

## **Appendix 1.25. Model Archive Summary for Fecal Coliform Bacteria Density at U.S. Geological Survey site 07143672; Little Arkansas River at Highway 50 near Halstead, Kansas, during March 2017 through December 2019**

This model archive summary summarizes the fecal coliform bacteria model developed to compute hourly or daily fecal coliform bacteria. Model development methods follow U.S. Geological Survey (USGS) guidance from Office of Surface Water/Office of Water Quality Technical Memoranda and USGS Techniques and Methods, book 3, chap. C4 (Rasmussen and others, 2009).

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: 07143672

Site Name: Little Arkansas River at Highway 50 near Halstead, Kansas

Location: Latitude 38°01'43", longitude 97°32'25" referenced to North American Datum of 1927, in NW 1/4 NE 1/4 NE 1/4 sec.28, T.23 S., R.2 W., Harvey County, Kansas, hydrologic unit 11030012.

Equipment: A Sutron Satlink II High Data Rate Collection Platform and a Design Analysis Water Log H350/355 nonsubmersible pressure transducer transfers real-time stage and water-quality data via satellite. The primary reference gage is a Type-A wire-weight gage located on the downstream bridge guardrail. Check-bar elevation is 33.396 feet. The orifice tube is enclosed in 1.25-inch steel conduit trenched into the ground down to the edge of water, where the orifice emerges from the bank and culminates in a 2-inch open-end orifice tethered to a steel fencepost near the left edge of water. Gage height was measured during May 1998 through December 2019. A YSI 6600 water-quality monitor equipped with water temperature, specific conductance, pH, dissolved oxygen, and turbidity (a YSI Model 6026 [December 1998 through December 2006] and YSI Model 6136 [July 2004 through December 2017]) sensors collected data during May 1998 through December 2017. A YSI EXO2 water-quality monitor equipped with water temperature, specific conductance, pH, dissolved oxygen, turbidity, and fluorescent dissolved organic matter sensors collected data during January 2017 through December 2019. A Hach Nitratex monitor collected nitrate data during February 2017 through December 2019.

Date model was developed: June 1, 2020

Model calibration data period: March 30, 2017 through December 10, 2019

### **Model Data**

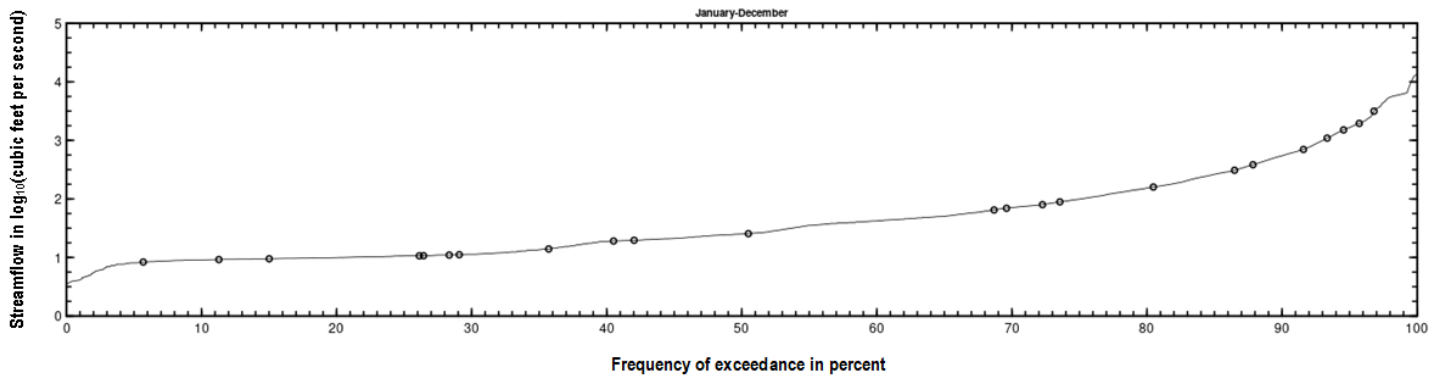
All data were collected using USGS protocols (U.S. Geological Survey, variously dated; Wagner and others, 2006; Sauer and Turnipseed, 2010; Turnipseed and Sauer, 2010) and are stored in the National Water Information System (NWIS) database (U.S. Geological Survey, 2021). Explanatory variables were evaluated individually and in combination. Potential explanatory variables included streamflow, water temperature, specific conductance, pH, dissolved oxygen, YSI EXO2 turbidity, nitrate, and fluorescent dissolved organic matter. Seasonal components (sine and cosine variables) also were evaluated as explanatory variables.

