Appendix 4. Model Archive Summary for Total Carbon at U.S. Geological Survey Station 06856600, Republican River at Clay Center, Kansas, during July 2018 through March 2021

This model archive summary summarizes the total carbon (totC) model developed to compute the 15-minute totC concentration from July 24, 2018, onward. This is the first model computing totC concentration that has been developed for this site. Model development was completed in accordance with Rasmussen and others (2009) and 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

Station number: 06856600

Site name: Republican River at Clay Center, Kansas

Location: Latitude 39°21'20", longitude 97°07'38", referenced to North American Datum of 1927, in Clay County, Kansas, hydrologic unit code 10250017.

Water-quality equipment: A YSI, Inc., EXO3 water-quality monitor (YSI, Inc., 2017) equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, and turbidity was installed July 24, 2018. The EXO3 was housed in a 4-inch-diameter metal pipe suspended from the downstream side of the bridge in the deepest and fastest flowing water. Measurements from the EXO3 were recorded every 15 minutes and transmitted hourly via satellite.

Date model was created: August 23, 2021

Model calibration data period: August 1, 2018, through March 25, 2021 (dataset consisted of 42 discrete water-quality samples collected).

Model application date: July 24, 2018, onward (date of continuous water-quality monitor installation).

Total Carbon Concentration Data

Equal-width-increment samples (as described in U.S. Geological Survey, 2006) were collected from the downstream side of the bridge using a Federal Interagency Sedimentation Project US DH–95 depth-integrated sampler with a Teflon bottle, cap, and nozzle and a manually operated reel. Subsamples from 10 equally spaced verticals were composited. During July 2018 through June 2020, discrete water-quality samples were collected at Clay Center biweekly during May through October and monthly during November through April. During July 2020 through March 2021, discrete water-quality samples were collected about monthly. Samples were analyzed at the U.S. Geological Survey (USGS) National Water Quality Laboratory in Lakewood, Colorado, in accordance with standard methods (U.S. Environmental Protection Agency, 2017).

Continuous Water-Quality Data

Continuously measured (15-minute) water-quality data collection at Clay Center began on July 24, 2018. During July 2018 through March 2021, a YSI, Inc., EXO3 multiparameter sonde measured water temperature, specific conductance, pH, dissolved oxygen, and turbidity (YSI, Inc., 2017). The water-quality monitor was suspended from the bridge deck of Kansas Highway 18 near the centroid of flow. The continuous water-quality monitor was operated and maintained according to standard USGS methods (Wagner and others, 2006; Bennett and others, 2014). All continuous water-quality data at Clay Center are available in near-real time (updated hourly) from the USGS National Water Information System database (U.S. Geological Survey, 2021) using the station number 06856600.

Model Data

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 (U.S. Geological Survey, 2021). The regression model is based on 42 concurrent measurements of totC and continuously measured turbidity collected from August 1, 2018, through March 25, 2021. The potential explanatory variables were interpolated from the continuous record and paired based on discrete sample

collection time. Daylight saving time was observed, so sample time was either central standard time or central daylight time. The maximum time span between two continuous data points used for interpolation was 5 hours. Samples were collected throughout the range of continuously observed hydrologic conditions. No samples had concentrations that were less than the laboratory reporting limit of 0.10 milligram per liter (mg/L). Summary statistics and the complete model-calibration dataset are provided below.

Potential outliers were identified as the data points for which both the studentized residual was greater than 3 or less than −3 and the Cook's D value exceeded the outlier test criteria, as described by Helsel and others (2020). This methodology resulted in one potential outlier: the sample collected on March 25, 2021. To investigate this potential outlier, the sample collection information sheets and laboratory reports were reviewed to check for data entry errors. Then, the linear relationship of totC versus turbidity at this site was examined. Because no data entry issues were discovered and the totC versus turbidity linear relationship was consistent with other samples collected at this site, the March 25, 2021, result was retained. No potential outliers were removed from the model calibration dataset.

Model Development

All continuously measured water-quality parameters and streamflow were considered as explanatory variables for estimating totC concentration using ordinary least squares regression. All models that predict totC and logarithm base 10 totC ($\log_{10}$[totC]) were evaluated from simple linear regression (SLR; single explanatory variable) and multiple linear regression (more than one explanatory variable) models. Potential regression models were evaluated based on normality and homoscedasticity in residual values. Residual values are the difference between the measured and predicted values. Homoscedastic plots are those in which the magnitude of residual values does not change substantially over the range of predicted values; that is, the magnitude of residual values neither increases nor decreases over the range of predicted values and the variance is constant. These comparisons led to the conclusion that the most appropriate and reliable model would be the SLR that estimated non-transformed totC.

Turbidity was selected as the best predictor of totC in an untransformed SLR model based on residual plots, fairly high adjusted coefficient of determination (R^2_{adj}) compared to other models considered, and fairly low model standard percentage error compared to other models considered. Additional explanatory variables in multiple linear regression models did not substantially increase the R^2_{adj} value. Values for the aforementioned statistics and metrics were computed and are included below along with all relevant sample data and more indepth statistical information.

