Model Archive Summary for Suspended-Sediment Concentration at U.S. Geological Survey Station 11312676; Middle River at Middle River, California

This model archive summary describes the suspended-sediment concentration (SSC) model developed to compute a 15--minute SSC time-series for the period of record: October 1, 2010 to January 12, 2015. This is the first suspended-sediment model developed for the site. The methods used follow U.S. Geological Survey (USGS) guidance as referenced in the Office of Surface Water/Office of Water Quality Technical Memorandum and USGS Techniques and Methods, book 3 chapter 4 (USGS, 2016; 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, California (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 6-series sonde began logging turbidity with a model 6136 sensor on December 3, 2009 and was removed on January 12, 2015.

Model number: 11312676.SSC.WY11.1
Model calibration data period: November 23, 2010 – December 19, 2014
Model application date: October 1, 2010 – January 12, 2015
Computed by: Anna Conlen, USGS, Sacramento, CA (aconlen@usgs.gov)
Reviewed by: Tara Morgan-King, USGS, Sacramento, CA (tamorgan@usgs.gov)

Physical Sampling Details and Sediment Data

All sediment data were collected using USGS protocols and are stored in the National Water Information System (NWIS) database: https://waterdata.usgs.gov/nwis (USGS, 2006). Discrete, boat-based samples were collected seasonally, spanning the range of site conditions and specifically targeting large sediment transport events.

Sample collection is consistent with approved field methods described in Edwards and Glysson (1999) and USGS (2006). Samples were collected using either the Equal Discharge Increment (EDI) or Equal Width Increment (EWI) method. The EWI method was used to collect two samples in the calibration dataset and the EDI method was used to collect the remaining 23 samples. Samples were predominantly collected using the EDI method because velocities were not always isokinetic due to the tidal nature of the site (from Table 4-5 of TWRI09A4; USGS, 2006). A boat-based discharge measurement was collected immediately before EDI sampling with an Acoustic Doppler Current Profiler (ADCP) to determine the location of each vertical. A Federal Interagency Sedimentation Project (FISP) US D-74 sampler was used to collect one sample and the remaining depth-integrated samples were collected using a FISP US D-96 bag sampler. The channel cross section can approach 28 feet deep in the thalweg with a mean

depth of approximately 16 feet. Velocities during the model calibration data period ranged from -1.494 ft/s to +1.383 ft/s. Sediment at this station is mostly fines (88% fines on average) and any potential sampling bias due to non-isokinetic sampling is considered minimal.

Samples collected before January 2012 were analyzed for SSC (mg/L) by the filtration method at the USGS Sediment Laboratory in Marina, California, while those collected after January 2012 were analyzed for SSC by the USGS Sediment Laboratory at its current location in Santa Cruz, California. Many samples were also analyzed for the percentage of fines (<0.063 mm), which can be used to identify outliers. EWI verticals were composited to a single container and the EDI sample collected on January 31, 2012 was composited. Each of the five verticals from the remaining EDI samples were analyzed individually by the lab for quality control purposes. The average SSC from these five verticals was computed and used in the calibration dataset. Sediment results are publicly available on NWIS.

All sediment data were reviewed and approved in the USGS NWIS Water-Quality System database (QWDATA) before being applied in the calibration model.

Surrogate Data

Continuous, 15-minute turbidity data, reported in Formazin Nephelometric Turbidity Units (FNU) and hourly, tidally-filtered discharge data (QFT), reported in cubic feet per second (cfs), were evaluated as explanatory variables for SSC. Turbidity and QFT time-series data were collected by the USGS California Water Science Center and are located at: https://waterdata.usgs.gov/usa/nwis/uv?site_no=11312676. Turbidity data were analyzed and approved per USGS guidelines (Wagner and others, 2006). QFT data were computed, analyzed and approved before using in the sediment calibration model. Methods to compute discharge (and thus QFT) follow Levesque and Oberg (2012).

