# Appendix 16. Model Archive Summary for Total Phosphorus Concentration at U.S. Geological Survey Station 07144790, Cheney Reservoir near Cheney, Kansas, during October 1, 2014, through September 30, 2021

This model archive summary summarizes the total phosphate concentration model developed to compute 15-minute, hourly, or daily total phosphate concentrations during October 1, 2014, onward. This model supersedes all prior models used during this period. The methods follow U.S. Geological Survey (USGS) guidance as referenced in relevant Office of Surface Water/Office of Water Quality Technical Memoranda and USGS Techniques and Methods, book 3, chapter C4 (Rasmussen and others, 2009; U.S. Geological Survey, 2016).

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: 07144790

Site name: Cheney Reservoir near Cheney, Kansas

Location: Lat 37°43'34", long 97°47'38" referenced to North American Datum of 1927, in SE 1/4 NE 1/4 NW 1/4 sec.06, T.27 S., R.4 W., Sedgwick County, KS, hydrologic unit 11030014, in control house structure at outlet works of Cheney Dam on North Fork Ninnescah River, 6.0 mi north of Cheney, and at mile 15.9.

Equipment: A YSI, Inc., EXO water-quality monitor (YSI, Inc., 2017) equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, turbidity, chlorophyll fluorescence, and phycocyanin fluorescence was installed November 14, 2015. The EXO monitor is suspended from the dam intake tower walkway. The monitor is at a depth that fluctuates between three to six feet depending on reservoir elevation. Measurements from the EXO were recorded every 15 minutes to hourly and transmitted hourly via satellite. Reservoir elevation was measured using a Design Analysis H–350 nonsubmersible pressure transducer and H–355 gas system.

Date model was created: August 9, 2022

Model calibration data period: February 17, 2016, through August 31, 2021 (dataset consisted of 45 discrete water-quality samples).

Model application date: November 14, 2015, onward (date of EXO continuous water-quality monitor installation).

Model developed by: Ariele Kramer, USGS, Lawrence, Kans. (akramer@usgs.gov)

## Model Calibration Dataset

All data were collected using USGS protocols (U.S. Geological Survey, 2006; Wagner and others, 2006; Bennett and others, 2014) and are stored in the USGS National Water Information System database (https://doi.org/10.5066/F7P55KJN; U.S. Geological Survey, 2022). Potential explanatory variables evaluated individually and in combination were water temperature, specific conductance, pH, dissolved oxygen, turbidity, chlorophyll fluorescence, phycocyanin fluorescence, seasonality (sine and cosine variables), and reservoir elevation.

The regression model is based on 45 concomitant values of discretely collected total phosphorus concentration and continuously measured water temperature and turbidity during February 17, 2016, through August 31, 2021. Discrete samples were collected throughout the range of continuously observed hydrologic conditions. No samples had total phosphorus concentrations that were less than laboratory minimum reporting level. All potential explanatory variables were time interpolated within the 15-minute to hourly continuous record based on the discrete sample time. The maximum time span between two continuous data points used for interpolation was 4 hours (to preserve the sample dataset, field monitor averages obtained during sample collection were used for model development data if no continuous data were available or if gaps larger than 4 hours in the continuous data record resulted in missing interpolated data). Summary statistics and the complete model-calibration dataset are provided below. Potential outliers were identified using the methods described in Rasmussen and others (2009) and Helsel and others (2020). All potential outliers were investigated by reviewing sample collection information sheets and laboratory reports; if there were no clear issues, explanations, or conditions that would cause a result to be invalid for model calibration, the sample was retained in the dataset. Two samples in the model calibration dataset were flagged as outliers but all were retained in the dataset after further review.

## Total Phosphorus Sampling Details

Discrete water-quality samples were collected primarily by depth-integrating through the photic-zone (depth at which light is approximately 1 percent of that at the surface) using a double check-valve bailer (Lane and others, 2003). Vertical water-quality profiles collected during sampling indicated that thermal stratification rarely occurs, and water-quality conditions are typically uniform throughout the water column. Samples were collected from the walkway on the dam intake tower. Discrete samples were collected on a semifixed to event-based schedule seven to eight times per year. All samples were collected between 9:15 a.m. and 12:20 p.m. Samples were analyzed for total phosphorus by the Wichita Municipal Water and Wastewater Laboratory in Wichita, Kans., according to standard methods (Eaton and others, 1995).

## Continuous Water-Quality Data

Water temperature and turbidity were continuously measured (15 minutes to hourly) using a YSI, Inc., EXO multiparameter sonde (YSI, Inc., 2017). The water-quality monitor was operated and maintained according to standard USGS methods (Wagner and others, 2006; Bennett and others, 2014). All continuous water-quality data at Cheney Reservoir near Cheney, Kans. are available in near-real time (updated hourly) from the USGS National Water Information System database

(https://doi.org/10.5066/F7P55KJN; U.S. Geological Survey, 2022) using the site number 07144790.

## Model Development

Ordinary least squares linear regression was used to develop surrogate regression models that relate continuous water-quality conditions to discretely sampled constituent concentrations. All regressions were computed using the R software environment (R Core Team, 2020). The data and subsequent regression equation must meet the five assumptions necessary to apply ordinary least squares regression: the dependent variable is linearly related to the explanatory variables, data used to fit the model are representative of the data of interest, the variance of the residuals is constant (homoscedastic), the residuals are independent of the explanatory variables, and the residuals are normally distributed (Helsel and others, 2020). Previously published explanatory variables also were considered for continuity.

Water temperature and turbidity were selected as good surrogates for total phosphorus concentration based on residual plots, coefficient of determination ($R^2$), and model standard percentage error (MSPE). Values for the aforementioned statistics were computed and are included below along with all relevant sample data and additional statistical information.