Model Summary

Summary of final regression analysis for totC at U.S. Geological Survey station 06856600.

TotC concentration model:

$$totC = 0.0566 \times TBY + 2.39,$$

where

totC = total carbon, in milligrams per liter, and

TBY = turbidity, YSI EXO3, in formazin nephelometric units.

Turbidity makes physical and statistical sense as an explanatory variable for totC. Turbidity makes sense physically because suspended solids (including some with attached carbon) in the water column scatter light and increase turbidity. The model selected was the simplest model (one explanatory variable) and the best statistically.

Total Carbon Concentration Record

The totC concentration record is computed using this regression model and stored at the National Real-Time Water Quality website. Model-estimated concentrations in the National Real-Time Water Quality website are computed at hourly intervals. The complete water-quality record is available at https://nrtwq.usgs.gov/ks.

Remarks

All regression models were developed using R software environment (R Core Team, 2021).

Computed by: Brianna Leiker

Reviewed by: Dawn McCausland (USGS Ohio-Kentucky-Indiana Water Science Center) and Tara Morgan-King (USGS California Water Science Center)

Model Statistics, Data, and Plots

Definitions for terms used in this output are provided at the end of this document.

Model

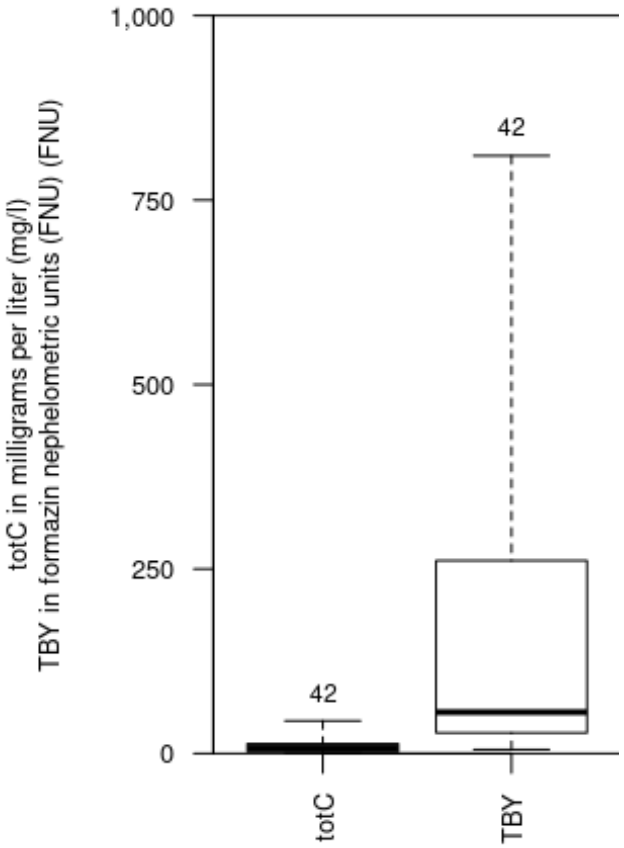
totC=+0.0566×TBY+2.39

Variable Summary Statistics

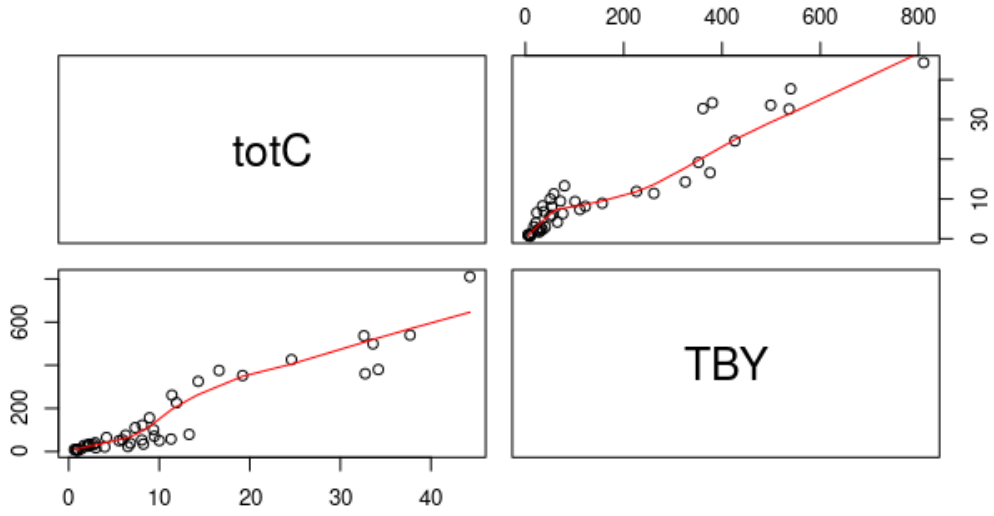
	totC	TBY
Minimum	0.62	5.47
First Quartile	2.77	28.00
Median	7.70	56.00
Mean	11.20	155.00
Third Quartile	13.30	261.00
Maximum	44.30	810.00

TotC cannot be extrapolated more than 10 percent outside the range of sample data used to fit the model (U.S. Geological Survey, 2016). Because the maximum observed continuous turbidity value in the calibration dataset was 810 formazin nephelometric units (FNU), the maximum turbidity value for which this model is valid is 891 FNU. Less than 1 percent (59 of 84,392 15-minute measurements) of continuous turbidity measurements during the study period exceeded 891 FNU. At a turbidity of 891 FNU, the model-estimated totC value is 52.8 mg/L.

