Model Archive Summary No. 2 for Turbidity Derived Suspended-Sediment Concentration at Station 11312676 Middle River at Middle River, CA

This model archive summary details the suspended-sediment concentration (SSC) model developed to compute 15-minute SSC beginning on January 12, 2015. This is the second suspended-sediment model developed for the site. The methods used follow U.S. Geological Survey (USGS) guidance as referenced in Office of Surface Water/Office of Water Quality Technical Memorandum 2016.10 (USGS, 2016) and USGS Techniques and Methods, book 3 section C, chapter 4 (Rasmussen and others, 2009). This summary and model archive are in accordance with Attachment A of Office of Water Quality Technical Memorandum 2015.01 (USGS, 2014).

Site and Model Information

Site number: 11312676
Site name: Middle River at Middle River, CA (MDM)
Location: Latitude 37°56’34", Longitude 121°31'59" referenced to North American Datum of 1927, San Joaquin County, CA, Hydrologic Unit 18040003.
Equipment: A YSI EXO sonde began logging turbidity on January 12, 2015.

Model number: 11312676.SSC.WY15.1
Model calibration data period: March 25, 2015 – January 14, 2020
Model application date: January 12, 2015 – September 30, 2020
Computed by: Tara Morgan-King, USGS, Sacramento, CA (tamorgan@usgs.gov)
Reviewed by: Anna Conlen, USGS, Sacramento, CA (aconlen@usgs.gov)

Physical Sampling Details and Sediment Data

Discrete, boat-based sample collection for SSC monitoring ideally occurs between 6-12 times per year. Sample collection spans the range of conditions targeting storm events during winter and spring flows as well as summer low flow conditions. Sample collection spanned 6 water years (WYs). A total of 28 sediment samples representative of the cross section were collected during WYs 2015-2020. Sample collection varied year to year, with an average of 5 samples collected per water year (WY). The minimum of 1 sample was collected during WY 2015 (a partial year) and the maximum of 8 samples was collected in WY 2018. WY 2020 was also a partial year.

Sample collection is consistent with approved field methods described in Edwards and Glysson (1999). Sediment samples represent the discharge-weighted concentrations of the stream cross section. Samples are collected using the equal-discharge-increment (EDI) method to establish five sampling verticals along the transect. Each of the five sections established using the EDI method represents 20% of the total flow. Samples are obtained at the centroid of each equal-discharge section. Due to the tidal nature of the site, the EDI method was used to collect discharge-weighted samples to represent the average cross section because velocities are typically not high enough to achieve isokinetic conditions (based on Table 4-5 from TWRI09A4, USGS 2006). A boat-based discharge (Q) measurement was collected using an ADCP immediately before sampling to determine the location of each vertical.

Trained USGS technicians collected samples at MDM approximately 75 ft downstream of the gage using a FISP US D-96 depth-integrated, suspended-sediment bag sampler. The channel cross section is roughly

28 ft deep in the thalweg with a mean depth of approximately 16 ft. Sampling depths range from roughly 15-29 feet depending on the tide and the season. Station velocities ranged from -1.7 to +1.7 ft/sec (less than the isokinetic transit-rate requirement for D-96 samplers). EDI sampling techniques are preferred in non-isokinetic conditions because they still produce a discharge-weighted sample. Sediment at this station was mostly fines (89% on average from sand/fine analysis) and potential bias of SSC due to non-isokinetic sampling is considered minimal.

Samples were analyzed by the USGS Sediment Laboratory in Santa Cruz, California. All samples were analyzed for suspended-sediment concentration (mg/L) by the filtration method and most samples were also analyzed for the percentage of fines (< 0.062 mm). The sand/fine break analysis can be used to identify dataset variability and potential outliers and shows that sediment at this station was composed of mostly fines (89% fines on average). Each vertical from the EDI set was analyzed individually by the lab. Individual analysis of each vertical is important for quality control purposes because of rapidly changing, tidal conditions. It can also help define potential channel variability and identify samples contaminated by bed sediment due to nozzle scooping and/or hitting the bed to hard. The set average SSC of the five verticals represents the cross-sectional average and was used in the calibration model dataset. The sediment lab automates the set average to the database. In rare occasions when the SSC at a vertical was deemed an outlier, a manual average was computed from fewer than 5 verticals. This occurred on 5/24/2016 and 10/17/2017 with notes applied to the database.

