Appendix 2.25. Model Archive Summary for Fecal Coliform Bacteria Density at U.S. Geological Survey site 07144100; Little Arkansas River near Sedgwick, Kansas, during December 2014 through October 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: 07144100

Site Name: Little Arkansas River near Sedgwick, Kansas

Location: Latitude 37°52'59", longitude 97°25'27" referenced to North American Datum of 1927, in NE 1/4 NW 1/4 NW 1/4 sec.15, T.25 S., R.1 W., Sedgwick 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 handrail. Check-bar elevation is 33.614 feet. The orifice is enclosed in a well-screen and attached to a concrete pier on the left downstream side of the bridge. Gage height was measured during December 2014 through October 2019. A YSI 6600 water-quality monitor equipped with water temperature, specific conductance, pH, dissolved oxygen, and turbidity (a YSI Model 6026 [September 1998 through December 2006] and YSI Model 6136 [July 2004 through March 2015]) sensors collected data during April 1998 through March 2015. 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 September 2014 through December 2019. A Hach Nitratax monitor collected nitrate data during March 2012 through December 2019.

Date model was developed: June 1, 2020

Model calibration data period: December 9, 2014 through October 9, 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 36 concomitant values of discretely collected fecal coliform bacteria and continuously measured turbidity during December 2014 through October 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. Two samples (collection dates February 25, 2015 and August 20, 2019) were not representative of the dataset and exceeded Cook's D and DFITS outlier criteria and were 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 1 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–81 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
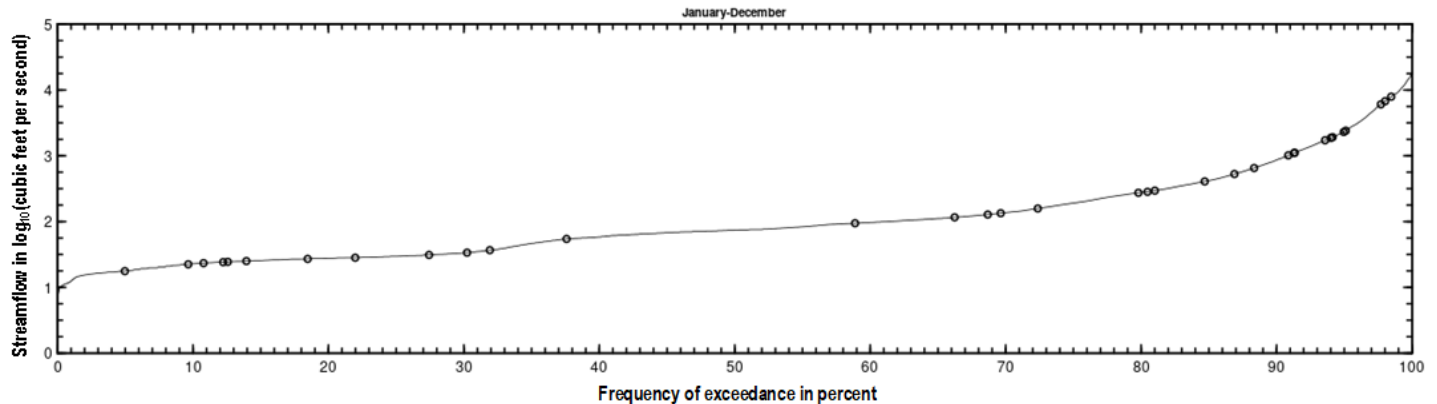
Fecal Coliform Bacteria Samples Plotted on YSI EXO Turbidity Duration Curve
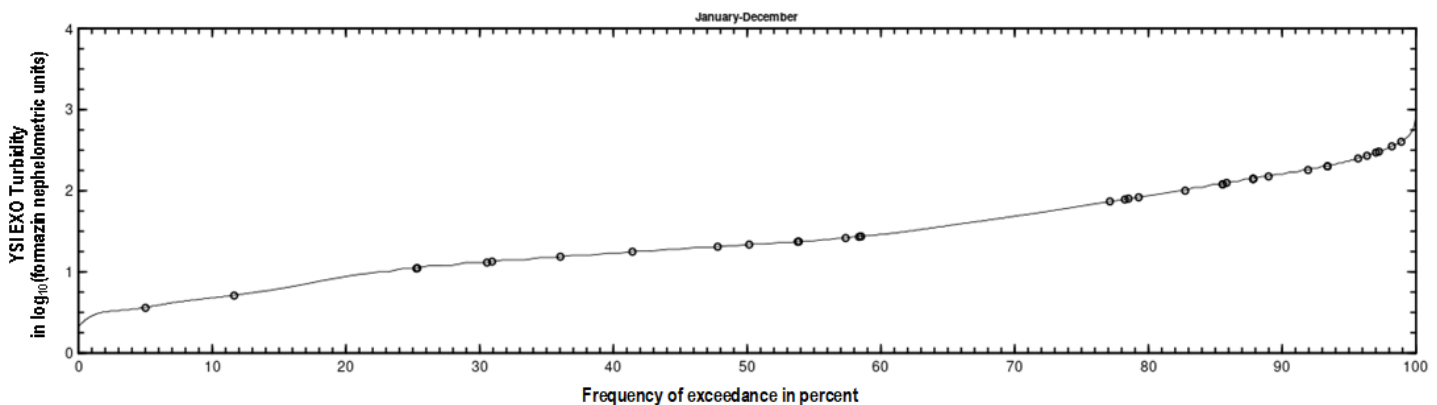
Continuous Data

