Model Archive Summary for Suspended-Sediment Concentration at U.S. Geological Survey Station 11455335; Sacramento River Deep Water Ship Channel near Rio Vista, 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 during August 4, 2008 to May 5, 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 2016.07/2016.10 and USGS Techniques and Methods, book 3 chapter 4 (U.S. Geological Survey, 2016; Rasmussen and others, 2009). This model archive summary is in accordance with Attachment A of Office of Water Quality Technical Memorandum 2015.01 (U.S. Geological Survey, 2014).

Site and Model Information

Site number: 11455335
Site name: Sacramento River Deep Water Ship Channel near Rio Vista, CA (DWS)
Location: Latitude 38°15'22", longitude 121°40'00" referenced to North American Datum of 1983, Solano County, CA. Hydrologic Unit 18020163, on Channel Marker 54.
Equipment: A YSI 6-series sonde began logging turbidity with a model 6136 sensor on August 4, 2008 and was removed on May 5, 2015.

Model number: 11455335.SSC.WY08.1
Model calibration data period: August 14, 2008 – March 24, 2015
Model application date: August 4, 2008 – May 5, 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 (U.S. Geological Survey, 2006). Discrete, boat-based samples were collected seasonally, spanning the range of site conditions, and specifically targeting high flow & sediment transport events.

Sample collection is consistent with approved field methods described in Edwards and Glysson (1999) and U.S. Geological Survey (2006). The equal-discharge-increment (EDI) method was used to determine the locations of five sampling verticals along the transect where discharge-weighted, suspended sediment samples were collected. The EDI method was used because velocities are not always isokinetic due to the tidal nature of the site (from Table 4-5 of TWRI09A4; U.S. Geological Survey, 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 sampling vertical. A Federal Interagency Sedimentation Project US D-96 bag sampler was used to collect depth-integrated samples. The channel cross section can reach depths up to 38 feet in the thalweg with an average depth of about 18 feet. Velocities during the model calibration data period ranged from -2.22 ft/s to +1.71 ft/s. Sediment at this station is mostly fines (97% fines on average) and any potential sampling bias due to non-isokinetic

sampling is considered minimal. Sampling bias can occur with the presence of sand but 95% of the samples analyzed for percent fines were more than 90%.

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. Each of the five EDI verticals 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/ca/nwis/uv/?site_no=11455335. Turbidity data were analyzed and approved following USGS guidelines (Wagner and others, 2006). QFT data were computed, reviewed, and approved following Levesque and Oberg (2012).

Model Calibration Dataset

The USGS Surrogate Analysis and Index Developer Tool (SAID) was used to pair concurrent continuous time-series data with discrete SSC data (Domanski and others, 2015). Concurrent turbidity and QFT values were selected for each discrete SSC sample by selecting the closest value within ± 15 minutes and ± 30 minutes, respectively.

Two EDI sets were collected on August 14, 2008 and January 12, 2009. The sampling time span for the two sets on each date exceeds one hour and the times of the two sample averages were over 45 minutes apart. Both samples were included in the calibration dataset for each date.

The final calibration dataset is compiled from 34 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 others (2020). Four 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) linear model with two explanatory variables (turbidity and QFT) and Model 4) $\log_{10}$-transformed model with two explanatory variables (turbidity and QFT).

Diagnostic statistics and plots for model review were computed using a combination of Matlab, SAID and the R environment (Matlab, 2019, Domanski and others, 2015, 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), adjusted R^2_a (a measure for comparing models with differing numbers of explanatory variables because it is adjusted

for the degrees of freedom), 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 four models and are included in the table below. The best SLR model was a log_{10}-transformed using turbidity as the explanatory variable (highlighted in table below). Adding QFT as a second explanatory variable did not improve the model significantly. Though the linear SLR had a slightly higher R^2 value than the log_{10}-transformed model, the R^2_a was essentially equivalent, and the log model had a slightly lower MSPE value. The residual plot and normal probability of residuals plot indicate the log model best fits the data. The normal probability plot of residuals for the log model was approximately linear with less skew than the linear model, and the plot of residual versus fitted values for the log model suggests that the variances of the residuals are more homoscedastic.