All sediment data were reviewed and marked as approved in the USGS National Water Information System (NWIS) Water-Quality System database (QWDATA) and made publicly available before being included in the sediment model. Sample results for SSC and % fines were stored in the USGS NWIS database. Publicly available field/lab sediment data and metadata can be found at: https://waterdata.usgs.gov/nwis/uv?site_no=11312676.

Surrogate Data

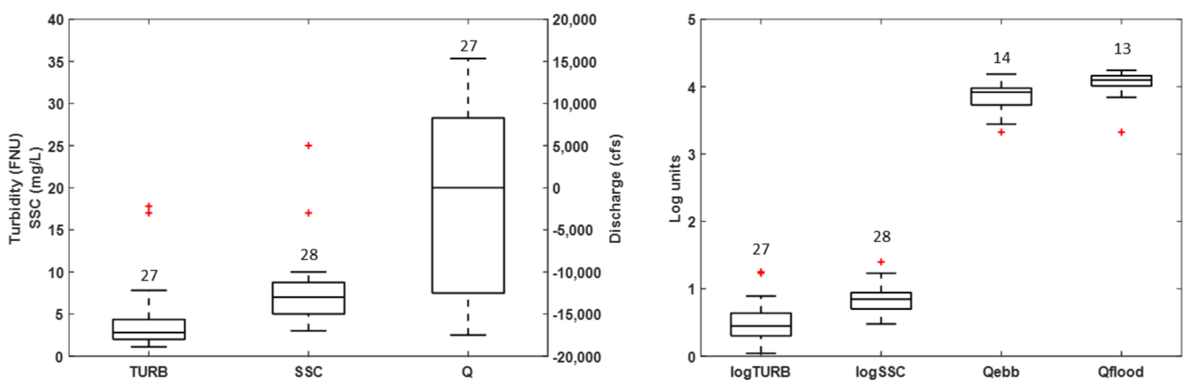
Continuous 15-minute turbidity data and discharge data were collected and computed by the USGS California Water Science Center and evaluated as possible explanatory variables for SSC. Turbidity data were measured using a YSI EXO2 sonde and reported in Formazin Nephelometric Turbidity Units (FNU). Turbidity from the EXO sensor began logging on 1/12/2015 at 14:15 PST. All surrogate turbidity data were computed, reviewed, and approved before using in the sediment calibration model per USGS guidelines (Wagner and others 2006). Discharge data were collected, computed, reviewed, and approved by the USGS California Water Science Center. Methods to compute discharge follow Levesque and Oberg (2012). The 15-minute discharge timeseries data are measured and reported in cubic feet per second (cfs). Time-series data are located at: https://waterdata.usgs.gov/usa/nwis/uv?site_no=11312676.

Model Calibration Dataset

The approved time-series turbidity and discharge data spanning the dates of the sediment constituent dataset were retrieved from NWIS-TS (Rasmussen and others 2009). The USGS Surrogate Analysis and Index Developer Tool (SAID) was used to pair the surrogate data with the discrete sediment data (Domanski and others 2015). Turbidity and discharge values were paired with each sediment sample observation from a matching max of +/- 15 minutes. The SAID manual is found at https://pubs.er.usgs.gov/publication/ofr20151177. Of the 28 cross-sectional average sediment samples collected during the deployment, one did not have corresponding turbidity values due to a data gap in the record, leaving a total of 27 observations in the calibration dataset. Summary statistics and the complete model-calibration dataset are provided in the following sections.

Regression Model Development

Multiple models were evaluated including simple linear regression (SLR) and multiple linear regression (MLR). The most common estimation technique is SLR, but MLR is an alternate tool for computing SSCs when the SLR model standard percentage error (*MSPE*) statistic is larger than 20 percent (Rasmussen and others, 2009). The calibration dataset is composed of 27 concurrent turbidity, SSC, and discharge measurements. Boxplots are shown below. Note that due to negative tidal discharge values during the flood tide, ebb and flood values are shown separately with the absolute values shown during flood tides. USGS (2016) *recommends* a minimum of 36 paired observations, while that guideline was not achieved, it is equally important that the paired observations span the range of conditions at a site. The dataset achieves that objective as it represents 97% of the conditions across the 15-min real-time turbidity measurement time-series, with near equal number of samples collected during ebb tide and flood tide.