Model Calibration Dataset

The USGS Surrogate Analysis and Index Developer Tool (SAID) was used to pair surrogate data with discrete sediment data (Domanski and others, 2015). Turbidity and QFT values were paired with each suspended sediment sample with a matching window of ± 15 minutes and ± 30 minutes, respectively. The SAID manual is available at: https://pubs.er.usgs.gov/publication/ofr20151177. Turbidity data was missing for two samples; there are data gaps in the turbidity time-series during the measurements on June 4, 2014 and September 10, 2014.

Two EDI sets (A and B) were collected on December 14, 2012. The average time between the two sets is less than 45 minutes, so the event average was used in the calibration model dataset.

The sample on February 19, 2014 was not included in the final calibration model dataset because: 1) the sample was flagged in SAID as an outlier, 2) the sample has a higher percentage of sand compared to other samples which may be an indication of sampling error and 3) the US

D-74 sampler was used to collect this sample, which is the incorrect choice as sampling depths were greater than 15 feet.

The final calibration dataset is compiled from 21 concurrent measurements of SSC, turbidity and QFT. Summary statistics and the complete model calibration dataset are provided in the following sections.

Model Development

Simple linear regression (SLR) models and multiple linear regression (MLR) models were assessed using methods described in Helsel and Hirsch (2002). Three models were evaluated: Model 1) linear model with one explanatory variable (turbidity), Model 2) $\log_{10}$-transformed model with one explanatory variable (turbidity) and Model 3) linear model with two explanatory variables (turbidity and QFT). The addition of $\log_{10}$-transformed QFT in a multi-log model reduced the number of samples to n=2, as most of the QFT values are negative and cannot be transformed. A $\log_{10}$-transformed MLR was therefore not considered.

Diagnostic statistics and plots for model review were output using a combination of Matlab, SAID, and the R environment (R Core Team, 2018). Table 3 in Rasmussen and others (2009) shows the best statistical diagnostics to help evaluate regression models. The best model was chosen based on residual plots, coefficient of determination (R^2), root-mean-squared error (RMSE), Mean Square Prediction Error (MSPE), significance tests (p-values) and prediction error sum of squares (PRESS) statistics. RMSE and PRESS statistics cannot be used to compare regressions with different response variable units, so R^2, MSPE values and residual plots were used as the main determinants of model strength when comparing $\log_{10}$-transformed and untransformed models. Values for these statistics were computed for three models and are included in the table below. The best SLR model is a log model with $\log_{10}$-transformed turbidity as the surrogate (Model 2 - highlighted in table below).