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 07144100:

Fecal coliform bacteria-based model:

$$\log_{10}(FC) = 1.22 \times \log_{10}(TBY) + 0.738$$

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 1.55. The retransformed model, accounting for BCF is:

$$FC = 8.479 \times TBY^{1.22}$$
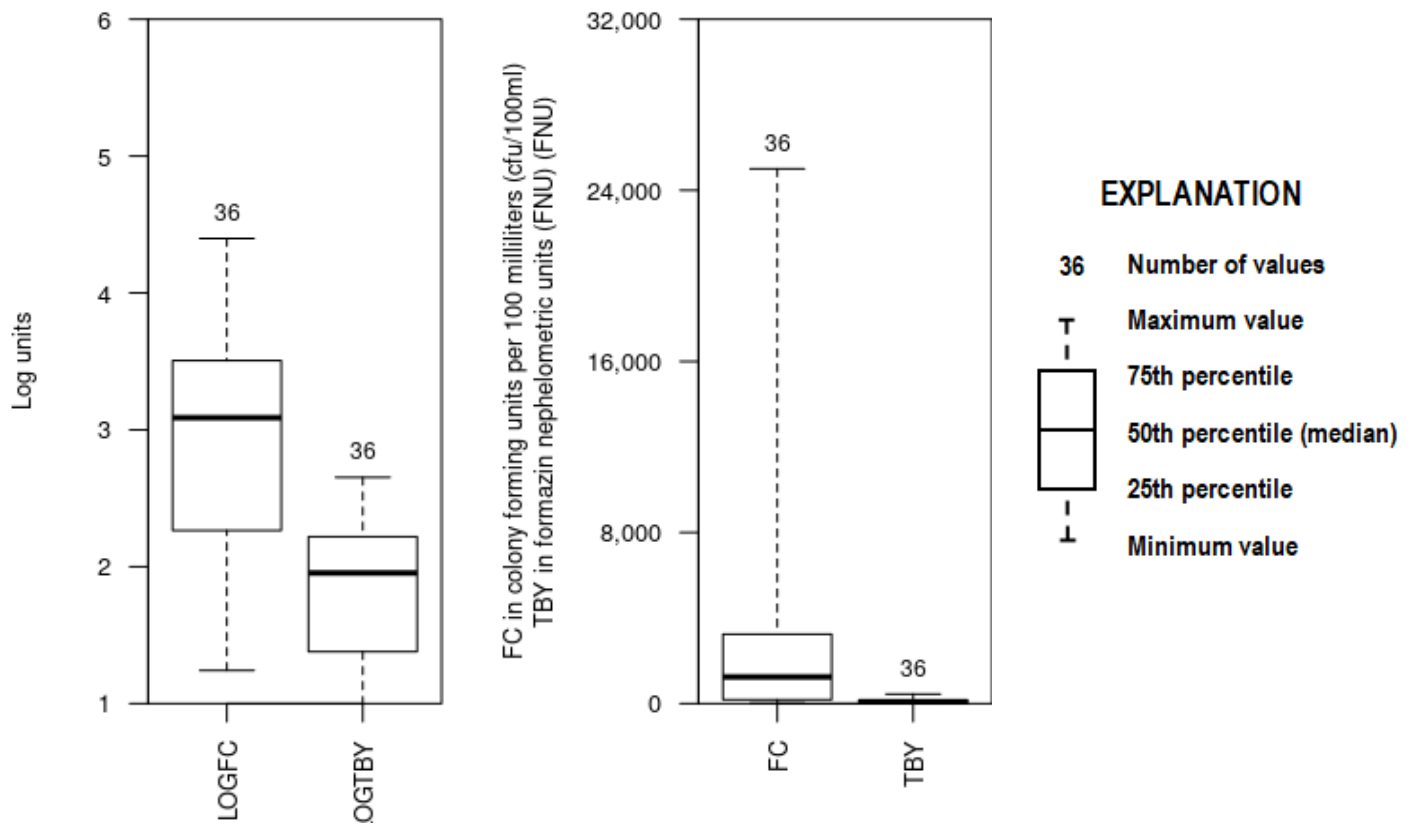
Model Statistics, Data, and Plots

