Model Archival Summary for Suspended-Sediment Concentration at U.S. Geological Survey Site 06889110, Soldier Creek near Corning, Kansas, during August 2021 through October 2024

This model archival summary summarizes the suspended-sediment concentration (SSC; U.S. Geological Survey [USGS] parameter code 80154) model developed to compute 15-minute SSC from August 26, 2021, onward. This is the first SSC model developed for USGS site 06889110, Soldier Creek near Corning, Kansas. The methods and techniques used in developing this model follow USGS guidelines as described by Rasmussen and others (2009) and Stone and others (2024).

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Site and Model Information

Site number: 06889110

Site name: Soldier Creek near Corning, Kansas

Location: Lat 39°36'34.37", Longitude 95°58'33.80" referenced to North American Datum of 1983, in sec.21, T.5 S., R. 13 E., Nemaha County, KS, Hydrologic Unit 10270102.

Equipment: A Xylem YSI EXO2 water-quality monitor (equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, and turbidity) and a Hach Nitratax plus sc sensor (2-millimeter path length nitrate plus nitrite sensor) were installed during July 2021 through October 2024. The Xylem YSI EXO2 water-quality monitor and Hach Nitratax plus sc sensor were housed in side-by-side 4-inch diameter galvanized steel and PVC pipes, respectively. Readings from the sensors were recorded every 15 minutes and transmitted by way of satellite, hourly.

Date model was created: March 24, 2025

Model-calibration data period: August 26, 2021, through October 31, 2024

Model-application date: August 26, 2021, onward.

Model computations are available at the USGS National Real-Time Water-Quality website (https://nrtwq.usgs.gov/ks/).

Suspended-Sediment Sampling Details

A total of 36 samples are included in the model calibration dataset: 2 non-isokinetic multiple vertical samples were collected from the downstream side of the bridge during high flow events when gage height was 8 feet or higher and discharge was at least 30 cfs or higher. 34 non-isokinetic single vertical grab samples were collected in stream, in low flow conditions. There are some limitations to the non-isokinetic sampling method when it comes to developing models. Non-isokinetic sampling does not represent average concentrations unless the stream is completely mixed laterally and vertically (U.S. Geological Survey, 2006).

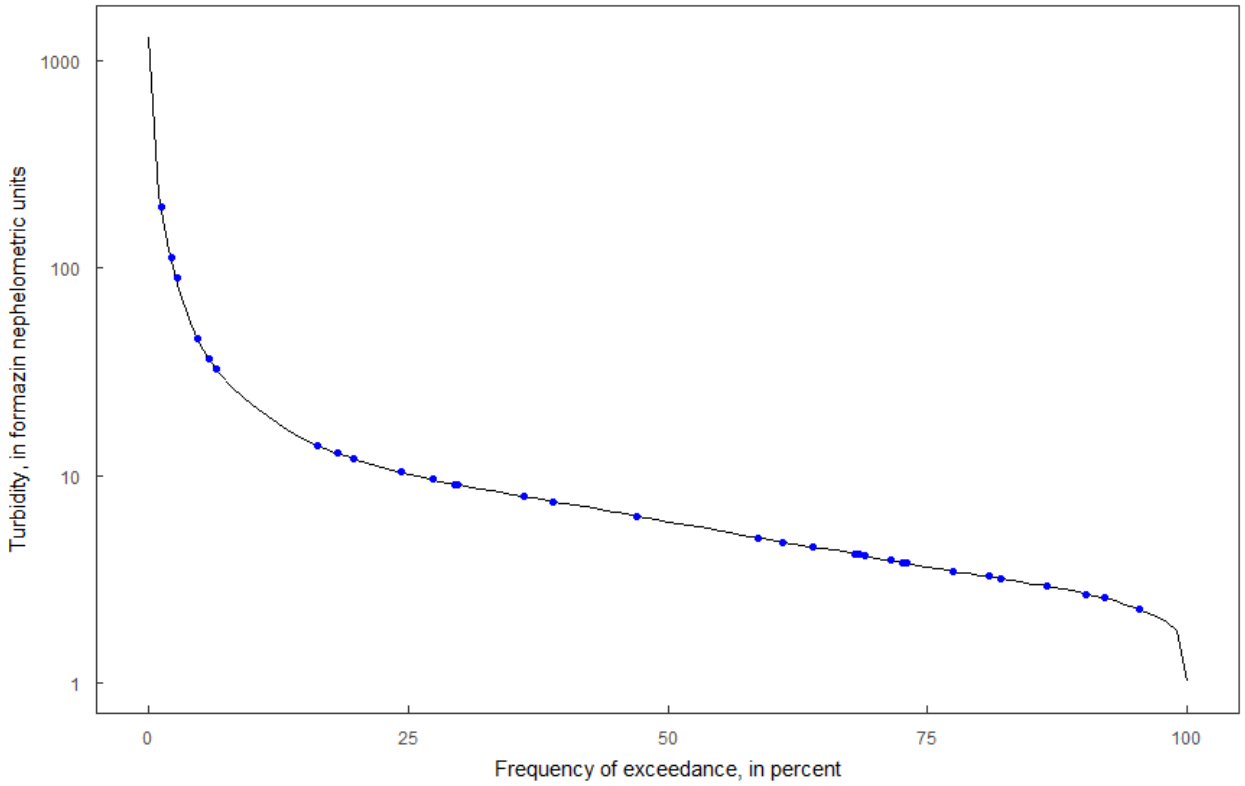

Figure 1. Turbidity duration curve and discrete water-quality samples collected at the Soldier Creek near Corning, Kansas streamgage (U.S. Geological Survey station 06889110) during August 2021 through October 2024.