The regression model is based on 23 concomitant values of discretely collected fecal coliform bacteria and continuously measured turbidity during March 2017 through December 2019. Discrete samples were collected over a range of streamflow and turbidity conditions. No samples had densities that were below laboratory detection limits. Summary statistics and the complete model-calibration dataset are provided below. Outliers and influential points were identified using studentized residuals, DFITS, Cook's D (Cook, 1977), and leverage. Outliers in previously published versions of this model (Christensen and others, 2003; Rasmussen and others, 2016) were examined and retained in the dataset if there were no clear issues, explanations, or conditions that would cause a result to be invalid for model calibration. One sample (collection date March 14, 2019) was not representative of the dataset and exceeded Cook's D and DFITS outlier criteria and was removed from the model dataset to avoid erroneous inflation of model-computed values at the upper range of surrogate relations. Removing data points based only on outlier criteria may only overestimate the certainty of the model.

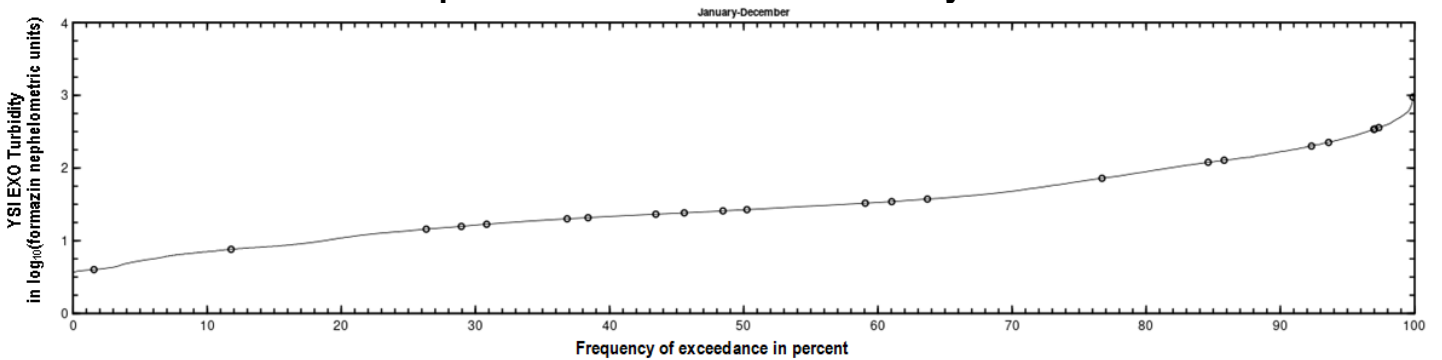
## Fecal Coliform Bacteria

Discrete samples were collected from the downstream side of the bridge or instream within 50 feet of the bridge using equal-width-increment, multi-vertical, single vertical or grab-dip methods following U.S. Geological Survey (variously dated) and Rasmussen and others (2014). Discrete samples were collected on a semifixed to event-based schedule ranging from 6 to 9 samples per year with a FISP US DH-95 or D-95 with a Teflon bottle, cap, and nozzle depth-integrating sampler, a DH-1 with a Teflon bottle, cap, and nozzle hand sampler or a grab sample with a Teflon bottle depending on sample location. Samples were analyzed for fecal coliform bacteria by the U.S. Geological Survey Kansas Water Science Center.

## Fecal Coliform Bacteria Samples Plotted on Streamflow Duration Curve



## Fecal Coliform Bacteria Samples Plotted on YSI EXO Turbidity Duration Curve



## Continuous Data

Concomitant turbidity values were time interpolated. If no concomitant continuous data were available within 2 hours of sample collection, the sample was not included in the dataset.

## Model Development

Ordinary least squares regression analysis was done using R (version 4.0.0) programming language (R Core Team, 2020) to relate discretely collected fecal coliform bacteria to turbidity and other continuously measured data. The distribution of residuals was examined for normality and plots of residuals (the difference between the measured and model-calculated values) compared to model-computed fecal coliform bacteria were examined for homoscedasticity (departures from zero did not change substantially over the range of model-calculated values). Previously published explanatory variables were also strongly considered for continuity; however, the best explanatory variable(s) were ultimately selected.

Turbidity was selected as the best predictor of fecal coliform bacteria based on residual plots, high coefficient of determination ( $R^2$ ), and low model standard percentage error (MSPE). Turbidity was positively correlated with fecal coliform bacteria because turbidity measures light scattered by particulates in water.

