

Appendix 1.31. Model Archive Summary for Total Organic Carbon Concentration at U.S. Geological Survey site 07143672; Little Arkansas River at Highway 50 near Halstead, Kansas, during March 2017 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).

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: 07143672

Site Name: Little Arkansas River at Highway 50 near Halstead, Kansas

Location: Latitude 38°01'43", longitude 97°32'25" referenced to North American Datum of 1927, in NW 1/4 NE 1/4 NE 1/4 sec.28, T.23 S., R.2 W., Harvey 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 guardrail. Check-bar elevation is 33.396 feet. The orifice tube is enclosed in 1.25-inch steel conduit trenched into the ground down to the edge of water, where the orifice emerges from the bank and culminates in a 2-inch open-end orifice tethered to a steel fencepost near the left edge of water. Gage height was measured during May 1998 through December 2019. A YSI 6600 water-quality monitor equipped with water temperature, specific conductance, pH, dissolved oxygen, and turbidity (a YSI Model 6026 [December 1998 through December 2006] and YSI Model 6136 [July 2004 through December 2017]) sensors collected data during May 1998 through December 2017. 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 January 2017 through December 2019. A Hach Nitratex monitor collected nitrate data during February 2017 through December 2019.

Date model was developed: June 1, 2020

Model calibration data period: March 30, 2017 through December 10, 2019

Model Data

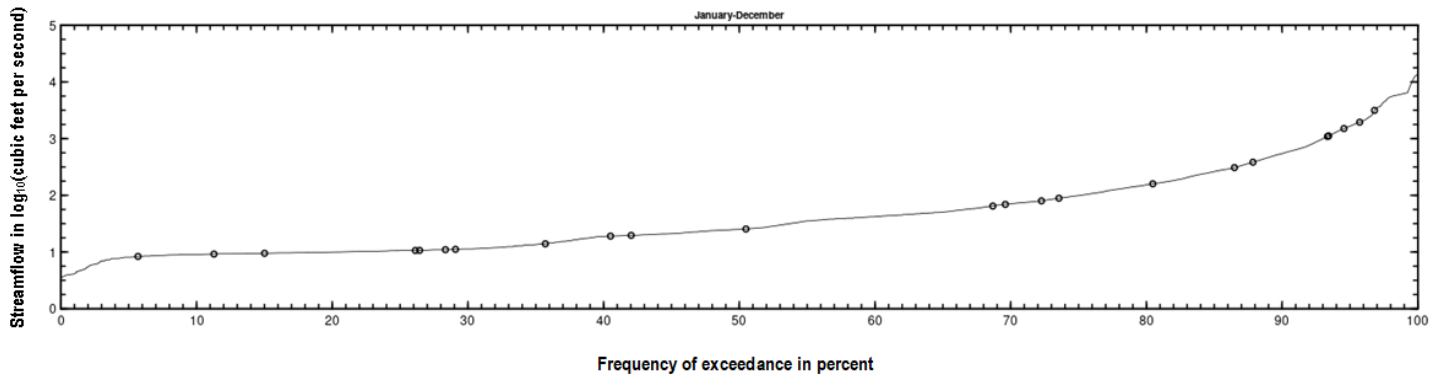
All data were collected using USGS protocols (U.S. Geological Survey, variously dated; Wagner and others, 2006; Sauer and Turnipseed, 2010; Turnipseed and Sauer, 2010) and are stored in the National Water Information System (NWIS) database (U.S. Geological Survey, 2021). Explanatory variables were evaluated 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 23 concomitant values of discretely collected total organic carbon and continuously measured turbidity during March 2017 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 dataset are provided below. Outliers and influential points were identified using studentized residuals, DFITS, Cook's D (Cook, 1977), 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.