Box Plots

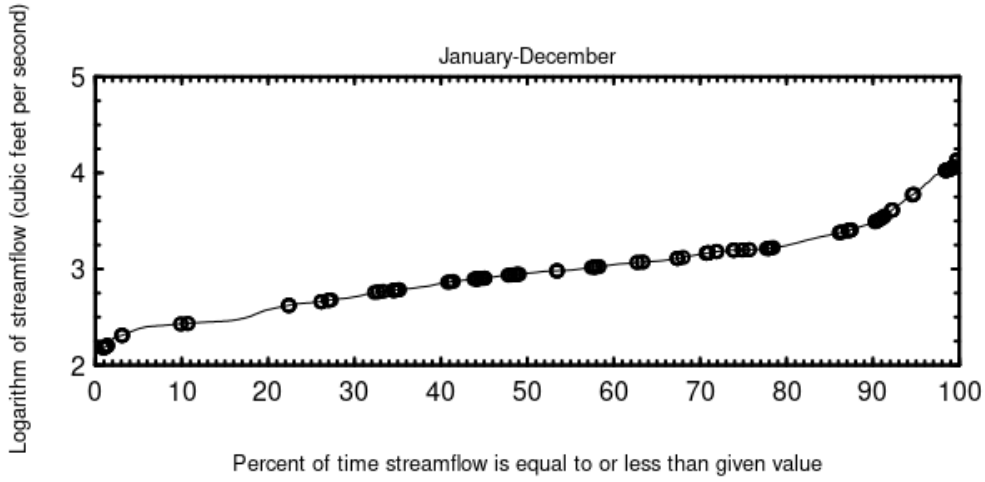


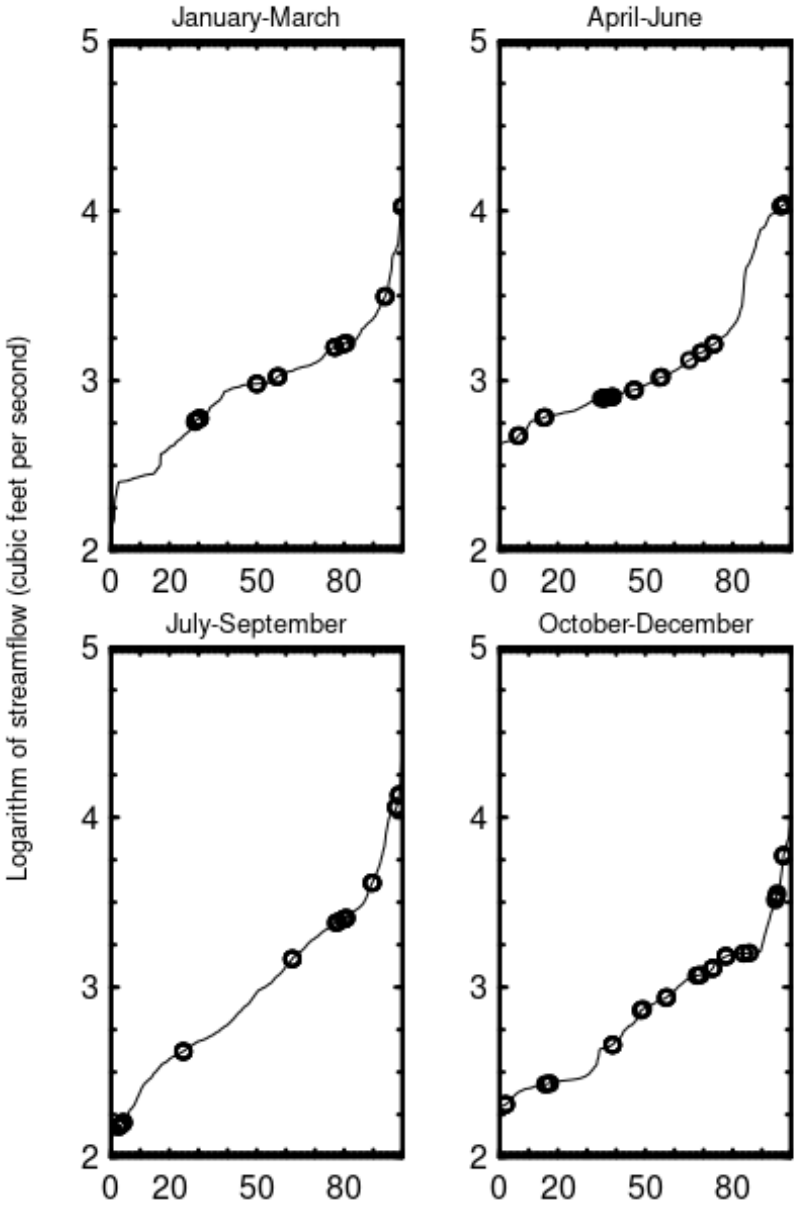
Exploratory Plots



Red lines show the locally weighted scatterplot smoothing (Cleveland, 1979; Helsel and others, 2020).

The x- and y-axis labels for a given bivariate plot are defined by the intersecting row and column labels.


January-December
Logarithm of streamflow (cubic feet per second)
5
4
3
2
0
10
20
30
40
50
60
70
80
90
100
Percent of time streamflow is equal to or less than given value


January-March
April-June
July-September
October-December
Logarithm of streamflow (cubic feet per second)
5
4
3
2
0
20
50
80