## Model Summary

Summary of final total phosphorus concentration regression analysis at USGS site 07144790:

Total phosphorus concentration-based model:

$$\log_{10}(TP) = (0.00517 \times TEMP) + (0.0196 \times TBY) - 1.45,$$

where,

*TP* = total phosphorus, in milligrams per liter as phosphorus (mg/L as P) (USGS parameter code 00665);

*TEMP* = water temperature, in degrees Celsius (°C) (USGS parameter code 00010);

*TBY* = turbidity, monochrome near infrared light-emitting diode light, 780-900 nanometers, detection angle 90 +-2.5 degrees, formazin nephelometric units (FNU) (USGS parameter code 63680); and

$\log_{10}$ = decimal logarithm.

The $\log_{10}$-transformed model may be retransformed to the original units so that total phosphorus concentration can be calculated directly. The retransformation introduces a negative bias in the retransformed calculated constituent (Helsel and others, 2020). This bias may be corrected using Duan's bias correction factor (BCF; Duan, 1983; Helsel and others, 2020). For this model, the calculated BCF was 1.03. The retransformed model, accounting for BCF, is as follows:

$$TP = (10^{0.00517 \times TEMP} \times 10^{0.0196 \times TBY} \times 10^{-1.45}) \times 1.03.$$

Total phosphorus can be associated with suspended sediment and runoff containing organic matter, as seen in similar studies in this region, and turbidity can be a good surrogate for

computing total phosphorus concentrations (Juracek and Rasmussen, 2008; Rasmussen and others, 2016).

Extrapolation, defined as computation beyond the range of the model calibration dataset, may be used to extrapolate no more than 10 percent outside the range of the calibration data used to fit the model and is therefore limited. The extrapolation limit for total phosphorus concentration using this model is 0.22 milligram per liter as phosphorus. Computed estimates outside that limit are not supported by the current model calibration dataset.

## Model Statistics, Data, and Plots

### Definitions

| Variable | Explanation |
|---|---|
| BCF | Bias Correction Factor, used to correct logarithmic bias (Duan 1983) |
| Cook's D | Cook's distance, a measure of influence (Helsel and others, 2020) |
| DFFITS | Difference in fits, a measure of influence (Helsel and others, 2020) |
| E.vars | Explanatory variables |
| Leverage | An outlier's measure in the x direction (Helsel and others, 2020) |
| LOESS | Local polynomial regression fitting (Helsel and others, 2020) |
| logTP | Total phosphorus, in milligrams per liter as phosphorus (mg/L as P) (USGS parameter code 00665), $\log_{10}$-transformed |
| MSE | Model standard error (Helsel and others, 2020) |
| MSPE | Model standard percentage error (Helsel and others, 2020) |
| *Pr*(>\|*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) |
| TBY | Turbidity, monochrome near infrared LED light, 780-900 nm, formazin nephelometric units (FNU) (USGS parameter code 63680) |
| TEMP | Water temperature, in degrees Celsius (°C) (USGS parameter code 00010) |
| TP | Total phosphorus, in milligrams per liter as phosphorus (mg/L as P) (USGS parameter code 00665) |
| t value | Student's *t* value; the coefficient divided by its associated standard error (Helsel and others, 2020) |

### Model

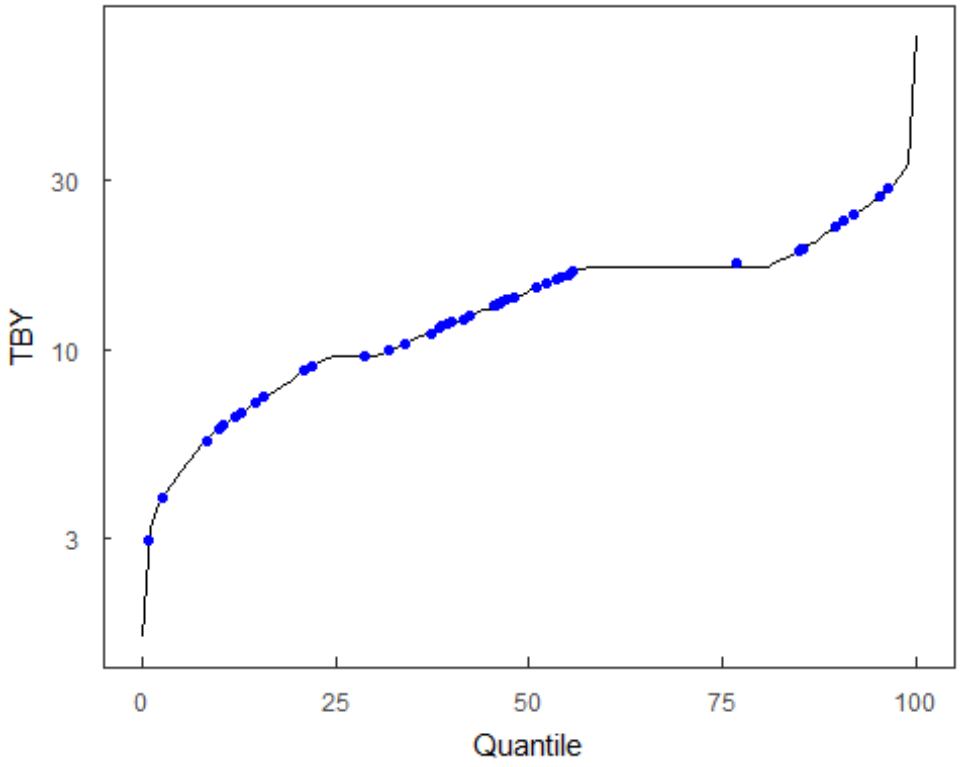
$$\log_{10}(TP) = (0.00517 \times TEMP) + (0.0196 \times TBY) - 1.45$$