Model

LOGFC = + 1.22 * LOGTBY + 0.738

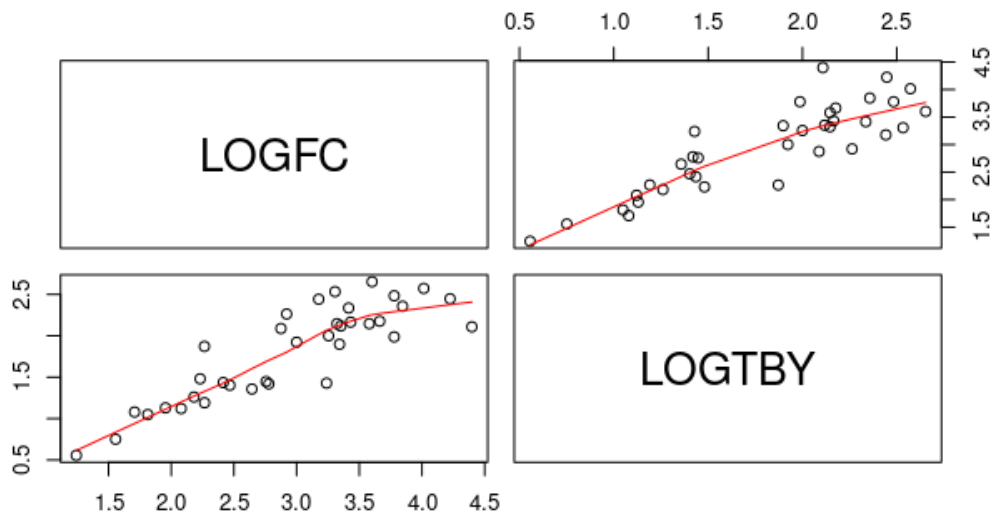### Variable Summary Statistics

	LOGFC	FC	LOGTBY	TBY
Minimum	1.24	17.5	0.556	3.6
1st Quartile	2.27	184.0	1.380	24.0
Median	3.09	1250.0	1.950	90.3
Mean	2.94	3010.0	1.800	120.0
3rd Quartile	3.51	3250.0	2.220	167.0
Maximum	4.40	25000.0	2.650	450.0

Box Plots

Exploratory Plots

Basic Model Statistics

```
Number of Observations                          36
Standard error (RMSE)                        0.386
Average Model standard percentage error (MSPE)  101
Coefficient of determination (R²)            0.768
Adjusted Coefficient of Determination (Adj. R²) 0.761
Bias Correction Factor (BCF)                  1.55
```

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.738	0.218	3.39	1.78e-03
LOGTBY	1.220	0.115	10.60	2.60e-12

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.955
E.vars	-0.955	1.000

Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.167	0.194	0.471

Flagged Observations

	LOGFC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
6/17/2016 12:10	4.4	3.32	1.08	2.85	3.22	0.0363	0.153	0.625