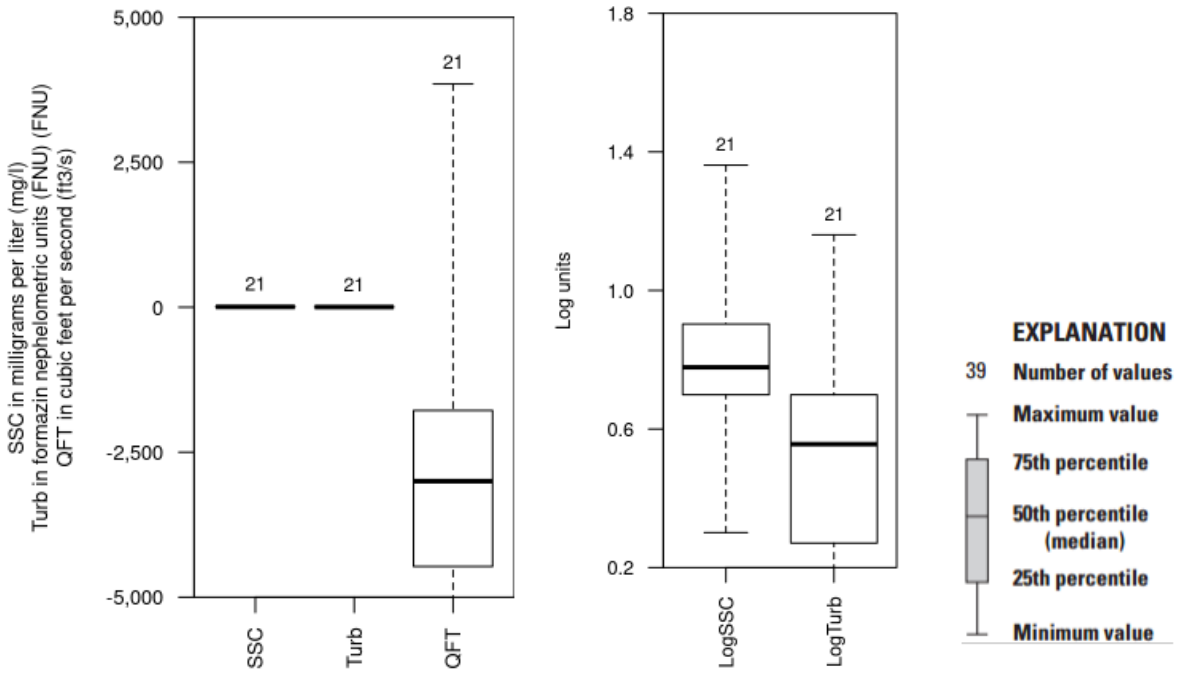
No.	R^2	R^2_a	RMSE	PRESS	MSPE	n	Type
Model 1	0.296	0.258	4.30	488.28	55.3	21	Linear
Model 2	0.488	0.461	0.18	0.73	42.5	21	Log
Model 3	0.330	0.256	4.30	609.91	55.4	21	Multi-linear

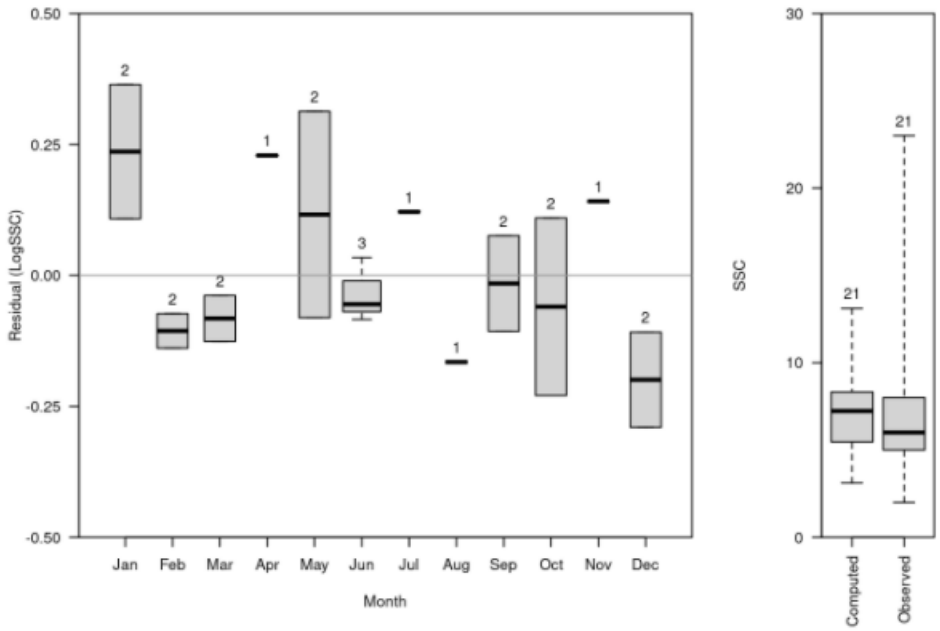
Flagged observations from the SAID outlier test criteria were evaluated. Studentized residuals from the models were inspected for values greater than three or less than negative three; values outside this range are considered potential extreme outliers. The studentized residuals were reviewed from the output reports and none of the samples were deemed to be extreme outliers. All 21 observations were retained in the model.