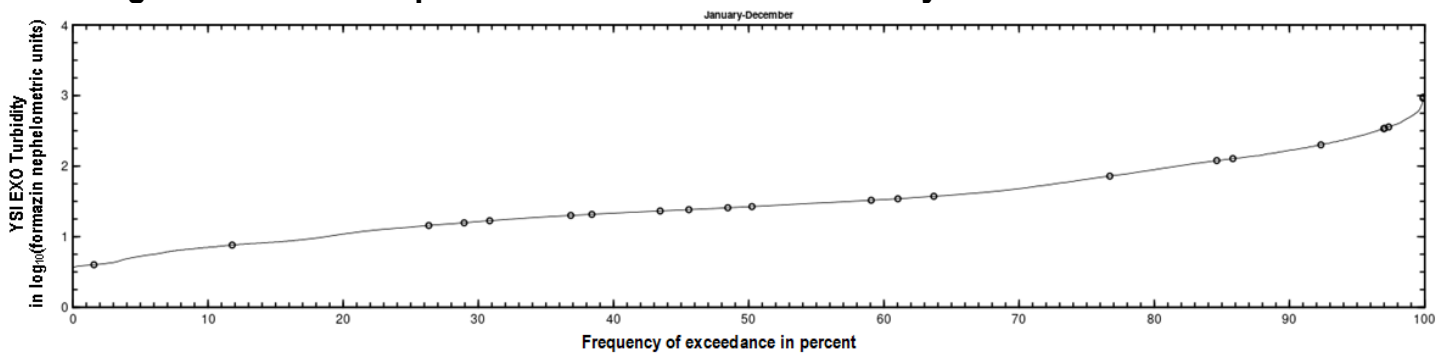
Total Organic Carbon

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 7 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 suspended solids 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:

$$\log_{10}(TOC) = 0.501 \times \log_{10}(TBY) + 0.113$$

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 1.03. The retransformed model, accounting for BCF is:

$$TOC = 1.336 \times TBY^{0.501}$$

Model Statistics, Data, and Plots

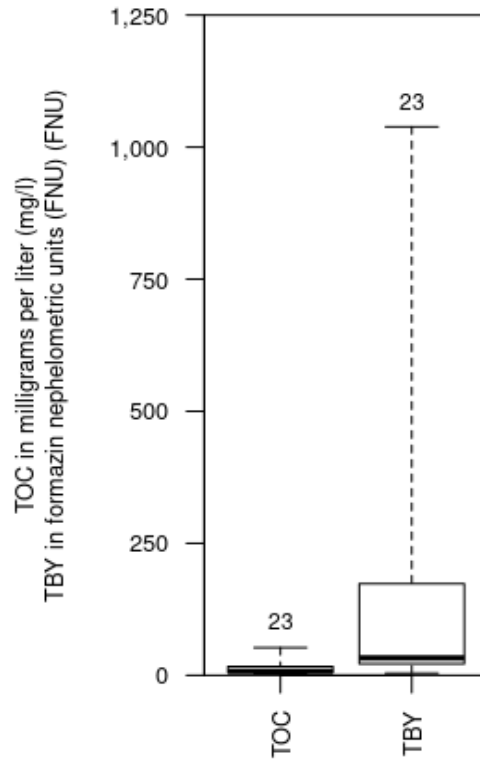
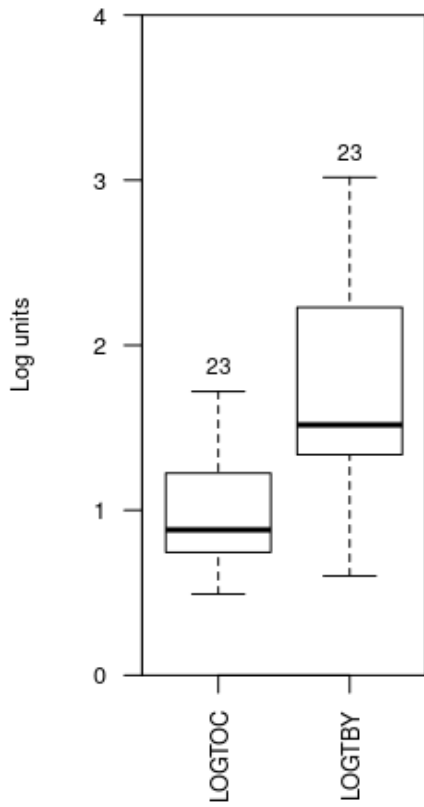
Model

$$\text{LOGTOC} = + 0.501 * \text{LOGTBY} + 0.113$$

Variable Summary Statistics

	LOGTOC	TOC	LOGTBY	TBY
Minimum	0.492	3.11	0.602	4.0
1st Quartile	0.729	5.35	1.300	20.0
Median	0.882	7.61	1.520	32.9
Mean	0.990	13.70	1.750	175.0
3rd Quartile	1.250	17.70	2.320	210.0
Maximum	1.720	52.50	3.020	1040.0