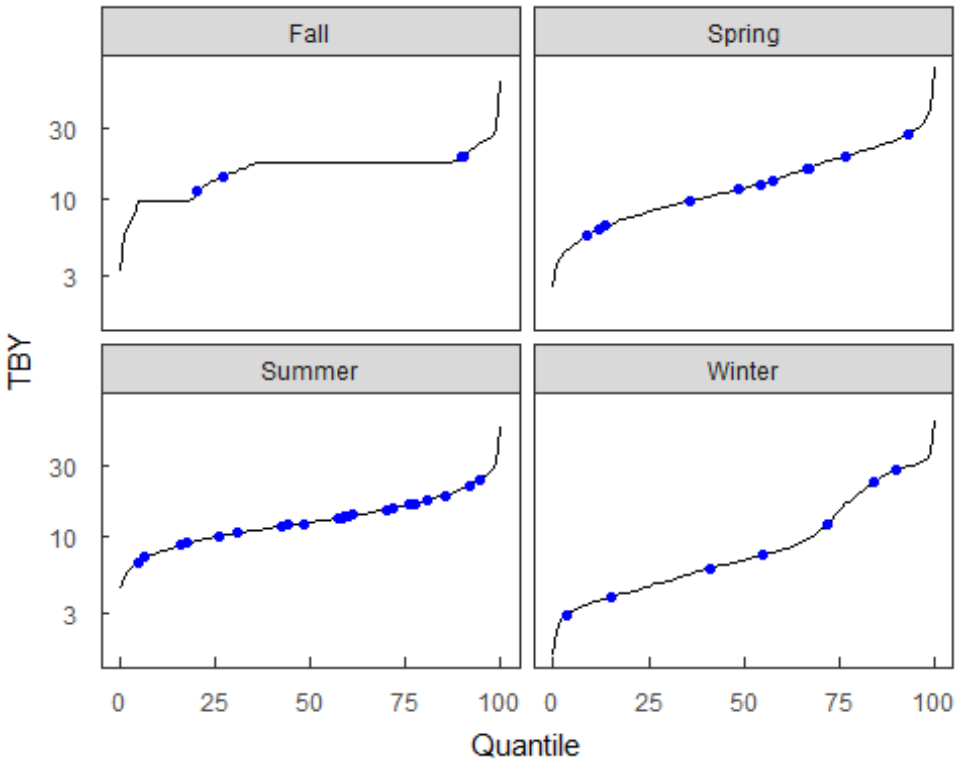
### Variable summary statistics

| Variable | Minimum | Q1 | Median | Mean | Q3 | Maximum |
|---|---|---|---|---|---|---|
| logTP | -1.4 | -1.22 | -1.05 | -1.09 | -1 | -0.699 |
| TBY | 2.99 | 9.73 | 13.3 | 13.5 | 16.3 | 28.4 |
| TEMP | 0.36 | 15.6 | 24.4 | 19.1 | 25.9 | 28.3 |

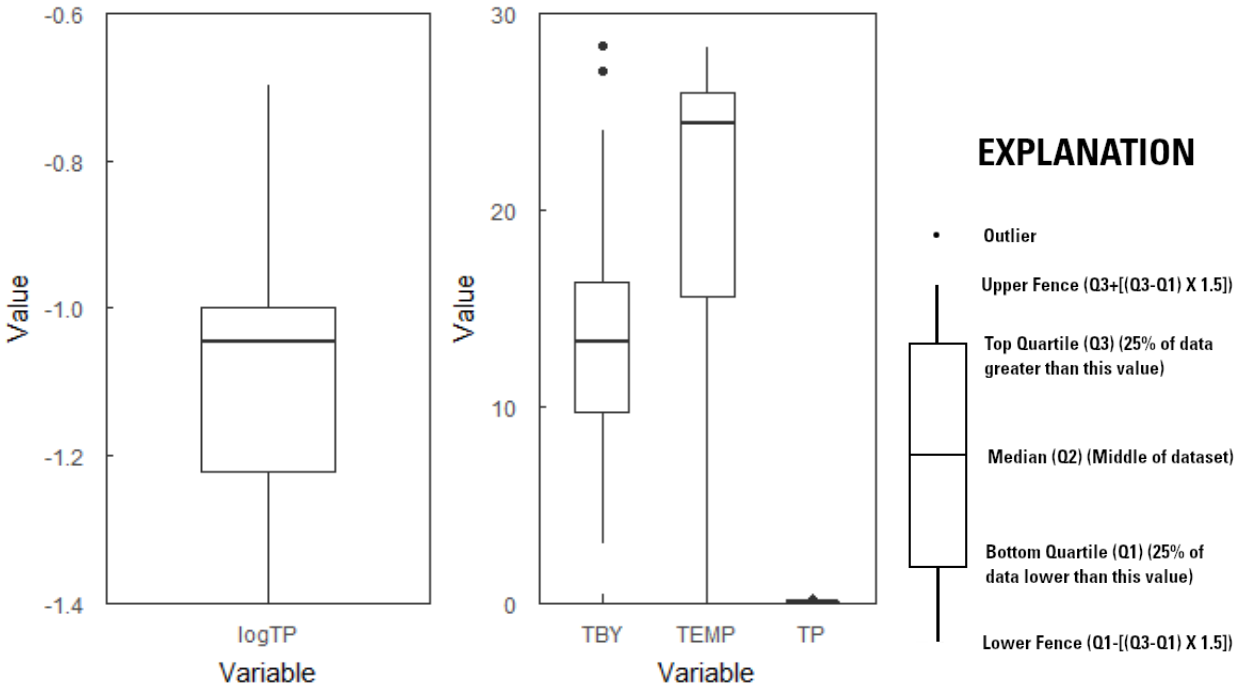
| Variable | Minimum | Q1 | Median | Mean | Q3 | Maximum |
|---|---|---|---|---|---|---|
| TP | 0.04 | 0.06 | 0.09 | 0.0871 | 0.1 | 0.2 |