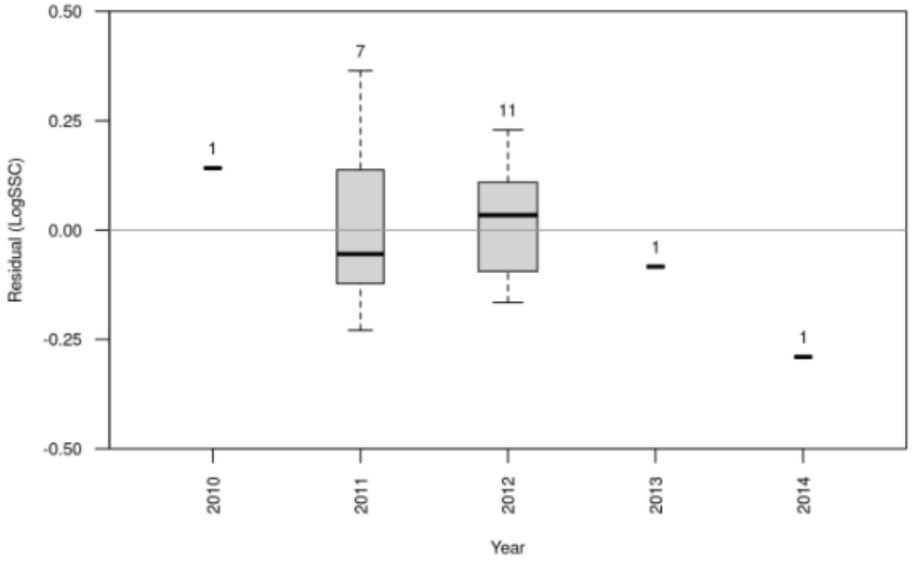
Plots

The following plots were generated using a R-based application (Version 1.0) developed by Patrick Eslick of the USGS Kansas Water Science Center. It is available at: http://kswsc.cr.usgs.gov:3838/peslick/ModelArchiveSummary/.

Boxplots of turbidity, QFT and SSC data show the range of measured data for each parameter. The second set of boxplots show SSC residuals of the SLR log model by month and water year.

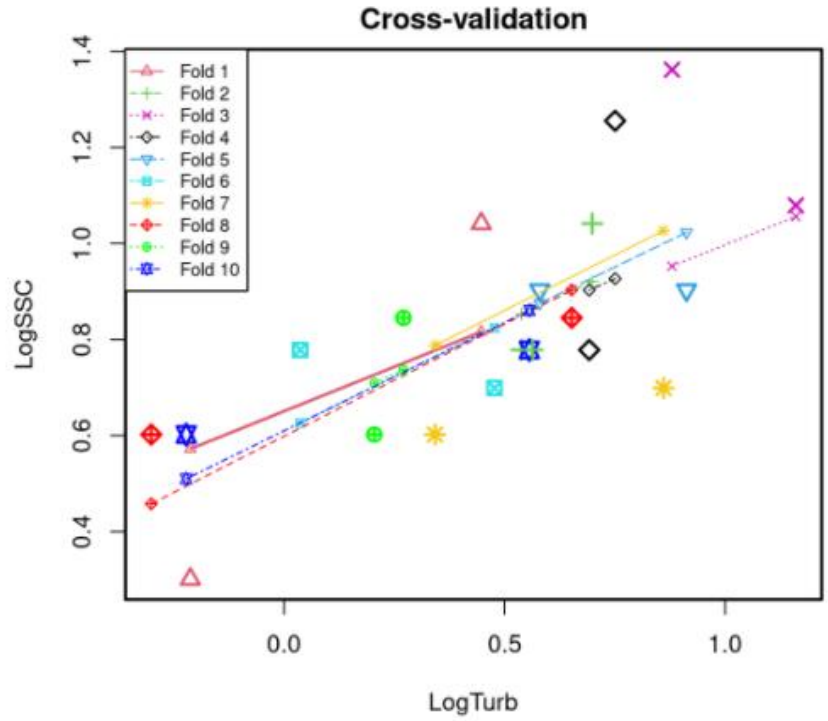



0.50
0.25
0.00
-0.25
-0.50
Residual (LogSSC)
2
2
2
1
2
3
1
1
2
2
1
2
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
30
20
10
0
SSC
21
21
Computed
Observed