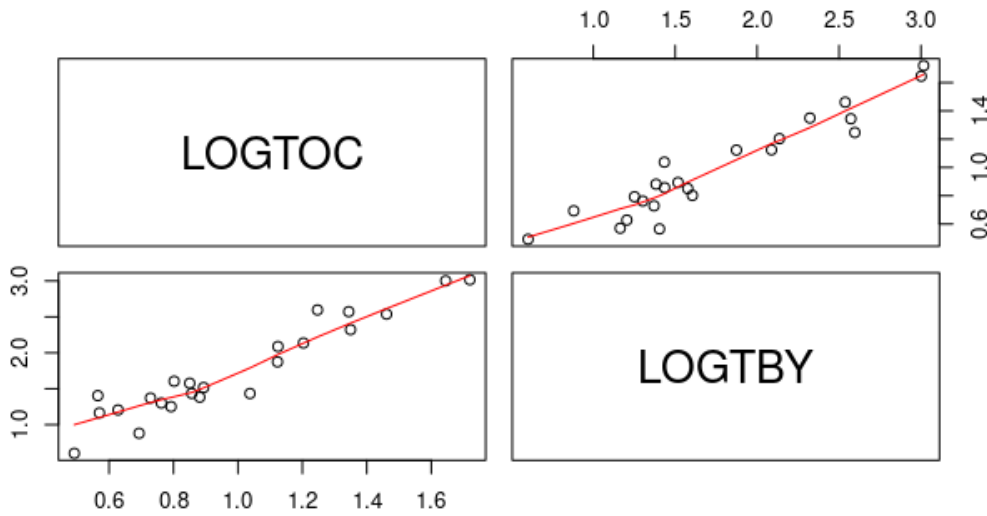
Box Plots



EXPLANATION

- 23 Number of values
- T Maximum value
- 75th percentile
- 50th percentile (median)
- 25th percentile
- Minimum value

Exploratory Plots



Basic Model Statistics

Number of Observations	23
Standard error (RMSE)	0.108
Average Model standard percentage error (MSPE)	25.1
Coefficient of determination (R ²)	0.908
Adjusted Coefficient of Determination (Adj. R ²)	0.904
Bias Correction Factor (BCF)	1.03

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(> t)
(Intercept)	0.113	0.0648	1.75	9.53e-02
LOGTBY	0.501	0.0347	14.40	2.29e-12

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.938
E.vars	-0.938	1.000

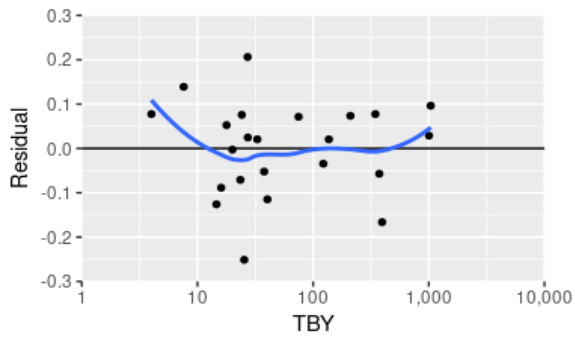
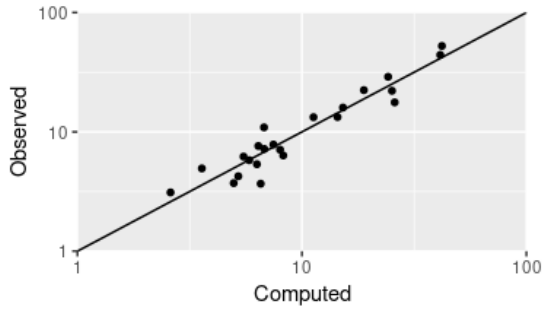
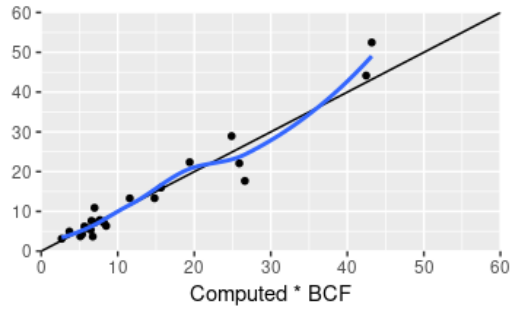
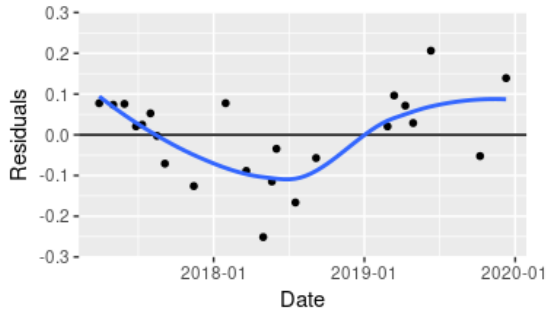
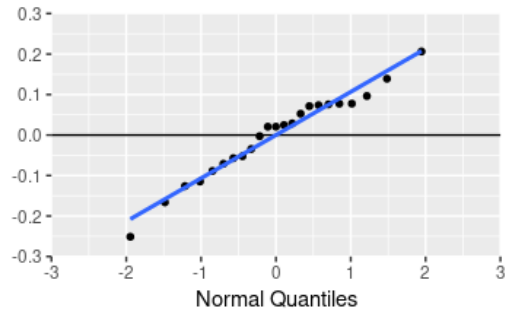
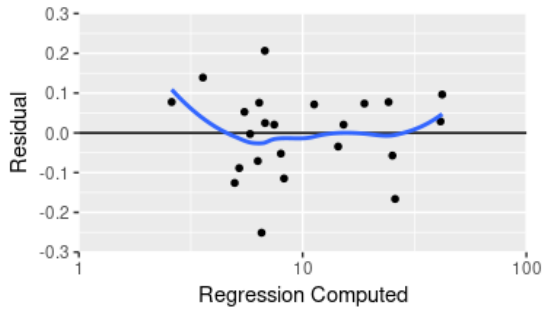
Outlier Test Criteria