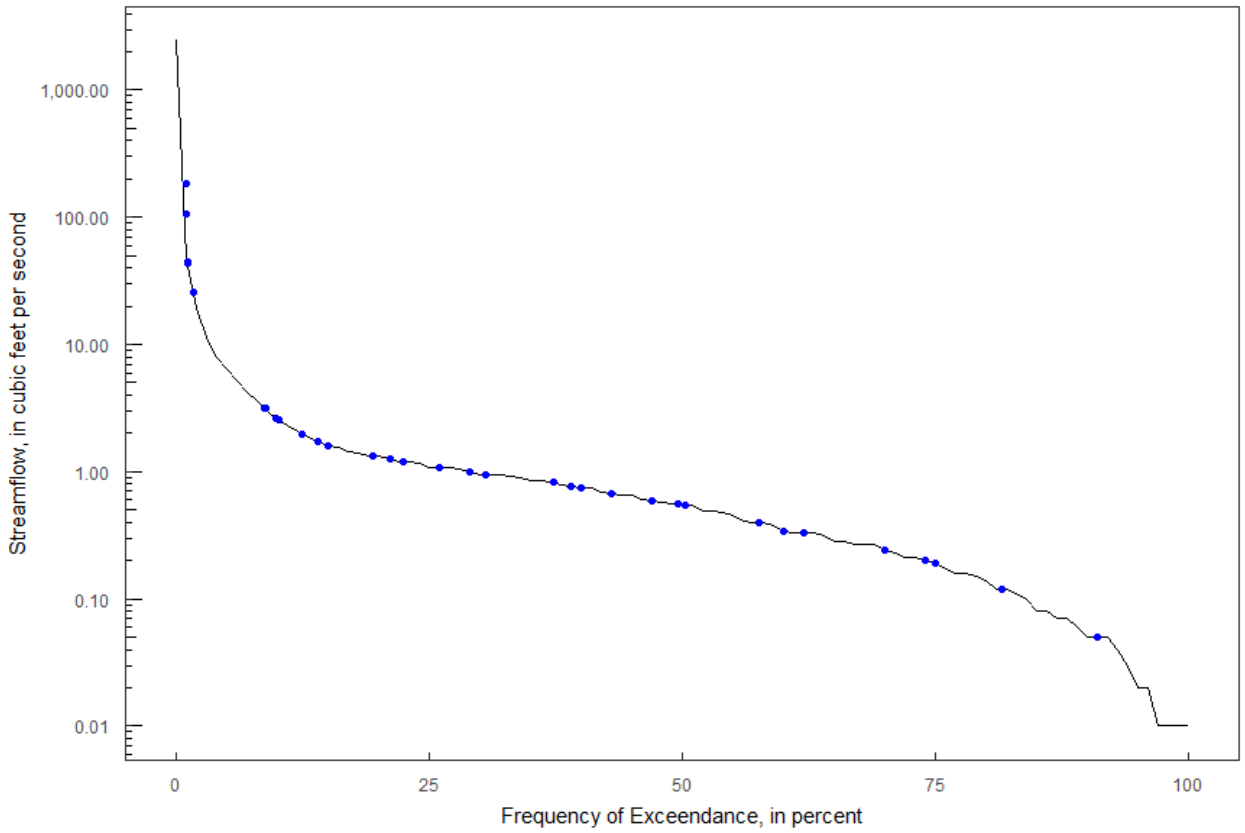

Figure 2. Streamflow duration curve and discrete water-quality samples collected at the Soldier Creek near Corning, Kansas streamgage (U.S. Geological Survey station 06889110) during August 2021 through October 2024. Zero streamflow values were excluded from the duration curve dataset.

Due to the low flow characteristics of this site, most of the grab samples were collected at one location in a narrow width flow channel (< 5 ft.) near the continuous water-quality monitor. As conditions warranted, with greater depth and a wider flow channel (> 5 ft.), open mouth grab samples were collected at three locations across the stream width and coded as multiple verticals.

There were several factors that limited the ability to collect samples using the isokinetic sampling methods at this site. First, the site typically has limited flow and depth making depth-integrated verticals impossible to obtain. To meet the guidelines for collecting isokinetic depth-integrated verticals requires at least 1.5 ft/s velocity and a minimum of 1 foot depth per isokinetic sampling charts (U.S. Geological Survey, variously dated). Most of the sampling conditions did not meet these requirements. Secondly, the site is very flashy, so the time required for an EWI makes it difficult to capture water-quality conditions at a specific sampling time. Finally, because of overgrowth, at times the sampleable channel is very narrow, less than 5 feet.

All grab and multiple vertical samples were composited for analysis as described in (U.S. Geological Survey, variously dated). During August 2021 through June 2023, multiple-vertical and single-vertical grab samples were collected monthly. From June 2023 through October 2024 the sampling frequency was reduced to 8 samples per year, with an emphasis on collecting event samples. Event samples would be defined as any significant increase in the discharge (> 30 cfs and gage height > 8 ft), in addition to targeting flows that have not been captured. A FISP DH-95 was used during event sampling, otherwise open mouth single and multiple verticals were collected due to the low flow, flashy nature of the stream. Samples were analyzed for SSC at the USGS Iowa Sediment Laboratory in Iowa City, Iowa, using the methods documented by Guy (1969). All suspended-sediment data are available in the USGS National Water Information System database (https://doi.org/10.5066/P13PXWPL; U.S. Geological Survey, 2025) using site number 06889110.

Model-Calibration Dataset

Careful efforts were made to follow U.S. Geological Survey sampling protocols and data analysis described in U.S. Geological Survey (variously dated) and Wagner and others (2006). All data are stored in the National Water Information System (U.S. Geological Survey, 2025) database and available to the public using site number 06889110. Additionally, water-quality sample data is also publicly available from the Water Quality Portal (https://www.waterqualitydata.us/).

Sensor-measured water temperature, specific conductance, dissolved oxygen, pH, turbidity, and nitrate plus nitrite, along with streamflow were considered as potential explanatory variables for SSC both individually and in combination. All continuous explanatory data were interpolated from the time series points immediately before and after the time of sample collection. This data was merged using R programming code (R Core Team, 2024).

The final selected regression model is based on 36 concurrent measurements of SSC and sensor measured TBY during August 26, 2021, through October 31, 2024. Samples were collected throughout the range of continuously observed hydrologic conditions. Summary statistics and the complete model-calibration dataset are provided below. Two missing values occurred in the continuous turbidity dataset on 1/31/2022 and 3/19/2024 when samples were collected. The missing turbidity value on 1/31/2022 was due to the water-quality monitor being removed during this period because of ice formation within the stream potentially damaging the sensors. The missing turbidity value on 3/19/2024 was due to suspected animal activity resulting in the deletion of data during that period. Therefore, the missing continuous turbidity values on 1/31/2022 and 3/19/2024 were replaced with values recorded by the field meter during sample collection.

Computations of studentized residuals, Cook's distance (Cook's D), difference in fits (DFFITS), and leverage were used for each data point to identify and investigate potential outliers and their effect on the final selected regression model (Cook, 1977; Helsel and others, 2020). All potential outliers were not found to have errors associated with collection, processing, or analysis, and were considered valid and included in the model calibration data set.

This model is specific to Soldier Creek near Corning, KS during this model-application period and cannot be applied to data collected from other sites on the Soldier Creek or data collected from other water bodies.

Quality-Assurance and Quality-Control

All SSC results collected during August 2021 through October 2024 were reviewed and approved following USGS guidance as documented in Quality-Assurance and Data-Management Plan for Water-Quality Activities in the Kansas Water Science Center (Rasmussen and others, 2014). Concurrent replicate samples were collected during approximately 6 percent of all SSC samples. Relative percentage differences (RPDs) were used to quantify variability between the environmental sample and concurrent replicate sample concentrations. All RPDs for the SSC concurrent replicate pairs were ≤ 10%.

All continuous water-quality data collected during August 2021 through October 2024 followed USGS guidance on reviewing and approving data on a quarterly basis (U.S Geological Survey, 2017). Due to low flow and shallow conditions, the monitor was regularly removed during the winter months to prevent damage to the equipment. Additionally, the site was also susceptible to sensor fouling, especially during the summertime when backwater conditions were more prominent. As a result, there were periods of data that were either corrected or deleted because of sensor fouling, sensor calibration drift, or equipment malfunction (Wagner and others, 2006).

Model Development

Ordinary least squares regression analysis was done using R programming language (R Core Team, 2024) to relate discretely-collected SSC to sensor-measured TBY. The distribution of residuals was examined for normality, and the plots of residuals (the difference between the measured and computed values) were examined for homoscedasticity (departures from zero did not change substantially over the range of computed values). Additionally, seasonal components (sine and cosine variables) were evaluated as potential explanatory variables.

TBY was selected as a surrogate for SSC based on residual plots, relatively high coefficient of determination (adjusted R^2) 0.905, and relatively low model standard percentage error (MSPE) 40.4 to 67.7. Additionally, TBY is a well-documented surrogate for SSC and makes physical sense as TBY is a measurement of the amount of light scattered by suspended particles.

Model Summary

Values for all the aforementioned statistics were computed and are included below along with all relevant sample data and additional statistical information.

The following is a summary of final regression model for SSC at USGS station number 06889110:

SSC-based model:

$$\log_{10} SSC = 1.06 \times \log_{10} TBY + 0.181$$

where

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

TBY = sensor-measured turbidity, in formazin nephelometric units (FNU).

The log-transformed model may be retransformed to the original units so that SSC can be calculated directly. The retransformation introduces a bias in the calculated constituent. This bias may be corrected using Duan's Bias Correction Factor (BCF; Duan, 1983). For this model, the calculated BCF is 1.18. The retransformed model, accounting for BCF is:

$$SSC = 1.18 \times (TBY^{1.06} \times 10^{0.181})$$

Previous Models

There are no previously published models at this site.

Model statistics, data, and plots

Definitions

MSE: Mean square error (Helsel and others, 2020).

MSPE: Model standard percentage error (Helsel and others, 2020).

Probability (>|t|): The probability that the independent variable has no effect on the dependent variable (Helsel and others, 2020).

RMSE: Root mean square error (Helsel and others, 2020).

Cook's D: Cook's distance (Helsel and others, 2020).

DFFITS: Difference in fits statistics (Helsel and others, 2020).

Leverage: An outlier's measure in the x direction (Helsel and others, 2020).

SSC: Suspended sediment concentration (SSC) in mg/l (80154).

TBY: Turbidity, in FNU (USGS parameter code 63680; USGS method code TS213).

t value: Student's t value; the coefficient divided by its associated standard error (Helsel and others, 2020).

BCF: Duan's bias correction factor (Duan, 1983).

LOESS: Local polynomial regression fitting, or locally estimated scatterplot smoothing (Helsel and others, 2020).

Log: Common logarithm with base 10.

Q1: The value at which 25 percent of the data fall under when data are arranged in ascending order (25th percentile).

Q2: The value at which 50 percent of the data fall under when data are arranged in ascending order (Median).

Q3: The value at which 75 percent of the data fall under when data are arranged in ascending order (75th percentile).

R^2: Coefficient of determination.

Model

logSSC = 0.181 + 1.06 * logTBY

Variable summary statistics

Variable	Minimum	Q1	Median	Mean	Q3	Maximum
logSSC	0.602	0.889	1.040	1.31	1.45	3.47
logTBY	0.204	0.646	0.848	1.07	1.23	2.84
SSC	4.000	7.750	11.000	157.00	28.20	2920.00
TBY	1.600	4.430	7.050	52.80	17.10	700.00

Duration plots

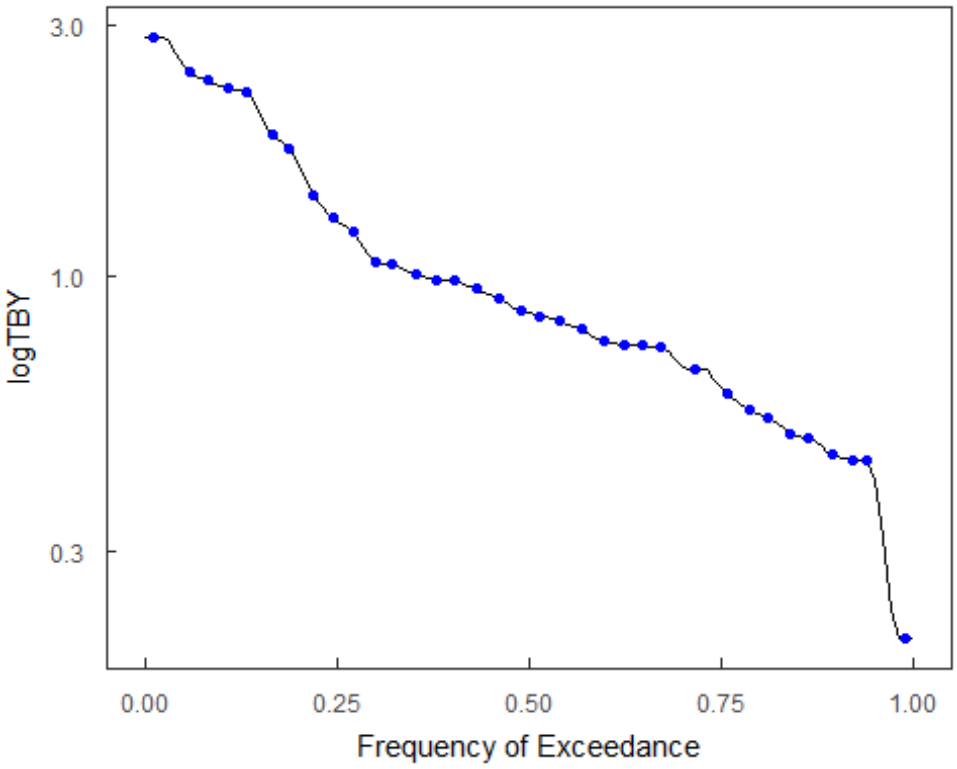

Figure 3. Duration curve of continuous log-scale turbidity (black line) and turbidity observations during discrete sample collection (blue dots) by frequency of exceedance.

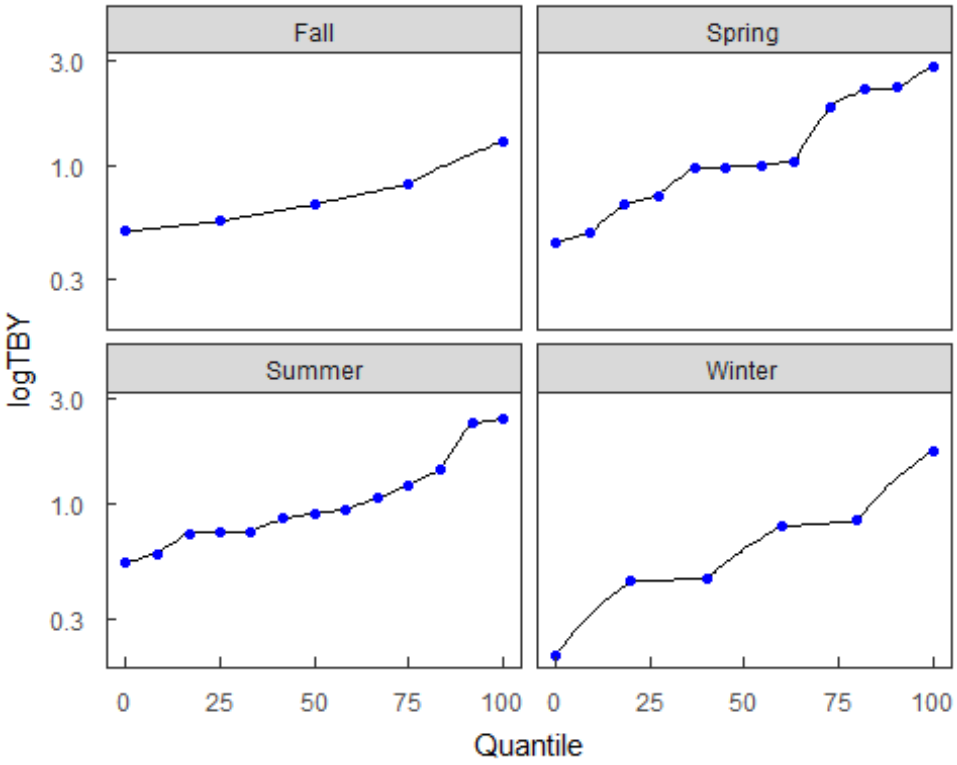

Figure 4. Seasonal duration plots of continuous log-scale turbidity (black line) and observed turbidity during discrete sample collection (blue dots) by quantile.

Box plots

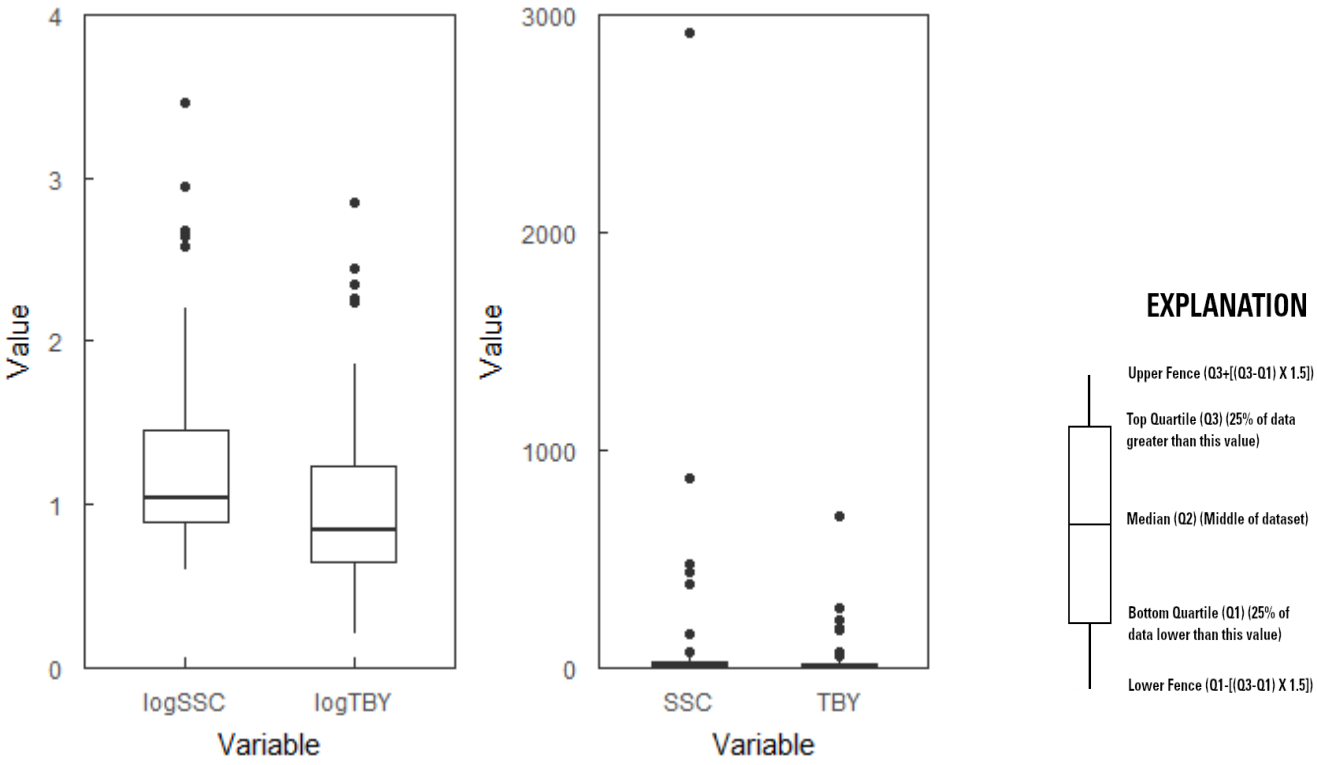

Figure 5. Box plots of log-transformed (left) and linear (right) SSC and turbidity values used in the calibration dataset.

Scatter plots

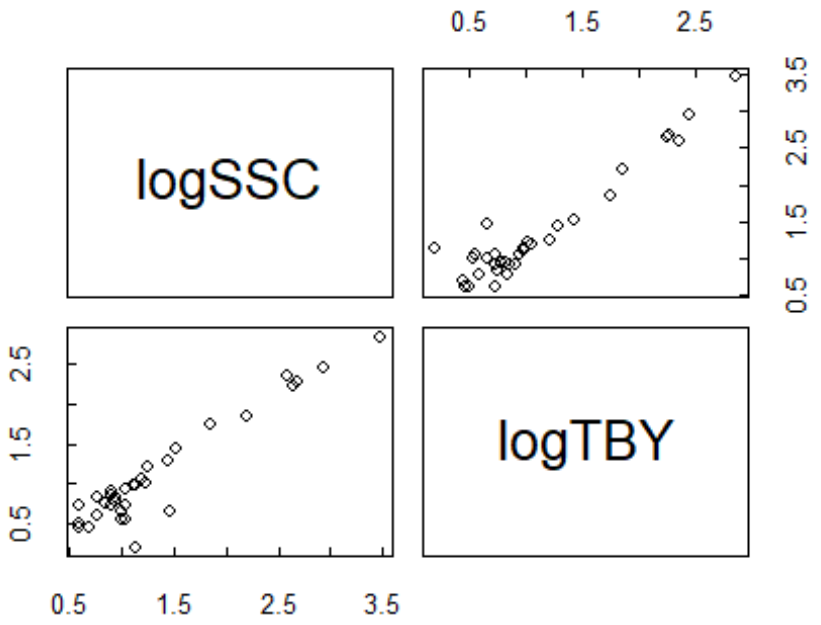

Figure 6. Scatter plot of log-transformed SSC and log-transformed turbidity.

Basic model statistics

Statistic	Value
Observations	36.000
adjusted R2	0.905
RMSE	0.225
Upper MSPE (90%)	67.700
Lower MSPE (90%)	40.400
BCF	1.180

Model coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.1806124	0.0721104	2.504663	0.0172147
logTBY	1.0556564	0.0577457	18.281122	0.0000000

Correlation matrix

	logSSC	logTBY
logSSC	1.0000000	0.9527113
logTBY	0.9527113	1.0000000

Outlier test criteria

Leverage	DFFITS	CooksD
0.1667	0.4714	0.1936

Flagged observations

datetime	logSSC	CooksD	DFFITS	Leverage	StudResidual	Flag*
2022-01-31 19:20:00	1.15	0.504	1.230	0.0770	4.26	CDS
2023-04-20 14:40:00	3.47	0.319	0.812	0.2370	1.46	CDL
2023-05-31 16:00:00	1.46	0.140	0.585	0.0386	2.92	D

*C: Cook’s distance; L: Leverage; D: Difference in fits statistic; S: Studentized residual

Statistical plots

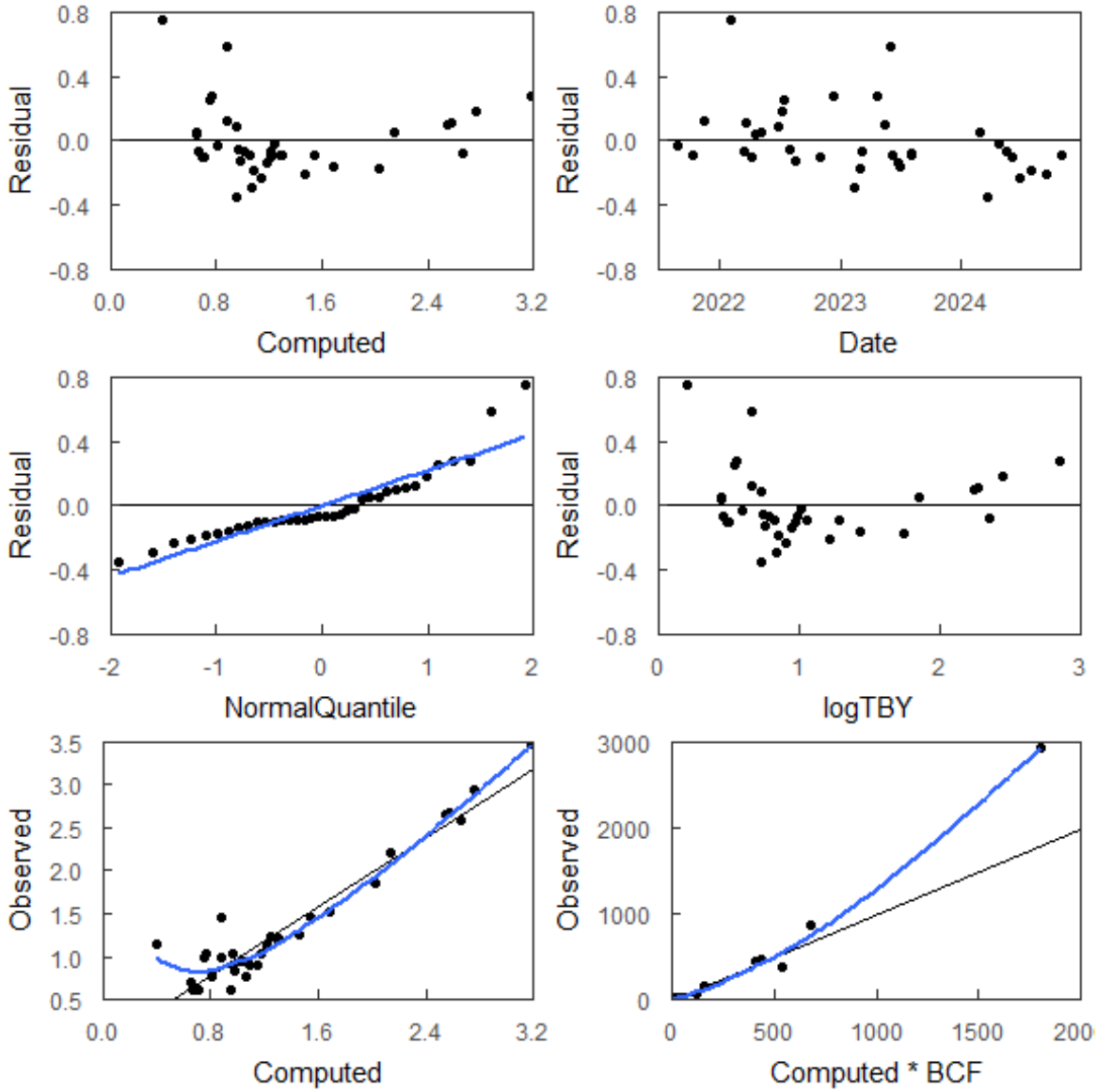

Figure 7. Statistical plots of model residuals and observed and computed SSC. Blue line shows the locally estimated scatterplot smoothing (LOESS).