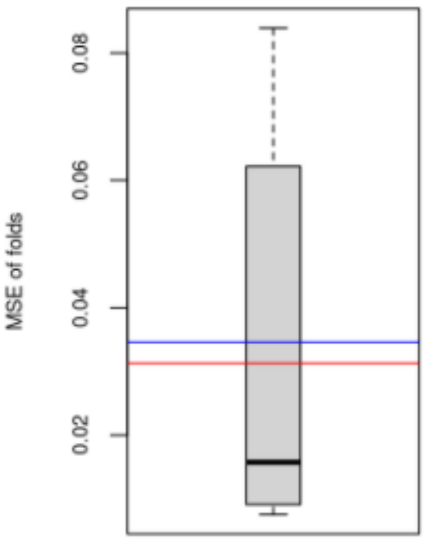

0.50
0.25
0.00
-0.25
-0.50
Residual (LogSSC)
1
7
11
1
1
2010
2011
2012
2013
2014
Year

Cross Validation

The cross-validation plot below shows a k-fold validation with k=10 for the final model. The points represent observations that were left out of each fold.



Minimum MSE of folds:	0.008
Mean MSE of folds:	0.035
Median MSE of folds:	0.016
Maximum MSE of folds:	0.084
(Mean MSE of folds) / (Model MSE):	1.11



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

Model Summary

The final SSC model at MDM is a $\log_{10}$-transformed SLR model based on 21 concurrent measurements of SSC and turbidity collected over approximately four water years. The model is shown below with basic model information, regression coefficients, correlation, summary statistics and Duan's bias correction factor (Duan, 1983):

Linear Regression Model	Coefficient of Determination (R^2)
$log_{10}SSC = 0.621 + 0.428 * log_{10}Turb$	0.488

where
SSC = suspended-sediment concentration, in milligrams per liter (mg/L) and
Turb = turbidity, in formazin nephelometric units

SSC was transformed during regression model development, so the computed prediction may be biased and needs to be multiplied by a non-parametric smearing bias correction factor (BCF) when it is retransformed, shown below.

Model	Start date	End date	Linear Regression Model	*BCF*
1	10/01/2010	01/12/2015	$SSC = 10^{0.621} \times Turb^{0.428} \times BCF$	1.082

The SSC timeseries is computed from USGS turbidity data. Minimum and maximum turbidity values for the model application period are listed below. SSC timeseries data exceeding extrapolation limits were removed. This model cannot be used to extrapolate more than 10% above or below the range of samples in the calibration dataset (USGS, 2016). The extrapolated, maximum computed SSC for this model is 25 mg/L. The original maximum, computed SSC was 20 mg/L.

Parameter	Minimum	Maximum
Computed SSC (mg/L)	0	25
Turbidity (FNU)	0	32

Suspended-Sediment Concentration Record

The SSC record is computed using this regression model on the USGS National Real-Time Water Quality (NRTWQ) website. The complete record can be found at: https://nrtwq.usgs.gov/ca.

Model

Log10SSC = 0.621 + 0.428Log10Turb

Variable Summary Statistics

	Turb	log10Turb	SSC	log10SSC
Minimum	0.50	-0.30	2	0.30
1st Quartile	1.80	0.25	4.75	0.67
Median	3.60	0.56	6	0.78
Mean	4.11	0.47	7.76	0.82
3rd Quartile	5	0.71	9	0.94
Maximum	14.47	1.16	23	1.36

Basic Model Statistics

Number of observations	21
Root Mean Squared Error (RMSE)	0.177
Model Standard Percentage Error (MSPE)	42.48
Coefficient of determination (R^2)	0.488
Adjusted R^2	0.461
Bias Correction Factor	1.082