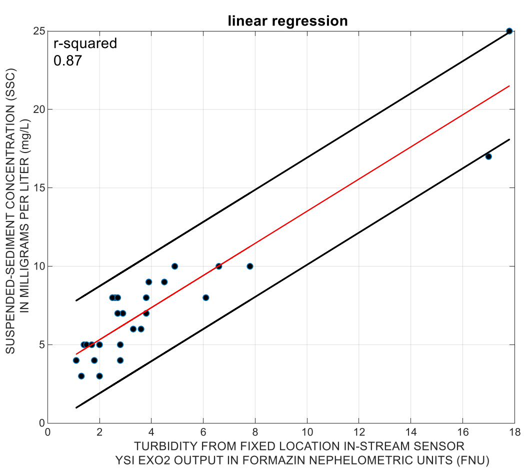
Model diagnostics and plots for model review were output using a variety of applications, including Matlab, SAID, and the R environment (R Core Team, 2018). An R-based application created by the USGS Kansas Water Science Center was also used to produce model statistics and plots for this model archive summary and is available at: https://patrickeslick.github.io/ModelArchiveSummary/. The regression methods used are described in Helsel and Hirsch (2002). Table 3 in Rasmussen and others (2009) shows the best statistical diagnostics to help evaluate the models. The best model was chosen based on residual plots, model standard error, R^2, significance tests (p-values), correlation of explanatory variables, variance inflation factor (VIF), and PRESS (prediction error sum of squares) statistics. Values for the statistics and metrics were computed for various models and are included below along with all relevant sample data and more in-depth statistical information.

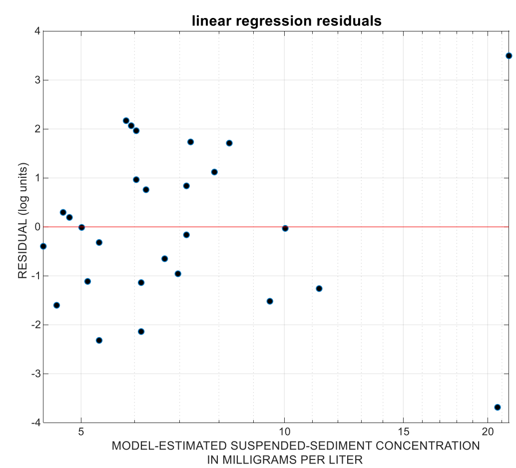
A variety of models were evaluated: Model 1) linear model with one explanatory variable (turbidity), Model 2) $\log_{10}$ transformed model with one explanatory variable (turbidity), Model 3) repeated medians method (Helsel and Hirsh, 2002) using one explanatory variable (turbidity), Model 4) linear model with two explanatory variables (turbidity and discharge), Model 5) log_{10} transformed model with two explanatory variables (turbidity and absolute discharge), and Model 6) log_{10} transformed model with two explanatory variables (turbidity and discharge). The number of observations in model 6 is reduced because this log transformation omits negative discharge values. Diagnostic statistics are summarized below for the six models evaluated. Discharge was not considered further as a second surrogate (in addition to turbidity) because it was not significant as a second variable to the model ($p > 0.05$). The site is variable; sediment can be transported from both upstream and downstream depending on storm tracks and wind patterns. High sediment concentrations coincide with fluvial events, but turbidity is also dependent on wind speeds and transport from shallow flooded regions in the region. Thus, discharge was not considered further in model development (Models 4-6).

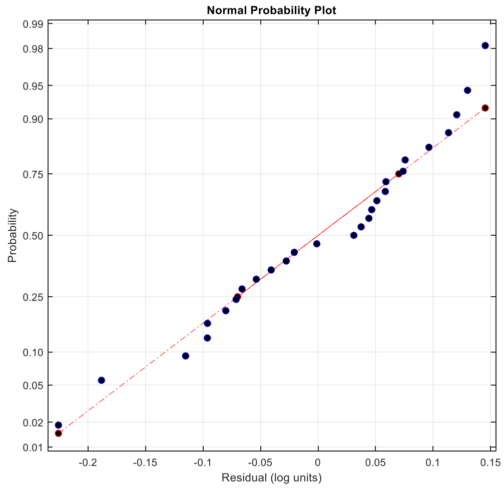
No.	R^2	R^2_a	RMSE	PRESS	MSPE	N	(type)
Model 1	0.87	0.87	1.7	124	21.2	27	linear
Model 2	0.79	0.78	0.1	0.3	22.8	27	log
Model 3	0.87	0.87	1.7	125	21.5	27	repeated median
Model 4	0.89	0.88	1.6	140	20.4	26	multi-linear
Model 5	0.79	0.77	0.10	29.6	23.4	26	ABS multi-log
Model 6	0.89	0.86	0.08	14.2	18.9	13	multi-log