QFT was not considered further as an explanatory variable because: 1) QFT was not significant in the log MLR model (p-value > 0.05), 2) the MLR model calibration datasets contain either 33 or 20 observations, though a total of 48 samples are recommended when a second explanatory variable is included (U.S. Geological Survey, 2016) and 3) including QFT in the final model would limit the computed time-series to an hourly record rather than a 15-minute record.

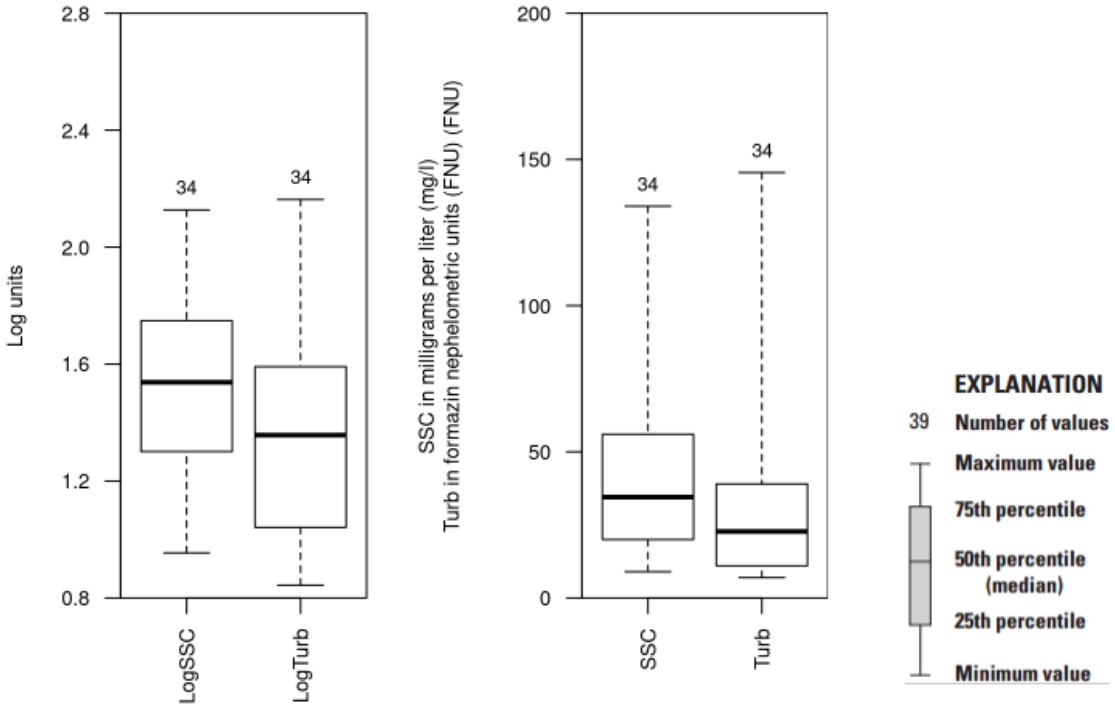
No.	Model Formula	R^2	R^2_a	RMSE	PRESS	MSPE	n	p-value Q	Model Type
Model 1	SSC ~ TURB	0.854	0.85	10.92	4575	26.4	34	na	SLR linear
Model 2	log_{10}SSC ~ log_{10}TURB	0.850	0.846	0.11	0.45	26.1	34	na	SLR log_{10}
Model 3	SSC ~ TURB + QFT	0.873	0.864	10.52	4261	25.2	33	0.05	MLR linear
Model 4	log_{10}SSC ~ log_{10}TURB + log_{10}QFT	0.874	0.860	0.12	0.31	27.8	20	0.98	MLR log_{10}