### Duration plots

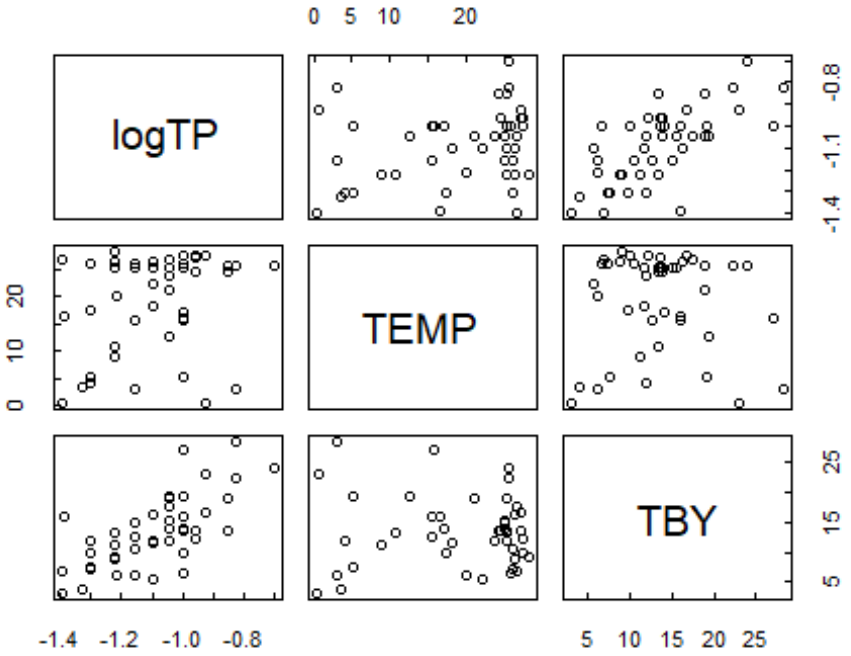



Fall
Spring
Summer
Winter
TBY
30
10
3
0
25
50
75
100
Quantile

## Box plots


Value
-0.6
-0.8
-1.0
-1.2
-1.4
logTP
Variable
30
20
10
0
TBY
TEMP
TP
Variable
EXPLANATION
Outlier
Upper Fence (Q3+[(Q3-Q1) X 1.5])
Top Quartile (Q3) (25% of data greater than this value)
Median (Q2) (Middle of dataset)
Bottom Quartile (Q1) (25% of data lower than this value)
Lower Fence (Q1-[(Q3-Q1) X 1.5])

### Scatter plots




The x- and y-axis labels for a given bivariate plot are defined by the intersecting row and column labels.

### Basic model statistics

| Statistic | Value |
|---|---|
| Observations | 45 |
| $R^2$ | 0.544 |
| Adjusted $R^2$ | 0.522 |
| RMSE | 0.115 |
| Upper MSPE (90%) | 30.3 |
| Lower MSPE (90%) | 23.2 |
| BCF | 1.03 |

### Model coefficients

| | Estimate | Standard error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -1.4539357 | 0.0576328 | -25.227580 | 0.0000000 |
| TEMP | 0.0051728 | 0.0019452 | 2.659345 | 0.0110344 |
| TBY | 0.0196153 | 0.0029692 | 6.606347 | 0.0000001 |

**Correlation matrix**

| | logTP | TEMP | TBY |
|---|---|---|---|
| logTP | 1.0000000 | 0.2653769 | 0.6836468 |
| TEMP | 0.2653769 | 1.0000000 | -0.0170260 |
| TBY | 0.6836468 | -0.0170260 | 1.0000000 |