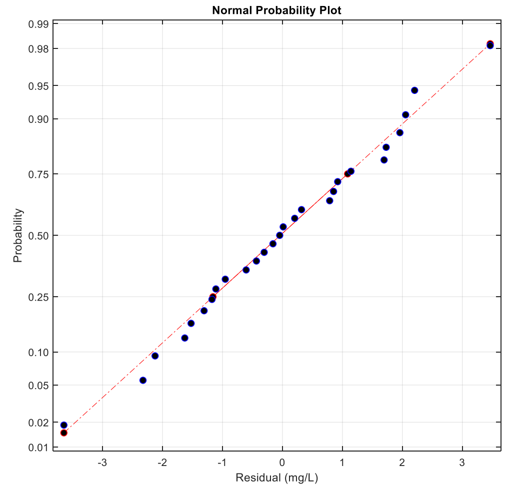
Flagged observations from the SAID outlier test criteria were evaluated. Standardized residuals from the models were inspected for values greater than 3 or less than negative 3. Values outside of the 3 to – 3 range are considered potential extreme outliers. The standardized residuals were reviewed from the SAID output reports and none of the samples were deemed as extreme outliers that should be removed from the model. All 27 observations were left in the model.

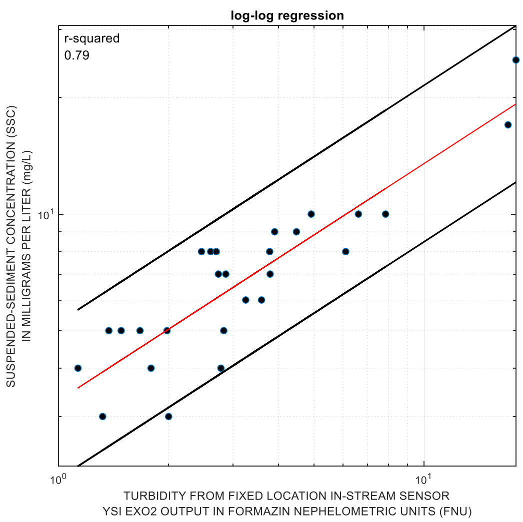
Of the SLR models, the un-transformed linear model had the highest adjusted R^2 and lowest MSPE. The normal probability plot looked better compared to the log-transformed option.

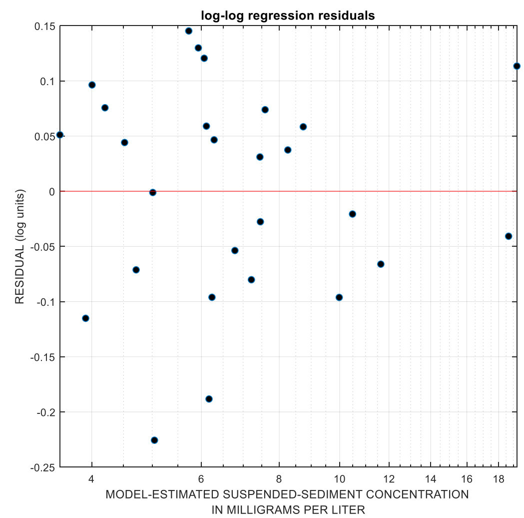





Normal Probability Plot
Probability
Residual (log units)


Normal Probability Plot
Probability
Residual (mg/L)


log-log regression
r-squared
0.79
SUSPENDED-SEDIMENT CONCENTRATION (SSC)
IN MILLIGRAMS PER LITER (mg/L)
TURBIDITY FROM FIXED LOCATION IN-STREAM SENSOR
YSI EXO2 OUTPUT IN FORMAZIN NEPHELOMETRIC UNITS (FNU)


log-log regression residuals
RESIDUAL (log units)
MODEL-ESTIMATED SUSPENDED-SEDIMENT CONCENTRATION
IN MILLIGRAMS PER LITER

Model Summary

The final regression model for computing suspended-sediment concentration at site number 11312676 is a simple non-transformed linear regression model based on 27 measurements of cross-sectional SSC samples and in situ turbidity values collected over approximately 5 years from March 25, 2015 to January 14, 2020. The regression model is shown below with basic model information, regression coefficients, correlation, and summary statistics.

Linear Regression Model	Coefficient of Determination (R^2)
$SSC = 3.28 + 1.02 * Turb$	0.872

Where

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

Turb = turbidity, in formazin nephelometric units.

Parameter	Minimum	Maximum
Turbidity (FNU) entire record	0.6	71
Computed SSC (mg/L)	4	*76/28

*Extrapolation, defined as computation beyond the range of the values in the model calibration dataset, may be used to extrapolate no more than 10 percent outside the range of the sample data used to fit the model. The original maximum computed SSC was 76 mg/L. However, following USGS guidelines, a threshold filter was applied to the time-series limiting the computation above 28 mg/L. Thus, the extrapolated, maximum computed SSC for this model is 28 mg/L. The portion of time-series data beyond the extrapolation limit is less than 1%.

Suspended-Sediment Concentration Record

The complete SSC record is computed using this regression model and can be found at https://nrtwq.usgs.gov/explore/dyplot?site_no=11312676 as well as the links to all the stations in the sediment network at http://nrtwq.usgs.gov/ca.

Model

SSC = + 1.02 * TURB + 3.28

Variable Summary Statistics

	SSC	TURB
Minimum	3	1.1
1st Quartile	5	2.0
Median	7	2.8
Mean	8	4.2
3rd Quartile	9	4.5
Maximum	25	17.8

Basic Model Statistics

Number of Observations	27
Standard error (RMSE)	1.66
Average Model standard percentage error (MSPE)	21.7
Coefficient of determination (R^2)	0.872
Adjusted Coefficient of Determination (Adj. R^2)	0.867

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	3.28	0.4610	7.1	1.92e-07
TURB	1.02	0.0785	13.0	1.19e-12

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.723
E.vars	-0.723	1.000

Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.222	0.193	0.544

Flagged Observations

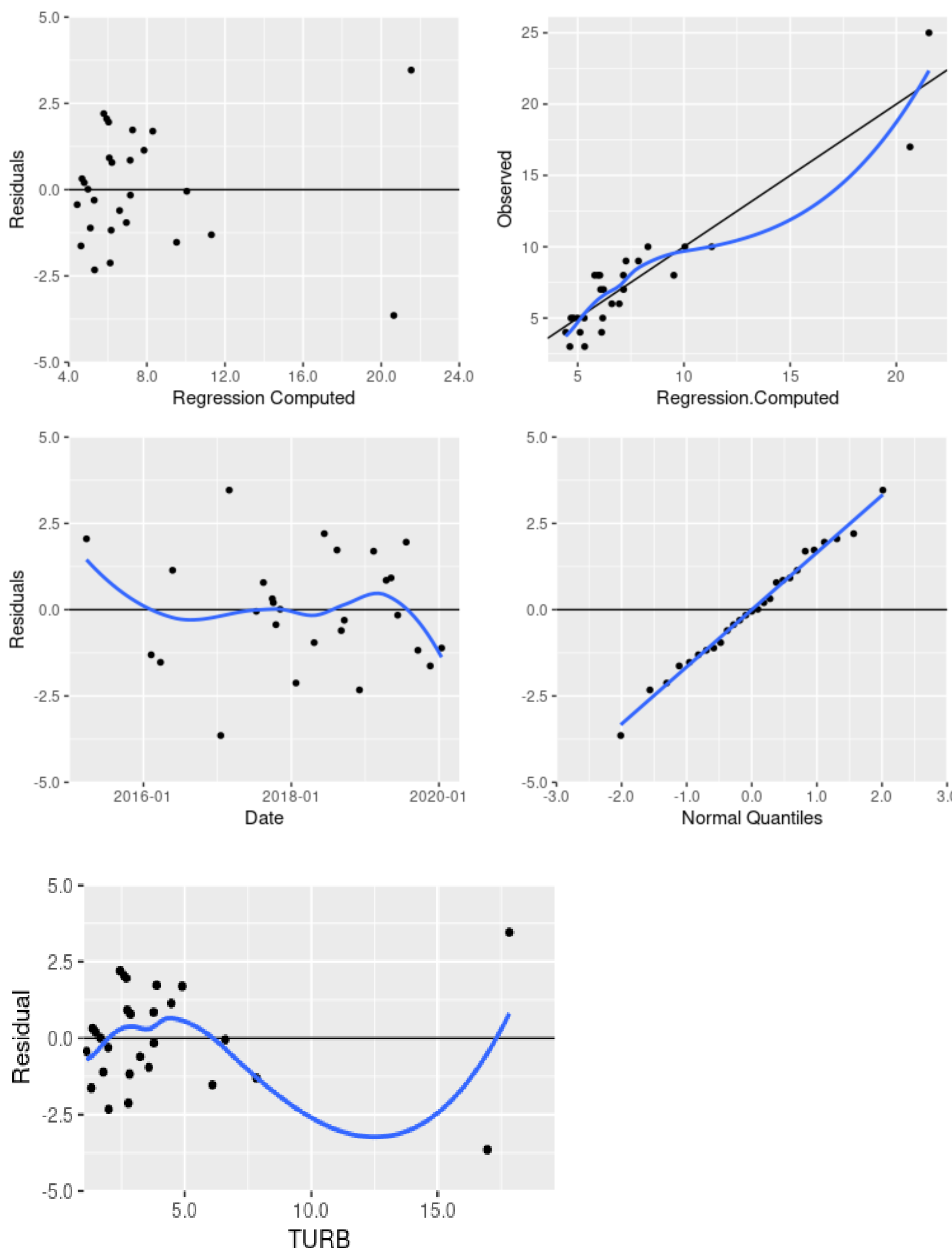
Date	Time	SSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
1/17/2017	13:30	17	20.6	-3.64	-2.84	-3.39	0.401	2.7	-2.77
2/28/2017	12:48	25	21.5	3.46	2.83	3.36	0.452	3.3	3.05

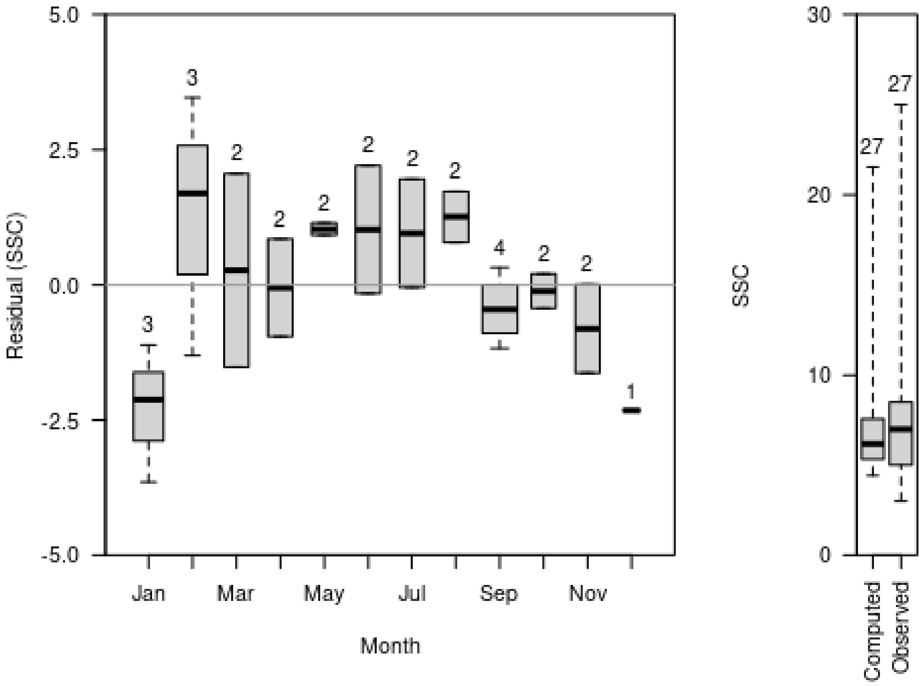
Plots of $\log_{10}$SSC and explanatory variables and residual diagnostic plots