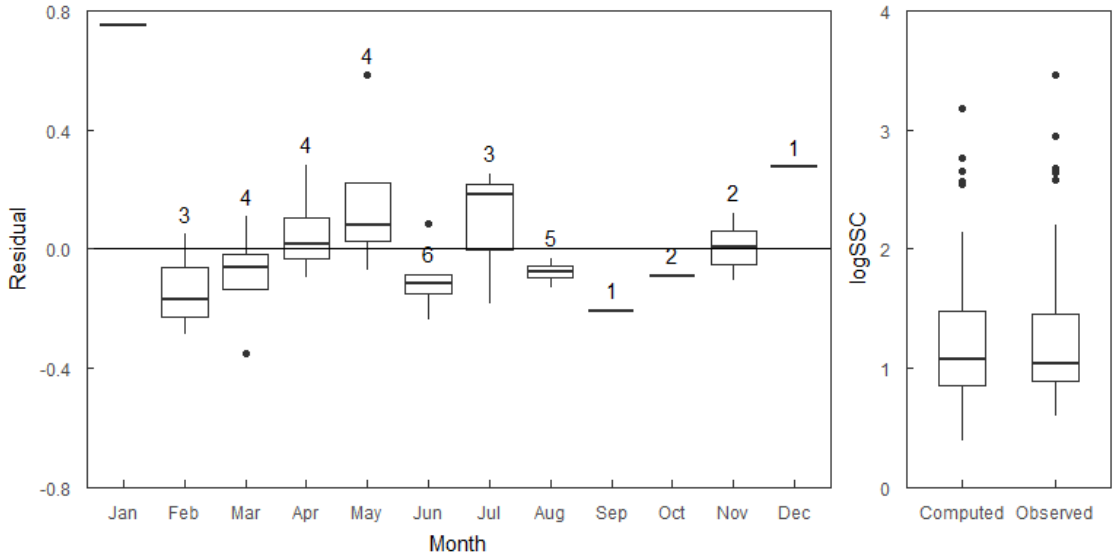

Figure 8. Box plots of monthly SSC model residuals.

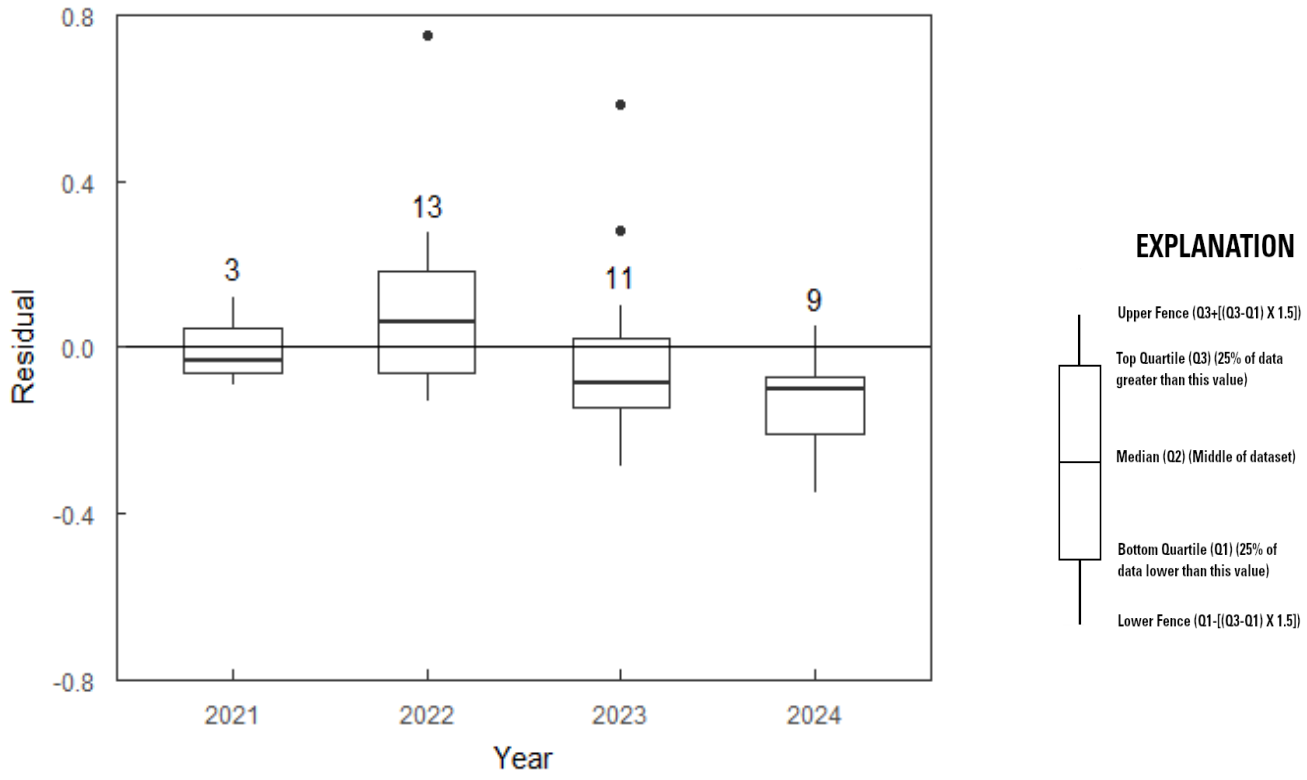

Figure 9. Box plots of SSC model residuals by year.

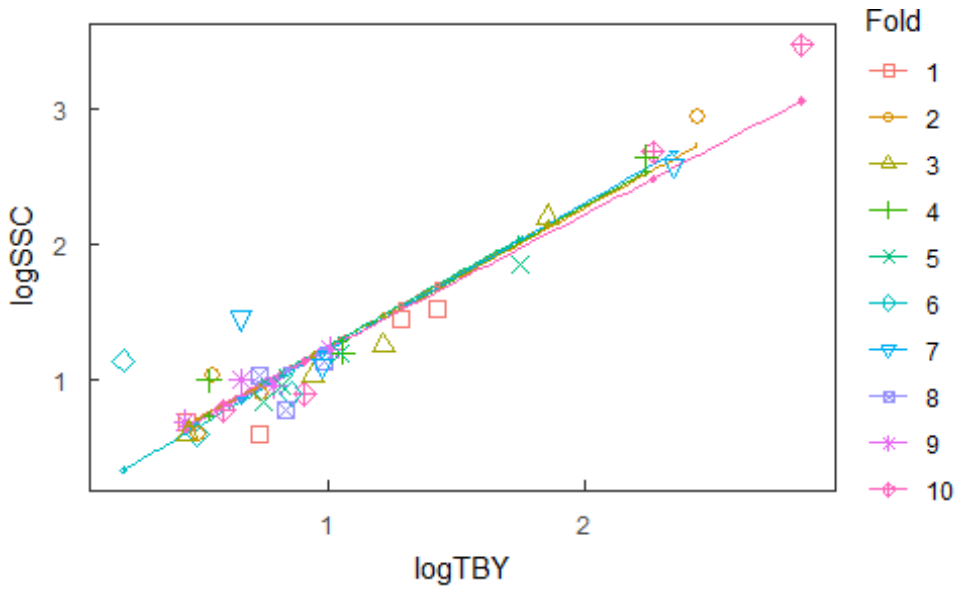

Figure 10. 10-fold cross validation plot.

Fold - equal partition of the data (10 percent of the data).

Large symbols – observed value of a data point removed in a fold.

Small symbols – recomputed value of a data point removed in a fold.

Recomputed regression lines – adjusted regression line with one-fold removed.

Statistic	Value
Minimum MSE of folds	0.0348
25th Percentile	0.0509
Median MSE of folds	0.0524
Mean MSE of folds	0.0505
75th percentile	0.0543
Maximum MSE of folds	0.0565
Model MSE	0.0505

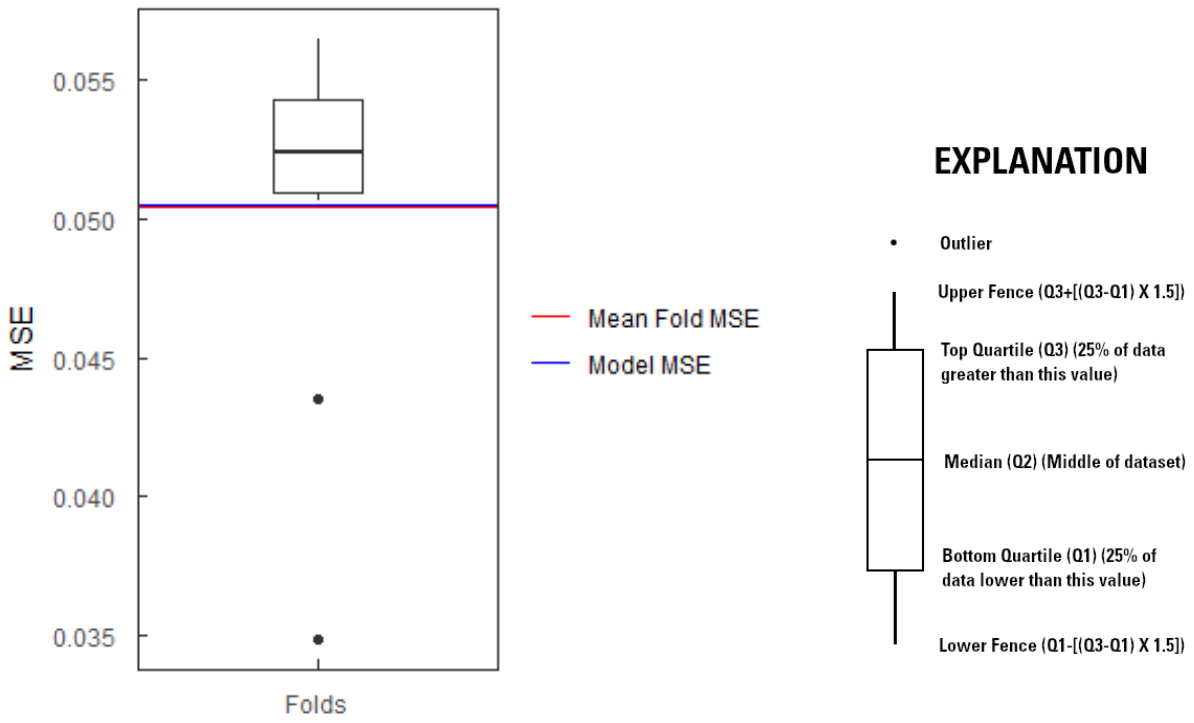

Figure 11. Mean square error of folds from cross validation.

Red line - Model MSE

Blue line - Mean MSE of folds

Model calibration dataset

datetime[1]	logSSC	logTBY	SSC	Computed logSSC	Retransformed SSC
2021-08-26 16:10:00	0.778	0.595	6	0.808	7.59
2021-10-13 15:00:00	1.450	1.290	28	1.540	40.70
2021-11-17 15:30:00	1.000	0.663	10	0.880	8.96
2022-01-31 19:20:00	1.150	0.204	14	0.396	2.94
2022-03-15 15:10:00	0.954	0.790	9	1.010	12.20
2022-03-22 16:30:00	2.680	2.260	478	2.570	440.00
2022-04-06 16:00:00	0.602	0.491	4	0.699	5.91
2022-04-20 15:30:00	0.699	0.447	5	0.653	5.31
2022-05-06 15:20:00	2.200	1.860	158	2.140	163.00
2022-06-28 17:00:00	1.040	0.736	11	0.958	10.70
2022-07-07 16:30:00	2.940	2.440	873	2.760	678.00
2022-07-13 17:50:00	1.000	0.538	10	0.748	6.61

