Appendix 16. 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 October 16, 2009

This model archive summary summarizes the total organic carbon (TOC) model developed to compute hourly or daily TOC from January 1, 1999, through October 16, 2009. This model supersedes all prior models used during this period. 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, chap. C4 (Rasmussen and others, 2009).

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 6600 Extended Deployment System water-quality monitor equipped with sensors for water temperature, specific conductance, pH, dissolved oxygen, and turbidity (a YSI Model 6026 turbidity sensor [November 9, 1998–December 1, 2010] and a YSI Model 6136 turbidity sensor [October 17, 2009, to November 12, 2015; March 31, 2017, to June 7, 2017]) (YSI Incorporated, 2007, 2012). The YSI 6600 water-quality monitor was in operation during November 9, 1998, through November 12, 2015.

Date model was developed: April 26, 2019

Model calibration data period: April 16, 1999, to August 25, 2010

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 60 concomitant values of discretely measured total organic carbon samples and continuously measured turbidity collected during April 16, 1999, through August 25, 2010. Discrete samples were collected over a range of continuously measured turbidity conditions. No samples were below laboratory detection limits. Summary statistics and the complete model-calibration data are provided in the dataset. 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). The sample collected on September 19, 2001, had large heterogeneity in the channel cross-sectional data during sample collection and was removed from the model calibration dataset. Outliers in previously published versions of this model (Stone and others, 2013) 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.

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 2 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

Turbidity was measured using a YSI model 6026 sensor installed during November 9, 1998, through December 1, 2010. Concomitant turbidity values were time interpolated. If continuous data were not available (2 or more hours of specific conductance values bracketing the sample collection time were missing) because of fouling, changes in equipment, or unsuitable site conditions, then the field monitor turbidity value measured during sampling was substituted. If no concomitant continuous data were 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.

Turbidity was selected as the best 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).

Model Summary

Summary of final TOC regression analysis at USGS site 07144780:

TOC-based model:

$\log_{10}(TOC) = 0.421 \times \log_{10}(TBY6026) + 0.138$,

where,

TOC = organic carbon, total, in milligrams per liter, and
$TBY6026$ = turbidity, YSI model 6026, in formazin nephelometric units.

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

$TOC = (TBY6026^{0.421} \times 10^{0.138}) \times 1.05$

Previous Models

Version	Model Equation	Reference
1.0	$\log_{10}(TOC) = 0.485 \times \log_{10}(TBY6026) + 0.008$	Stone and others (2013)