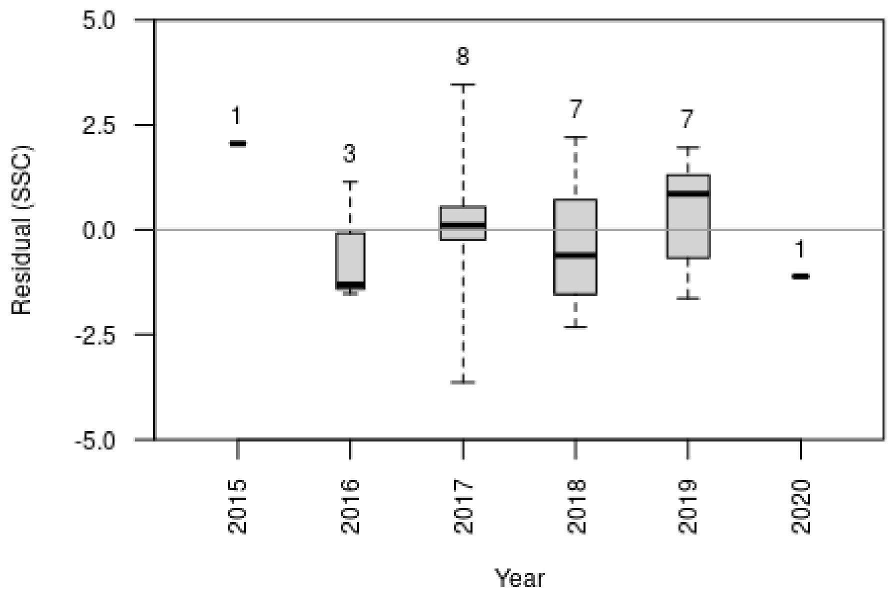
The following plots were generated online using a specialized R-Script developed by Patrick Eslick of the KSWSC and is located at the following address:

https://patrickeslick.github.io/ModelArchiveSummary/

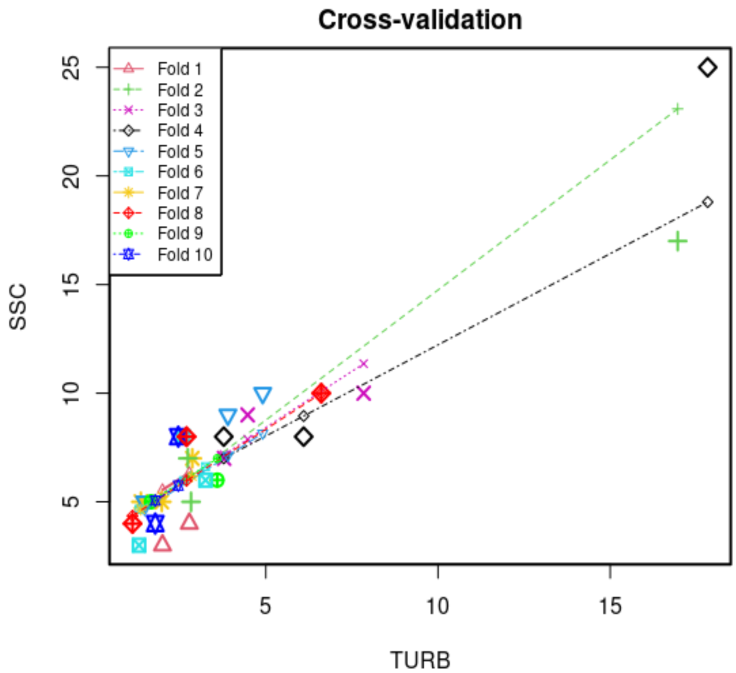
Statistical Plots



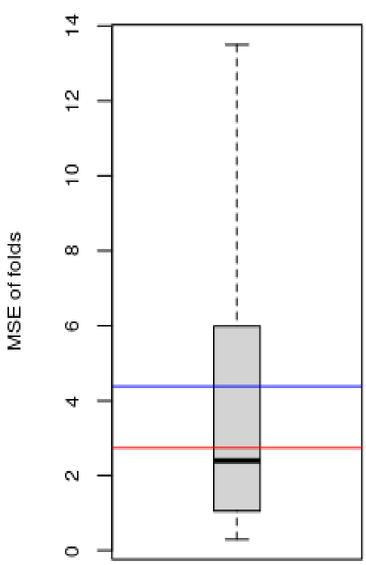

Residual (SSC)
Month
Jan
Mar
May
Jul
Sep
Nov
SSC
Computed
Observed


Residual (SSC)
Year
2015
2016
2017
2018
2019
2020

The graph below shows a k-fold cross validation with k=10 and the large points represent observations that were left out of each fold and are identified by the color and shape.




```
                Minimum MSE of folds:   0.294
                   Mean MSE of folds:   4.370
                 Median MSE of folds:   2.390
                Maximum MSE of folds:  13.500
 (Mean MSE of folds) / (Model MSE):     1.600
```



Red line - Model MSE
Blue line - Mean MSE of folds

Calibration Dataset

Observation Number	DateTime	SSC	TURB	Computed SSC	Residual	Normal Quantile	Censored Values
1	3/25/2015 9:37	8	2.6	6	2.05	1.31	--
2	2/8/2016 12:38	10	7.8	11	-1.31	-0.82	--
3	3/25/2016 12:45	8	6.1	10	-1.52	-0.96	--
4	5/24/2016 15:51	9	4.5	8	1.14	0.70	--
5	1/17/2017 13:30	17	17.0	21	-3.64	-2.01	--
6	2/28/2017 12:48	25	17.8	22	3.46	2.01	--
7	7/12/2017 12:14	10	6.6	10	-0.05	0.00	--
8	8/16/2017 12:48	7	2.9	6	0.79	0.38	--
9	9/29/2017 11:45	5	1.4	5	0.32	0.28	--
10	10/4/2017 10:01	5	1.5	5	0.20	0.19	--
11	10/17/2017 12:44	4	1.1	4	-0.43	-0.28	--
12	11/7/2017 12:01	5	1.7	5	0.01	0.09	--
13	1/24/2018 11:44	4	2.8	6	-2.12	-1.31	--
14	4/24/2018 12:43	6	3.6	7	-0.95	-0.48	--
15	6/13/2018 10:29	8	2.5	6	2.20	1.56	--
16	8/15/2018 10:47	9	3.9	7	1.73	0.96	--
17	9/5/2018 12:11	6	3.3	7	-0.61	-0.38	--
18	9/20/2018 11:31	5	2.0	5	-0.30	-0.19	--
19	12/4/2018 15:00	3	2.0	5	-2.33	-1.56	--
20	2/12/2019 13:20	10	4.9	8	1.69	0.82	--
21	4/15/2019 11:30	8	3.8	7	0.85	0.48	--
22	5/9/2019 12:10	7	2.7	6	0.92	0.58	--
23	6/11/2019 10:26	7	3.8	7	-0.16	-0.09	--
24	7/23/2019 11:59	8	2.7	6	1.96	1.12	--
25	9/19/2019 13:52	5	2.8	6	-1.18	-0.70	--
26	11/19/2019 11:43	3	1.3	5	-1.63	-1.12	--
27	1/14/2020 12:02	4	1.8	5	-1.11	-0.58	--

Definitions

SSC: Suspended sediment concentration (SSC) in mg/L (80154)
TURB: Turbidity in FNU (63680)

App Version 1.0

References

Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate Analysis and Index Developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., https://pubs.usgs.gov/of/2015/1177/ofr20151177.pdf.

Edwards TK and Glysson GD. 1999. Field methods for measurement of fluvial sediment: U.S. Geological Survey Techniques of Water-Resources Investigations. Book 3, Chap. C2. 89 p. Available from: https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI_3-C2.pdf.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources-Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources investigations, book 4, chap. A3, 510 p.

Levesque, V.A., and Oberg, K.A., 2012, Computing discharge using the index velocity method: U.S. Geological Survey Techniques and Methods 3-A23, 148 p. (Also available at http://pubs.usgs.gov/tm/3a23/.)

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Available from: https://www.R-project.org/.

Rasmussen P, Gray JR, Glysson GD, Ziegler AC. 2009. Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data. Book 3 Applications of Hydraulics, Section C. 52 p. Available from: https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf.

[USGS] U.S. Geological Survey. 2006. National field manual for the collection of water quality data: U.S. Geological Survey Techniques of Water-Resources Investigations. Book 9, Chapter A4. Available from: https://pubs.usgs.gov/twri/twri9a4/twri9a4_Chap4_v2.pdf.

[USGS] U.S. Geological Survey, 2014, Policy and guidelines for archival of surface-water, groundwater, and water-quality model applications: Office of Groundwater Technical Memorandum 2015.02, Office of Surface Water Technical Memorandum 2015.01, Office of Water Quality Technical Memorandum 2015.01, Available from: https://water.usgs.gov/admin/memo/SW/sw2015.01.pdf

[USGS] U.S. Geological Survey. 2016. Policy and guidance for approval of surrogate regression models for computation of time series suspended-sediment concentrations and loads: Office of Surface Water Technical Memorandum 2016.07. Available from: https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf.

Wagner RJ, Boulger RW, Jr, Oblinger CJ, Smith BA. 2006. Guidelines and standard procedures for continuous waterquality monitors: station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1-D3. Available from: https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf.