Appendix 7. Model Archive Summary for Suspended-Sediment Concentration at U.S. Geological Survey Site 07182390, Neosho River at Neosho Rapids, Kansas, during November 13, 2015, through December 31, 2019

This model archive summary summarizes the suspended-sediment concentration (SSC) model developed to compute hourly or daily SSC during November 13, 2015, through December 19, 2019. 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, chapter C4, and the policy and guidance for approval of surrogate regression models for computation of time series SSCs and loads (Rasmussen and others, 2009; U.S. Geological Survey, 2016).

Site and Model Information

Site number: 07182390
Site name: Neosho River near Neosho Rapids, Kansas
Location: Lat 38°22'05", long 96°00'00" referenced to North American Datum of 1927, in SW 1/4 NW 1/4 sec.29, T.19 S., R.13 E., Lyon County, Kans., hydrologic unit 11070201, on right upstream side of bridge, 0.75 mile west of the intersection of Kansas Highway 130 and South Street at Neosho Rapids, and at mile 370.7.

Equipment: A YSI EXO2 water-quality monitor equipped with sensors for water temperature, specific conductance, and turbidity. The YSI EXO2 water-quality monitor recorded readings every 15 minutes and transmitted hourly via satellite using a Sutron Satlink 2 high data rate collection platform. The EXO2 water-quality monitor began operation during November 13, 2015.

Date model was created: January 16, 2020
Model calibration data period: December 1, 2015, through March 14, 2019

Model Data

All data were collected using approved USGS protocols (Wagner and others, 2006; Sauer and Turnipseed 2010; Turnipseed and Sauer, 2010; U.S. Geological Survey, variously dated) and are stored in the National Water Information System (NWIS) database (https://doi.org/10.5066/F7P55KJN) and is available for public use (U.S. Geological Survey, 2020). Explanatory variables were evaluated individually and in combination. Potential

explanatory variables included streamflow, water temperature, specific conductance and turbidity.

The regression model is based on 17 measurements of discretely collected SSC samples and continuously measured turbidity collected during November 13, 2015, through December 31, 2019. Samples were collected over a range of streamflow and turbidity conditions. No samples had concentrations below laboratory detection limits. Identification of potential outliers included any values that exceeded the Cook's D test (Cook, 1977) and any point for which the studentized residual was greater than 3 or less than −3. Two samples were removed from the dataset; the January 14, 2016, sample was impacted by ice and March 3, 2017, because no water quality data received from laboratory for comparison.

Suspended-Sediment Sampling Details

Discrete samples were collected from the downstream side of the bridge or instream within 350 feet of the bridge using equal-width-increment, multiple vertical, single vertical, or grab-dip 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 three to six samples per year with a Federal Interagency Sediment Project U.S. DH–95 or D–95 with a Teflon bottle, cap, and nozzle depth-integrating sampler, a D–96 bag sampler, or a DH–81 with a Teflon bottle, cap, and nozzle hand sampler. Samples were analyzed for SSC, loss on ignition, and occasionally five-point grain size by the USGS Sediment Laboratory in Iowa City, Iowa.

Continuous Data

Continuously monitored turbidity was measured using a YSI EXO turbidity sensor installed during November 13, 2015, through December 31, 2019 (U.S. Geological Survey, 2018). Concomitant turbidity values were time interpolated. If continuous data were not available (2 or more hours of turbidity values bracketing the sample collected 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 neither concomitant continuous data nor field monitor data were available, the sample was not included in the dataset. The range of continuous turbidity data of the YSI EXO2 sensor (in formazin nephelometric units) was as follows: maximum 1,190; minimum 1.80; mean 76.6; median 27.2.

Model Development

Ordinary least squares regression analysis was done using R programming language (R Core Team, 2019) to relate discretely collected SSC 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 calculated SSC were examined for homoscedasticity (departures from zero did not change substantially over the range of model calculated values).

Turbidity was selected as the best predictors of logarithm base 10 ($\log_{10}$) (SSC) based on residual plots, relatively high coefficient of determination (R^2), and relatively low model standard percentage error (MSPE).

Model Summary

Summary of SSC regression analysis at site 07182390:
SSC-based model:

$$\text{Log}_{10}(SSC) = 1.13 \times \text{Log}_{10}(TurbEXO) + 0.105$$

where

SSC = suspended-sediment concentration, in milligrams per liter, and
*Turb*EXO = turbidity, YSI model EXO, in formazin nephelometric units.

The use of turbidity as an explanatory variable is appropriate physically and statistically. In a physical sense, particles comprised of suspended solids scatter light which affects turbidity. In a statistical sense using turbidity resulted in a model with a low standard error and high R^2 values. The relation between turbidity and SSC can vary given varying concentrations of organic suspended particles that increase turbidity but are not included in the SSC analysis.

The log-transformed model may be retransformed to the original units to calculate SSC directly. A bias is introduced in the calculated constituent during retransformation and may be corrected using the Duan's bias correction factor (BCF; Duan, 1983). The calculated BCF is 1.02 for this model and the formula for the retransformed model accounting for BCF is as follows:

$$SSC = 1.30 \times TurbEXO^{1.13}$$

Suspended-Sediment Concentration Record

The SSC record that is being used in this regression model is stored at the National Real-Time Water Quality (NRTWQ) website (https://nrtwq.usgs.gov/ks).

Previously Published Model

No previously published model

Model Statistics, Data, and Plots

Model

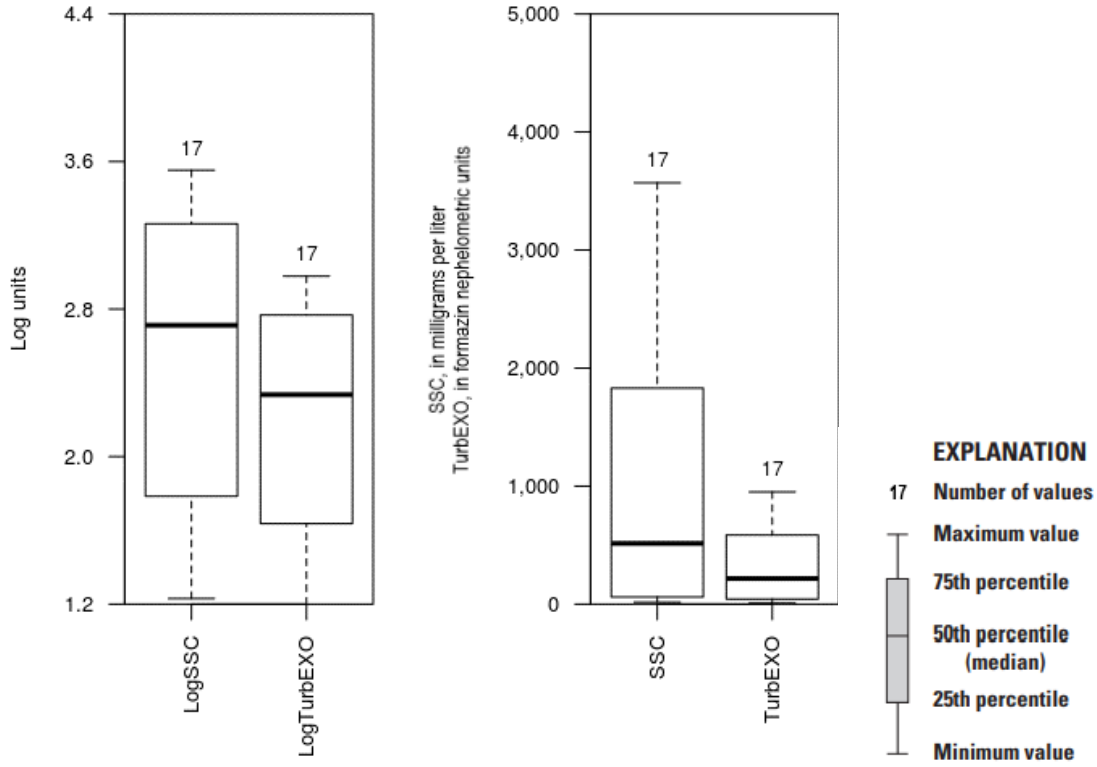
$$\text{Log}(SSC) = + 1.13 * \text{Log}(Turb\text{EXO}) + 0.105$$

Variable Summary Statistics

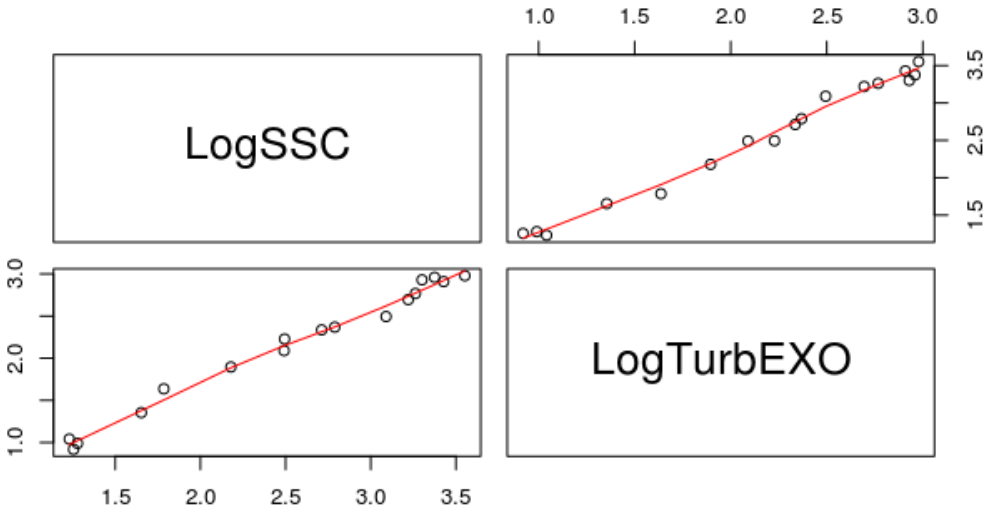
	LogSSC	SSC	LogTurbEXO	TurbEXO
Minimum	1.23	17	0.919	8.3

1st Quartile	1.79	61	1.640	43.3
Median	2.71	515	2.340	217.0
Mean	2.54	1020	2.150	343.0
3d Quartile	3.26	1830	2.770	586.0
Maximum	3.55	3570	2.980	951.0

Box Plots



Exploratory Plots



Basic Model Statistics

Number of Observations	17
Standard error (RMSE)	0.0931
Average Model standard percentage error (MSPE)	21.6
Coefficient of determination (R^2)	0.988
Adjusted Coefficient of Determination (Adj. R^2)	0.987
Bias Correction Factor (BCF)	1.02

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.105	0.0725	1.45	1.67e-01
LogTurbEXO	1.130	0.0320	35.30	7.53e-16

Correlation Matrix

	Intercept	E.vars
Intercept	1.00	-0.95
E.vars	-0.95	1.00

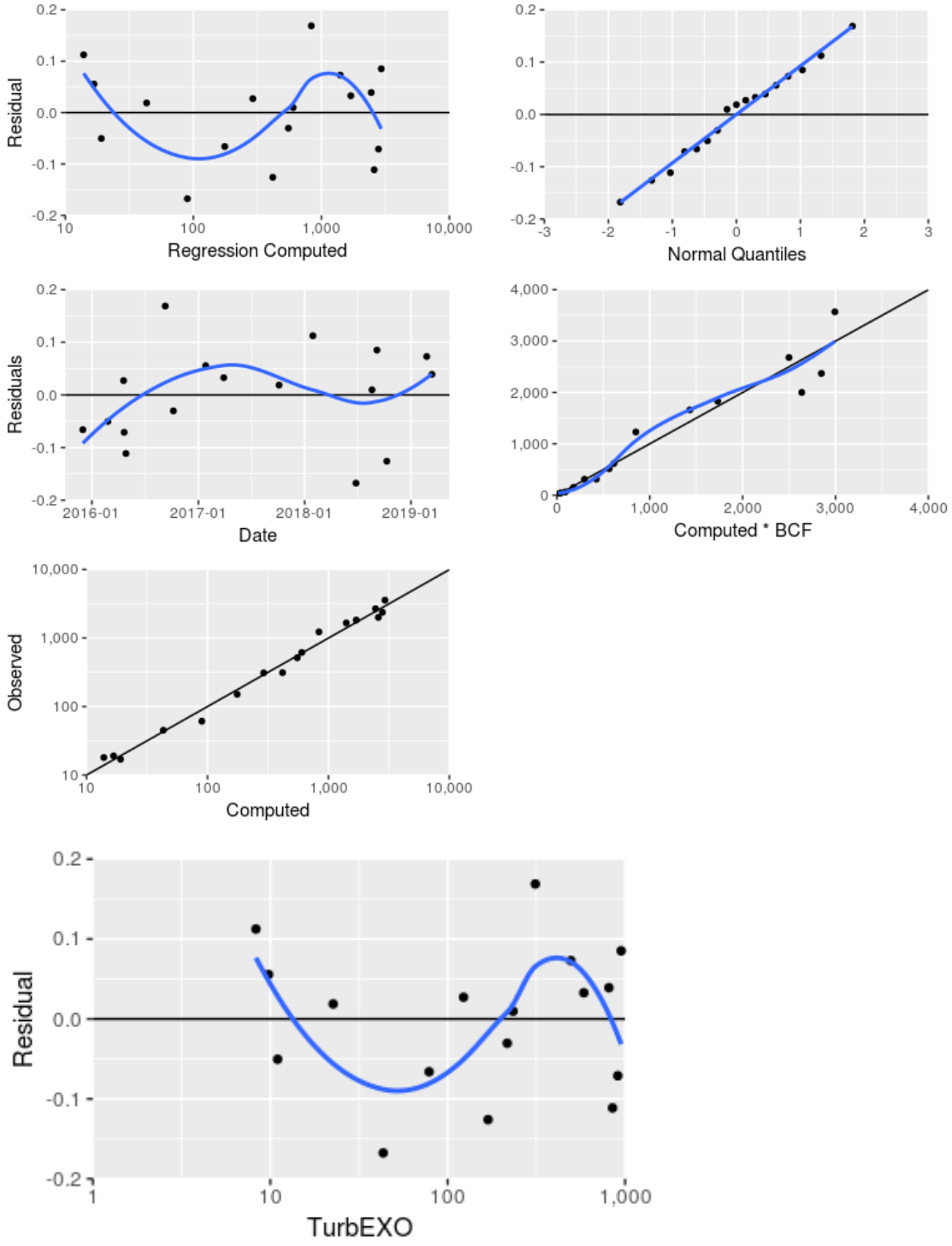
Outlier Test Criteria

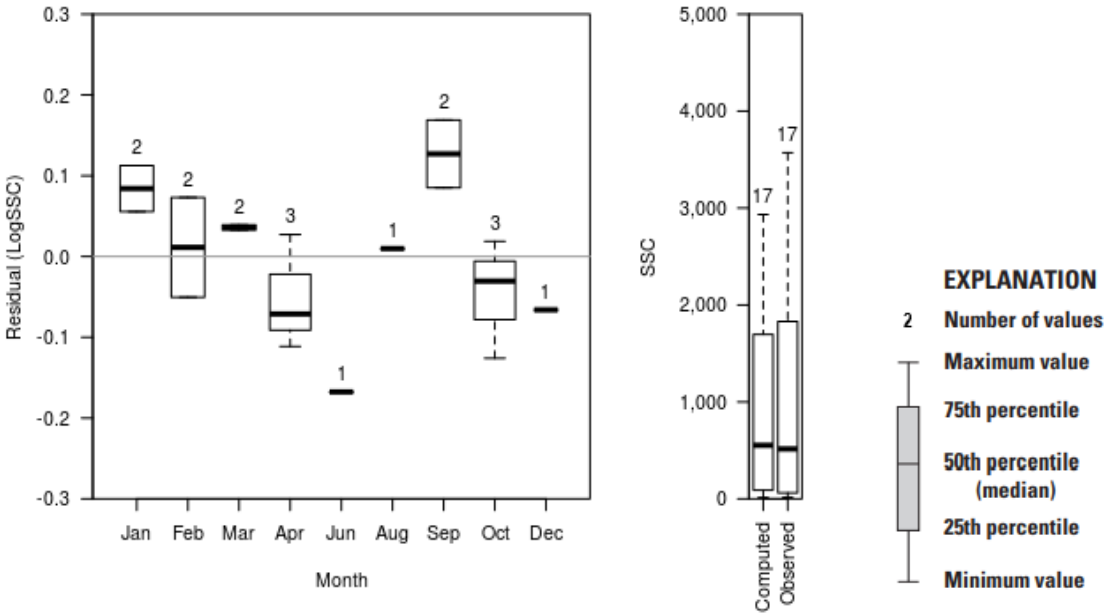
Leverage	Cook's D	DFFITS
0.353	0.192	0.686

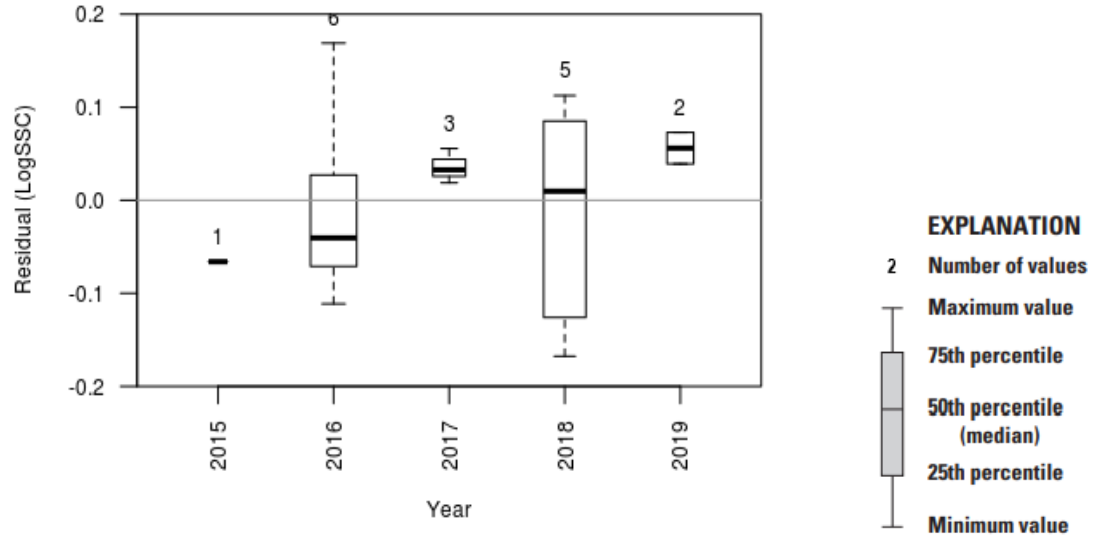
Flagged Observations

	LogSSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
1/29/2018 12:40	1.26	1.14	0.113	1.39	1.43	0.239	0.301	0.802

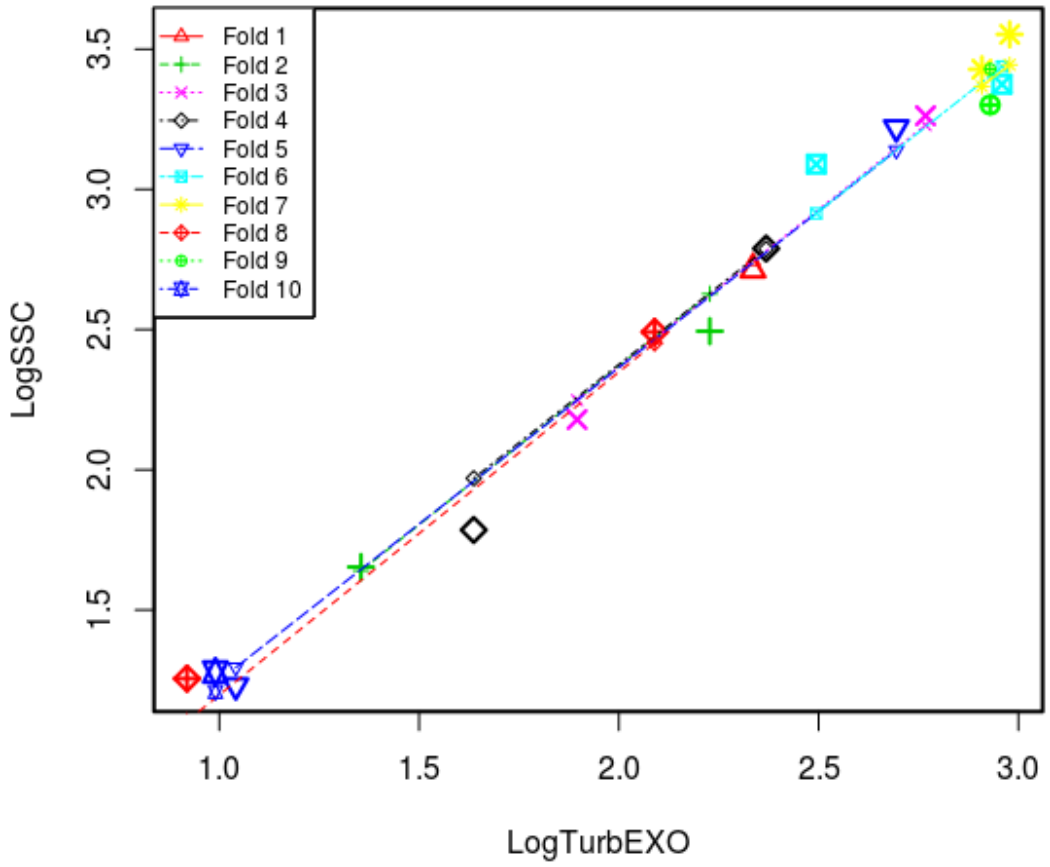
Statistical Plots



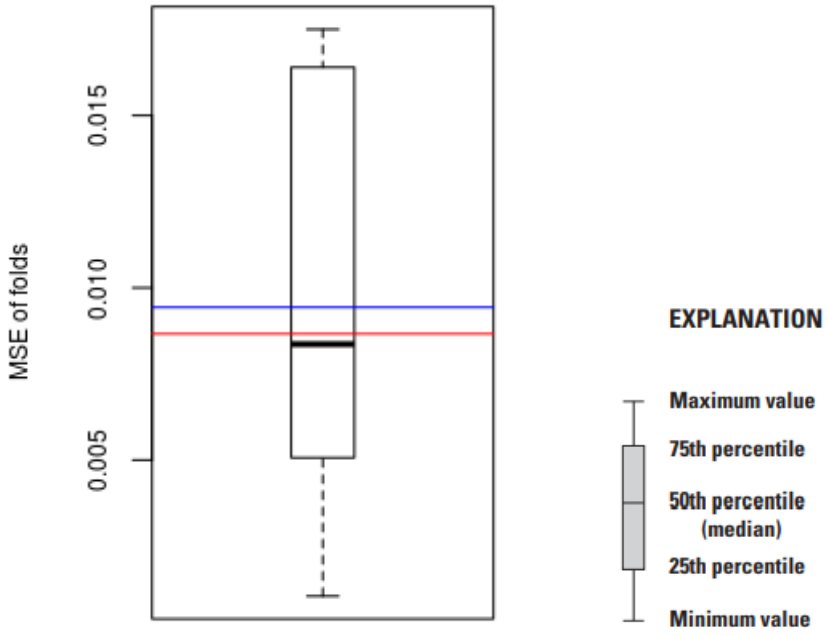

Residual (LogSSC)
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
2
2
2
3
1
1
2
3
1
Jan
Feb
Mar
Apr
Jun
Aug
Sep
Oct
Dec
Month
SSC
5,000
4,000
3,000
2,000
1,000
0
17
17
Computed
Observed
EXPLANATION
2 Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value


Residual (LogSSC)
0.2
0.1
0.0
-0.1
-0.2
1
6
3
5
2
2015
2016
2017
2018
2019
Year
EXPLANATION
2 Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value

Cross Validation




```
Minimum mean squared error (MSE) of folds:  0.00105
                         Mean MSE of folds:  0.00943
                       Median MSE of folds:  0.00837
                      Maximum MSE of folds:  0.01750
       (Mean MSE of folds) / (Model MSE):  1.09000
```



Red line - Model MSE

Blue line - Mean MSE of folds

Model-Calibration Dataset

	Date	LogSSC	LogTurbEXO	SSC	TurbEXO	Computed LogSSC	Computed SSC	Residual	Normal Quantiles	Censored Values
1	2015-12-01	2.18	1.9	151	78.6	2.25	179	-0.0661	-0.621	--
2	2016-02-25	1.23	1.04	17	11	1.28	19.5	-0.0504	-0.452	--
3	2016-04-19	2.49	2.09	310	123	2.46	297	0.0272	0.146	--
4	2016-04-21	3.37	2.96	2370	910	3.45	2850	-0.0711	-0.809	--
5	2016-04-27	3.3	2.93	2000	850	3.41	2640	-0.111	-1.03	--
6	2016-09-09	3.09	2.49	1230	312	2.92	851	0.169	1.81	--
7	2016-10-07	2.71	2.34	515	217	2.74	564	-0.0304	-0.296	--
8	2017-01-26	1.28	0.99	19	9.78	1.22	17.1	0.0556	0.621	--
9	2017-03-29	3.26	2.77	1830	586	3.23	1730	0.0328	0.296	--
10	2017-10-05	1.65	1.35	45	22.6	1.63	44	0.0188	0	--
11	2018-01-29	1.26	0.919	18	8.3	1.14	14.2	0.113	1.32	--
12	2018-06-26	1.79	1.64	61	43.3	1.95	91.6	-0.168	-1.81	--
13	2018-08-20	2.79	2.37	615	234	2.78	614	0.00972	-0.146	--
14	2018-09-06	3.55	2.98	3570	951	3.47	2990	0.0853	1.03	--
15	2018-10-10	2.49	2.23	312	169	2.62	425	-0.126	-1.32	--
16	2019-02-24	3.22	2.69	1660	495	3.15	1430	0.073	0.809	--
17	2019-03-14	3.43	2.91	2680	810	3.39	2500	0.0391	0.452	--

Definitions

Adj R^2: Adjusted coefficient of determination

```
BCF: Bias correction factor

DFFITS: Studentized difference in fits

Log: logarithm base 10

MSE: Mean squared error

MSPE: Model standard percentage error

R2: Coefficient of determination

RMSE: Root mean square error

SSC: Suspended-sediment concentration, in milligrams per liter (80154)

TurbEXO: Turbidity, 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

Cook, R.D., 1977, Detection of influential observations in linear regression: Technometrics, v. 19, no. 1, p. 15–18. [Also available at https://doi.org/10.2307/1268249.]

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, 2019, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, accessed September 2019 at https://www.Rproject.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 September 2019 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, 54 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 September 2019 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., accessed September 2019 at https://doi.org/10.3133/tm3A8.

U.S. Geological Survey, 2006, Collection of water samples (ver. 2.0): U.S. Geological Survey Techniques of Water Resources Investigations, book 9, chap. A4 [variously paged]. [Also available at https://pubs.water.usgs.gov/ twri9A4/.]

U.S. Geological Survey, 2016, Policy and guidance for approval of surrogate regression models for computation of time series suspended-sediment concentrations and loads: U.S. Geological Survey, Office of Water Quality Technical Memorandum 2016.10, 40 p., accessed November 20, 2020, at https://water.usgs.gov/admin/memo/SW/sw.2016.07+wq.2016.10.pdf.

U.S. Geological Survey, 2018, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed January 2017 at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey, 2020, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed January 2020, 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, 51 p. [Also available at https://doi.org/10.3133/tm1D3.]