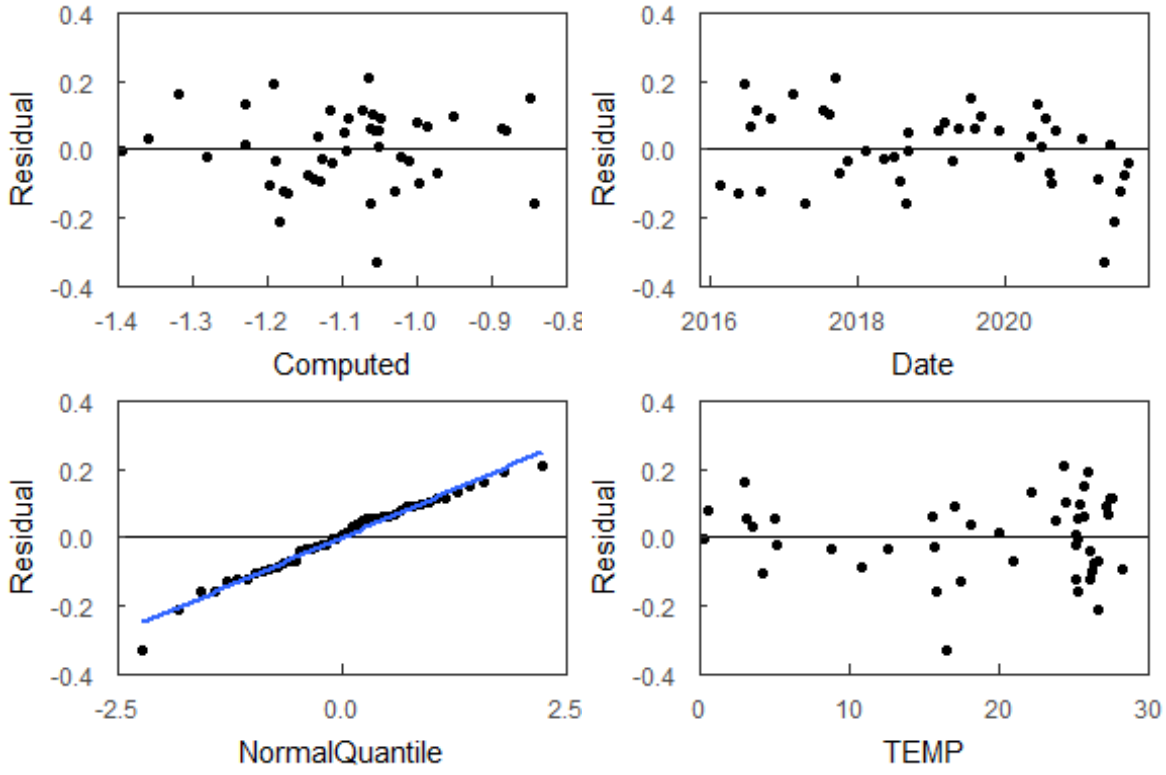
**Outlier test criteria**

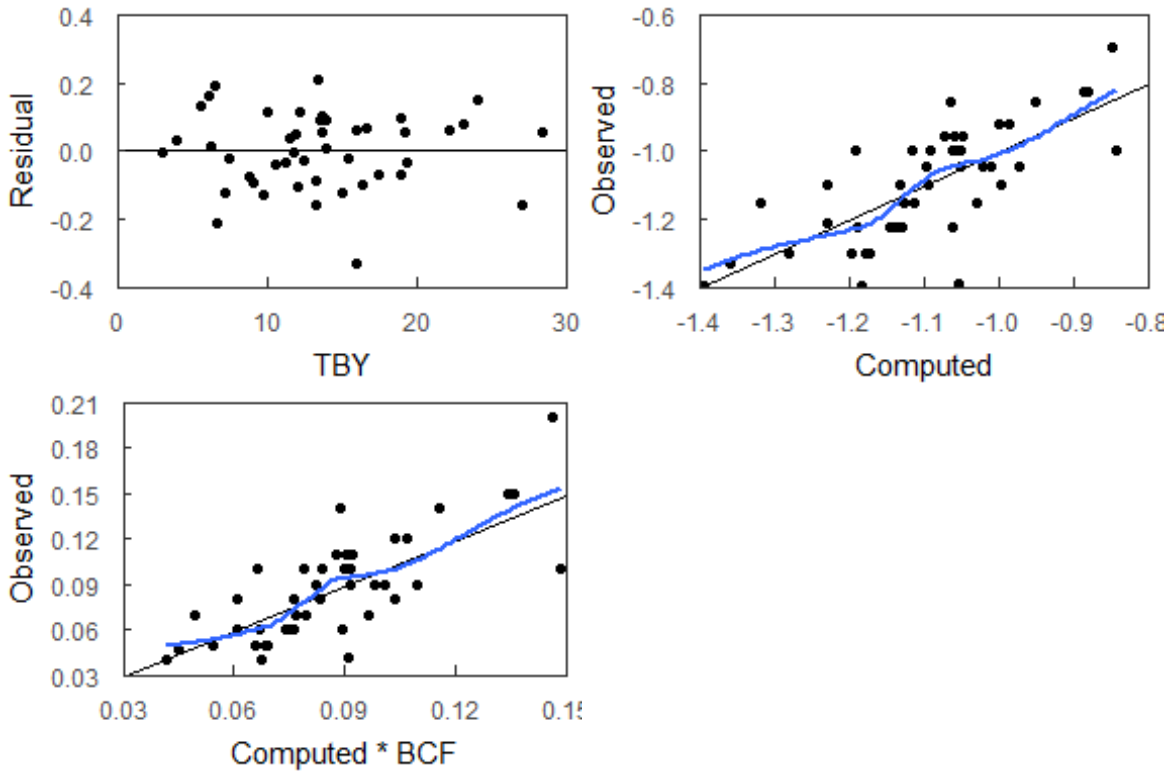
| Leverage | DFFITS | CooksD |
|---|---|---|
| 0.2 | 0.5164 | 0.263 |

**Flagged observations**

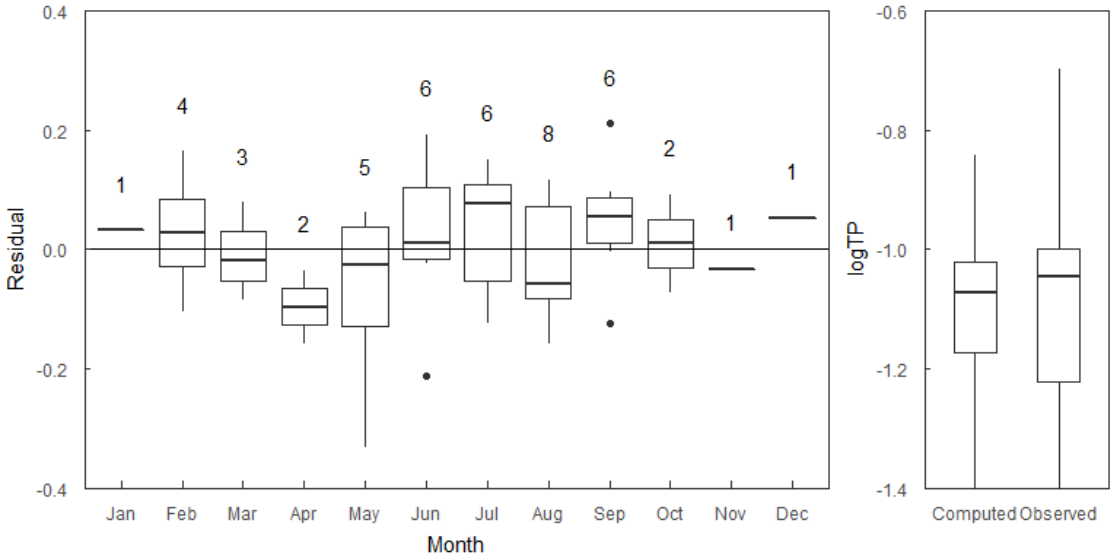
| datetime | logTP | CooksD | DFFITS | Leverage | Studentized Residual |
|---|---|---|---|---|---|
| 2017-02-09 10:40:00 | -1.15 | 0.122 | 0.616 | 0.135 | 1.56 |
| 2019-02-05 11:20:00 | -0.824 | 0.0348 | 0.32 | 0.24 | 0.57 |