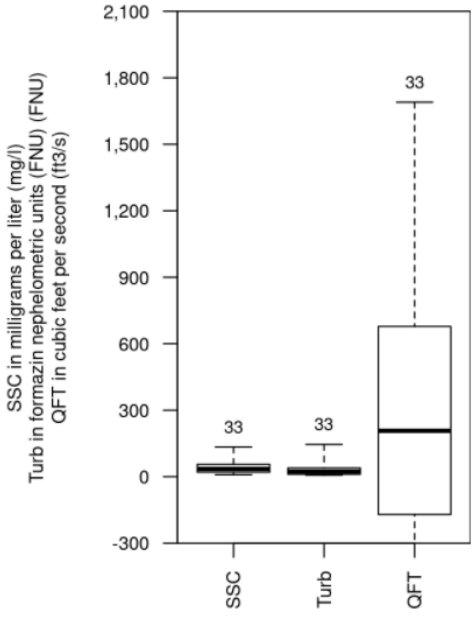
Flagged observations from the SAID outlier test criteria were evaluated. Studentized residuals from the models were inspected for values greater than three or less than negative three; values outside this range are considered potential extreme outliers. The studentized residuals were reviewed from the output reports and none of the samples were deemed to be extreme outliers. All 34 observations were retained in the model.

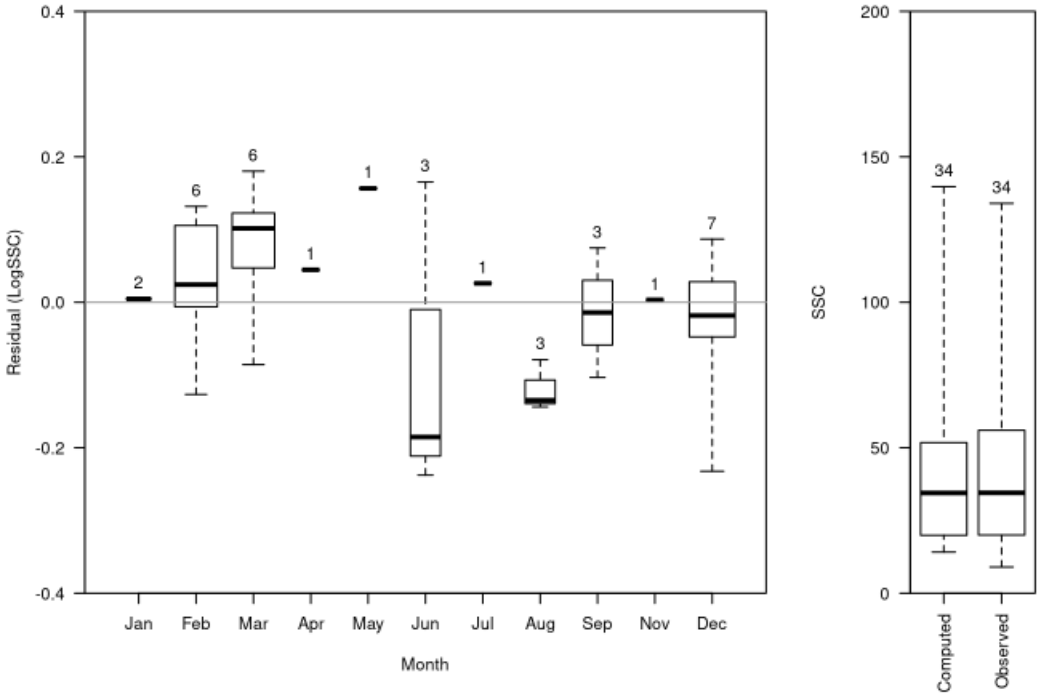
Plots

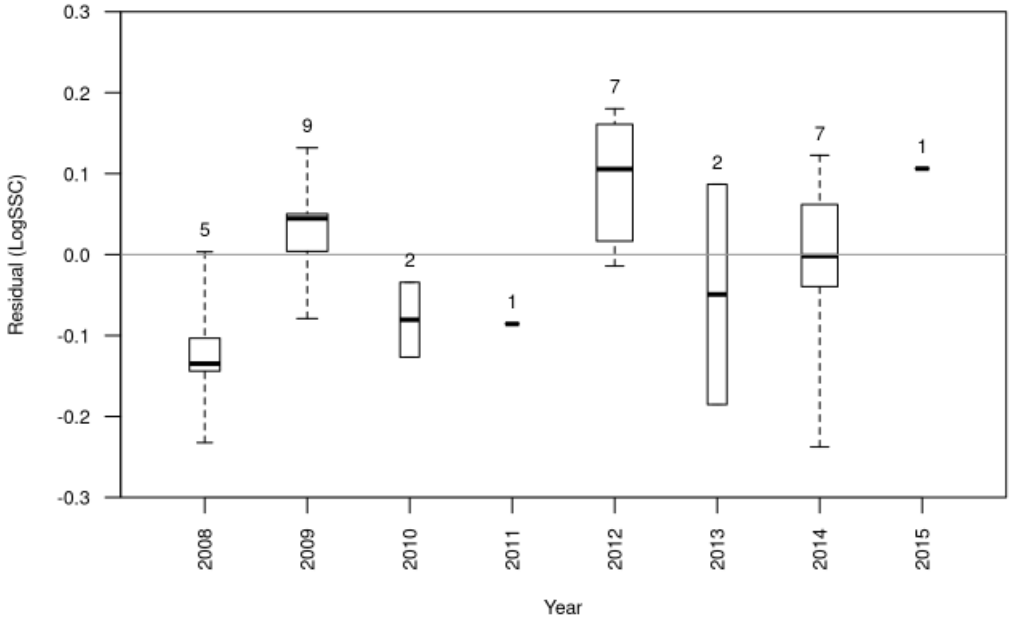
The following plots were generated using a R-based application (Version 1.0) developed by Patrick Eslick of the USGS Kansas Water Science Center. It is available internally for USGS personnel at: http://kswsc.cr.usgs.gov:3838/peslick/ModelArchiveSummary/.

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


Log units
2.8
2.4
2.0
1.6
1.2
0.8
34
34
LogSSC
LogTurb
SSC in milligrams per liter (mg/l)
Turb in formazin nephelometric units (FNU) (FNU)
200
150
100
50
0
34
34
SSC
Turb
EXPLANATION
39 Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value

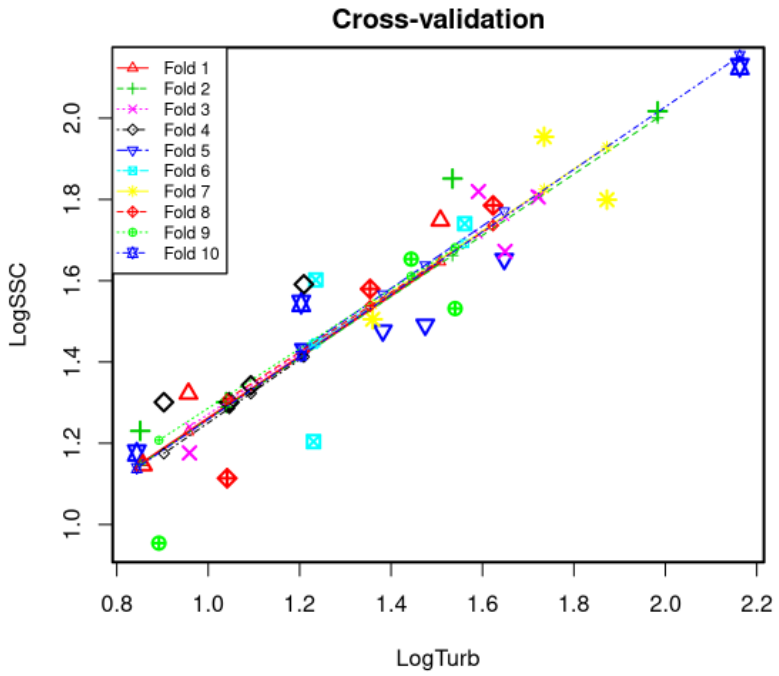

SSC in milligrams per liter (mg/l)
Turb in formazin nephelometric units (FNU) (FNU)
QFT in cubic feet per second (ft3/s)
2,100
1,800
1,500
1,200
900
600
300
0
-300
33
33
33
SSC
Turb
QFT