## Model Summary

Summary of final fecal coliform bacteria regression analysis at USGS site number 07143672:

Fecal coliform bacteria-based model:

$$\log_{10}(FC) = 1.13 \times \log_{10}(TBY) + 0.91$$

where,

$\log_{10}$  = logarithm base 10;

$FC$  = fecal coliform bacteria, in colony forming units per 100 milliliters (cfu/100 mL); and

$TBY$  = turbidity, in formazin nephelometric units (FNU)

The log-transformed model may be retransformed to original units so that FC 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 1.25. The retransformed model, accounting for BCF is:

$$FC = 10.16 \times TBY^{1.13}$$

## Model Statistics, Data, and Plots

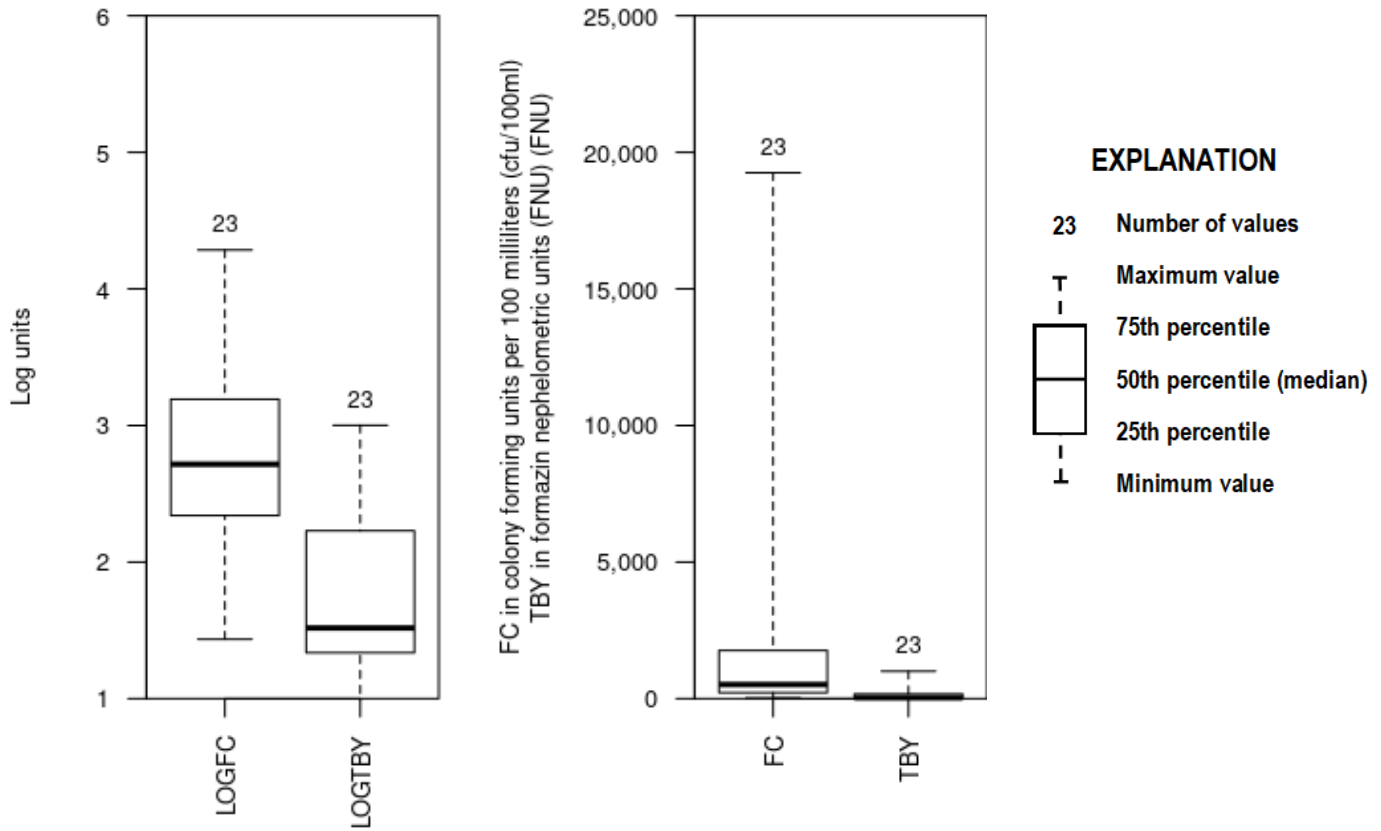
### Model

$$\text{LOGFC} = + 1.13 * \text{LOGTBY} + 0.91$$

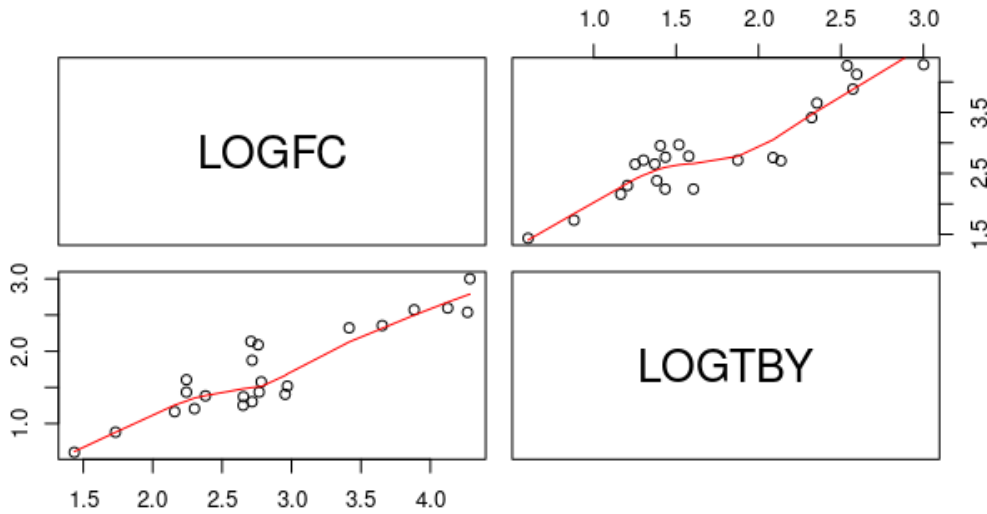
### Variable Summary Statistics

|              | LOGFC | FC      | LOGTBY | TBY    |
|--------------|-------|---------|--------|--------|
| Minimum      | 1.44  | 27.3    | 0.602  | 4.0    |
| 1st Quartile | 2.30  | 200.0   | 1.300  | 20.0   |
| Median       | 2.72  | 520.0   | 1.520  | 32.9   |
| Mean         | 2.86  | 3170.0  | 1.720  | 139.0  |
| 3rd Quartile | 3.41  | 2600.0  | 2.320  | 210.0  |
| Maximum      | 4.28  | 19300.0 | 3.000  | 1000.0 |

## Box Plots



## Exploratory Plots



## Basic Model Statistics

|                                                     |       |
|-----------------------------------------------------|-------|
| Number of Observations                              | 23    |
| Standard error (RMSE)                               | 0.317 |
| Average Model standard percentage error (MSPE)      | 79.6  |
| Coefficient of determination ( $R^2$ )              | 0.836 |
| Adjusted Coefficient of Determination (Adj. $R^2$ ) | 0.828 |
| Bias Correction Factor (BCF)                        | 1.25  |

## Explanatory Variables

|             | Coefficients | Standard Error | t value | Pr(> t ) |
|-------------|--------------|----------------|---------|----------|
| (Intercept) | 0.91         | 0.200          | 4.56    | 1.72e-04 |
| LOGTBY      | 1.13         | 0.109          | 10.40   | 1.05e-09 |

## Correlation Matrix

|           | Intercept | E.vars |
|-----------|-----------|--------|
| Intercept | 1.000     | -0.944 |
| E.vars    | -0.944    | 1.000  |

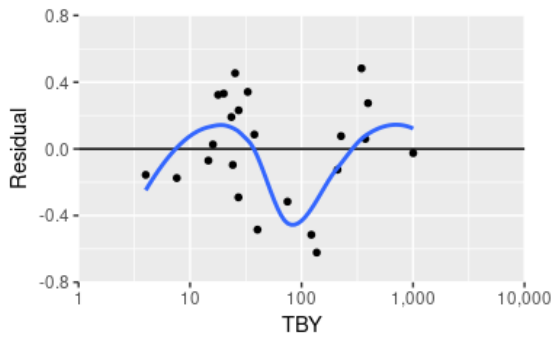
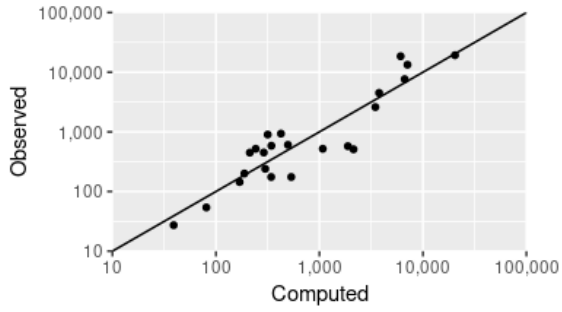
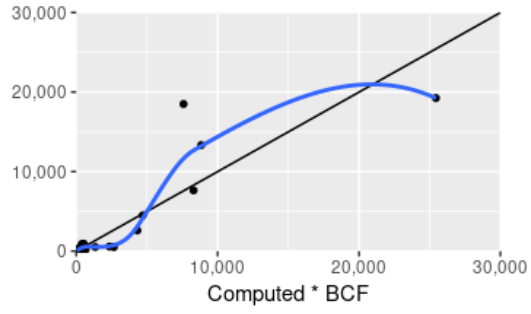
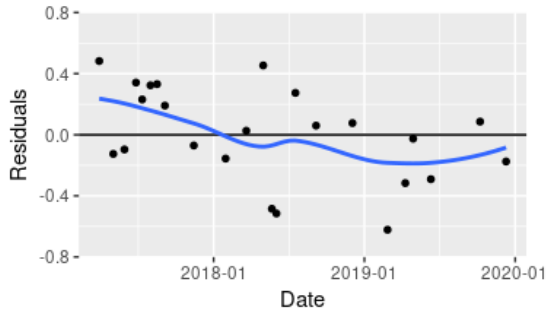
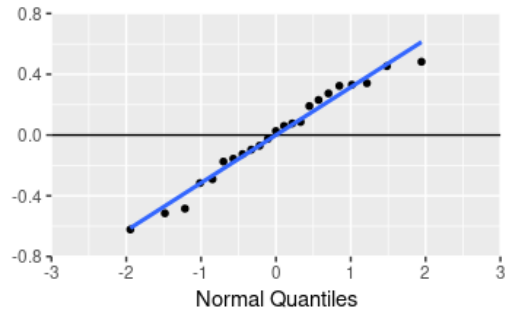
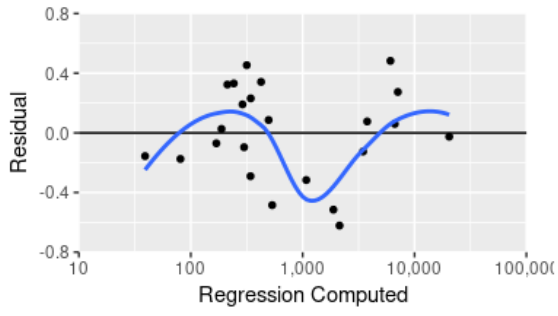
## Outlier Test Criteria

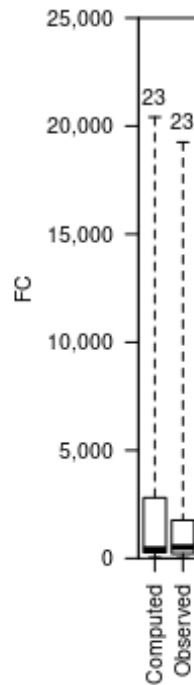
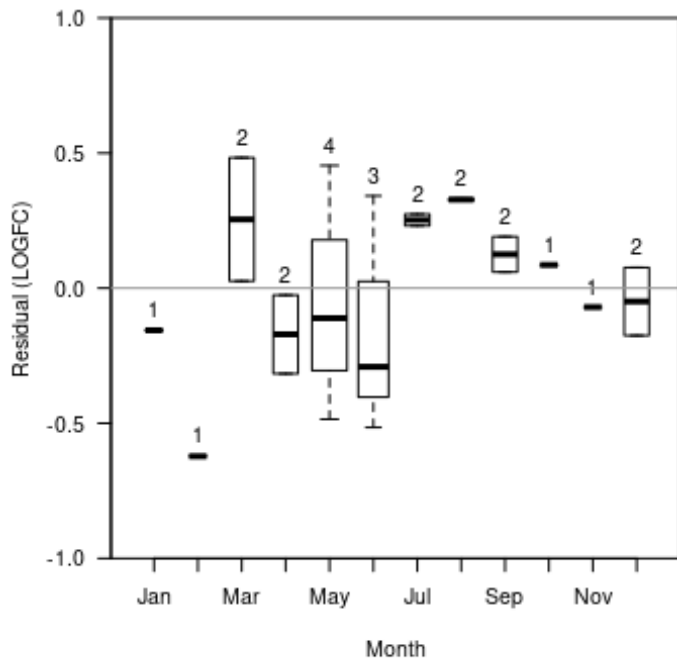
| Leverage | Cook's D | DFFITS |
|----------|----------|--------|
| 0.261    | 0.193    | 0.590  |

## Flagged Observations

|                 | LOGFC | Estimate | Residual | Standard Residual | Studentized Residual | Leverage | Cook's D | DFFITS |
|-----------------|-------|----------|----------|-------------------|----------------------|----------|----------|--------|
| 3/30/2017 10:30 | 4.27  | 3.78     | 0.483    | 1.63              | 1.7                  | 0.123    | 0.185    | 0.634  |

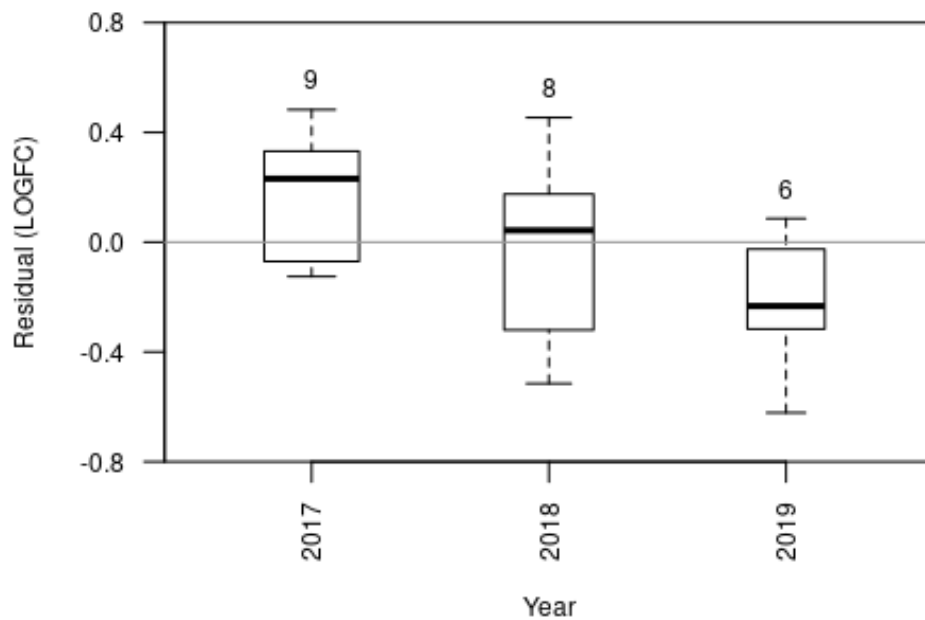
# Statistical Plots





### EXPLANATION

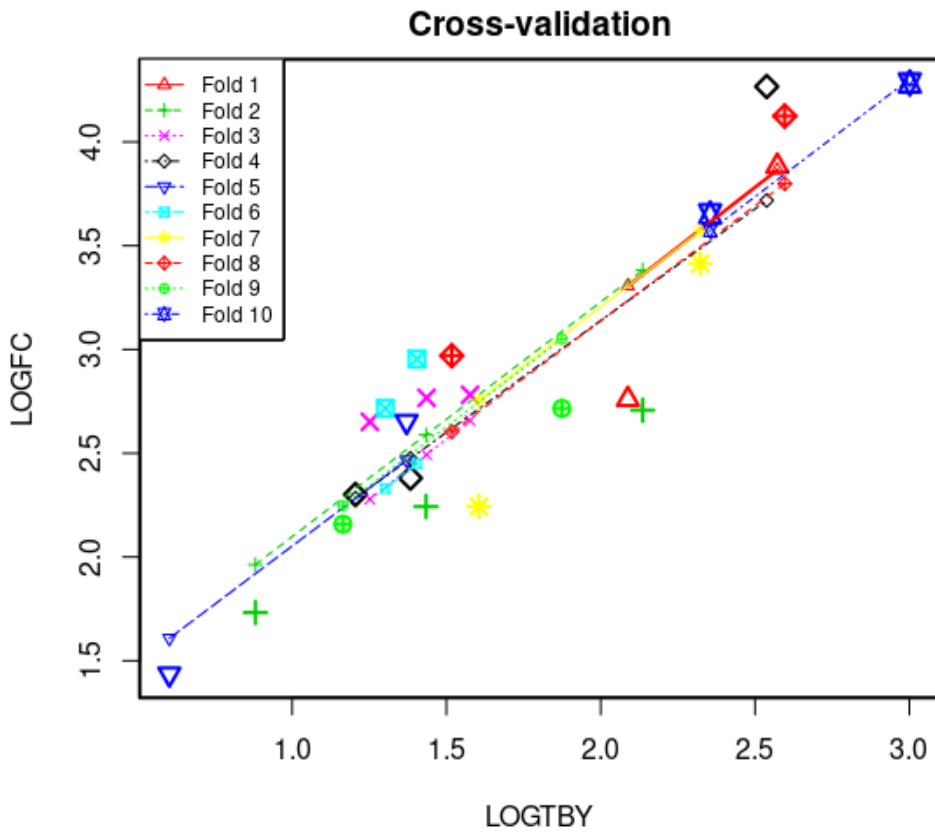
- 23 Number of values
- T Maximum value
- 75th percentile
- 50th percentile (median)
- 25th percentile
- Minimum value



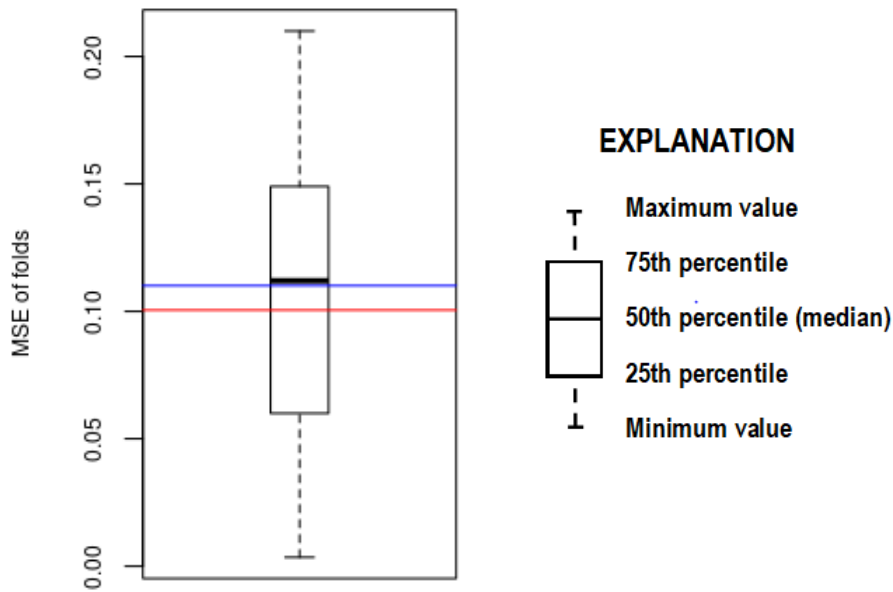
### EXPLANATION

- 9 Number of values
- T Maximum value
- 75th percentile
- 50th percentile (median)
- 25th percentile
- Minimum value

# Cross Validation



Minimum MSE of folds: 0.00347  
Mean MSE of folds: 0.11000  
Median MSE of folds: 0.11200  
Maximum MSE of folds: 0.21000  
(Mean MSE of folds) / (Model MSE): 1.10000



Red line - Model MSE

Blue line - Mean MSE of folds



## Model-Calibration Dataset

|    | Date       | LOGFC | LOGTBY | FC    | TBY  | Computed<br>LOGFC | Computed<br>FC | Residual | Normal<br>Quantiles |
|----|------------|-------|--------|-------|------|-------------------|----------------|----------|---------------------|
| 1  | 3/30/2017  | 4.27  | 2.54   | 18500 | 345  | 3.78              | 7590           | 0.483    | 1.95                |
| 2  | 5/3/2017   | 3.41  | 2.32   | 2600  | 210  | 3.54              | 4320           | -0.125   | -0.447              |
| 3  | 5/30/2017  | 2.38  | 1.38   | 240   | 24.2 | 2.48              | 373            | -0.0963  | -0.33               |
| 4  | 6/27/2017  | 2.97  | 1.52   | 933   | 32.9 | 2.63              | 530            | 0.342    | 1.22                |
| 5  | 7/12/2017  | 2.77  | 1.44   | 585   | 27.3 | 2.54              | 428            | 0.231    | 0.57                |
| 6  | 8/1/2017   | 2.65  | 1.25   | 448   | 17.8 | 2.33              | 265            | 0.324    | 0.848               |
| 7  | 8/17/2017  | 2.72  | 1.3    | 520   | 20   | 2.38              | 302            | 0.331    | 1.01                |
| 8  | 9/5/2017   | 2.65  | 1.37   | 450   | 23.5 | 2.46              | 361            | 0.191    | 0.447               |
| 9  | 11/14/2017 | 2.16  | 1.16   | 144   | 14.6 | 2.23              | 211            | -0.0703  | -0.218              |
| 10 | 1/30/2018  | 1.44  | 0.602  | 27.3  | 4    | 1.59              | 48.7           | -0.156   | -0.57               |
| 11 | 3/21/2018  | 2.3   | 1.21   | 200   | 16   | 2.27              | 235            | 0.0263   | 0                   |
| 12 | 5/1/2018   | 2.95  | 1.4    | 900   | 25.4 | 2.5               | 394            | 0.454    | 1.48                |
| 13 | 5/22/2018  | 2.24  | 1.6    | 175   | 40.3 | 2.73              | 666            | -0.485   | -1.22               |
| 14 | 6/2/2018   | 2.76  | 2.09   | 575   | 122  | 3.27              | 2350           | -0.515   | -1.48               |
| 15 | 7/18/2018  | 4.12  | 2.6    | 13300 | 395  | 3.85              | 8840           | 0.274    | 0.702               |
| 16 | 9/6/2018   | 3.88  | 2.57   | 7640  | 373  | 3.82              | 8300           | 0.0598   | 0.108               |
| 17 | 12/3/2018  | 3.65  | 2.35   | 4500  | 226  | 3.58              | 4700           | 0.0764   | 0.218               |
| 18 | 2/26/2019  | 2.71  | 2.14   | 510   | 137  | 3.33              | 2660           | -0.622   | -1.95               |
| 19 | 4/10/2019  | 2.72  | 1.87   | 520   | 74.8 | 3.03              | 1340           | -0.316   | -1.01               |
| 20 | 4/29/2019  | 4.28  | 3      | 19300 | 1000 | 4.31              | 25400          | -0.0255  | -0.108              |
| 21 | 6/11/2019  | 2.24  | 1.43   | 175   | 27.2 | 2.53              | 426            | -0.291   | -0.848              |
| 22 | 10/8/2019  | 2.78  | 1.58   | 605   | 37.7 | 2.7               | 619            | 0.086    | 0.33                |
| 23 | 12/10/2019 | 1.73  | 0.881  | 54    | 7.6  | 1.91              | 101            | -0.175   | -0.702              |

## Definitions

FC: Fecal coliforms in cfu/100mL (31625)

TBY: Turbidity in FNU (63680)

## References Cited

- Christensen, V.G., Ziegler, A.C., Rasmussen P.P., and Jian X., 2003, Continuous real-time water-quality monitoring of Kansas streams, *in* Proceedings of 2003 Spring Specialty Conference on Agricultural Hydrology and Water Quality, Kansas City, Mo., May 12–14, 2003: Middleburg, Va., American Water Resources Association Technical Publication Series No. TPS-03-1, compact disc. [Also available at <https://nrtwq.usgs.gov/ks/methods/christensen2003>.]
- Cook, D.R., 1977, Detection of influential observation in linear regression: *Technometrics*, v. 19, no. 1, p. 15–18. [Also available at [https://www.jstor.org/stable/1268249?seq=4#metadata\\_info\\_tab\\_contents](https://www.jstor.org/stable/1268249?seq=4#metadata_info_tab_contents).]
- 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>.]
- R Core Team, 2020, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, version 4.0.0. [Also available at <https://www.r-project.org>.]
- 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, 16 p. [Also available at <https://doi.org/10.3133/ofr20161057>.]
- 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, 53 p. [Also available at <https://doi.org/10.3133/tm3C4>.]

- 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. [Also available at <https://doi.org/10.3133/ofr20141233>.]
- Sauer, V.B., and Turnipseed, D.P., 2010, Stage measurement at gaging stations: U.S. Geological Survey Techniques and Methods, book 3, chap. A7, 45 p. [Also available at <https://doi.org/10.3133/tm3A7>.]
- Turnipseed, D.P., and Sauer, V.B., 2010, Discharge measurements at gaging stations: U.S. Geological Survey Techniques and Methods, book 3, chap. A8, 87 p. [Also available at <https://doi.org/10.3133/tm3A8>.]
- U.S. Geological Survey, 2021, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed December 8, 2021, at <https://doi.org/10.5066/F7P55KJN>.
- 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]. [Also available at <https://water.usgs.gov/owq/FieldManual/>.]
- 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, 96 p. [Also available at <https://doi.org/10.3133/tm1D3>.]