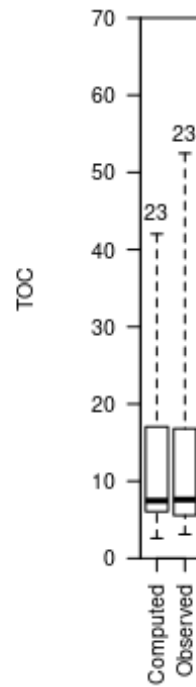
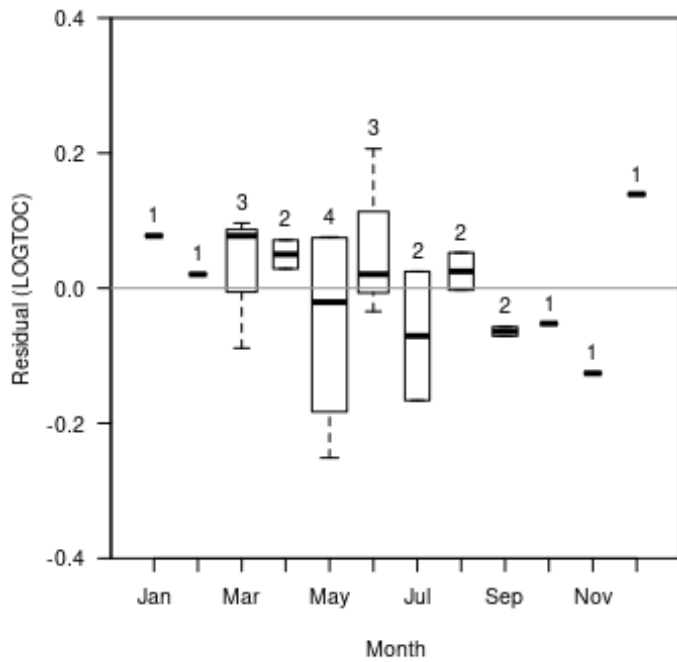
Leverage	Cook's D	DFFITS
0.261	0.193	0.590

Flagged Observations

	LOGTOC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
5/1/2018 11:10	0.565	0.816	-0.251	-2.4	-2.74	0.0559	0.17	-0.668
7/18/2018 10:20	1.25	1.41	-0.166	-1.64	-1.71	0.117	0.179	-0.625

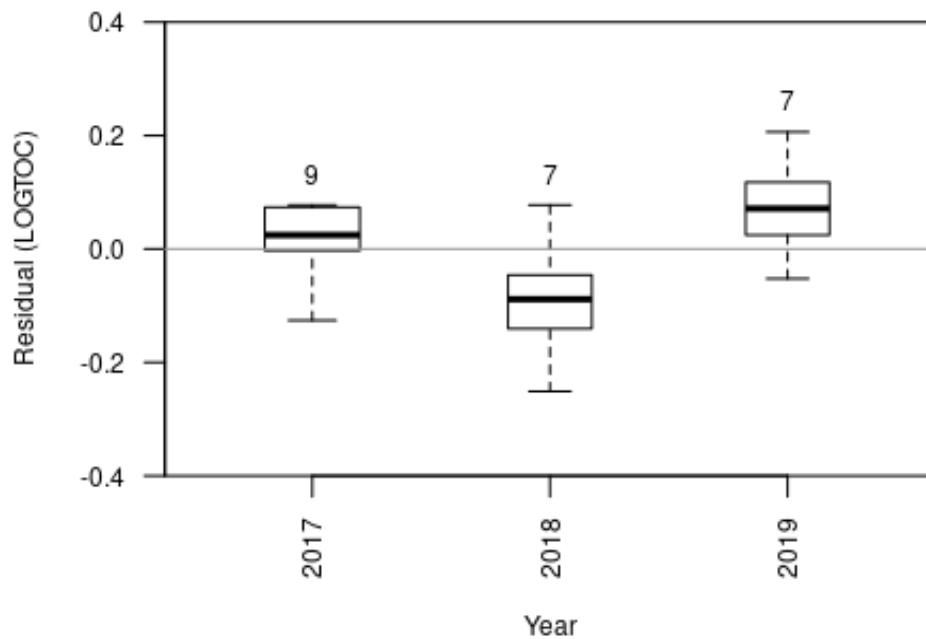
Statistical Plots





EXPLANATION

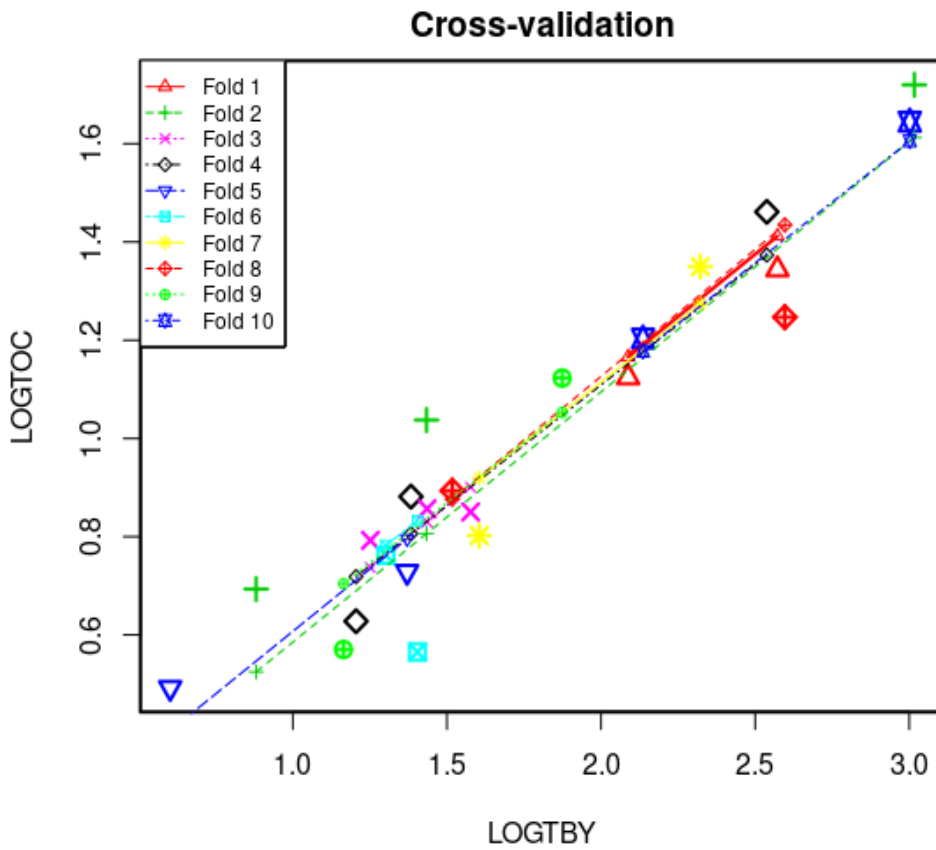
- 23 Number of values
- T Maximum value
- 75th percentile
- 50th percentile (median)
- 25th percentile
- ⊥ Minimum value



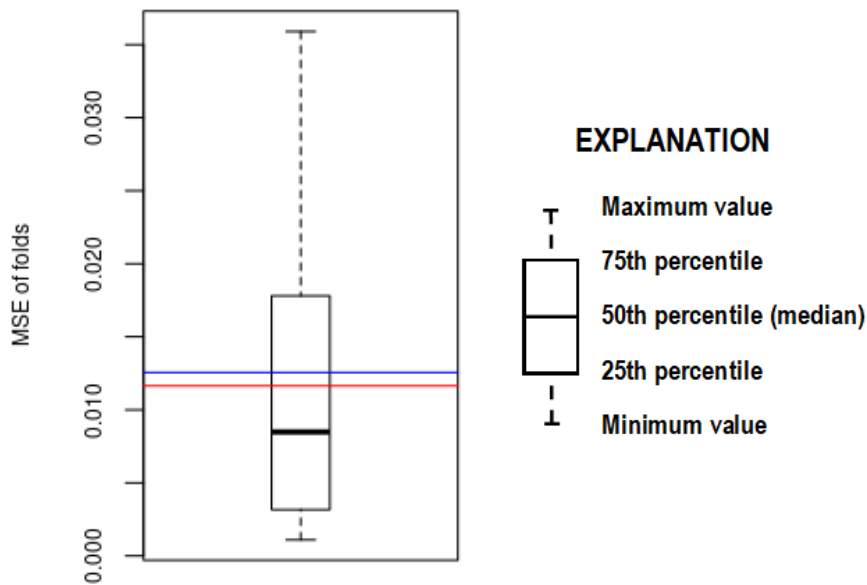
EXPLANATION

- 9 Number of values
- T Maximum value
- 75th percentile
- 50th percentile (median)
- 25th percentile
- ⊥ Minimum value

Cross Validation



Minimum MSE of folds: 0.00110
Mean MSE of folds: 0.01260
Median MSE of folds: 0.00848
Maximum MSE of folds: 0.03590
(Mean MSE of folds) / (Model MSE): 1.08000



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	3/30/2017	1.46	2.54	28.9	345	1.38	24.9	0.0774	0.848
2	5/3/2017	1.35	2.32	22.4	210	1.28	19.4	0.0736	0.57
3	5/30/2017	0.882	1.38	7.61	24.2	0.806	6.57	0.0757	0.702
4	6/27/2017	0.894	1.52	7.83	32.9	0.873	7.67	0.0206	0
5	7/12/2017	0.857	1.44	7.19	27.3	0.832	6.98	0.0247	0.108
6	8/1/2017	0.792	1.25	6.2	17.8	0.74	5.65	0.0525	0.33
7	8/17/2017	0.762	1.3	5.79	20	0.765	5.98	-0.00282	-0.218
8	9/5/2017	0.729	1.37	5.35	23.5	0.8	6.48	-0.0709	-0.702
9	11/14/2017	0.57	1.16	3.72	14.6	0.696	5.11	-0.126	-1.22
10	1/30/2018	0.492	0.602	3.11	4	0.415	2.67	0.0775	1.01
11	3/21/2018	0.628	1.21	4.25	16	0.717	5.35	-0.0887	-0.848
12	5/1/2018	0.565	1.4	3.67	25.4	0.816	6.73	-0.251	-1.95
13	5/22/2018	0.802	1.6	6.34	40.3	0.917	8.48	-0.115	-1.01
14	6/2/2018	1.12	2.09	13.3	122	1.16	14.8	-0.0345	-0.33
15	7/18/2018	1.25	2.6	17.7	395	1.41	26.6	-0.166	-1.48
16	9/6/2018	1.34	2.57	22.1	373	1.4	25.9	-0.0571	-0.57
17	2/26/2019	1.2	2.14	16	137	1.18	15.7	0.0205	-0.108
18	3/14/2019	1.72	3.02	52.5	1040	1.62	43.2	0.0964	1.22
19	4/10/2019	1.12	1.87	13.3	74.8	1.05	11.6	0.0714	0.447
20	4/29/2019	1.64	3	44.2	1000	1.62	42.5	0.0288	0.218
21	6/11/2019	1.04	1.43	10.9	27.2	0.831	6.97	0.206	1.95
22	10/8/2019	0.85	1.58	7.09	37.7	0.903	8.21	-0.0523	-0.447
23	12/10/2019	0.693	0.881	4.93	7.6	0.554	3.68	0.139	1.48

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>.]