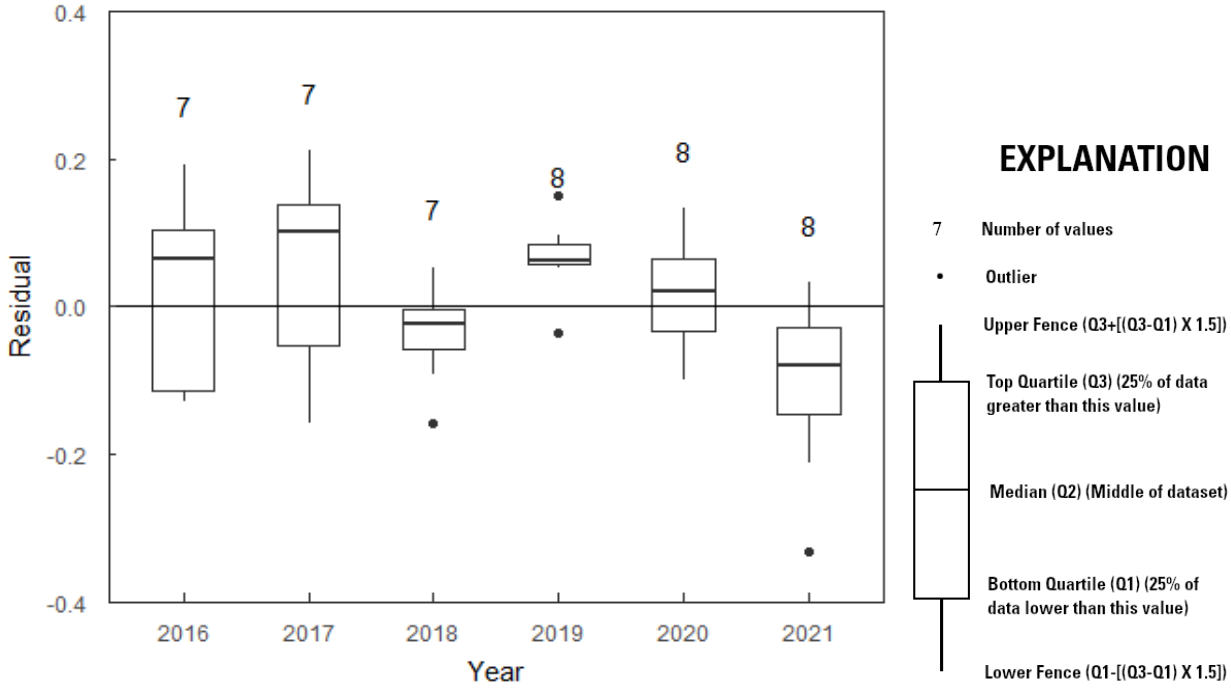
**Statistical plots**



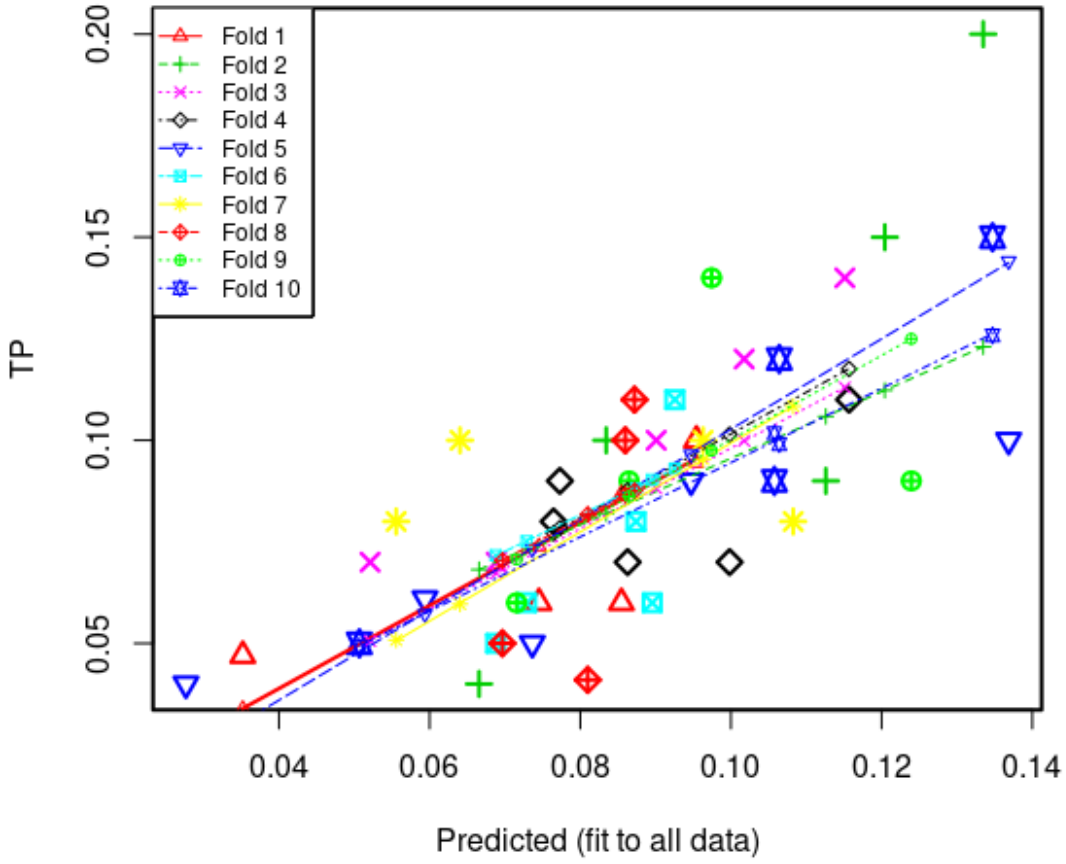



The blue line shows the locally estimated scatterplot smoothing (LOESS). The black dots correspond to observed values. The black line represents the 1:1 line.






## Cross Validation



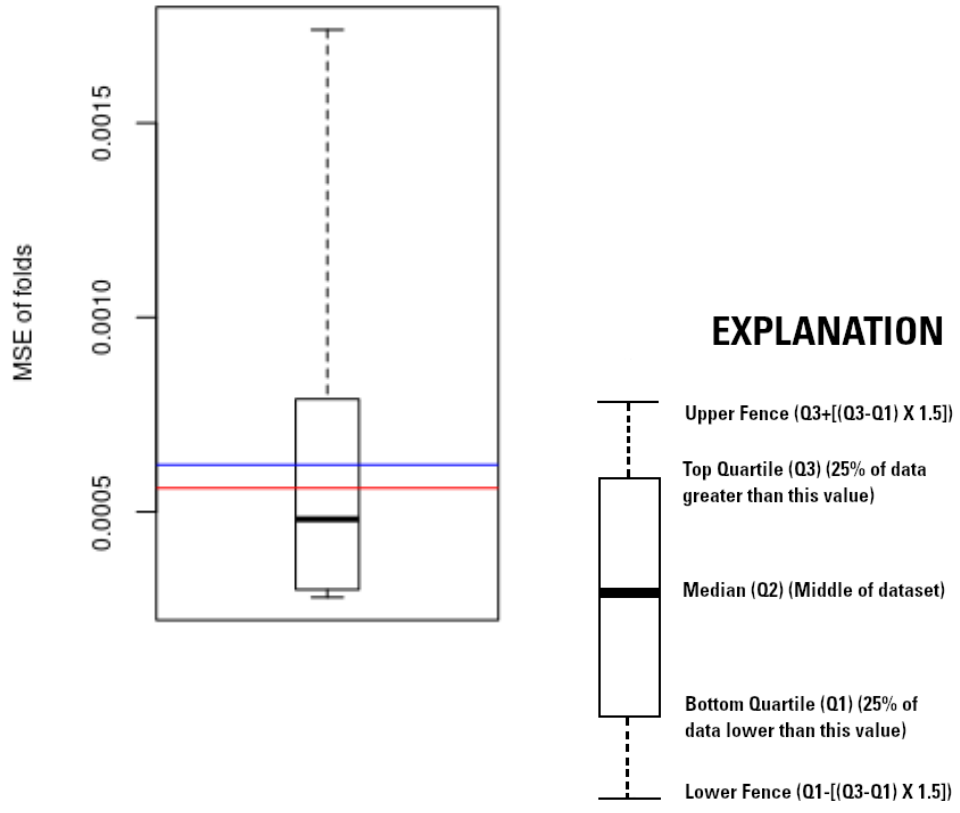

Fold - equal partition of the data (10 percent of the data).

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 | 0.000280 |
| Median MSE of folds | 0.000481 |
| Mean MSE of folds | 0.000621 |
| Maximum MSE of folds | 0.001740 |
| (Mean MSE of folds) / (Model MSE) | 1.110000 |




Red line - Model MSE

Blue line - Mean MSE of folds

**Model calibration dataset**

| datetime | logTP | TEMP | TBY | TP | Computed | Retransformed |
|---|---|---|---|---|---|---|
| 2016-02-17 10:45:00 | -1.3 | 4.20 | 12.0 | 0.05 | -1.2 | 0.0656 |
| 2016-05-17 10:20:00 | -1.3 | 17.4 | 9.73 | 0.05 | -1.17 | 0.0693 |
| 2016-06-15 09:15:00 | -1 | 25.9 | 6.55 | 0.10 | -1.19 | 0.0664 |
| 2016-07-18 10:40:00 | -0.921 | 27.3 | 16.7 | 0.12 | -0.986 | 0.107 |
| 2016-08-15 10:30:00 | -1 | 27.5 | 10.0 | 0.10 | -1.12 | 0.0791 |
| 2016-09-06 10:40:00 | -1.15 | 25.1 | 15.0 | 0.07 | -1.03 | 0.0964 |
| 2016-10-25 10:15:00 | -1 | 17.1 | 14.0 | 0.10 | -1.09 | 0.0837 |
| 2017-02-09 10:40:00 | -1.15 | 2.97 | 6.10 | 0.07 | -1.32 | 0.0495 |
| 2017-04-17 10:30:00 | -1 | 15.8 | 27.0 | 0.10 | -0.843 | 0.148 |
| 2017-07-10 11:40:00 | -0.959 | 27.4 | 12.2 | 0.11 | -1.07 | 0.0874 |
| 2017-08-15 10:00:00 | -0.959 | 24.4 | 13.7 | 0.11 | -1.06 | 0.0902 |
| 2017-09-07 10:00:00 | -0.854 | 24.4 | 13.4 | 0.14 | -1.06 | 0.0889 |
| 2017-10-03 10:20:00 | -1.05 | 21.0 | 19.0 | 0.09 | -0.973 | 0.11 |
| 2017-11-13 12:00:00 | -1.22 | 8.80 | 11.2 | 0.06 | -1.19 | 0.0668 |
| 2018-02-13 10:40:00 | -1.4 | 0.36 | 2.99 | 0.04 | -1.39 | 0.0417 |
| 2018-05-08 10:30:00 | -1.15 | 15.6 | 12.5 | 0.07 | -1.13 | 0.0769 |
| 2018-06-25 12:00:00 | -1.05 | 25.2 | 15.4 | 0.09 | -1.02 | 0.0982 |
| 2018-07-26 11:40:00 | -1.22 | 28.3 | 9.03 | 0.06 | -1.13 | 0.0764 |
| 2018-08-29 11:00:00 | -1.22 | 25.2 | 13.3 | 0.06 | -1.06 | 0.0893 |
| 2018-09-05 12:20:00 | -1.1 | 25.3 | 11.7 | 0.08 | -1.09 | 0.0833 |
| 2018-09-11 09:40:00 | -1.05 | 23.7 | 11.9 | 0.09 | -1.1 | 0.0824 |
| 2019-02-05 11:20:00 | -0.824 | 3.11 | 28.4 | 0.15 | -0.881 | 0.136 |
| 2019-03-07 10:50:00 | -0.921 | 0.55 | 23.0 | 0.12 | -0.999 | 0.103 |
| 2019-04-09 10:30:00 | -1.05 | 12.6 | 19.3 | 0.09 | -1.01 | 0.101 |
| 2019-05-14 11:10:00 | -1 | 15.6 | 15.9 | 0.10 | -1.06 | 0.0896 |
| 2019-07-09 10:15:00 | -0.699 | 25.7 | 24.0 | 0.20 | -0.85 | 0.146 |
| 2019-08-06 11:00:00 | -0.824 | 25.7 | 22.1 | 0.15 | -0.887 | 0.134 |
| 2019-09-03 10:40:00 | -0.854 | 25.4 | 18.9 | 0.14 | -0.951 | 0.116 |
| 2019-12-04 10:50:00 | -1 | 5.10 | 19.1 | 0.10 | -1.05 | 0.0915 |
| 2020-03-04 11:00:00 | -1.3 | 5.22 | 7.43 | 0.05 | -1.28 | 0.054 |
| 2020-05-06 10:30:00 | -1.1 | 18.1 | 11.6 | 0.08 | -1.13 | 0.0759 |
| 2020-06-03 10:20:00 | -1.1 | 22.2 | 5.61 | 0.08 | -1.23 | 0.0609 |
| 2020-06-25 11:30:00 | -1.05 | 25.1 | 13.9 | 0.09 | -1.05 | 0.0917 |
| 2020-07-15 10:00:00 | -0.959 | 27.2 | 13.5 | 0.11 | -1.05 | 0.0922 |
| 2020-08-04 11:30:00 | -1.05 | 26.6 | 17.5 | 0.09 | -0.973 | 0.11 |

| datetime | logTP | TEMP | TBY | TP | Computed | Retransformed |
|---|---|---|---|---|---|---|
| 2020-08-18 11:40:00 | -1.1 | 26.3 | 16.3 | 0.08 | -0.998 | 0.104 |
| 2020-09-01 10:50:00 | -1 | 25.3 | 13.6 | 0.10 | -1.06 | 0.0906 |
| 2021-01-13 10:30:00 | -1.33 | 3.53 | 3.90 | 0.047 | -1.36 | 0.0451 |
| 2021-03-31 10:30:00 | -1.22 | 10.8 | 13.3 | 0.06 | -1.14 | 0.0752 |
| 2021-05-04 10:20:00 | -1.39 | 16.5 | 16.0 | 0.041 | -1.05 | 0.0909 |
| 2021-06-03 10:30:00 | -1.21 | 20.1 | 6.20 | 0.061 | -1.23 | 0.061 |
| 2021-06-21 11:00:00 | -1.4 | 26.6 | 6.73 | 0.04 | -1.18 | 0.0675 |
| 2021-07-20 10:40:00 | -1.3 | 26.0 | 7.20 | 0.05 | -1.18 | 0.0685 |
| 2021-08-10 10:00:00 | -1.22 | 26.3 | 8.80 | 0.06 | -1.15 | 0.0739 |
| 2021-08-31 11:40:00 | -1.15 | 26.1 | 10.5 | 0.07 | -1.11 | 0.0796 |

## References Cited


Bennett, T.J., Graham, J.L., Foster, G.M., Stone, M.L., Juracek, K.E., Rasmussen, T.J., and Putnam, J.E., 2014, U.S. Geological Survey quality-assurance plan for continuous water-quality monitoring in Kansas, 2014: U.S. Geological Survey Open-File Report 2014–1151, 34 p. plus appendixes, accessed September 7, 2022, at https://doi.org/10.3133/ofr20141151.

Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: Journal of the American Statistical Association, v. 78, n. 383 p. 605–610. [Also available at https://doi.org/10.1080/01621459.1983.10478017.]

Eaton, A.D., Clesceri, L.S., and Greenberg, A.E., eds., 1995, Standard methods for the examination of water and wastewater (19th ed.): New York, American Public Health Association, 905 p.

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. [Also available at https://doi.org/10.3133/tm4A3.] [Supersedes USGS Techniques of Water-Resources Investigations, book 4, chap. A3, version 1.1.]

Juracek, K.E., and Rasmussen, P.P., 2008, Sediment quality and comparison to historical water quality, Little Arkansas River Basin, south-central Kansas, 2007: U.S. Geological Survey Scientific Investigations Report 2008–5187, 47 p., accessed July 13, 2022, at https://doi.org/10.3133/sir20085187.

Lane, S.L., Flanagan, S., and Wilde, F.D., 2003, Selection of equipment for water sampling (ver. 2.0): U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A2, accessed September 2022 at http://pubs.water.usgs.gov/twri9A2/

R Core Team, 2020, R—A language and environment for statistical computing: R Foundation for Statistical Computing software release (version 4.0.2), accessed September 7, 2022, at https://www.R-project.org/.

Rasmussen, P.P., Gray, J.R., Glysson, G.D., and 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, 52 p. [Also available at https://doi.org/10.3133/tm3C4.]

Rasmussen, P.P., Eslick, P.J., and Ziegler, A.C., 2016, Relations between continuous real-time physical properties and discrete water-quality constituents in the Little Arkansas River, south-central Kansas, 1998-2014: U.S. Geological Survey Open-File Report 2016–1057, 20 p., https://doi.org/10.3133/ofr20161057.

U.S. Geological Survey, 2006, Collection of water samples (ver. 2.0, September 2006): U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A4 [variously paged]. [Also available at https://doi.org/10.3133/twri09A4.]

U.S. Geological Survey, 2016, Policy and guidance for approval of surrogate regression models for computation of time series suspended-sediment concentration and loads: U.S. Geological Survey Office of Surface Water Technical Memorandum 2016.07, Office of Water Quality Technical Memorandum 2016.10, 40 p., accessed September 7, 2022, at https://water.usgs.gov/water-resources/memos/memo.php?id=467.

U.S. Geological Survey, 2022, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed September 7, 2022, at https://doi.org/10.5066/F7P55KJN.

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. plus 8 attachments. [Also available at https://doi.org/10.3133/tm1D3.]

YSI, Inc., 2017, EXO user manual—Advanced water quality monitoring platform (rev. G): Yellow Springs, Ohio, YSI, Inc., 154 p., accessed September 7, 2022, at https://www.ysi.com/file%20library/documents/manuals/exo-user-manual-web.pdf.