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.621	0.061	10.200	4.04E-09
log10Turb	0.428	0.101	4.252	4.31E-04

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.775
E.vars	-0.775	1.000

Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.286	0.193	0.617

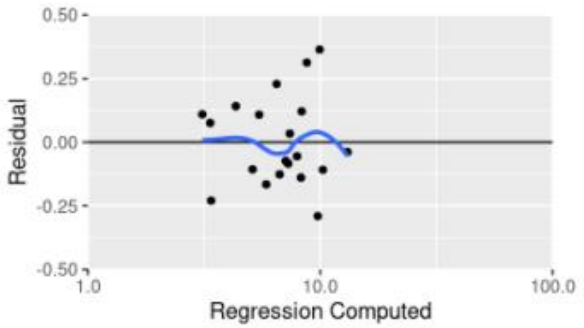
Flagged Observations

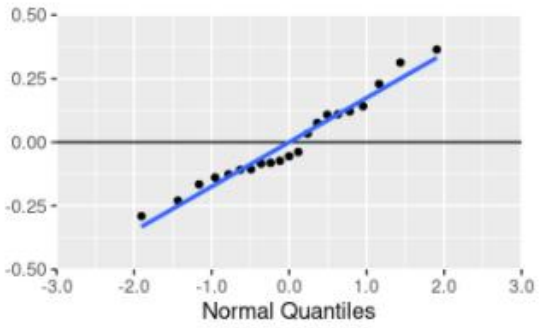
Date	Time	LogSSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
1/19/2011	12:05	1.36	0.997	0.364	2.170	2.440	0.102	0.267	0.821
10/19/2011	12:20	0.30	0.53	-0.229	-1.450	-1.490	0.199	0.26	-0.744

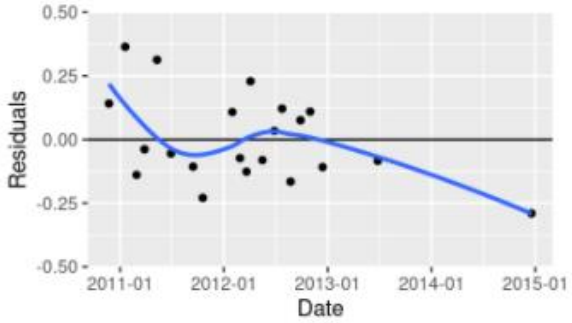
Residual diagnostic plots

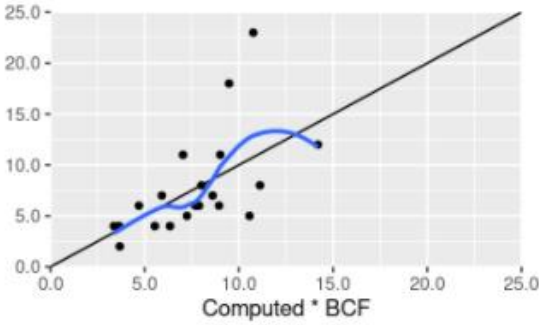
Plots were generated using the model archive summary application developed by Patrick Eslick of the USGS Kansas Water Science Center.