Percent of time streamflow is equal to or less than given value

Basic Model Statistics

Number of observations	42
Standard error (root mean square error [RMSE])	3.87
Average model standard percentage error (MSPE)	34.6
Coefficient of determination (R^2)	0.891
Adjusted coefficient of determination (R^2_{adj})	0.888

Explanatory Variables

	Coefficients	Standard Error	t value	$Pr(>\|t\|)$
(Intercept)	2.3900	0.77000	3.1	3.55e−03
TBY	0.0566	0.00313	18.1	7.54e−21

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	−0.631
E.vars	−0.631	1.000

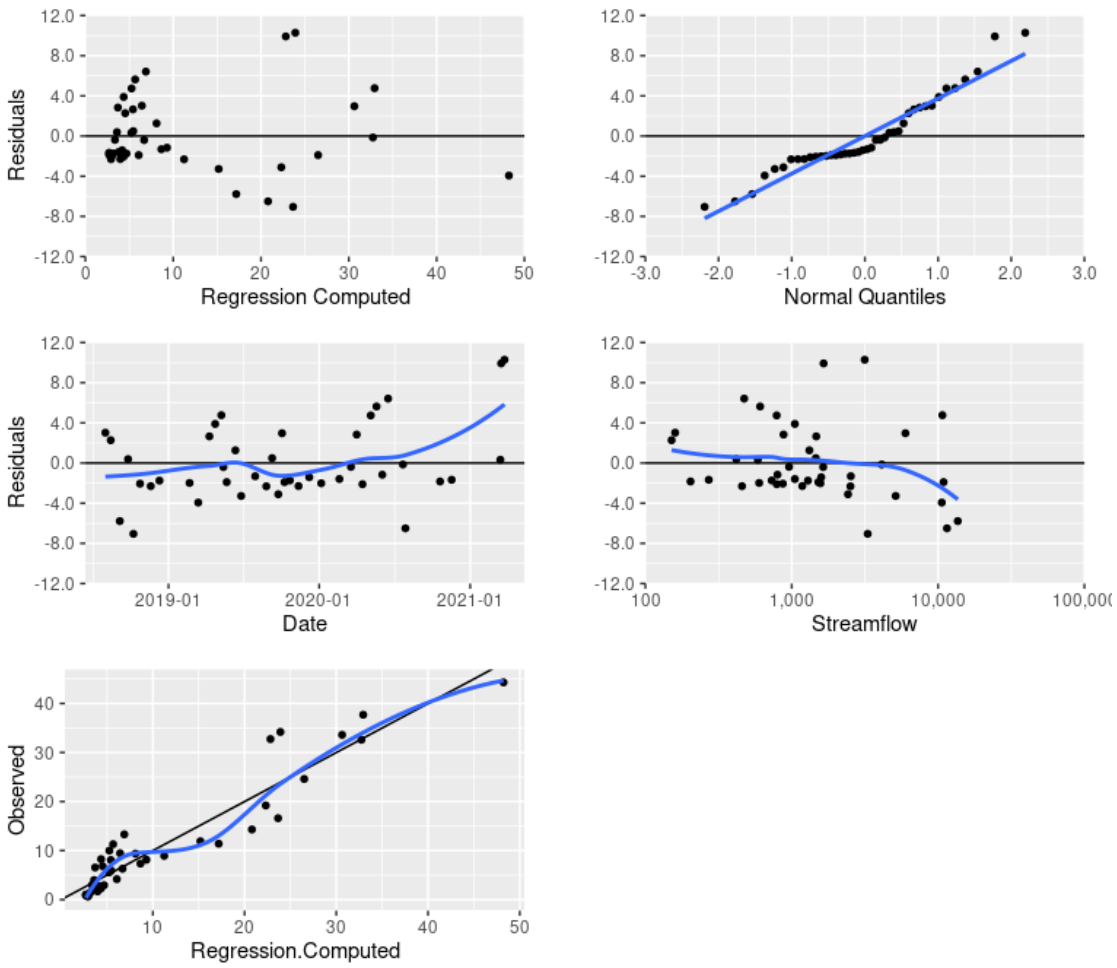
Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.143	0.194	0.436

Flagged Observations

Date	totC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
10/8/2018 13:10	16.6	23.6	−7.05	−1.87	−1.94	0.0556	0.104	−0.470
3/14/2019 12:10	44.3	48.2	−3.93	−1.22	−1.23	0.3050	0.326	−0.812
5/9/2019 10:20	37.7	32.9	4.77	1.31	1.33	0.1210	0.118	0.491
3/17/2021 12:30	32.7	22.8	9.92	2.63	2.86	0.0515	0.188	0.666
3/25/2021 13:40	34.2	23.9	10.30	2.74	3.00	0.0570	0.226	0.737

Statistical Plots



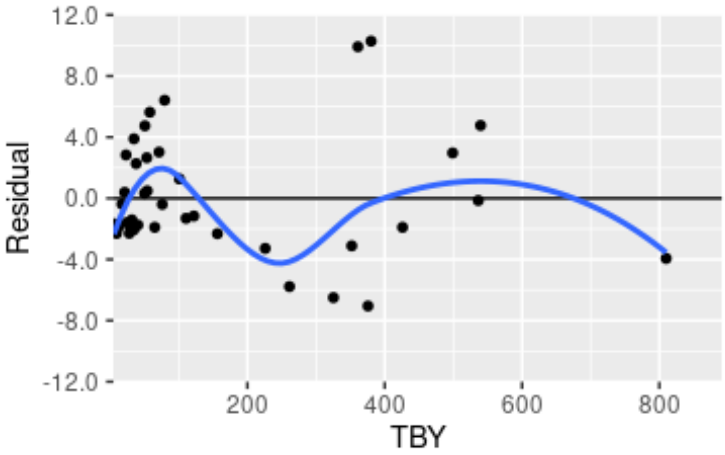
First row (left): residual totC (in milligrams per liter) related to regression-computed totC (in milligrams per liter).

First row (right): residual totC (in milligrams per liter) related to the corresponding normal quantile (unitless) of the residual with simple linear regression indicated by the blue line.

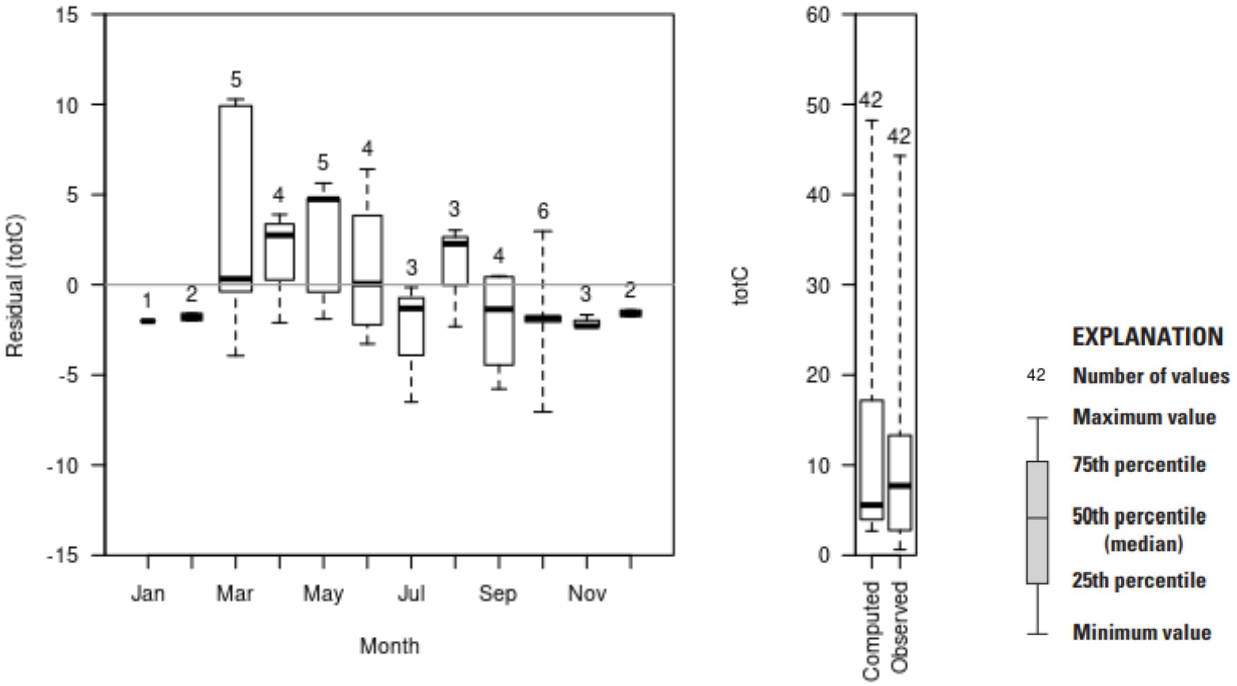
Second row (left): residual totC (in milligrams per liter) related to date with locally estimated scatterplot smoothing (LOESS) indicated by the blue line.

Second row (right): residual totC (in milligrams per liter) related to streamflow (in cubic feet per second) with LOESS indicated by the blue line.

Third row: observed totC (in milligrams per liter) related to regression-computed totC (in milligrams per liter) with LOESS indicated by the blue line.

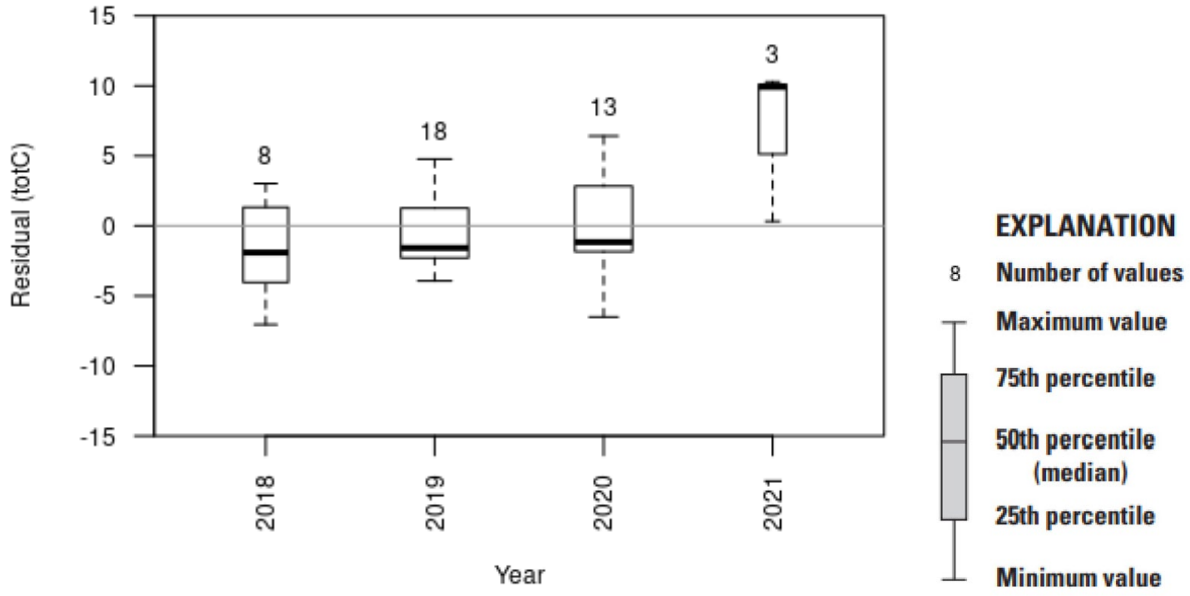


Residual totC (in milligrams per liter) related to TBY (in formazin nephelometric units) with LOESS indicated by the blue line



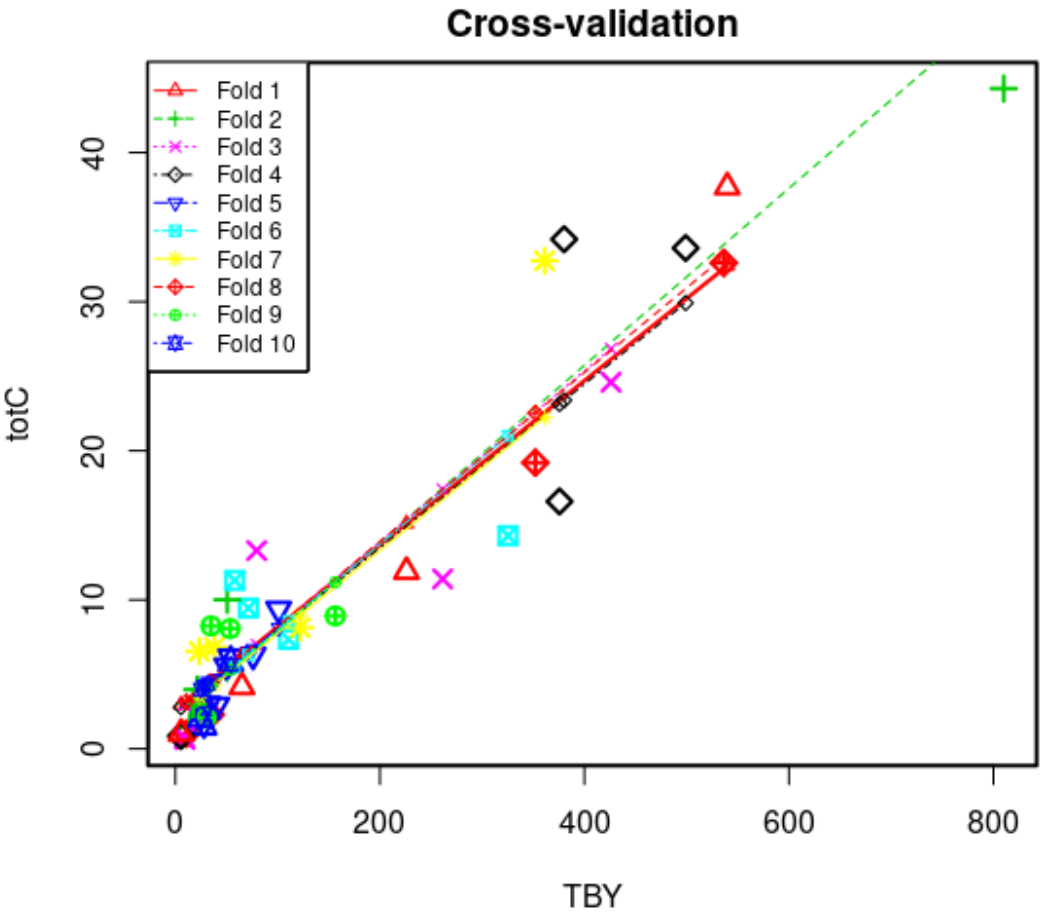
Left: residual totC (in milligrams per liter) by month.

Right: totC (in milligrams per liter) in regression-computed and observed values.



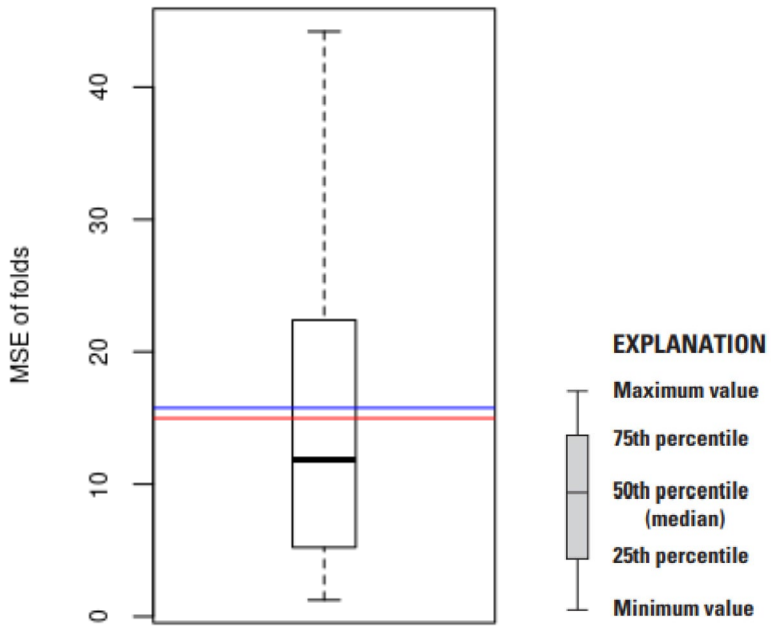
Residual totC (in milligrams per liter) by year.

Cross Validation



TotC (in milligrams per liter) related to TBY (in formazin nephelometric units) for 10 folds of data. Each fold is an equal partition of the data (10 percent of the data). Large symbols are the observed values of data points removed in a fold and small symbols are the recomputed values of data points removed in a fold. Recomputed regression lines are adjusted regression lines with one fold removed.

```
             Minimum MSE of folds:   1.24
                Mean MSE of folds:  15.80
              Median MSE of folds:  11.80
             Maximum MSE of folds:  44.20
(Mean MSE of folds) / (Model MSE):   1.05     (MSE ratio)
```




```
Red line - Model MSE (unitless)
Blue line - Mean MSE of folds (unitless)
```

Model-Calibration Dataset

	Date	totC	TBY	Computed totC	Residual	Normal Quantiles	Censored Values
1	2018-08-01	9.45	71.4	6.43	3.02	0.915	--
2	2018-08-14	6.81	38	4.54	2.27	0.598	--
3	2018-09-05	11.4	261	17.2	-5.78	-1.54	--
4	2018-09-25	3.98	21.3	3.59	0.388	0.396	--
5	2018-10-08	16.6	376	23.6	-7.05	-2.19	--
6	2018-10-24	2.29	34.6	4.34	-2.05	-0.671	--
7	2018-11-19	0.62	9.43	2.92	-2.3	-0.915	--
8	2018-12-10	2.93	40.5	4.68	-1.75	-0.271	--
9	2019-02-21	1.01	10.7	2.99	-1.98	-0.528	--
10	2019-03-14	44.3	810	48.2	-3.93	-1.37	--
11	2019-04-10	8.07	53.5	5.41	2.66	0.671	--

12	2019-04-24	8.25	34.8	4.36	3.89	1.01	--
13	2019-05-09	37.7	540	32.9	4.77	1.23	--
14	2019-05-14	6.29	76	6.69	−0.398	0.149	--
15	2019-05-22	24.6	426	26.5	−1.9	−0.461	--
16	2019-06-12	9.37	101	8.1	1.27	0.528	--
17	2019-06-26	11.9	226	15.2	−3.28	−1.23	--
18	2019-07-30	7.34	111	8.65	−1.31	0.0297	--
19	2019-08-26	8.92	156	11.2	−2.31	−1.01	--
20	2019-09-09	5.94	54.2	5.45	0.488	0.461	--
21	2019-09-24	19.2	352	22.3	−3.11	−1.11	--
22	2019-10-03	33.6	499	30.6	2.97	0.828	--
23	2019-10-09	4.17	65.1	6.07	−1.9	−0.396	--
24	2019-10-22	1.45	14.1	3.18	−1.73	−0.209	--
25	2019-11-12	1.68	28	3.97	−2.29	−0.828	--
26	2019-12-08	2.77	31.7	4.18	−1.41	−0.0297	--
27	2020-01-06	2.07	30	4.08	−2.01	−0.598	--
28	2020-02-19	2.19	24.6	3.78	−1.59	−0.0892	--
29	2020-03-18	2.99	17.3	3.37	−0.377	0.209	--
30	2020-04-01	6.54	23.3	3.7	2.84	0.747	--
31	2020-04-15	0.78	8.83	2.89	−2.11	−0.747	--
32	2020-05-05	10	50.7	5.26	4.74	1.11	--
33	2020-05-19	11.3	57.9	5.67	5.63	1.37	--
34	2020-06-02	8.13	122	9.29	−1.16	0.0892	--
35	2020-06-16	13.3	79.5	6.88	6.42	1.54	--
36	2020-07-22	32.6	536	32.7	−0.144	0.271	--
37	2020-07-28	14.3	325	20.8	−6.5	−1.78	--
38	2020-10-20	0.86	5.47	2.7	−1.84	−0.333	--
39	2020-11-17	1.03	5.5	2.7	−1.67	−0.149	--
40	2021-03-15	5.56	50.3	5.23	0.327	0.333	--
41	2021-03-17	32.7	361	22.8	9.92	1.78	--
42	2021-03-25	34.2	380	23.9	10.3	2.19	--

Definitions

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: a data point’s distance from the middle (mean) value in the x direction (Helsel and others, 2020).

LOESS: local polynomial regression fitting, or locally estimated scatterplot smoothing (Helsel and others, 2020).

MSE: model standard error, also known as standard error of the regression (Helsel and others, 2020).

MSPE: model standard percentage error (Helsel and others, 2020).

Pr(>|*t*|): 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).

```
t value: Student’s t value; the coefficient divided by its associated standard error (Hel
sel and others, 2020).

TBY: Turbidity, water, unfiltered, monochrome near infrared light-emitting diode, 780–900
nm, detection angle 90±2.5 degrees, in formazin nephelometric units (U.S. Geological Surv
ey parameter code 63680) (U.S. Geological Survey method code TS213).

TotC: Total carbon; carbon [inorganic plus organic], suspended sediment, total, in millig
ram per liter (U.S. Geological Survey parameter code 00694) (U.S. Geological Survey metho
d code COMB6).
```

App Version 1.0

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 June 17, 2021 at https://doi.org/10.3133/ofr20141151.

Cleveland, W.S., 1979, Robust locally weighted regression and smoothing scatterplots: Journal of the American Statistical Association, v. 74, no. 368, p. 829-836.

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, chapter A3, 458 p., https://doi.org/10.3133/tm4a3. [Supersedes USGS Techniques of Water-Resources Investigations, book 4, chapter A3, version 1.1.]

R Core Team, 2021, R—A language and environment for statistical computing, version 4.0.3 (Bunny-Wunnies Freak Out): Vienna, Austria, R Foundation for Statistical Computing, accessed June 17, 2021 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 concentration 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://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf].

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, 166 p., accessed June 17, 2021 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 June 17, 2021, at https://water.usgs.gov/water-resources/memos/memo.php?id=467.

U.S. Geological Survey, 2021, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed June 8, 2021, at https://doi.org/10.5066/F7P55KJN.

U.S. Environmental Protection Agency, 2017, Method 440.0: Determination of carbon and nitrogen in sediments and particulates of estuarine/coastal waters using elemental analysis, revision 1.4, 11 p.

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, accessed June 17, 2021 at https://doi.org/10.3133/tm1D3. [Supersedes USGS Water-Resources Investigations Report 2000–4252.]

YSI, Inc., 2017, EXO user manual—Advanced water quality monitoring platform (rev. G): YSI, Inc., 154 p., accessed December 21, 2019, at https://www.ysi.com/file%20library/documents/manuals/exo-user-manual-web.pdf.