Residual (LogSSC)
0.4
0.2
0.0
-0.2
-0.4
2
6
6
1
1
3
1
3
3
1
7
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Nov
Dec
Month
SSC
200
150
100
50
0
34
34
Computed
Observed

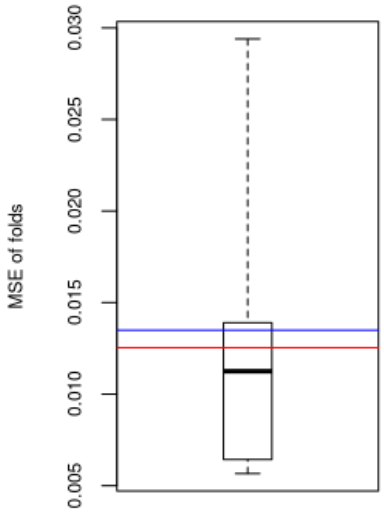

Residual (LogSSC)
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
5
9
2
1
7
2
7
1
2008
2009
2010
2011
2012
2013
2014
2015
Year

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:	0.00566
Mean MSE of folds:	0.0135
Median MSE of folds:	0.0112
Maximum MSE of folds:	0.0294
(Mean MSE of folds) / (Model MSE):	1.08



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

Model Summary

The selected SSC model at DWS was a $\log_{10}$-transformed SLR model based on 34 concurrent measurements of SSC and turbidity collected over eight 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)
$log_{10}SSC = 0.513 + 0.754 * log_{10}Turb$	0.850

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

The log-transformed model may be retransformed to the original units to calculate SSC directly. A bias is introduced during retransformation and can be corrected using a non-parametric smearing bias correction factor (BCF), Duan's BCF (Duan, 1983).

Model	Start date	End date	Linear Regression Model	*BCF*
1	08/04/2008	05/05/2015	$SSC = 10^{0.513} \times Turb^{0.754} \times BCF$	1.031

The SSC time-series is computed from USGS turbidity data. Minimum and maximum turbidity values for the model application period are listed below. SSC time-series data exceeding extrapolation limits were removed from public display using a threshold. This model cannot be used to extrapolate more than 10% above or below the range of samples in the calibration dataset (U.S. Geological Survey, 2016). The extrapolated, maximum computed SSC for this model is 147 mg/L. The original maximum, computed SSC was 304 mg/L. The higher SSC values primarily occur from short lived peaks caused by vessel traffic and less than 1% of the time-series exceeds the extrapolation threshold.

Parameter	Minimum	Maximum
Computed SSC (mg/L)	0	147
Turbidity (FNU)	0	391

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 = 0.513 + 0.754log10Turb

Variable Summary Statistics

	Turb	log10Turb	SSC	log10SSC
Minimum	6.98	0.84	9	0.95
1st Quartile	11	1.04	20	1.30
Median	22.75	1.36	34.5	1.54
Mean	30.62	1.34	41.44	1.53
3rd Quartile	39	1.59	56	1.75
Maximum	146	2.16	134	2.13

Basic Model Statistics

Number of observations	34
Root Mean Squared Error (RMSE)	0.11
Model Standard Percentage Error (MSPE)	26.2
Coefficient of determination (R^2)	0.850
Adjusted R^2	0.846
Bias Correction Factor	1.031

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.513	0.08	6.61	1.88E-07
log10Turb	0.754	0.06	13.48	9.71E-15

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.969
E.vars	-0.969	1.000

Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.176	0.194	0.485

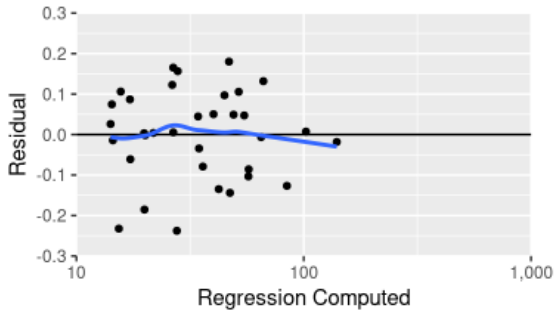
Flagged Observations

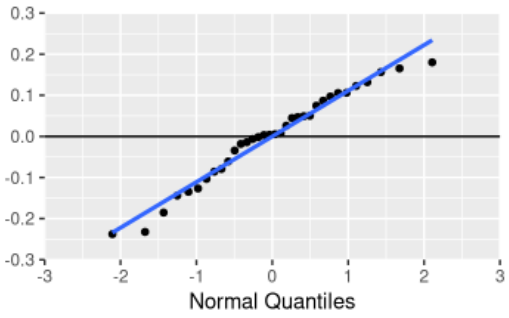
Date	Time	LogSSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
12/4/2008	15:55	0.954	1.19	-0.232	-2.160	-2.310	0.081	0.205	-0.683
12/16/2014	13:16	2.130	2.15	-0.018	-0.182	-0.179	0.197	0.004	-0.089

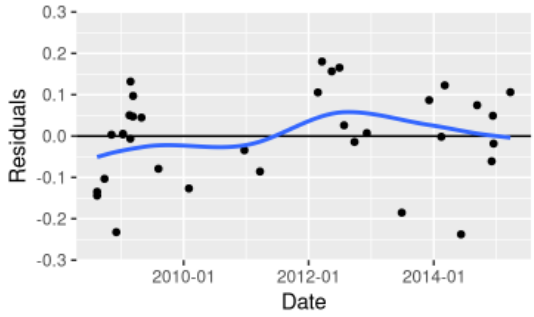
Residual diagnostic plots

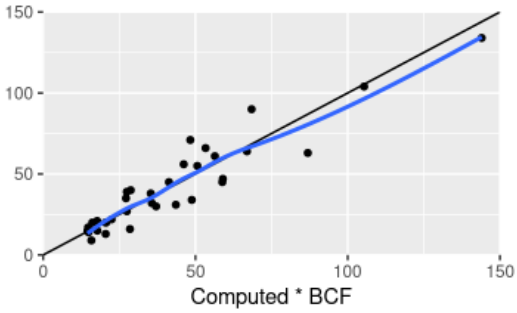
Plots were generated using the model archive summary application developed by Patrick Eslick of the USGS Kansas Water Science Center.

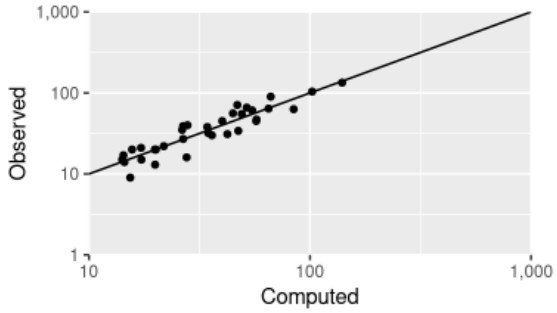
Statistical Plots

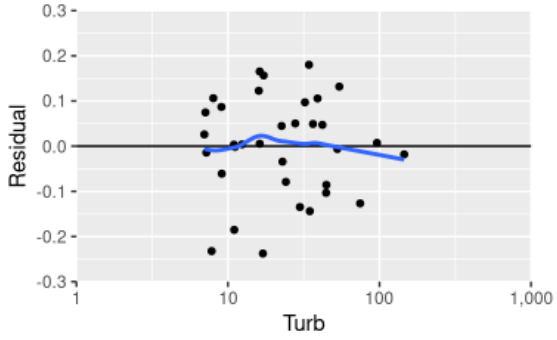












Model-Calibration Dataset

	Date & Time	LogSSC	LogTurb	SSC	Turb	Computed LogSSC	Computed SSC	Residual	Normal Quantiles	Censored Values
0										
1	8/14/2008 11:26	1.49	1.47	31	29.8	1.63	43.6	-0.135	-1.11	--
2	8/14/2008 13:25	1.53	1.54	34	34.7	1.68	48.8	-0.144	-1.25	--
3	9/26/2008 12:36	1.65	1.65	45	44.4	1.76	58.8	-0.103	-0.867	--
4	11/6/2008 17:03	1.3	1.04	20	10.9	1.3	20.5	0.00334	-0.11	--
5	12/4/2008 15:55	0.954	0.892	9	7.8	1.19	15.8	-0.232	-1.68	--
6	1/12/2009 13:12	1.34	1.09	22	12.4	1.34	22.5	0.004	-0.0367	--
7	1/12/2009 14:08	1.43	1.21	27	16.2	1.43	27.5	0.0054	0.0367	--
8	2/19/2009 11:31	1.65	1.44	45	27.8	1.6	41.3	0.0503	0.496	--
9	2/24/2009 15:05	1.81	1.72	64	52.7	1.81	66.9	-0.00632	-0.259	--
10	2/25/2009 14:44	1.95	1.73	90	54.3	1.82	68.5	0.132	1.25	--
11	3/11/2009 14:17	1.79	1.62	61	42	1.74	56.4	0.0472	0.336	--
12	3/12/2009 15:05	1.75	1.51	56	32.2	1.65	46.2	0.0971	0.765	--
13	4/30/2009 13:15	1.58	1.35	38	22.6	1.54	35.3	0.0447	0.259	--
14	8/7/2009 14:51	1.48	1.38	30	24.1	1.56	37.1	-0.079	-0.67	--
15	2/1/2010 11:44	1.8	1.87	63	74.5	1.93	86.9	-0.127	-0.979	--
16	12/22/2010 9:42	1.51	1.36	32	22.9	1.54	35.7	-0.0343	-0.496	--
17	3/23/2011 10:27	1.67	1.65	47	44.6	1.76	59	-0.0857	-0.765	--
18	2/24/2012 11:05	1.82	1.59	66	39	1.71	53.3	0.106	0.867	--
19	3/19/2012 12:53	1.85	1.53	71	34.2	1.67	48.3	0.18	2.11	--
20	5/15/2012 11:20	1.6	1.24	40	17.2	1.45	28.7	0.157	1.43	--
21	6/29/2012 11:07	1.59	1.21	39	16.2	1.43	27.5	0.165	1.68	--
22	7/26/2012 11:41	1.18	0.844	15	6.98	1.15	14.6	0.026	0.184	--
23	9/25/2012 9:40	1.15	0.857	14	7.2	1.16	14.9	-0.0142	-0.336	--
24	12/6/2012 11:33	2.02	1.98	104	96.2	2.01	105	0.00726	0.11	--
25	6/28/2013 10:03	1.11	1.04	13	11	1.3	20.5	-0.185	-1.43	--
26	12/5/2013 11:32	1.32	0.957	21	9.05	1.24	17.7	0.0869	0.67	--
27	2/14/2014 11:25	1.3	1.05	20	11.1	1.3	20.7	-0.00171	-0.184	--
28	3/6/2014 11:51	1.54	1.2	35	16	1.42	27.2	0.123	1.11	--
29	6/9/2014 8:24	1.2	1.23	16	17	1.44	28.5	-0.238	-2.11	--
30	9/11/2014 10:20	1.23	0.851	17	7.1	1.16	14.8	0.0747	0.581	--
31	12/5/2014 13:18	1.18	0.959	15	9.1	1.24	17.8	-0.0609	-0.581	--
32	12/12/2014 13:34	1.74	1.56	55	36.4	1.69	50.6	0.0491	0.415	--
33	12/16/2014 13:16	2.13	2.16	134	146	2.15	144	-0.0182	-0.415	--
34	3/24/2015 11:51	1.3	0.903	20	8	1.19	16.1	0.106	0.979	--

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

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/.

MATLAB, 2019, version 9.6.0 (R2019a), Natick, Massachusetts: The MathWorks, Inc.

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.

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.

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.

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.