Appendix 18. Model Archive Summary for Total Organic Carbon at U.S. Geological Survey Site 07144780, North Fork Ninnescah River above Cheney Reservoir, Kansas, during January 1, 1999, through December 31, 2019

This model archive summary summarizes the total organic carbon (TOC) model developed to compute hourly or daily TOC from January 1, 1999, through December 31, 2019. This model is used concomitantly with other models during this period to calculate concentrations when other explanatory variables are not available for the purposes of load and concentration model calculations. The methods used 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), and other standard USGS methods (Sauer and Turnipseed, 2010; Turnipseed and Sauer, 2010).

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 Sutron Satlink 2 High Data Rate Collection Platform and a Design Analysis Water Log H350/355 nonsubmersible pressure transducer transfers real-time stage, precipitation, 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 21.804 feet. The orifice is enclosed in 1 1/4-inch pipe, which runs from the gage house, under the bridge, and along an I-beam where it is attached to the concrete pier closest to the left edge of water.

Date model was developed: April 26, 2019

Model calibration data period: April 16, 1999, to September 28, 2017

Model Data

All data were collected using USGS protocols (U.S. Geological Survey, 2006; Wagner and others, 2006; Sauer and Turnipseed, 2010; Turnipseed and Sauer, 2010) and are stored in the National Water Information System (NWIS) database (https://doi.org/10.5066/F7P55KJN; U.S. Geological Survey, 2020). Explanatory variables were evaluated individually and in combination. Potential explanatory variables included streamflow, water temperature, specific conductance, pH, dissolved oxygen, and turbidity. Seasonal components (sine and cosine variables) were also evaluated as explanatory variables.

The regression model is based on 84 concomitant values of discretely collected TOC samples and continuously measured streamflow collected during April 16, 1999, through September 28, 2017. Discrete samples were collected over a range of streamflows. No samples were less than laboratory detection limits. Summary statistics and the complete model-calibration data are provided below. Outliers were identified using studentized residuals (for values greater than 3 or less than −3). None of the samples in this dataset were deemed outliers or removed from the model calibration 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, multiple vertical, single vertical, or grab methods following U.S. Geological Survey (2006) 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 Federal Interagency Sedimentation Project U.S. 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 TOC by the Wichita Municipal Water and Wastewater Laboratory in Wichita, Kans., according to standard methods (American Public Health Association and others, 1995).

Continuous Data

Streamflow was measured using a nonsubmersible pressure transducer during January 1, 1999, through December 31, 2019. The continuous streamflow data used were time interpolated values from the continuous time series. If the continuous data were not available, the sample was not included in the dataset.

Model Development

Ordinary least squares regression analysis was done using R programming language (R Core Team, 2019) to relate discretely collected TOC concentrations 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 calculated TOC 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) was ultimately selected.

Streamflow was selected as a good predictor of logarithm base 10 ($\log_{10}$) (TOC) based on residual plots, relatively high coefficient of determination (R^2), and relatively low model standard percentage error (MSPE). This model was developed with the sole purpose to fill in gaps of missing data of the primary model for concentration and load estimations.

Model Summary

Summary of final TOC regression analysis at USGS site 07144780:

TOC-based model:

$\log_{10}(TOC) = 0.257 \times \log_{10}(Q) + 0.331$,

where,

TOC = organic carbon, total, in milligrams per liter, and
Q = streamflow, in cubic feet per second.

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 1.08. The retransformed model, accounting for BCF, is as follows:

$TOC = (Q^{0.257} \times 10^{0.331}) \times 1.08$

Model Statistics, Data, and Plots

Definitions for terms used in this output can be found at the end of this document.

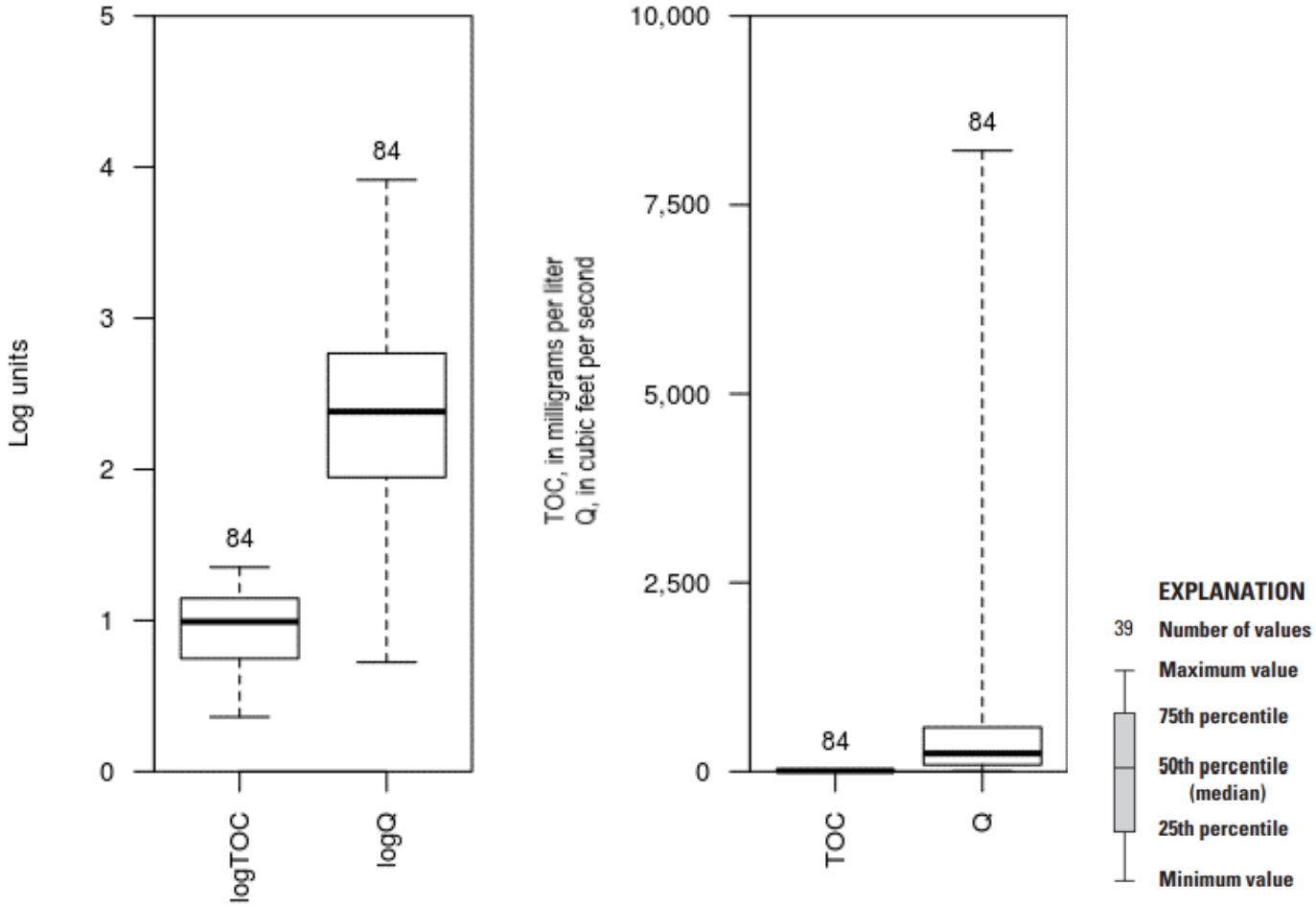
Model

$\log_{10}(TOC) = 0.257 \times \log_{10}(Q) + 0.331$,

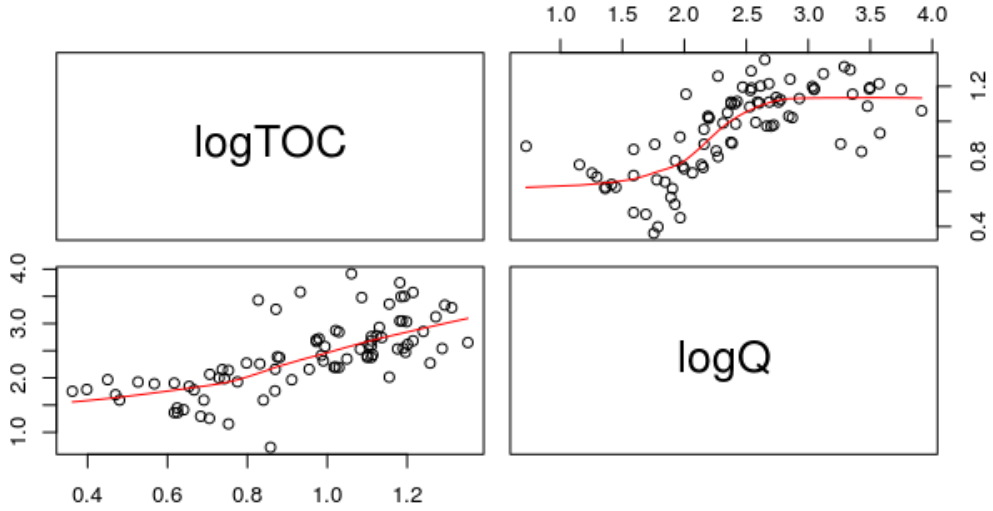
Variable Summary Statistics

	logTOC	TOC	logQ	Q
Minimum	0.362	2.30	0.724	5.29
1st Quartile	0.748	5.59	1.950	88.50
Median	0.992	9.83	2.380	241.00
Mean	0.946	10.20	2.400	742.00
3d Quartile	1.150	14.00	2.770	587.00
Maximum	1.350	22.50	3.910	8220.00

Box Plots



Exploratory Plots



Red line shows the locally weighted scatterplot smoothing (LOWESS).

Basic Model Statistics

For a detailed definition and explanation of the terms used below, refer to Helsel and Hirsch (2002).

```
Number of Observations                              84
Standard error (RMSE)                            0.178
Average Model standard percentage error (MSPE)    42.1
Coefficient of determination (R²)                 0.48
Adjusted Coefficient of Determination (Adj. R²) 0.473
Bias Correction Factor (BCF)                      1.08
```

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.331	0.0734	4.5	2.19e-05
logQ	0.257	0.0295	8.7	2.90e-13

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.964
E.vars	-0.964	1.000

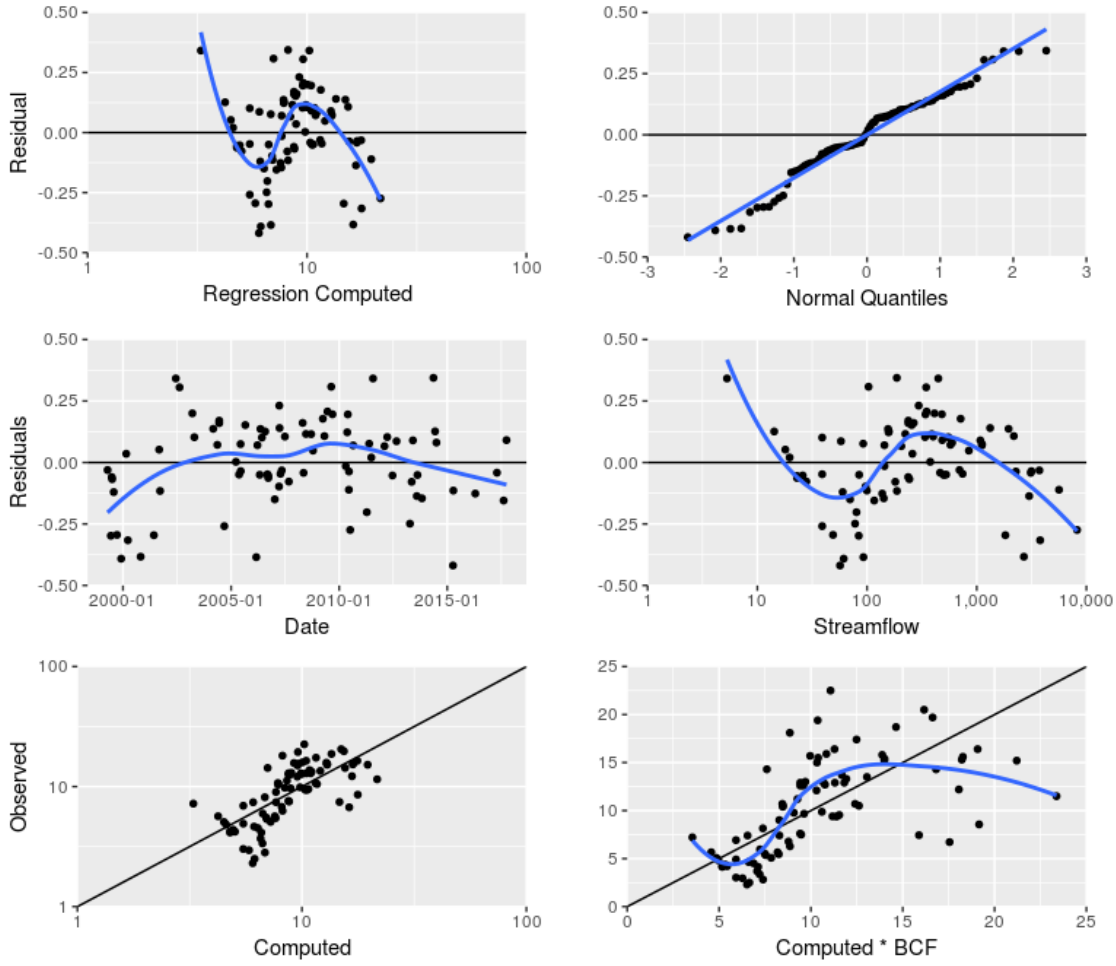
Outlier Test Criteria

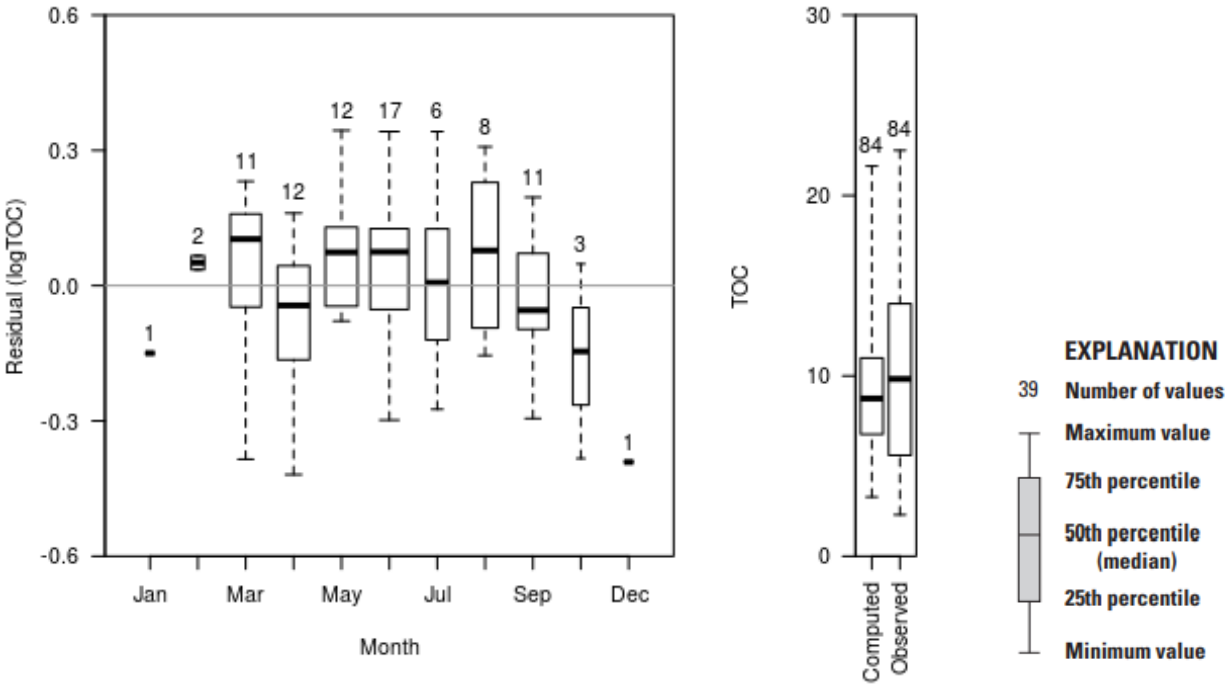
Leverage	Cook's D	DFFITS
0.0714	0.1943	0.3086

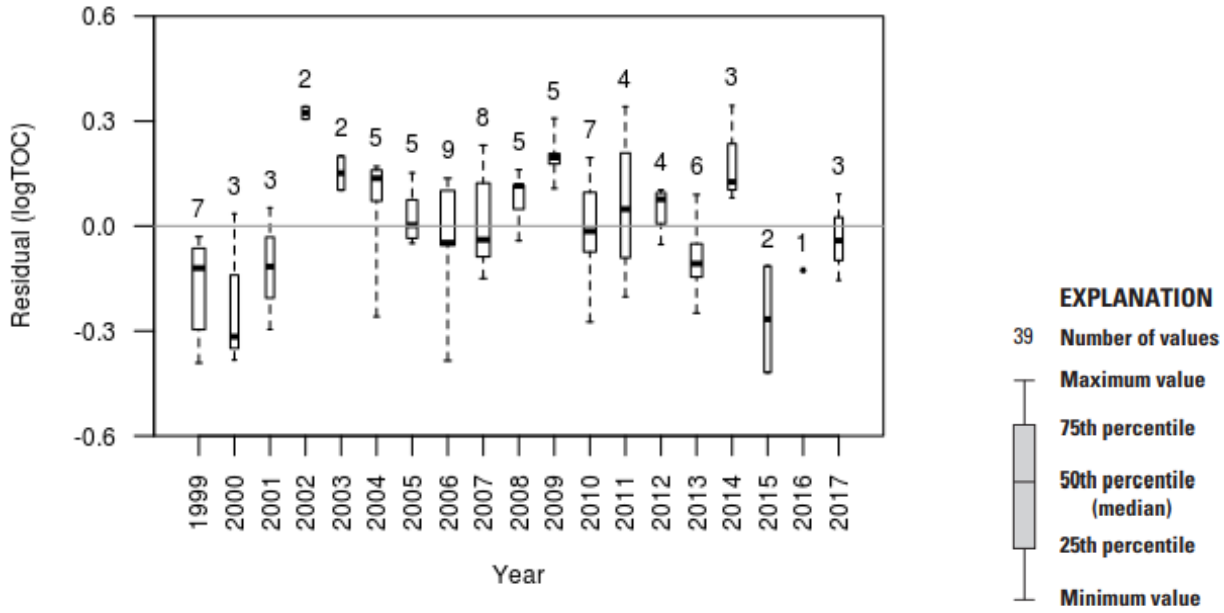
Flagged Observations

	logTOC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
12/2/1999 10:35	0.398	0.789	-0.391	-2.23	-2.28	0.0222	0.0564	-0.344
3/24/2000 13:50	0.932	1.250	-0.316	-1.83	-1.85	0.0502	0.0882	-0.426
10/26/2000 10:50	0.827	1.210	-0.383	-2.20	-2.26	0.0412	0.1040	-0.468
6/6/2001 11:35	0.872	1.170	-0.296	-1.69	-1.71	0.0324	0.0479	-0.313
7/6/2010 10:30	1.060	1.340	-0.274	-1.61	-1.62	0.0752	0.1050	-0.462
7/27/2011 11:20	0.858	0.516	0.342	2.02	2.05	0.0894	0.1990	0.644
4/8/2015 9:45	0.362	0.781	-0.419	-2.39	-2.46	0.0234	0.0682	-0.380

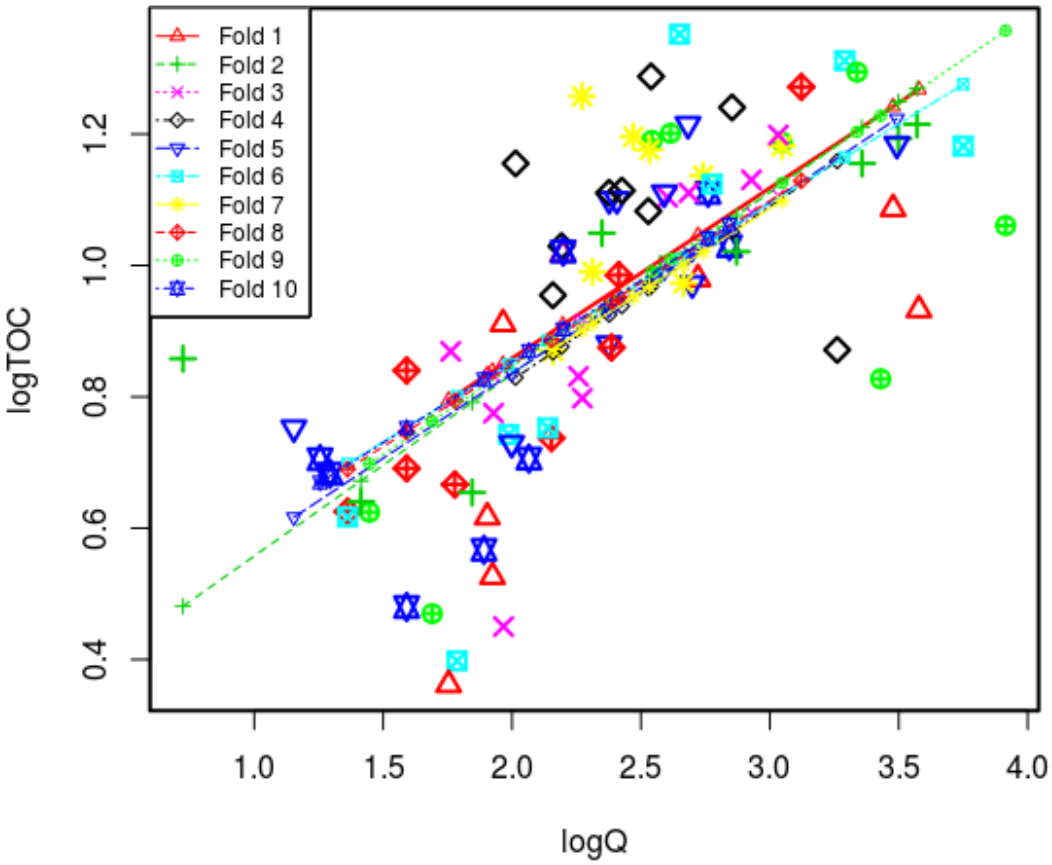
Statistical Plots







Cross Validation



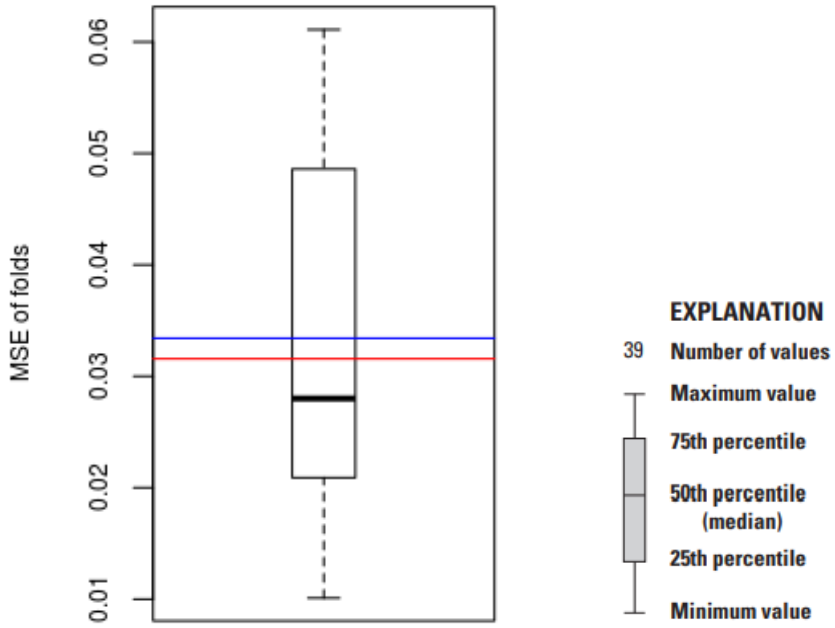
Fold – equal partition of the data (10 percent of the data)

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


              Minimum MSE of folds:  0.0101
                 Mean MSE of folds:  0.0334
               Median MSE of folds:  0.0280
              Maximum MSE of folds:  0.0611
 (Mean MSE of folds) / (Model MSE):  1.0600
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Model-Calibration Data Set

	Date	logTOC	logQ	TOC	Q	Computed logTOC	Computed TOC	Residual	Normal Quantiles	Censored Values
1	1999-04-16	1.03	2.84	10.7	697	1.06	12.4	-0.0306	-0.0745	--
2	1999-06-10	0.526	1.92	3.36	84	0.824	7.21	-0.298	-1.5	--
3	1999-06-25	0.881	2.38	7.6	237	0.94	9.41	-0.0592	-0.479	--
4	1999-07-02	0.875	2.39	7.5	243	0.943	9.47	-0.0678	-0.547	--
5	1999-07-29	0.667	1.78	4.64	60	0.787	6.61	-0.12	-0.81	--
6	1999-09-22	0.47	1.69	2.95	49	0.764	6.28	-0.294	-1.34	--
7	1999-12-02	0.398	1.79	2.5	61.2	0.789	6.65	-0.391	-2.07	--
8	2000-02-25	0.985	2.41	9.67	259	0.95	9.63	0.0355	0.0447	--
9	2000-03-24	0.932	3.58	8.56	3790	1.25	19.2	-0.316	-1.6	--
10	2000-10-26	0.827	3.43	6.72	2690	1.21	17.5	-0.383	-1.72	--
11	2001-06-06	0.872	3.26	7.44	1820	1.17	15.9	-0.296	-1.42	--
12	2001-09-04	0.705	1.26	5.07	18	0.653	4.86	0.0523	0.104	--
13	2001-09-19	0.798	2.27	6.28	187	0.914	8.86	-0.116	-0.769	--
14	2002-06-12	1.35	2.65	22.5	446	1.01	11.1	0.342	2.07	--