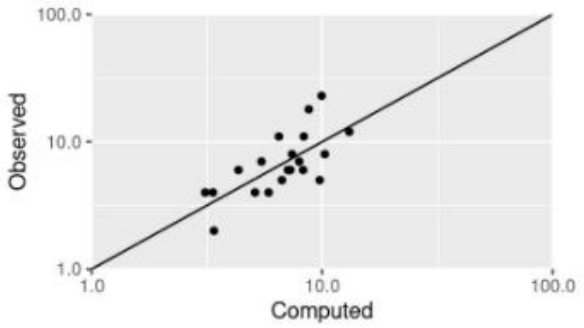
Statistical Plots

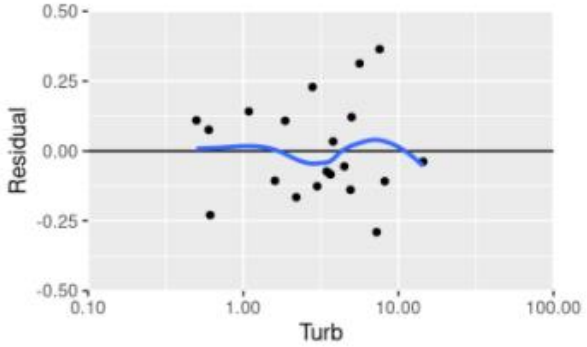












Model-Calibration Dataset

	Date	LogSSC	LogTurb	SSC	Turb	Computed LogSSC	Computed SSC	Residual	Normal Quantiles	Censored Values
0										
1	11/23/2010 12:00	0.778	0.0366	6	1.09	0.637	4.69	0.141	0.955	--
2	1/19/2011 12:05	1.36	0.88	23	7.58	0.997	10.8	0.364	1.91	--
3	2/28/2011 14:00	0.778	0.692	6	4.92	0.917	8.94	-0.139	-0.955	--
4	3/28/2011 10:53	1.08	1.16	12	14.5	1.12	14.2	-0.0384	0.119	--
5	5/11/2011 16:25	1.26	0.751	18	5.63	0.942	9.48	0.313	1.44	--
6	6/29/2011 12:42	0.845	0.652	7	4.49	0.9	8.6	-0.0551	0	--
7	9/15/2011 11:28	0.602	0.205	4	1.6	0.709	5.54	-0.107	-0.492	--
8	10/19/2011 12:20	0.301	-0.212	2	0.613	0.53	3.67	-0.229	-1.44	--
9	1/31/2012 11:47	0.845	0.271	7	1.87	0.737	5.91	0.108	0.492	--
10	2/27/2012 12:43	0.778	0.538	6	3.45	0.851	7.69	-0.0731	-0.119	--
11	3/21/2012 11:35	0.699	0.477	5	3	0.825	7.24	-0.126	-0.782	--
12	4/4/2012 11:48	1.04	0.447	11	2.8	0.812	7.03	0.229	1.16	--
13	5/16/2012 10:34	0.778	0.556	6	3.6	0.859	7.82	-0.0809	-0.239	--
14	6/27/2012 10:44	0.903	0.58	8	3.8	0.869	8.01	0.0339	0.239	--
15	7/24/2012 11:35	1.04	0.699	11	5	0.92	9.01	0.121	0.782	--
16	8/23/2012 11:54	0.602	0.342	4	2.2	0.768	6.34	-0.166	-1.16	--
17	9/27/2012 11:35	0.602	-0.222	4	0.6	0.526	3.64	0.0758	0.362	--
18	10/31/2012 11:37	0.602	-0.301	4	0.5	0.492	3.36	0.11	0.631	--
19	12/14/2012 11:25	0.903	0.913	8	8.18	1.01	11.1	-0.108	-0.631	--
20	6/26/2013 12:34	0.778	0.563	6	3.66	0.862	7.88	-0.0839	-0.362	--
21	12/19/2014 13:53	0.699	0.86	5	7.25	0.989	10.6	-0.29	-1.91	--

Definitions

SSC: Suspended sediment concentration (SSC) in mg/l (80154)
Turb: Turbidity in FNU (63680)

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.

Duan, Naihua, 1983, Smearing estimate – A nonparametric retransformation method: Journal of the American Statistical Association, Volume 78-383, 605-610 p.

Edwards, T.K. and Glysson, G.D., 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. Available from: https://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.P., Gray J.R., Glysson G.D., 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, 53 p., Available from: https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf.

[USGS] 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. 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, Office of Water Quality

Technical Memorandum 2016.10, Available from: https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf.

Wagner RJ, Boulger RW, Oblinger CJ, Smith BA, 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, Available from: https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf.