

Model Archive Summary for Suspended Sediment Concentration at Station 01478245; White Clay Creek near Strickersville, Pennsylvania [2020 Version]

This model archive summary describes the regression model developed to estimate continuous instantaneous (15-minute) suspended sediment concentrations starting in 2011. The model was updated in June 2020, using all available calibration data. The new model is used to compute suspended sediment concentrations for entire period of record, beginning November 15, 2011.

Site and Model Information

U.S. Geological Survey (USGS) station number: 01478245

Station name: White clay Creek near Strickersville, Pennsylvania

Lat 39°44'51", long 75°46'15" referenced to North American Datum of 1927, Chester County, PA, Hydrologic Unit 02040205, on right bank, 0.1 mi downstream from West Branch White Clay Creek, in the White Clay Creek State Preserve, and 1.5 mi northeast of Strickersville.

Date regression model was created: June 2020.

Period of data for model calibration: January 12, 2012 – November 1, 2019.

Model application date: October 2019 onward.

Database Reviewed by: Elizabeth A. Hittle, May 2020

Computed by: Matthew C. Gyves, June 2020.

Approved by: Joseph W. Duris, Water-Quality Specialist, June 2020.

Equipment

A Yellow Spring Instrument (YSI) 600OMS monitor equipped with sensors for temperature, specific conductance, and turbidity (6136 sensor) is deployed at this site. The monitor is housed in a 3-inch perforated plastic pipe placed in the stream about 2 ft from the right bank. Readings from the sensors are recorded every 15 minutes and transmitted hourly by way of GOES satellite.

Model Calibration Dataset

All data were collected using standard USGS protocols and are stored in the National Water Information System (NWIS) database. Linear regression models were developed using RStudio Version 1.2.1335. Primary packages included (dataRetrieval, v2.7.5; dplyr, v0.8.3; plotly v4.9.0, and the Model Archive Summary Shiny App (available at <https://github.com/PatrickEslick/ModelArchiveSummary>). Explanatory variables evaluated as inputs to linear regression were turbidity and streamflow.

The final regression model is based on 80 concurrent measurements of suspended sediment and turbidity concentrations. Suspended sediment concentrations were determined from analysis of discrete samples, and turbidity concentrations were determined from continuous record of 15-minute values, interpolated when necessary to correspond with collection time of the discrete sample for sediment analysis. Suspended sediment samples were collected over a range of hydrologic conditions throughout each year during sampling period. Studentized residuals for final model were inspected and considered for potential removal as outliers if residual values were greater than 3 or less than -3. Several samples met these criteria; and other regression diagnostics show their influence on model performance significant. Sand is being transported under certain flow conditions, which has a significant effect on total mass within a suspended sediment sample. Additional modeling looking at sand/fine splits is warranted. Because sand is considered part of the total sediment load, these samples are included in this model.

Data Collection

Discrete samples for suspended sediment analysis were collected using an ISCO automatic sampler which was set to sample based on stage and turbidity thresholds to provide data for a broad range of turbidity values. Over time, these thresholds were changed to collect samples under different conditions, covering the range of measured values. Water was pulled by the sampler peristaltic pump through 0.5-inch tubing from an intake located near the turbidity probe and stored in 1-liter polypropylene bottles. Once collected, sample times were recorded in the field during bottle retrieval as reported by the ISCO. Samples were analyzed for suspended sediment concentration (SSC) in the laboratory at the USGS office in Exton, Pennsylvania and the USGS Kentucky Sediment Laboratory in Louisville, Kentucky using standard USGS methods. Suspended sediment concentrations ranged from 19 to 9,530 mg/L in the 80 analyzed samples.

Turbidity data were collected according to USGS Techniques and Methods 1-D3 (Wagner and Others, 2006). Turbidity concentrations ranged from 12 to 1,420 formazin nephelometric units (FNU) during collection of concurrent discrete suspended sediment samples used for model calibration.

Streamflow data were collected according to USGS Water-Supply Paper 2175 (Rantz and others, 1982). Streamflow ranged from 153 to 7,640 cubic feet per second (cfs) during the collection of concurrent discrete suspended sediment and turbidity samples used for model calibration.

Model Development

Regression analysis was done using R by examining turbidity (*Turb*) and streamflow (*Q*) as explanatory variables for SSC. A variety of linear regression models that predict *SSC* and $\log_{10}(\text{SSC})$ were evaluated and computed using methods described by Rasmussen and others, 2011.

The model with $\log_{10}(\text{Turb})$ and $\log_{10}(Q)$ as explanatory variables was selected as the best predictor of $\log_{10}(\text{SSC})$ on the basis of residual plots, maximizing adjusted coefficient of determination (adjusted R^2), and minimizing model residual standard error (or root mean square error, *RMSE*) and standard percentage error (*MSPE*). The addition of discharge as an explanatory variable significantly improved model performance.

Prior Model

This model supersedes the previous equation used to compute suspended sediment that was never published. All data in NWIS have been revised based on this model.

Note 80155 (SS discharge ton/d) is present in QWDATA through 2019??? This shouldn't be in the database unless it's from the new model. We don't have to reject the old data in this case, but we do need to publish the model that derived the values.

Model Summary

Final regression model for suspended sediment concentration (SSC) at site number 01478245:

$$\log_{10} \text{SSC} = .9 \log_{10} \text{Turb} + .582 \log_{10} Q - 1$$

where

SSC = suspended sediment concentration, in milligrams per liter (mg/L) (parameter 80154);

Turb = turbidity, in formazin nephelometric units (FNU) (parameter 63680);

Q = discharge, in cubic feet per second (ft³/s) (parameter 00060)

Turb and *Q* makes physical and statistical sense as explanatory variables for *SSC* because previous studies showed suspended sediment concentrations were directly related to turbidity concentrations and streamflow. The transformed model may be retransformed to the original units so that *SSC* concentrations can be calculated

directly. A potential bias that is introduced because of retransformation can be corrected using Duan's bias correction factor (BCF). For this model the BCF is 1.26. The retransformed model, using the BCF, is:

$$SSC = .1 Turb^{\cdot 9} \times Q^{\cdot 582} \times 1.26$$

References

Rasmussen, P.P., Gray, J.R., Glysson, G.D., and Ziegler, A.C., 2009, revised 2011, 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.

Rantz, S.E., and others, 1982, Measurement and computation of streamflow, v. 2: U.S. Geological Survey Water-Supply Paper 2175, v. 2, 631 p. (Also available at http://pubs.usgs.gov/wsp/wsp2175/pdf/WSP2175_vol2a.pdf).

Sloto, R.A., and Olson, L.E., Estimated suspended-sediment loads and yields in the French and Brandywine Creek Basins, Chester County, Pennsylvania, water years 2008–09: U.S. Geological Survey Scientific Investigations Report 2011–5109, 31 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 1–D3, 51 p. + 8 attachments; accessed April 10, 2006, at <http://pubs.water.usgs.gov/tm1d3>

Model Statistics, Data, and Plots

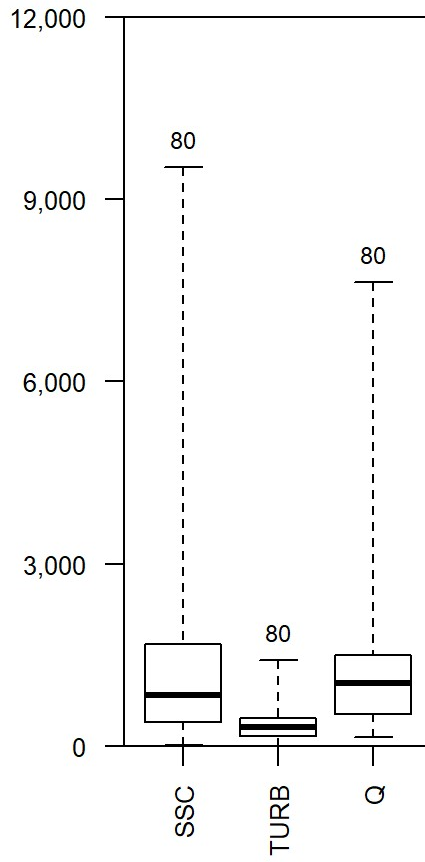
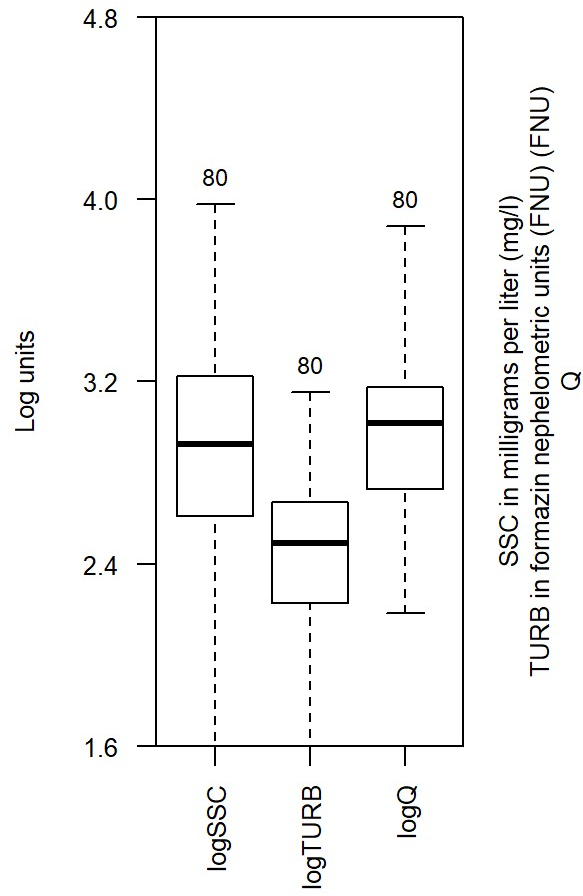
Model

$\log SSC = + 0.9 * \log TURB + 0.582 * \log Q - 1$

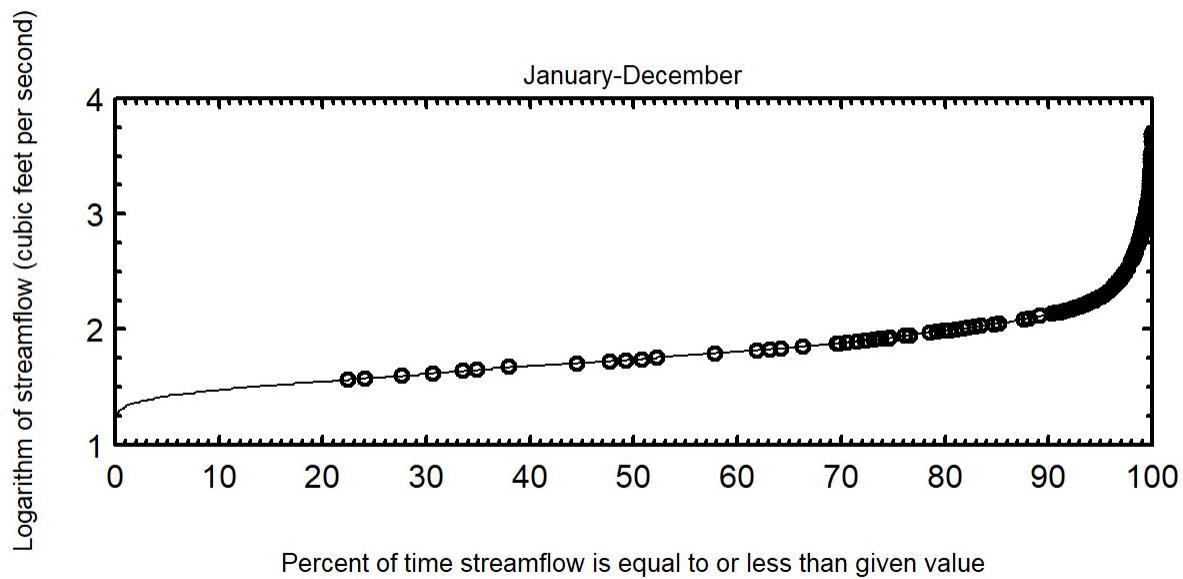
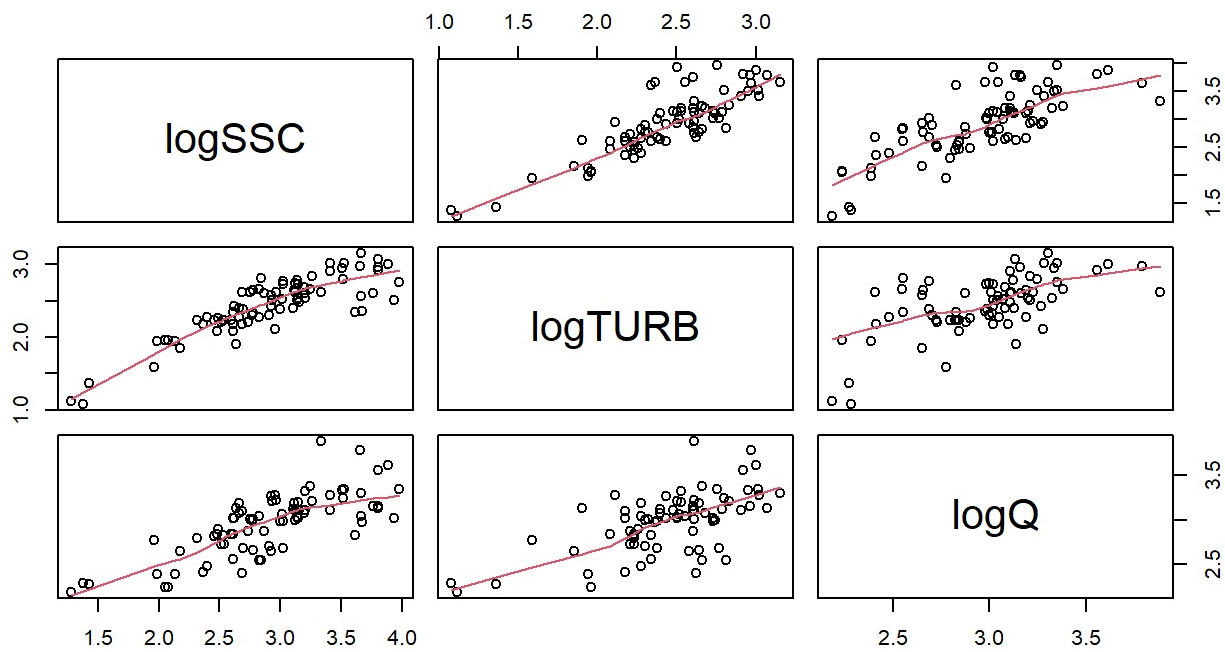
Variable Summary Statistics

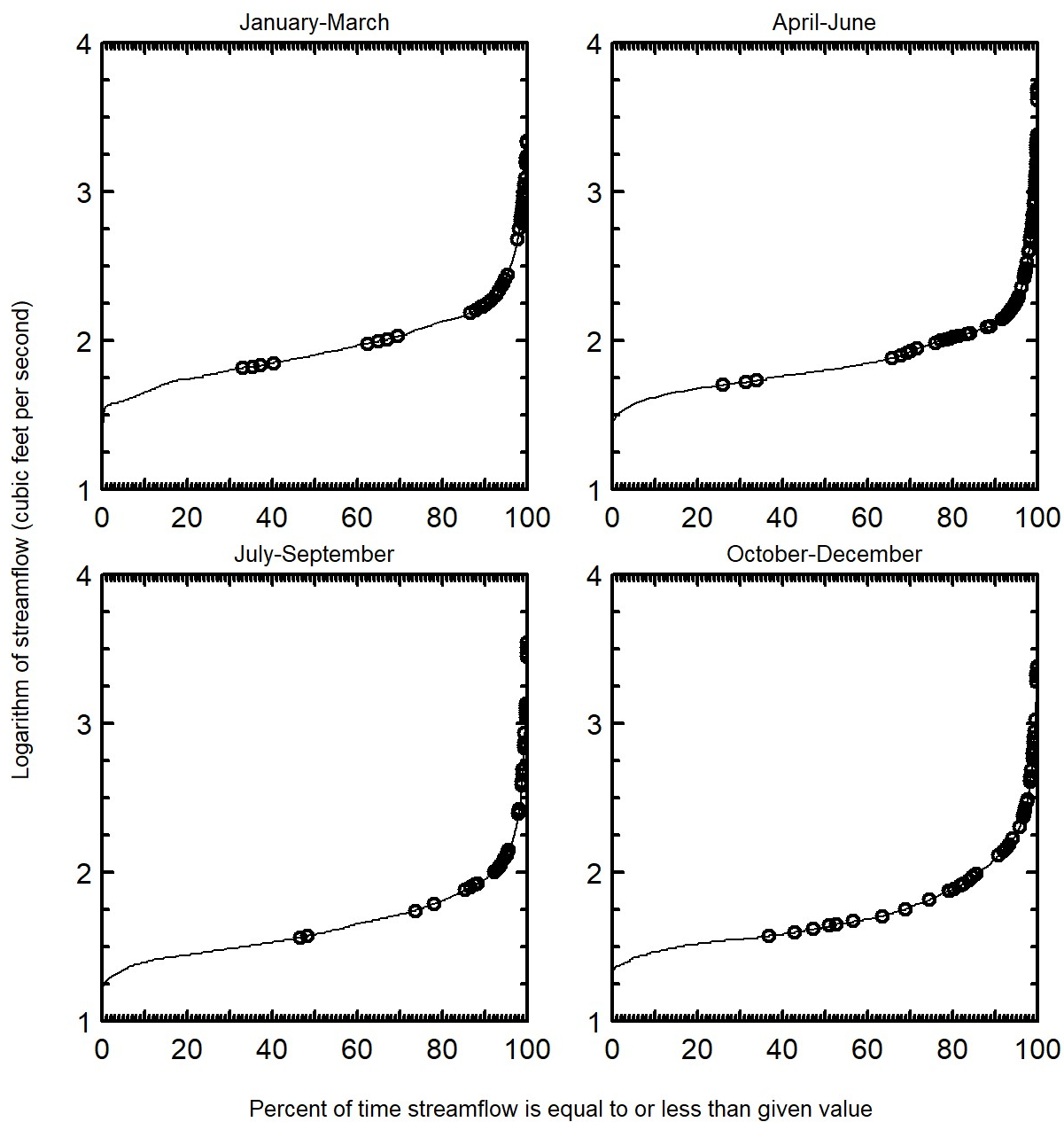
	logSSC	SSC	logTURB	logQ	TURB	Q
Minimum	1.28	19	1.08	2.18	12	153
1st Quartile	2.61	406	2.23	2.73	170	535
Median	2.92	840	2.49	3.02	310	1050
Mean	2.91	1670	2.43	2.96	376	1240
3rd Quartile	3.22	1670	2.67	3.17	470	1490
Maximum	3.98	9530	3.15	3.88	1420	7640

Box Plots



Exploratory Plots





Basic Model Statistics

Number of Observations	80
Standard error (RMSE)	0.25
Average Model standard percentage error (MSPE)	60.7
Coefficient of determination (R^2)	0.816
Adjusted Coefficient of Determination (Adj. R^2)	0.811
Bias Correction Factor (BCF)	1.26
Variance Inflation Factors (VIF)	
logTURB	logQ

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(> t)
(Intercept)	-1.000	0.2400	-4.17	8.04e-05
logTURB	0.900	0.0917	9.82	3.18e-15
logQ	0.582	0.1060	5.49	4.93e-07

Correlation Matrix

	Intercept	logTURB	logQ
Intercept	1.0000	-0.0766	-0.702
logTURB	-0.0766	1.0000	-0.651
logQ	-0.7020	-0.6510	1.000

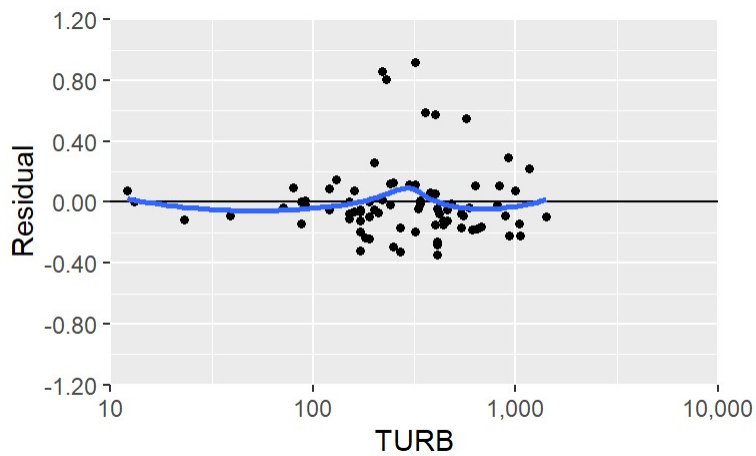
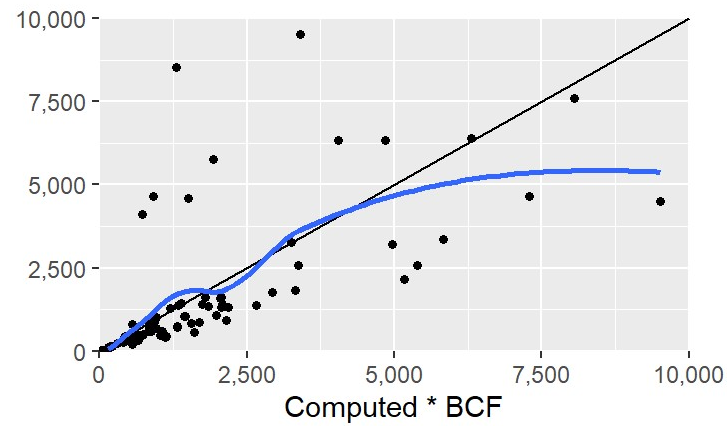
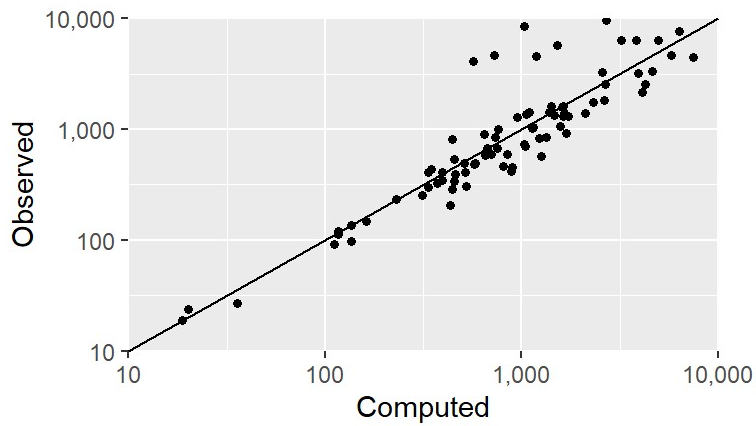
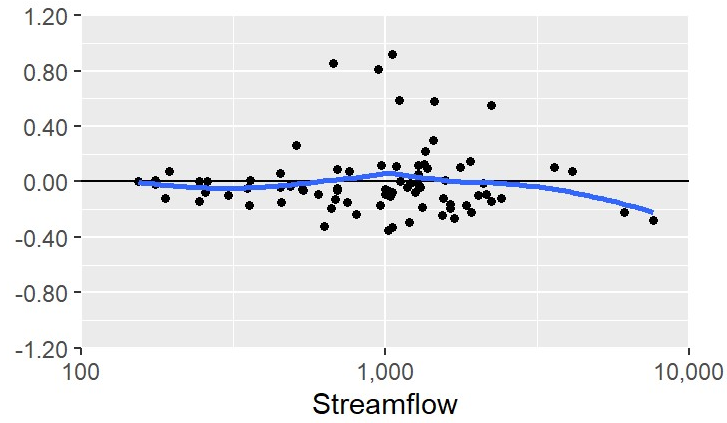
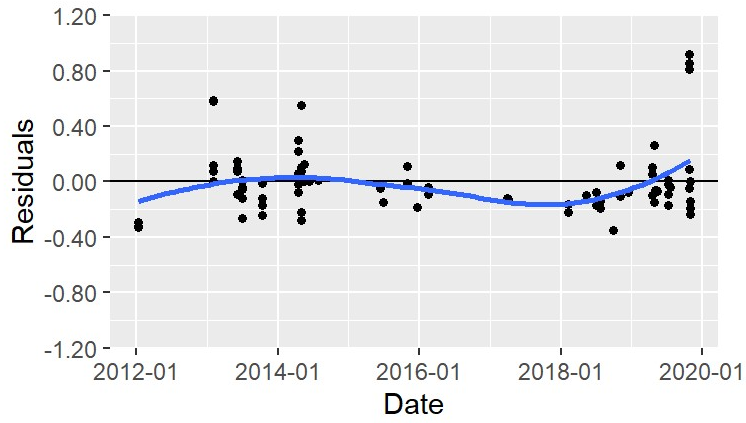
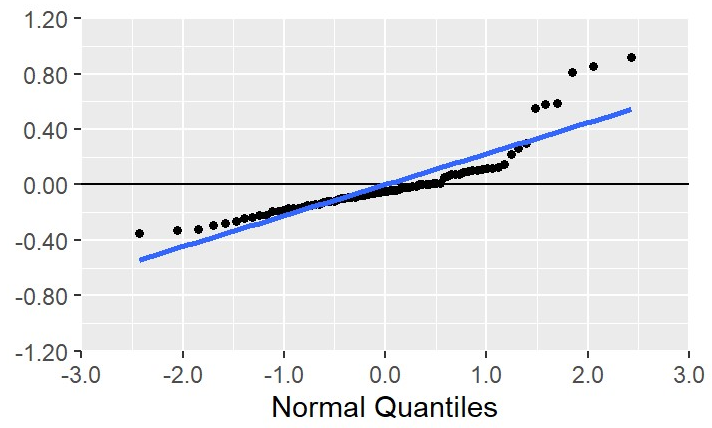
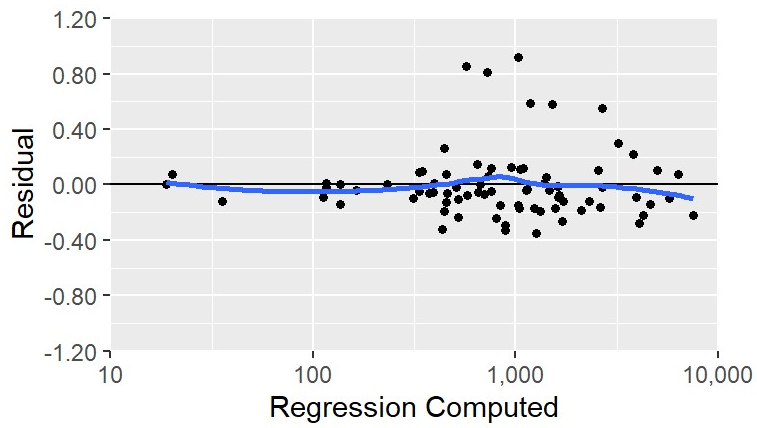
Outlier Test Criteria

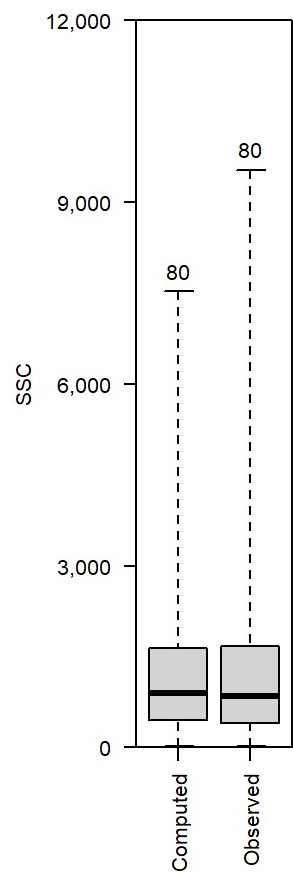
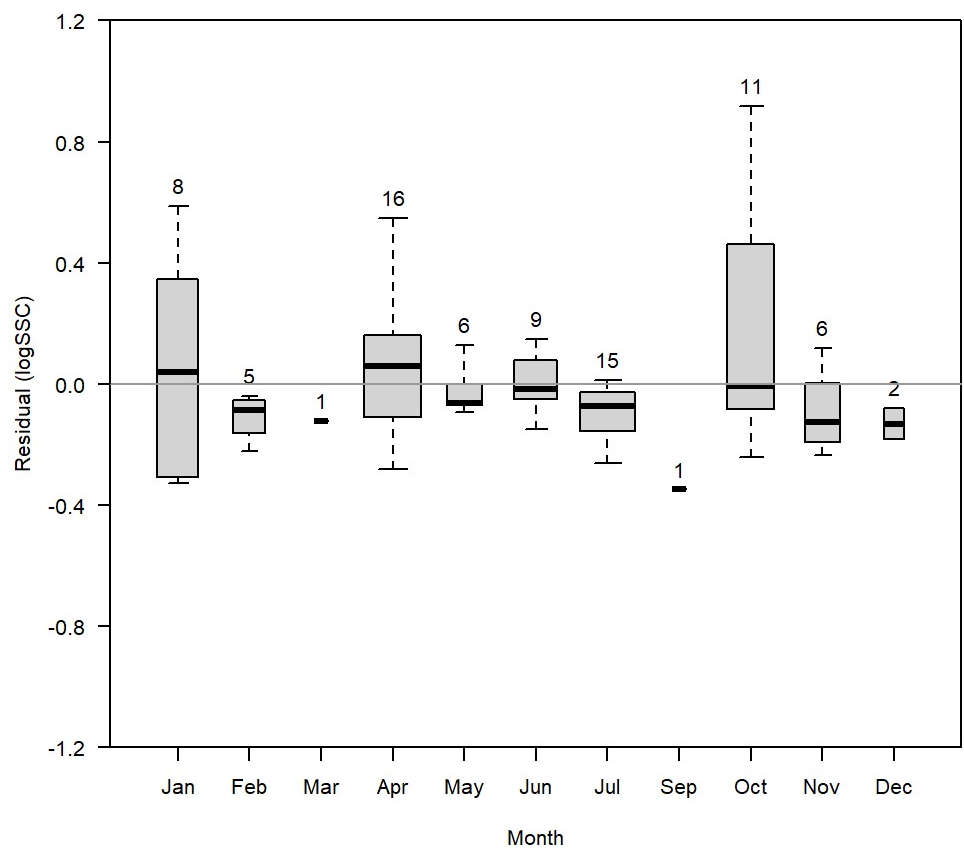
Leverage	Cook's D	DFFITS
0.112	0.264	0.387

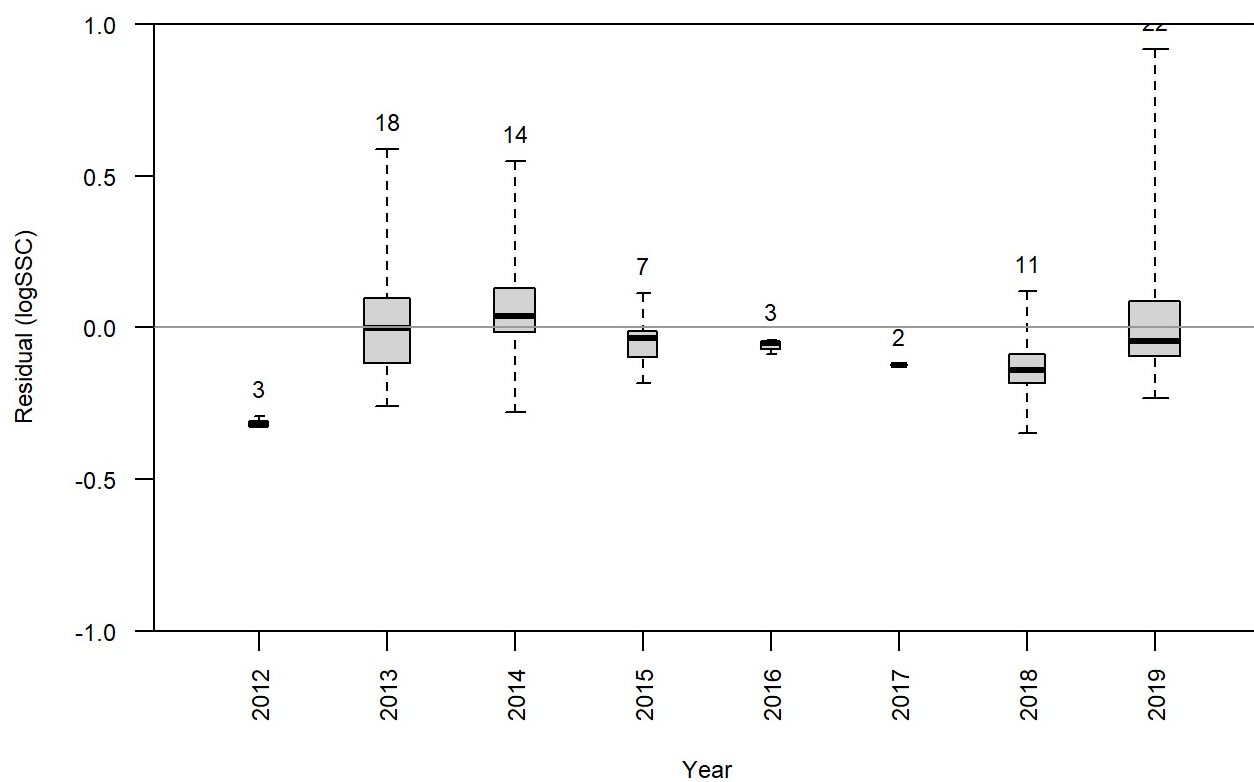
Flagged Observations

	logSSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
2013-01-30 23:00	1.28	1.27	0.00511	0.0222	0.022	0.1470	2.82e-05	0.00914
2013-06-08 13:45	1.38	1.30	0.07780	0.3390	0.337	0.1550	7.04e-03	0.14500
2014-04-30 14:30	3.98	3.43	0.54900	2.2300	2.290	0.0286	4.86e-02	0.39200
2014-04-30 21:00	3.33	3.61	-0.28000	-1.2000	-1.210	0.1370	7.66e-02	-0.48100
2019-10-27 13:30	3.61	2.75	0.85800	3.4600	3.740	0.0143	5.78e-02	0.45000
2019-10-27 14:30	3.93	3.01	0.91800	3.7000	4.050	0.0130	6.02e-02	0.46600
2019-10-27 15:30	3.67	2.86	0.80900	3.2600	3.490	0.0135	4.84e-02	0.40800

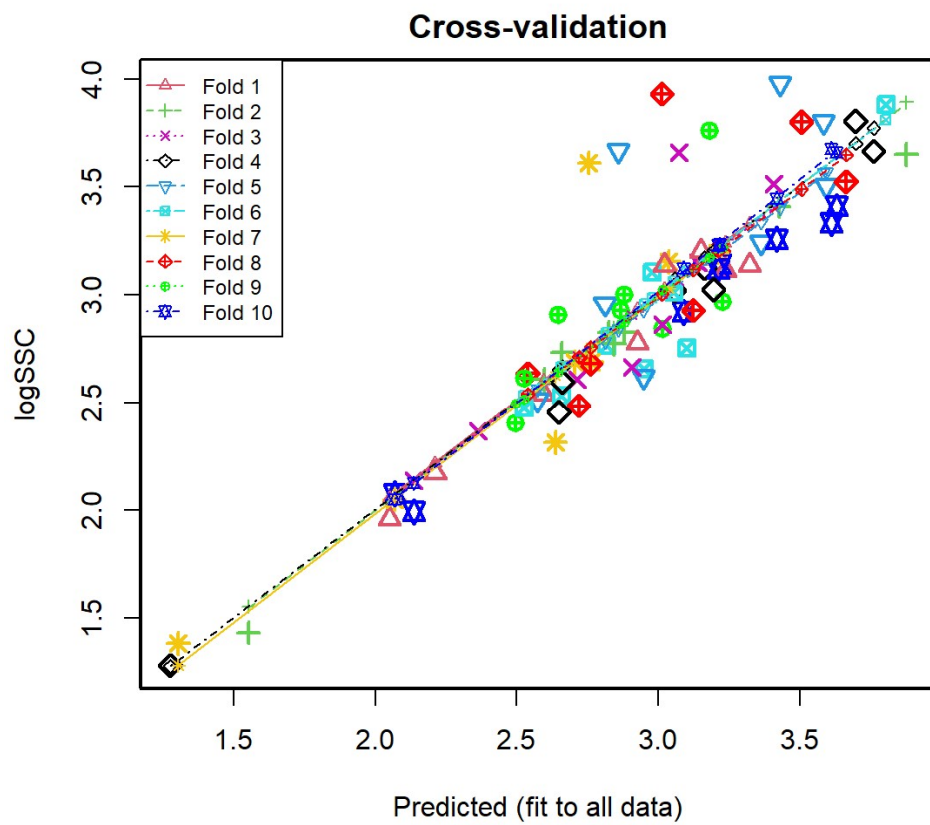
Statistical Plots







Cross Validation



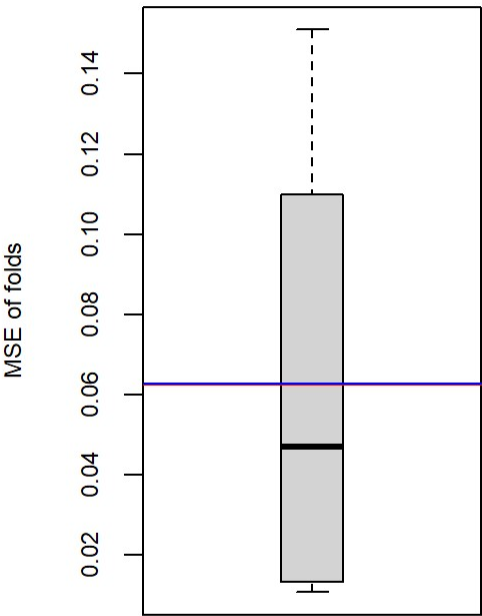
Minimum MSE of folds: 0.0107

Mean MSE of folds: 0.0628

Median MSE of folds: 0.0469

Maximum MSE of folds: 0.1510

(Mean MSE of folds) / (Model MSE): 1.0100



Red line - Model MSE

Blue line - Mean MSE of folds

Model-Calibration Data Set

	Date	logSSC	logTURB	logQ	SSC	TURB	Q	Computed	Computed	Residual	Normal	Censored
0								logSSC	SSC		Quantiles	Values
1	2012-01-12	2.66	2.4	3.08	454	250	1200	2.95	1130	-0.293	-1.7	--
2	2012-01-12	2.62	2.43	3.02	417	270	1050	2.95	1120	-0.327	-2.05	--
3	2012-01-12	2.31	2.23	2.8	206	170	628	2.64	546	-0.322	-1.85	--
4	2013-01-31	1.28	1.11	2.18	19	13	153	1.27	23.7	0.00511	0.435	--
5	2013-01-31	2.73	2.2	2.88	541	160	758	2.66	577	0.0734	0.653	--

6	2013-01-31	3.16	2.48	3.11	1430	300	1280	3.04	1380	0.117	1.01	--
7	2013-01-31	3.76	2.6	3.16	5760	400	1450	3.18	1920	0.578	1.58	--
8	2013-01-31	3.66	2.56	3.05	4580	360	1110	3.07	1500	0.588	1.7	--
9	2013-06-08	2.96	2.11	3.28	909	130	1900	2.81	818	0.148	1.18	--
10	2013-06-08	2.64	1.9	3.14	432	80	1370	2.54	437	0.097	0.817	--
11	2013-06-08	1.96	1.59	2.78	91	39	599	2.05	141	-0.0893	-0.334	--
12	2013-06-08	1.38	1.08	2.29	24	12	194	1.3	25.4	0.0778	0.733	--
13	2013-07-01	2.97	2.61	3.23	929	410	1680	3.23	2140	-0.261	-1.48	--
14	2013-07-01	3.15	2.53	3.2	1410	340	1580	3.14	1750	0.00892	0.505	--
15	2013-07-01	2.76	2.3	3	582	200	1000	2.82	829	-0.0522	-0.0469	--
16	2013-07-01	2.17	1.85	2.65	149	71	451	2.21	205	-0.0376	0.11	--
17	2013-07-02	1.43	1.36	2.28	27	23	189	1.55	44.9	-0.119	-0.505	--
18	2013-10-11	3.2	2.53	3.32	1600	340	2100	3.21	2060	-0.00814	0.301	--
19	2013-10-11	3.24	2.66	3.38	1750	460	2400	3.36	2920	-0.121	-0.541	--
20	2013-10-11	2.92	2.43	3.27	833	270	1850	3.09	1550	-0.169	-0.861	--
21	2013-10-11	2.66	2.28	3.19	461	190	1540	2.91	1020	-0.243	-1.39	--
22	2014-04-15	2.93	2.58	2.65	847	380	451	2.87	930	0.0612	0.615	--
23	2014-04-15	3.41	2.91	3.11	2570	810	1280	3.43	3370	-0.0165	0.205	--
24	2014-04-15	3.14	2.73	3.02	1390	540	1050	3.22	2090	-0.0748	-0.237	--
25	2014-04-15	3.8	2.96	3.16	6330	920	1440	3.51	4050	0.295	1.39	--
26	2014-04-15	3.8	3.07	3.13	6330	1170	1350	3.58	4850	0.218	1.24	--
27	2014-04-30	3.98	2.76	3.35	9530	570	2240	3.43	3410	0.549	1.48	--
28	2014-04-30	3.81	2.92	3.56	6390	830	3610	3.7	6310	0.107	0.908	--
29	2014-04-30	3.88	3	3.61	7610	1000	4120	3.8	8060	0.0769	0.692	--
30	2014-04-30	3.65	2.97	3.79	4490	930	6130	3.88	9510	-0.224	-1.24	--
31	2014-05-01	3.33	2.61	3.88	2150	410	7640	3.61	5170	-0.28	-1.58	--
32	2014-05-16	3.11	2.4	3.13	1280	250	1340	2.98	1200	0.129	1.12	--
33	2014-05-16	2.83	2.28	3.05	671	190	1120	2.83	846	0.000974	0.334	--
34	2014-06-10	2.37	2.18	2.41	232	150	260	2.36	292	0.00154	0.368	--
35	2014-07-28	2.61	2.34	2.56	407	220	360	2.6	498	0.0136	0.541	--
36	2015-06-15	3.02	2.77	2.69	1050	590	486	3.06	1440	-0.0364	0.141	--
37	2015-06-15	2.83	2.66	2.55	674	460	352	2.88	956	-0.05	-0.0156	--
38	2015-06-21	2.69	2.38	2.69	490	240	486	2.71	642	-0.0157	0.237	--
39	2015-06-28	2.78	2.64	2.66	599	440	455	2.93	1070	-0.149	-0.733	--
40	2015-10-29	3.2	2.68	3.08	1570	480	1210	3.21	2040	-0.0117	0.269	--
41	2015-10-29	3.14	2.51	3.04	1370	320	1090	3.02	1330	0.114	0.956	--
42	2015-12-24	3.14	2.79	3.12	1380	610	1320	3.32	2660	-0.183	-1.01	--
43	2016-02-16	3.51	2.95	3.33	3210	890	2150	3.59	4970	-0.0879	-0.301	--
44	2016-02-17	3.12	2.61	3.11	1330	410	1300	3.16	1840	-0.0403	0.0782	--
45	2016-02-17	2.54	2.23	2.73	348	170	532	2.59	496	-0.0524	-0.0782	--
46	2017-04-01	3.11	2.64	3.19	1300	440	1550	3.24	2180	-0.122	-0.578	--

47	2017-04-06	2.53	2.23	2.84	341	170	684	2.66	574	-0.125	-0.615	--
48	2018-02-11	3.41	3.02	3.28	2570	1050	1920	3.63	5400	-0.221	-1.18	--
49	2018-02-11	3.26	2.83	3.21	1810	680	1630	3.42	3320	-0.161	-0.817	--
50	2018-05-14	2.4	2.28	2.48	252	190	304	2.5	396	-0.0945	-0.401	--
51	2018-07-04	2.84	2.81	2.55	699	650	356	3.02	1310	-0.172	-0.956	--
52	2018-07-04	2.69	2.62	2.41	489	420	256	2.76	732	-0.0733	-0.205	--
53	2018-07-23	3.53	3.02	3.35	3350	1040	2230	3.66	5840	-0.14	-0.653	--
54	2018-07-23	2.93	2.51	3.21	850	320	1630	3.12	1680	-0.195	-1.12	--
55	2018-09-28	2.75	2.61	3.01	568	410	1020	3.1	1600	-0.348	-2.43	--
56	2018-11-03	3	2.38	2.99	1000	240	970	2.88	960	0.119	1.06	--
57	2018-11-06	2.61	2.18	3.02	406	150	1040	2.71	655	-0.106	-0.47	--
58	2018-12-16	2.68	2.18	3.1	481	150	1250	2.76	729	-0.079	-0.269	--
59	2019-04-20	3.67	3.15	3.31	4640	1420	2020	3.76	7290	-0.0948	-0.435	--
60	2019-04-20	3.52	2.8	3.25	3280	630	1770	3.41	3250	0.106	0.861	--
61	2019-04-20	3.2	2.6	3.11	1600	400	1280	3.15	1790	0.0535	0.578	--
62	2019-04-27	2.87	2.6	2.87	733	400	749	3.02	1310	-0.15	-0.774	--
63	2019-04-27	2.91	2.3	2.71	807	200	508	2.65	559	0.261	1.31	--
64	2019-05-05	2.6	2.23	2.84	395	170	695	2.66	580	-0.065	-0.141	--
65	2019-05-08	2.51	2.2	2.73	324	160	537	2.57	472	-0.0621	-0.11	--
66	2019-05-12	2.78	2.32	3.01	596	210	1030	2.84	882	-0.0684	-0.173	--
67	2019-07-12	3.12	2.74	3	1320	550	1000	3.21	2060	-0.0921	-0.368	--
68	2019-07-12	3.03	2.73	2.98	1060	540	961	3.2	1980	-0.17	-0.908	--
69	2019-07-12	2.05	1.96	2.24	112	91	174	2.07	147	-0.0177	0.173	--
70	2019-07-12	2.08	1.96	2.24	119	91	174	2.07	147	0.00858	0.47	--
71	2019-07-18	3.01	2.52	3.07	1030	330	1180	3.05	1430	-0.042	0.0469	--
72	2019-10-27	3.61	2.34	2.83	4100	220	674	2.75	718	0.858	2.05	--
73	2019-10-27	3.93	2.51	3.02	8540	320	1050	3.01	1300	0.918	2.43	--
74	2019-10-27	3.67	2.36	2.98	4660	230	951	2.86	914	0.809	1.85	--
75	2019-10-27	2.61	2.08	2.84	410	120	695	2.53	424	0.0874	0.774	--
76	2019-10-27	2.48	2.08	2.84	300	120	695	2.53	424	-0.0483	0.0156	--
77	2019-11-01	2.49	2.26	2.9	306	180	799	2.72	662	-0.233	-1.31	--
78	2019-11-01	2.46	2.23	2.82	286	170	661	2.65	563	-0.193	-1.06	--
79	2019-11-01	1.99	1.94	2.39	98	87	244	2.13	172	-0.144	-0.692	--
80	2019-11-01	2.14	1.94	2.39	137	87	244	2.13	172	0.00181	0.401	--

Definitions

SSC: Suspended sediment concentration (SSC), in mg/l (80154)
TURB: Turbidity, in FNU (63680)
Q: Discharge, in ft³/s (00060)