15	2002-08-14	1.29	2.54	19.4	346	0.982	10.4	0.306	1.6	--
16	2003-03-18	1.2	2.61	15.9	412	1	10.8	0.2	1.34	--
17	2003-04-21	1.14	2.74	13.7	551	1.03	11.7	0.103	0.479	--
18	2004-03-05	1.31	3.29	20.5	1950	1.17	16.2	0.137	0.852	--
19	2004-05-14	1.11	2.76	12.9	576	1.04	11.8	0.0718	0.225	--
20	2004-06-14	1.1	2.38	12.6	238	0.94	9.42	0.16	0.988	--
21	2004-06-14	1.11	2.38	12.9	238	0.94	9.42	0.17	1.09	--
22	2004-09-08	0.48	1.59	3.02	39	0.739	5.92	-0.259	-1.21	--
23	2005-03-24	0.994	2.58	9.87	377	0.992	10.6	0.00276	-0.0149	--
24	2005-05-16	0.98	2.72	9.54	526	1.03	11.5	-0.0492	-0.318	--
25	2005-06-10	1.19	3.05	15.4	1120	1.11	14	0.0749	0.256	--
26	2005-06-13	1.19	3.5	15.6	3150	1.23	18.3	-0.0351	-0.134	--
27	2005-08-29	1.1	2.41	12.6	255	0.948	9.59	0.152	0.941	--
28	2006-03-02	0.45	1.97	2.82	92.7	0.835	7.39	-0.385	-1.87	--
29	2006-03-22	0.955	2.16	9.01	144	0.884	8.28	0.0703	0.195	--
30	2006-05-01	0.775	1.93	5.96	84.8	0.825	7.23	-0.0501	-0.349	--
31	2006-05-12	1.03	2.19	10.7	155	0.893	8.45	0.136	0.81	--
32	2006-06-05	0.84	1.59	6.92	39	0.739	5.92	0.101	0.446	--
33	2006-07-31	0.753	1.15	5.66	14.2	0.627	4.57	0.126	0.73	--
34	2006-09-07	0.691	1.59	4.91	39	0.739	5.92	-0.0478	-0.287	--
35	2006-09-21	0.625	1.36	4.22	23	0.68	5.17	-0.0547	-0.446	--
36	2006-09-21	0.617	1.36	4.14	23	0.68	5.17	-0.063	-0.513	--
37	2007-01-09	0.654	1.85	4.51	70	0.804	6.88	-0.15	-0.988	--
38	2007-03-22	0.743	1.99	5.53	97	0.84	7.48	-0.0976	-0.654	--
39	2007-03-26	1.2	2.47	15.7	296	0.965	9.96	0.231	1.5	--
40	2007-03-31	1.27	3.12	18.7	1320	1.13	14.6	0.14	0.895	--
41	2007-04-16	1.02	2.87	10.5	744	1.07	12.6	-0.0461	-0.256	--
42	2007-05-07	1.21	3.57	16.4	3720	1.25	19.1	-0.032	-0.104	--
43	2007-06-29	1.1	2.6	12.7	401	0.999	10.8	0.105	0.547	--
44	2007-09-04	0.624	1.45	4.21	28	0.702	5.44	-0.0776	-0.582	--
45	2008-04-24	1.11	2.43	13	267	0.953	9.7	0.161	1.04	--
46	2008-05-09	1.18	3.49	15.3	3110	1.23	18.2	-0.0421	-0.225	--
47	2008-06-19	1.05	2.35	11.2	223	0.933	9.27	0.116	0.654	--
48	2008-09-15	1.11	2.59	12.9	390	0.995	10.7	0.115	0.618	--
49	2008-10-16	1.13	2.93	13.5	848	1.08	13.1	0.0484	0.0745	--
50	2009-03-31	1.24	2.85	17.4	713	1.06	12.5	0.178	1.15	--
51	2009-04-27	1.29	3.34	19.7	2180	1.19	16.6	0.107	0.582	--
52	2009-06-17	1.19	2.54	15.5	349	0.983	10.4	0.207	1.42	--
53	2009-08-20	1.16	2.01	14.3	103	0.847	7.6	0.308	1.72	--
54	2009-09-10	1.21	2.68	16.4	481	1.02	11.3	0.196	1.27	--
55	2010-04-23	0.87	2.16	7.41	144	0.885	8.29	-0.0148	-0.0447	--
56	2010-05-17	1.02	2.2	10.4	157	0.894	8.47	0.123	0.691	--
57	2010-05-27	1.18	2.53	15	341	0.981	10.3	0.196	1.21	--
58	2010-06-14	1.18	3.75	15.2	5630	1.29	21.2	-0.111	-0.691	--
59	2010-06-16	1.16	3.36	14.3	2280	1.19	16.8	-0.0367	-0.164	--
60	2010-07-06	1.06	3.91	11.5	8220	1.34	23.4	-0.274	-1.27	--
61	2010-08-25	1.18	3.05	15.2	1110	1.11	14	0.0695	0.164	--
62	2011-04-13	0.617	1.9	4.14	80.2	0.819	7.12	-0.202	-1.09	--
63	2011-05-23	0.911	1.96	8.15	92.2	0.835	7.39	0.0764	0.287	--
64	2011-06-28	0.683	1.29	4.82	19.7	0.663	4.97	0.0204	0.0149	--
65	2011-07-27	0.858	0.724	7.21	5.29	0.516	3.55	0.342	1.87	--
66	2012-02-06	0.99	2.31	9.78	205	0.924	9.07	0.0664	0.134	--
67	2012-03-23	1.08	2.53	12.1	338	0.979	10.3	0.103	0.513	--
68	2012-06-20	0.64	1.41	4.37	25.9	0.693	5.33	-0.0529	-0.413	--
69	2012-08-27	0.869	1.76	7.4	57.8	0.783	6.55	0.0865	0.349	--

```
70 2013-04-11  0.567  1.89 3.69 77.8    0.816      7.07   -0.249     -1.15       --
71 2013-05-10  0.831  2.26 6.78  181     0.91      8.78  -0.0786    -0.618       --
72 2013-05-31    1.2  3.03 15.8 1080     1.11      13.9   0.0899     0.381       --
73 2013-08-05   1.09  3.48 12.2 3010     1.22      18.1   -0.137    -0.895       --
74 2013-08-16  0.972   2.7 9.38  500     1.02      11.4  -0.0509    -0.381       --
75 2013-10-31  0.737  2.15 5.46  142    0.883      8.25   -0.146    -0.941       --
76 2014-05-13   1.26  2.27 18.1  187    0.913      8.85    0.344      2.45       --
77 2014-06-10   1.02   2.2 10.5  158    0.895      8.48    0.126     0.769       --
78 2014-07-02   1.12  2.78 13.3  597     1.04      11.9   0.0809     0.318       --
79 2015-04-08  0.362  1.75  2.3 56.8    0.781      6.52   -0.419     -2.45       --
80 2015-04-14  0.729     2 5.36 99.7    0.843      7.53   -0.114     -0.73       --
81 2016-04-19  0.753  2.14 5.66  138    0.879      8.18   -0.126    -0.852       --
82 2017-04-20  0.973  2.67  9.4  463     1.01      11.2  -0.0415    -0.195       --
83 2017-08-11  0.706  2.07 5.08  116    0.861      7.84   -0.155     -1.04       --
84 2017-09-28   1.11  2.69 12.9  485     1.02      11.3   0.0909     0.413       --
```

Definitions

```
TOC: organic carbon, total, in milligrams per liter (00680)
Q: streamflow, instantaneous, in cubic feet per second (00061)
```

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

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., Graham, J.L., Milligan, C.R., Pope, L.M., and Ziegler, A.C., 2006, Water quality and relation to taste-and-odor compounds in the North Fork Ninnescah River and Cheney Reservoir, south-central Kansas, 1997–2003: U.S. Geological Survey Scientific Investigations Report 2006–5095, 43 p. [Also available at https://doi.org/10.3133/sir20065095.]

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.]

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 522 p. [Also available at https://doi.org/10.3133/tm4A3.]

R Core Team, 2019, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, accessed August 2019 at https://www.R-project.org/.

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 April 2020 at https://doi.org/10.3133/ofr20141233.

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.]

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., accessed April 2020 at https://doi.org/10.3133/tm3A7.

Stone, M.L., Graham, J.L., and Gatotho, J.W., 2013, Continuous real-time water-quality monitoring and regression analysis to compute constituent concentrations and loads in the North Fork Ninnescah River upstream from Cheney Reservoir, south-central Kansas, 1999–2012: U.S. Geological Survey Scientific Investigations Report 2013–5071, 44 p., accessed July 2020 at https://doi.org/10.3133/sir20135071.

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 April 2020 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, 2020, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 20, 2020, 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.]