Model Statistics, Data, and Plots

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

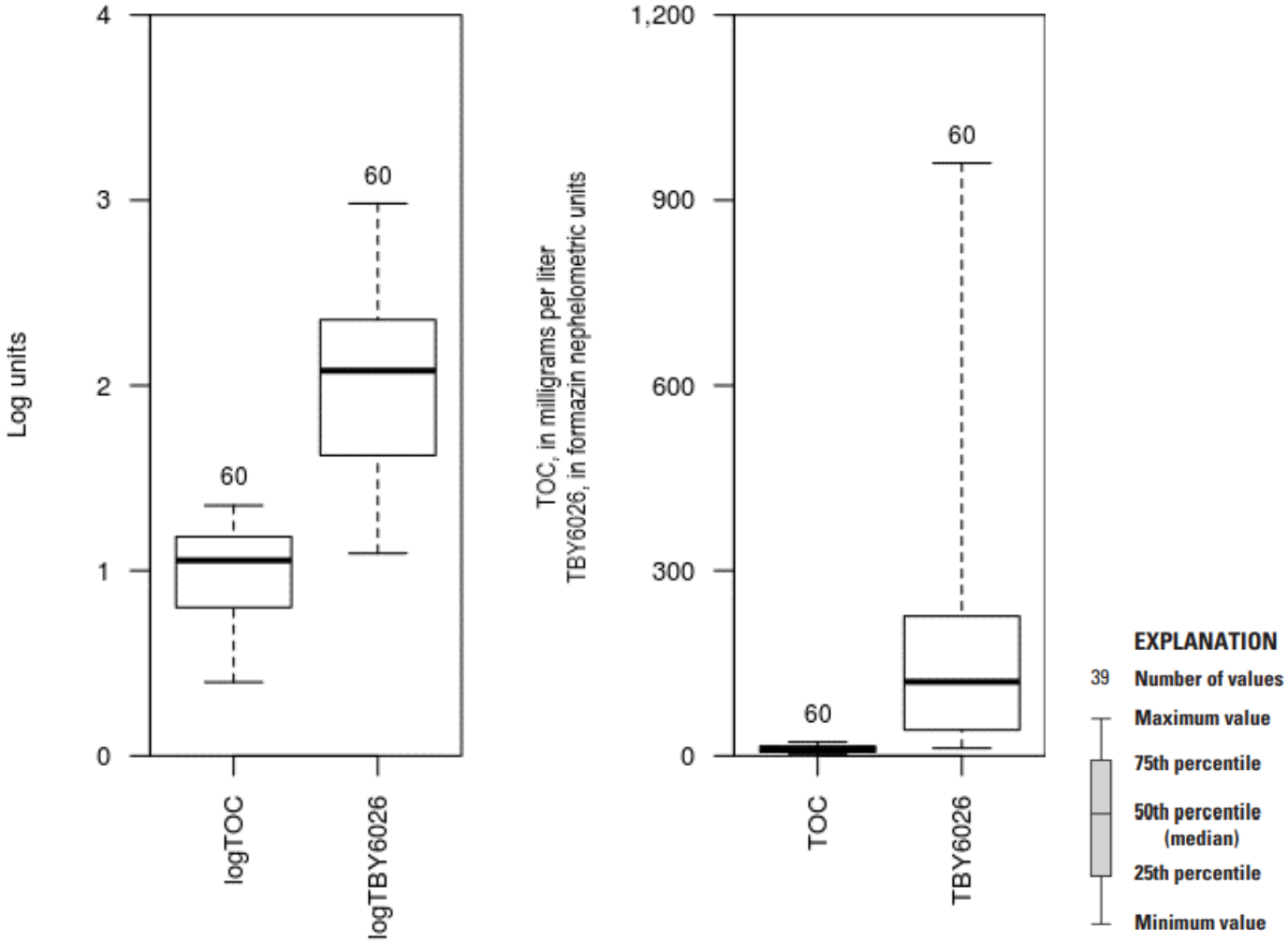
Model

$\log_{10}(TOC) = 0.421 \times \log_{10}(TBY6026) + 0.138$

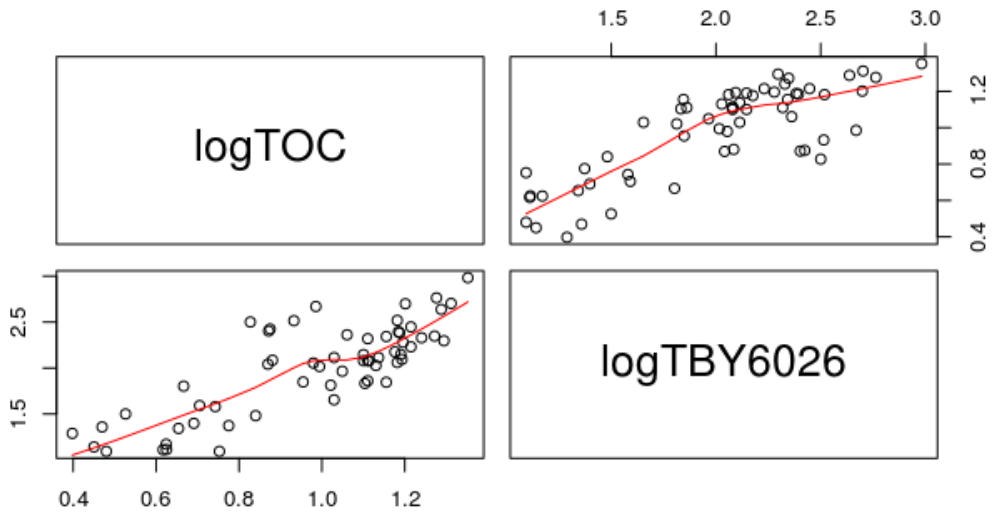
Variable Summary Statistics

	logTOC	TOC	logTBY6026	TBY6026
Minimum	0.398	2.50	1.09	12.4
1st Quartile	0.801	6.34	1.62	42.1
Median	1.050	11.30	2.08	120.0
Mean	0.981	11.00	2.00	168.0
3d Quartile	1.180	15.30	2.35	226.0
Maximum	1.350	22.50	2.98	960.0

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                              60
Standard error (RMSE)                            0.149
Average Model standard percentage error (MSPE)      35
Coefficient of determination (R²)                0.656
Adjusted Coefficient of Determination (Adj. R²)   0.65
Bias Correction Factor (BCF)                      1.05
```

Explanatory Variables

```
            Coefficients Standard Error t value Pr(>|t|)
(Intercept)        0.138         0.0825    1.67 9.96e-02
logTBY6026         0.421         0.0401   10.50 4.78e-15
```

Correlation Matrix

```
           Intercept E.vars
Intercept      1.000 -0.972
E.vars        -0.972  1.000
```

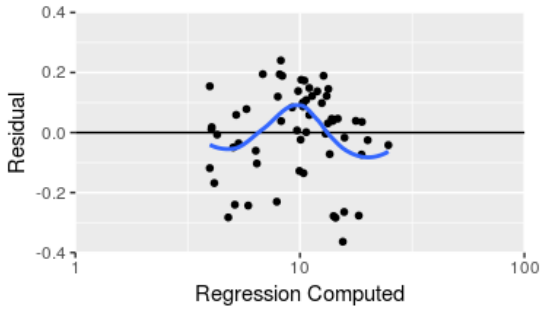
Outlier Test Criteria

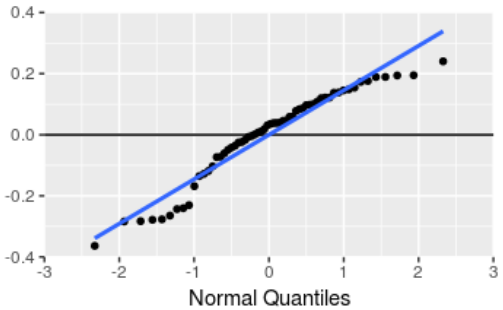
```
Leverage Cook's D   DFFITS
   0.100    0.194    0.365
```

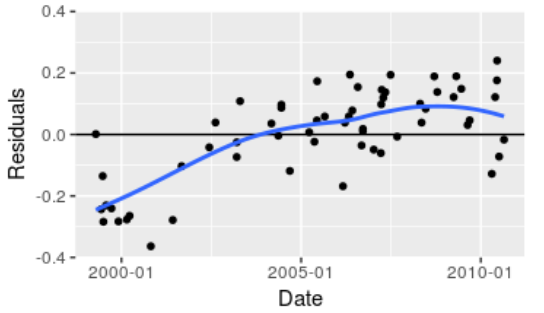
Flagged Observations

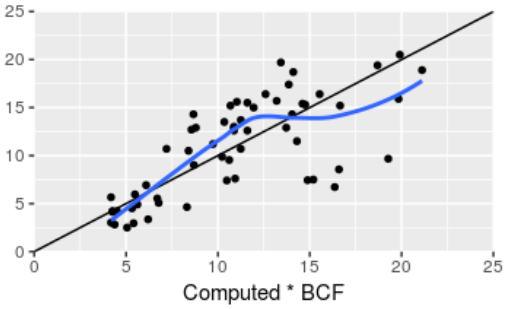
```
                 logTOC Estimate Residual Standard Residual Studentized Residual Leverage Cook's D DFFITS
9/22/1999 11:20   0.470    0.710   -0.240            -1.65                -1.68   0.0465   0.0665 -0.370
12/2/1999 10:35   0.398    0.681   -0.283            -1.95                -2.00   0.0534   0.1070 -0.475
2/25/2000 10:40   0.985    1.260   -0.277            -1.90                -1.95   0.0488   0.0928 -0.441
10/26/2000 10:50  0.827    1.190   -0.364            -2.48                -2.60   0.0346   0.1100 -0.492
```

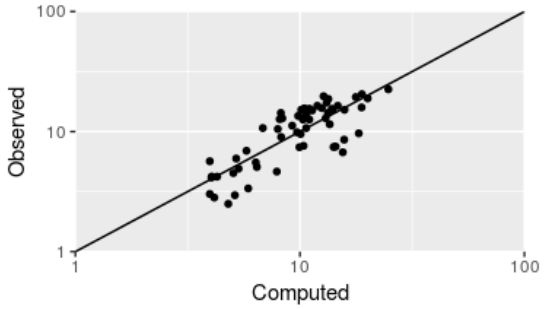
Statistical Plots

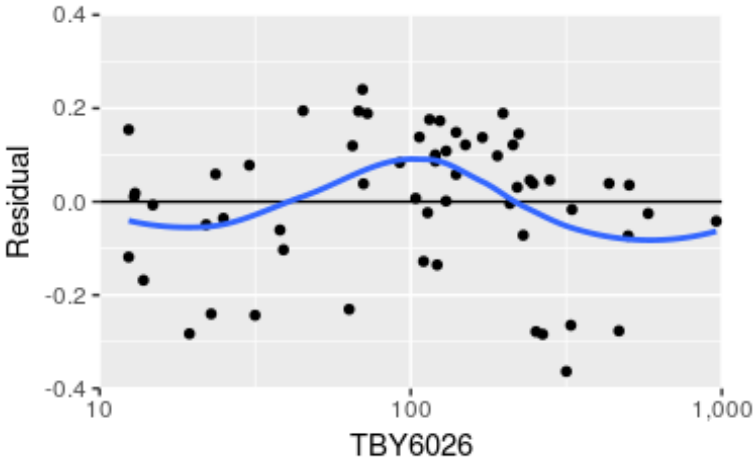


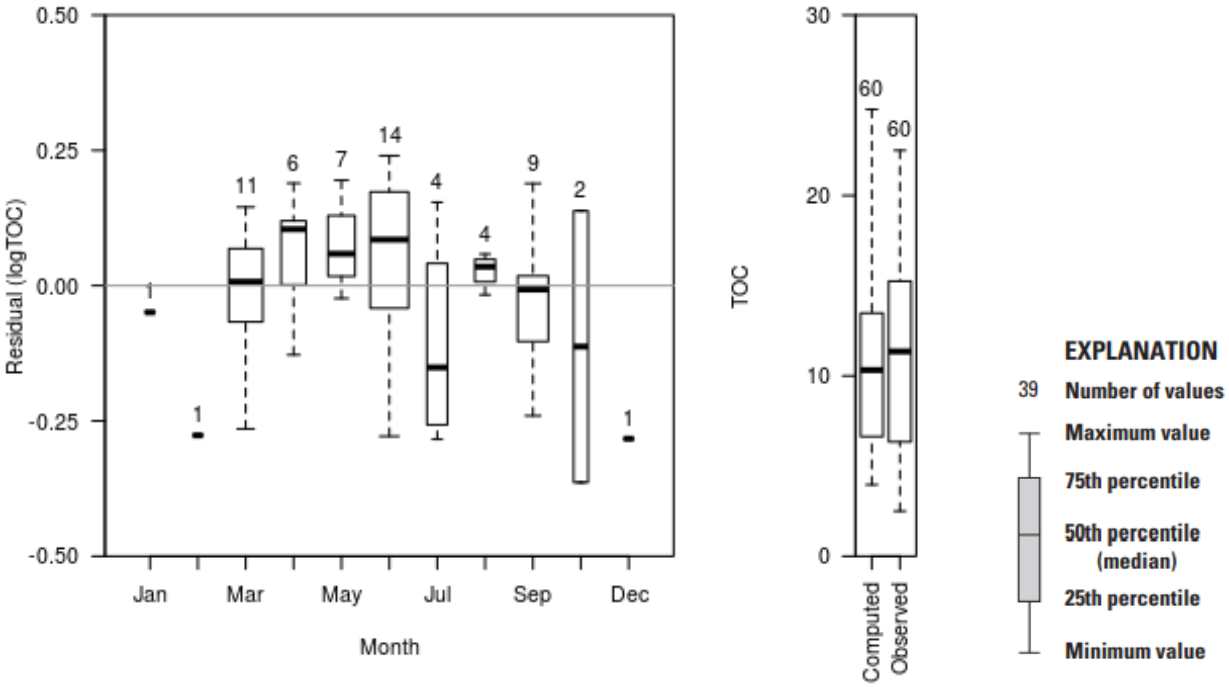


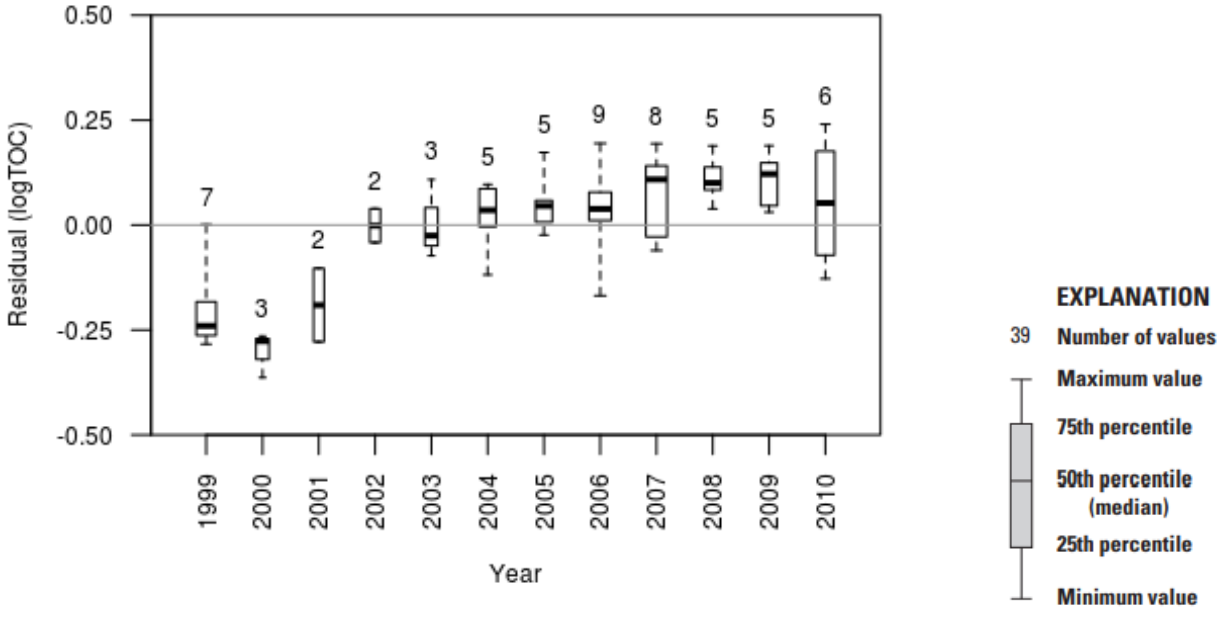




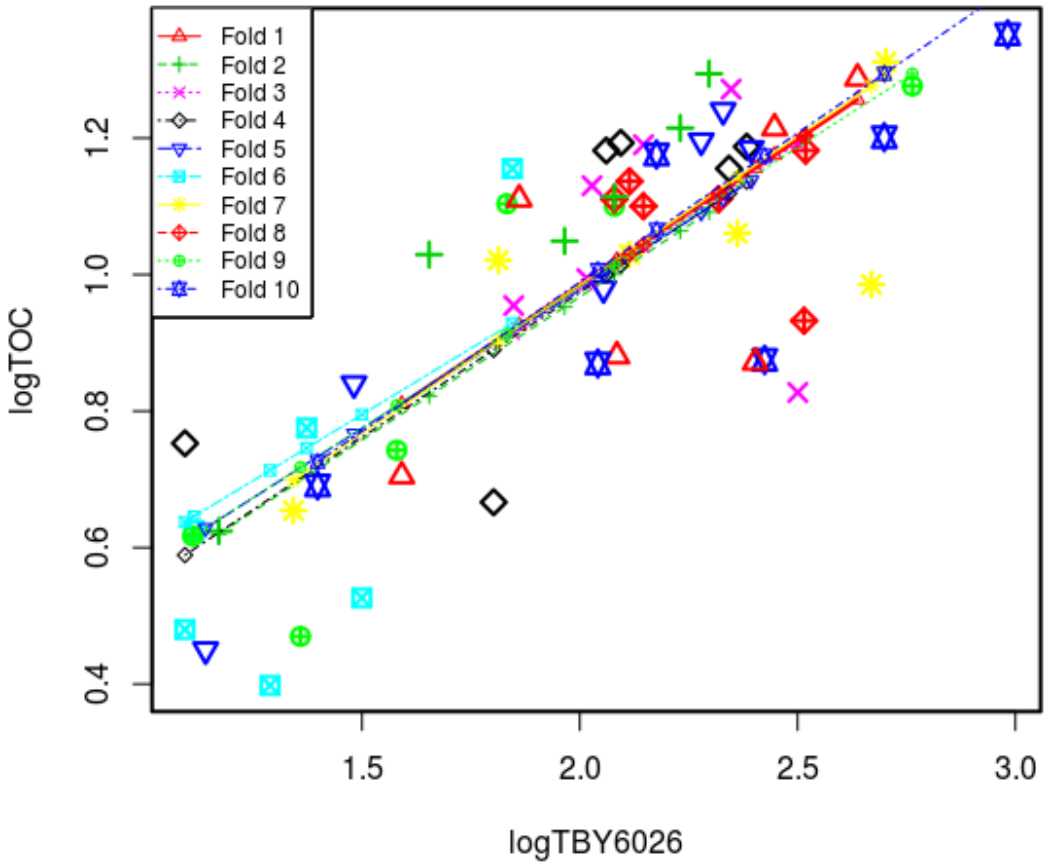





0.50
0.25
0.00
-0.25
-0.50
Residual (logTOC)
11
6
7
14
4
4
9
2
1
1
Jan
Mar
May
Jul
Sep
Dec
Month
30
20
10
0
TOC
60
60
Computed
Observed
EXPLANATION
39
Number of values
Maximum value
75th percentile
50th percentile
(median)
25th percentile
Minimum value


0.50
0.25
0.00
-0.25
-0.50
Residual (logTOC)
7
3
2
2
3
5
5
9
8
5
5
6
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
Year
EXPLANATION
39
Number of values
Maximum value
75th percentile
50th percentile
(median)
25th percentile
Minimum value

Cross Validation



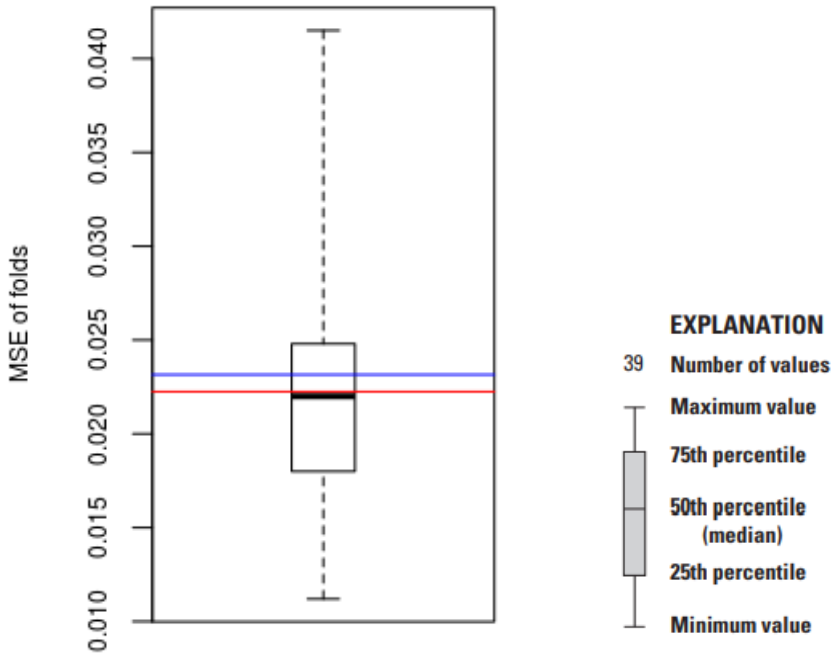

```
Fold – equal partition of the data (10 percent of the data)

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:  0.0112
                 Mean MSE of folds:  0.0232
               Median MSE of folds:  0.0220
              Maximum MSE of folds:  0.0415
 (Mean MSE of folds) / (Model MSE):  1.0400
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Model-Calibration Data Set

	Date	logTOC	logTBY6026	TOC	TBY6026	Computed logTOC	Computed TOC	Residual	Normal Quantiles	Censored Values
1	1999-04-16	1.03	2.11	10.7	130	1.03	11.2	0.00126	-0.188	--
2	1999-06-10	0.526	1.5	3.36	31.6	0.77	6.2	-0.243	-1.23	--
3	1999-06-25	0.881	2.09	7.6	122	1.02	10.9	-0.135	-0.93	--
4	1999-07-02	0.875	2.42	7.5	266	1.16	15.2	-0.284	-1.93	--
5	1999-07-29	0.667	1.8	4.64	63.4	0.897	8.31	-0.23	-1.07	--
6	1999-09-22	0.47	1.36	2.95	22.8	0.71	5.41	-0.24	-1.14	--
7	1999-12-02	0.398	1.29	2.5	19.4	0.681	5.05	-0.283	-1.71	--
8	2000-02-25	0.985	2.67	9.67	467	1.26	19.3	-0.277	-1.43	--
9	2000-03-24	0.932	2.51	8.56	327	1.2	16.6	-0.265	-1.32	--
10	2000-10-26	0.827	2.5	6.72	316	1.19	16.4	-0.364	-2.33	--
11	2001-06-06	0.872	2.4	7.44	253	1.15	14.9	-0.278	-1.56	--
12	2001-09-04	0.705	1.59	5.07	39	0.808	6.77	-0.103	-0.752	--
13	2002-06-12	1.35	2.98	22.5	960	1.39	26.1	-0.0417	-0.499	--
14	2002-08-14	1.29	2.64	19.4	434	1.25	18.7	0.039	0.146	--
15	2003-03-18	1.2	2.7	15.9	500	1.27	19.8	-0.0732	-0.698	--
16	2003-03-19	1.28	2.76	18.9	580	1.3	21.1	-0.0253	-0.406	--
17	2003-04-21	1.14	2.11	13.7	130	1.03	11.2	0.108	0.646	--
18	2004-03-05	1.31	2.7	20.5	504	1.28	19.9	0.0356	0.0208	--
19	2004-05-14	1.11	2.32	12.9	208	1.11	13.7	-0.00392	-0.231	--
20	2004-06-14	1.1	2.08	12.6	120	1.01	10.9	0.0867	0.452	--
21	2004-06-14	1.11	2.08	12.9	120	1.01	10.9	0.097	0.499	--
22	2004-09-08	0.48	1.09	3.02	12.4	0.599	4.18	-0.119	-0.809	--
23	2005-03-24	0.994	2.02	9.87	104	0.987	10.2	0.0073	-0.146	--
24	2005-05-16	0.98	2.05	9.54	113	1	10.6	-0.0236	-0.362	--
25	2005-06-10	1.19	2.38	15.4	242	1.14	14.6	0.0459	0.188	--
26	2005-06-13	1.19	2.09	15.6	124	1.02	11	0.173	1.23	--
27	2005-08-29	1.1	2.15	12.6	140	1.04	11.6	0.0586	0.274	--

28	2006-03-02	0.45	1.14	2.82	13.8	0.619	4.38	-0.168	-0.997	--
29	2006-03-22	0.955	1.85	9.01	70.5	0.916	8.69	0.0384	0.0625	--
30	2006-05-01	0.775	1.37	5.96	23.6	0.716	5.48	0.059	0.318	--
31	2006-05-12	1.03	1.65	10.7	45.1	0.835	7.2	0.195	1.93	--
32	2006-06-05	0.84	1.48	6.92	30.3	0.762	6.09	0.0782	0.362	--
33	2006-07-31	0.753	1.09	5.66	12.4	0.599	4.18	0.154	1.14	--
34	2006-09-07	0.691	1.4	4.91	25	0.727	5.62	-0.0357	-0.452	--
35	2006-09-21	0.625	1.11	4.22	13	0.607	4.27	0.0181	-0.0625	--
36	2006-09-21	0.617	1.11	4.14	12.9	0.606	4.25	0.011	-0.104	--
37	2007-01-09	0.654	1.34	4.51	22	0.703	5.32	-0.0492	-0.547	--
38	2007-03-22	0.743	1.58	5.53	38	0.803	6.7	-0.0606	-0.596	--
39	2007-03-26	1.2	2.28	15.7	190	1.1	13.2	0.0982	0.547	--
40	2007-03-31	1.27	2.35	18.7	222	1.13	14.1	0.145	0.997	--
41	2007-04-16	1.02	1.81	10.5	65	0.901	8.4	0.12	0.698	--
42	2007-05-07	1.21	2.23	16.4	170	1.08	12.6	0.138	0.868	--
43	2007-06-29	1.1	1.83	12.7	68	0.91	8.56	0.194	1.71	--
44	2007-09-04	0.624	1.17	4.21	14.8	0.631	4.51	-0.00701	-0.274	--
45	2008-04-24	1.11	2.08	13	120	1.01	10.9	0.1	0.596	--
46	2008-05-09	1.18	2.39	15.3	247	1.15	14.8	0.0387	0.104	--
47	2008-06-19	1.05	1.96	11.2	92.3	0.966	9.74	0.0837	0.406	--
48	2008-09-15	1.11	1.86	12.9	72.6	0.922	8.8	0.189	1.43	--
49	2008-10-16	1.13	2.03	13.5	107	0.992	10.3	0.138	0.93	--
50	2009-03-31	1.24	2.33	17.4	213	1.12	13.9	0.122	0.809	--
51	2009-04-27	1.29	2.3	19.7	198	1.11	13.4	0.189	1.56	--
52	2009-06-17	1.19	2.15	15.5	140	1.04	11.6	0.149	1.07	--
53	2009-08-20	1.16	2.34	14.3	220	1.12	14	0.0309	-0.0208	--
54	2009-09-10	1.21	2.45	16.4	280	1.17	15.5	0.0463	0.231	--
55	2010-04-23	0.87	2.04	7.41	110	0.998	10.5	-0.128	-0.868	--
56	2010-05-27	1.18	2.18	15	150	1.05	11.9	0.122	0.752	--
57	2010-06-14	1.18	2.06	15.2	115	1.01	10.7	0.176	1.32	--
58	2010-06-16	1.16	1.85	14.3	70	0.915	8.67	0.24	2.33	--
59	2010-07-06	1.06	2.36	11.5	230	1.13	14.3	-0.0719	-0.646	--
60	2010-08-25	1.18	2.52	15.2	330	1.2	16.7	-0.0168	-0.318	--

Definitions

```
TOC: organic carbon, total, in milligrams per liter (00680)
TBY6026: turbidity, YSI model 6026, in formazin nephelometric units (63680)
```

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.]

YSI Incorporated, 2007, YSI 6136 turbidity sensor: YSI Incorporated, 2 p., accessed November 2019 at https://www.ysi.com/File%20Library/Documents/Specification%20Sheets/E56-6136-Turbidity-Sensor.pdf.

YSI Incorporated, 2012, 6-series multiparameter water quality sondes—User manual, revision J: YSI Incorporated, 379 p., accessed November 2019 at https://www.ysi.com/File%20Library/Documents/Manuals/069300-YSI-6-Series-Manual-RevJ.pdf.