| 23 | 9/6/2017 | 0.579 | 1.26 | 3.79 | 18.2 | 0.754 | 5.8 | -0.174 | -1.49 |
| 24 | 11/15/2017 | 0.61 | 1.13 | 4.08 | 13.5 | 0.695 | 5.07 | -0.0847 | -0.845 |
| 25 | 1/31/2018 | 0.529 | 0.556 | 3.38 | 3.6 | 0.44 | 2.82 | 0.0893 | 0.845 |
| 26 | 3/22/2018 | 0.643 | 1.05 | 4.4 | 11.2 | 0.659 | 4.67 | -0.0157 | -0.232 |
| 27 | 5/2/2018 | 0.601 | 1.12 | 3.99 | 13.2 | 0.691 | 5.03 | -0.09 | -0.943 |
| 28 | 5/23/2018 | 0.764 | 1.4 | 5.81 | 25.3 | 0.817 | 6.71 | -0.0529 | -0.59 |
| 29 | 7/19/2018 | 1.2 | 2.65 | 15.9 | 450 | 1.37 | 24.2 | -0.172 | -1.32 |
| 30 | 9/6/2018 | 1.37 | 2.57 | 23.6 | 373 | 1.34 | 22.3 | 0.0354 | 0.0986 |
| 31 | 2/27/2019 | 1.12 | 1.87 | 13.1 | 74.5 | 1.03 | 10.9 | 0.0923 | 1.05 |
| 32 | 3/14/2019 | 1.41 | 2.53 | 25.4 | 341 | 1.32 | 21.4 | 0.0847 | 0.67 |
| 33 | 4/11/2019 | 1.13 | 1.92 | 13.4 | 83.7 | 1.05 | 11.4 | 0.079 | 0.59 |
| 34 | 5/1/2019 | 1.37 | 2.48 | 23.7 | 304 | 1.3 | 20.3 | 0.076 | 0.514 |
| 35 | 6/12/2019 | 0.943 | 1.42 | 8.76 | 26.3 | 0.824 | 6.83 | 0.118 | 1.73 |
| 36 | 8/20/2019 | 1.44 | 2.68 | 27.8 | 479 | 1.39 | 24.9 | 0.0584 | 0.369 |
| 37 | 10/9/2019 | 0.773 | 1.44 | 5.93 | 27.2 | 0.831 | 6.94 | -0.0586 | -0.755 |
| 38 | 12/11/2019 | 0.581 | 0.627 | 3.81 | 4.24 | 0.471 | 3.03 | 0.11 | 1.32 |

## Definitions

```
TOC: Organic carbon in mg/l (00680)
TBY: Turbidity in FNU (63680)
```

## References Cited


American Public Health Association, American Water Works Association, and Water Environment Federation, 1995, Standard methods for the examination of water and wastewater (19th ed.): Washington D.C., American Public Health Association, 905 p.

Christensen, V.G., Ziegler, A.C., Rasmussen P.P., and Jian X., 2003, Continuous real-time water-quality monitoring of Kansas streams, *in* Proceedings of 2003 Spring Specialty Conference on Agricultural Hydrology and Water Quality, Kansas City, Mo., May 12–14, 2003: Middleburg, Va., American Water Resources Association Technical Publication Series No. TPS–03–1, compact disc. [Also available at https://nrtwq.usgs.gov/ks/methods/christensen2003.]

Cook, D.R., 1977, Detection of influential observation in linear regression: Technometrics, v. 19, no. 1, p. 15–18. [Also available at https://www.jstor.org/stable/1268249?seq=4#metadata_info_tab_contents.]

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.]

R Core Team, 2020, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, version 4.0.0. [Also available at https://www.r-project.org.]

Rasmussen, P.P., Eslick, P.J., and Ziegler, A.C., 2016, Relations between continuous real-time physical properties and discrete water-quality constituents in the Little Arkansas River, south-central Kansas, 1998–2014: U.S. Geological Survey Open-File Report 2016–1057, 16 p. [Also available at https://doi.org/10.3133/ofr20161057.]

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., 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. [Also available at https://doi.org/10.3133/ofr20141233.]

Sauer, V.B., and Turnipseed, D.P., 2010, Stage measurement at gaging stations: U.S. Geological Survey Techniques and Methods, book 3, chap. A7, 45 p. [Also available at https://doi.org/10.3133/tm3A7.]

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. [Also available at https://doi.org/10.3133/tm3A8.]

U.S. Geological Survey, 2021, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed December 8, 2021, 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]. [Also available 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, 96 p. [Also available at https://doi.org/10.3133/tm1D3.]