datetime[1]	logSSC	logTBY	SSC	Computed logSSC	Retransformed SSC
2022-08-02 15:50:00	0.903	0.739	8	0.961	10.80
2022-08-17 15:30:00	0.845	0.752	7	0.975	11.10
2022-11-01 15:50:00	0.602	0.501	4	0.709	6.04
2022-12-08 16:30:00	1.040	0.556	11	0.768	6.92
2023-02-13 16:50:00	0.778	0.839	6	1.070	13.70
2023-02-28 16:20:00	1.860	1.750	72	2.030	125.00
2023-03-07 17:10:00	0.602	0.457	4	0.663	5.44
2023-04-20 14:40:00	3.470	2.840	2920	3.180	1800.00
2023-05-16 16:00:00	2.650	2.240	442	2.540	412.00
2023-05-31 16:00:00	1.460	0.663	29	0.880	8.96
2023-06-08 15:30:00	1.200	1.050	16	1.290	22.90
2023-06-22 15:00:00	1.040	0.942	11	1.180	17.70
2023-06-29 15:20:00	1.530	1.430	34	1.690	57.50
2023-08-02 13:50:00	1.200	1.060	16	1.300	23.50
2023-08-04 16:00:00	2.580	2.350	383	2.660	537.00
2024-02-27 16:50:00	0.699	0.445	5	0.650	5.27
2024-03-19 16:00:00	0.602	0.732	4	0.954	10.60
2024-04-25 19:50:00	1.230	1.010	17	1.240	20.70
2024-05-16 15:30:00	1.150	0.982	14	1.220	19.50
2024-06-05 15:30:00	1.110	0.978	13	1.210	19.30
2024-06-27 16:00:00	0.903	0.908	8	1.140	16.30
2024-07-31 14:30:00	0.903	0.857	8	1.090	14.40
2024-09-17 15:10:00	1.260	1.220	18	1.460	34.40
2024-10-31 16:30:00	0.954	0.820	9	1.050	13.10

[1]Dates are formatted as “year-month-day” and times are formatted as “hours:minutes:seconds.”

Report metadata

R Environment Information:

```
## R version 4.4.2 (2024-10-31 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 11 x64 (build 22631)
##
## Matrix products: default
##
##
```

```
## locale:
## [1] LC_COLLATE=English_United States.utf8  LC_CTYPE=English_United States.
utf8    LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C                           LC_TIME=English_United States.u
tf8
##
## time zone: America/Chicago
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] pandoc_0.2.0         gridExtra_2.3        plotly_4.10.4        ggplot
2_3.5.1        DT_0.33
##  [6] shinydashboard_0.7.2 shiny_1.10.0         wren_0.0.1           testth
at_3.2.2       devtools_2.4.5
## [11] usethis_3.1.0        dplyr_1.1.4          dataRetrieval_2.7.17
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.2.1  viridisLite_0.4.2 farver_2.1.2      fastmap_1.2.0
lazyeval_0.2.2    promises_1.3.2
##  [7] digest_0.6.37     timechange_0.3.0  mime_0.12         lifecycle_1.0.4
ellipsis_0.3.2    magrittr_2.0.3
## [13] compiler_4.4.2    rlang_1.1.4       sass_0.4.9        tools_4.4.2
yaml_2.3.10       data.table_1.16.4
## [19] knitr_1.49        labeling_0.4.3    htmlwidgets_1.6.4 bit_4.5.0.1
pkgbuild_1.4.5    curl_6.0.1
## [25] xml2_1.3.6        pkgload_1.4.0     miniUI_0.1.1.1    withr_3.0.2
purrr_1.0.2       desc_1.4.3
## [31] grid_4.4.2        urlchecker_1.0.1  profvis_0.4.0     xtable_1.8-4
colorspace_2.1-1  scales_1.3.0
## [37] MASS_7.3-64       cli_3.6.3         rmarkdown_2.29    crayon_1.5.3
ragg_1.3.3        generics_0.1.3
## [43] remotes_2.5.0     rstudioapi_0.17.1 httr_1.4.7        tzdb_0.4.0
sessioninfo_1.2.2 cachem_1.1.0
## [49] splines_4.4.2     parallel_4.4.2    vctrs_0.6.5       Matrix_1.7-1
jsonlite_1.8.9    hms_1.1.3
## [55] bit64_4.5.2       systemfonts_1.1.0 crosstalk_1.2.1   fontawesome_0.5
.3 tidyr_1.3.1       jquerylib_0.1.4
## [61] glue_1.8.0        lubridate_1.9.4   gtable_0.3.6      later_1.4.1
munsell_0.5.1     tibble_3.2.1
## [67] pillar_1.10.0     rappdirs_0.3.3    htmltools_0.5.8.1 brio_1.1.5
R6_2.5.1          textshaping_0.4.1
## [73] rprojroot_2.0.4   lattice_0.22-6    vroom_1.6.5       evaluate_1.0.1
readr_2.1.5       backports_1.5.0
## [79] memoise_2.0.1     httpuv_1.6.15     bslib_0.8.0       Rcpp_1.0.13-1
nlme_3.1-166      checkmate_2.3.2
## [85] mgcv_1.9-1        xfun_0.49         fs_1.6.5          pkgconfig
```

```
_2.0.3
```

References Cited

Cook, R.D., 1977, Detection of influential observations in linear regression: Technometrics, v. 19, no. 1, p. 15–18, accessed May 2025 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.]

Guy, H.P., 1969, Laboratory theory and methods for sediment analysis: U.S. Geological Survey Techniques of Water-Resources Investigations, book 5, chap. C1, 58 p. [Also available at https://doi.org/10.3133/twri05C1.]

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.]

R Core Team, 2024, R——A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing. [Also available at https://www.R-project.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 May 2025 at https://dx.doi.org/10.3133/ofr20141233.

Rasmussen, P.R., 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.4, 64 p., https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf.

Stone, M.L., Lee, C.J., Rasmussen, T.J., Williams, T.J., Kramer, A.R., and Klager, B.J., 2024, Methods for Computing Water-Quality Concentrations and Loads at Sites Operated by the U.S. Geological Survey Kansas Water Science Center: U.S. Geological Survey Open-File Report 2024–1049, 18 p., https://pubs.usgs.gov/of/2024/1049/ofr20241049.pdf.

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].

U.S. Geological Survey, 2025, National Water Information System—Web interface: accessed May 2025 at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey, 2017, Procedures for processing, approving, publishing, and auditing time-series records for water data: U.S. Geological Survey Office of Water Quality Technical Memorandum no. 2017.07, accessed February 2025 at https://water.usgs.gov/admin/memo/QW/qw2017.07.pdf

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. + 8 attachments, accessed April 10, 2025 at https://pubs.water.usgs.gov/tm1d3.