Statistical Plots

Cross Validation

Cross-validation plot with LOGTBY on the x-axis and LOGFC on the y-axis, showing Fold 1 through Fold 10.

```
         Minimum MSE of folds:   0.0017
            Mean MSE of folds:   0.1560
          Median MSE of folds:   0.1240
         Maximum MSE of folds:   0.5520
 (Mean MSE of folds) / (Model MSE):   1.0500
```
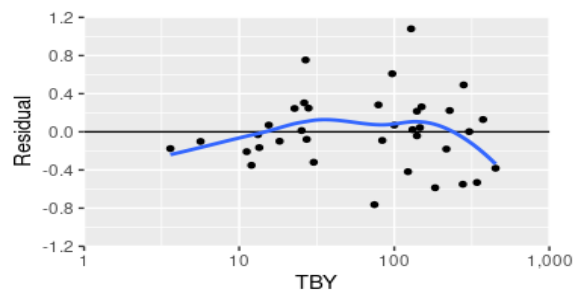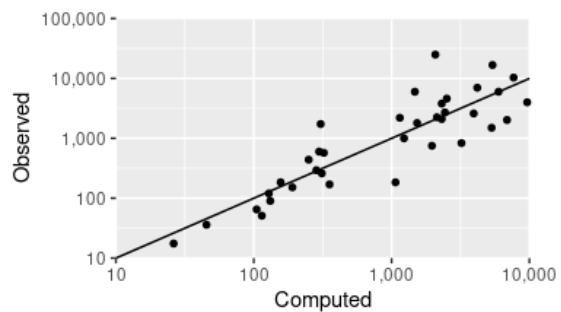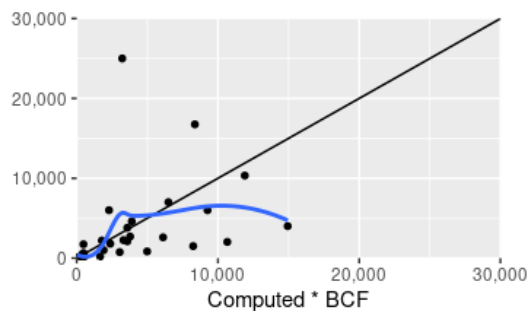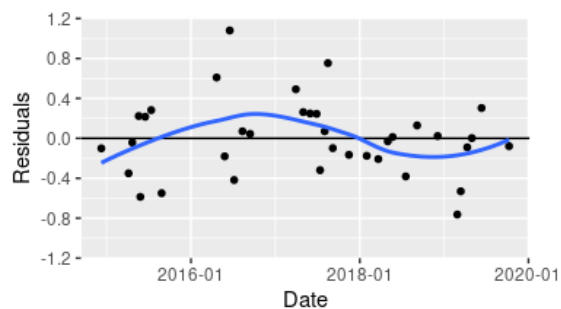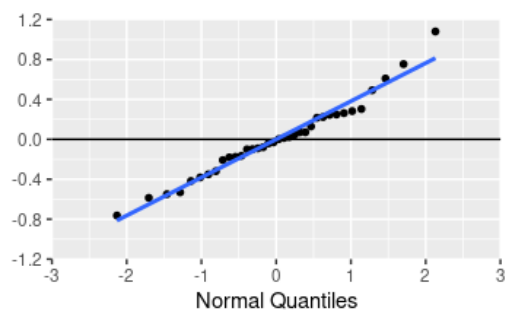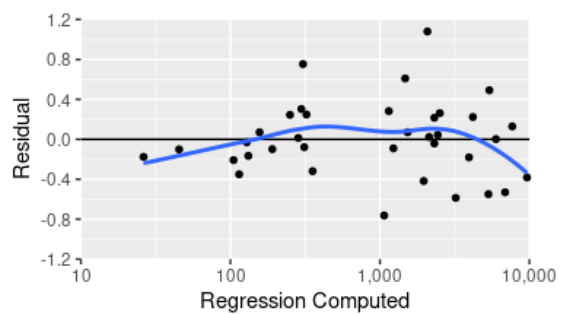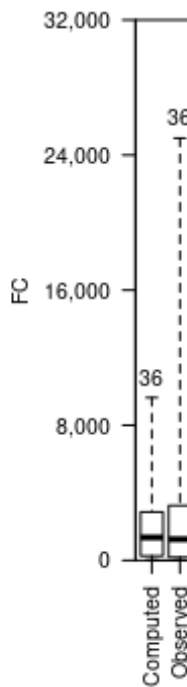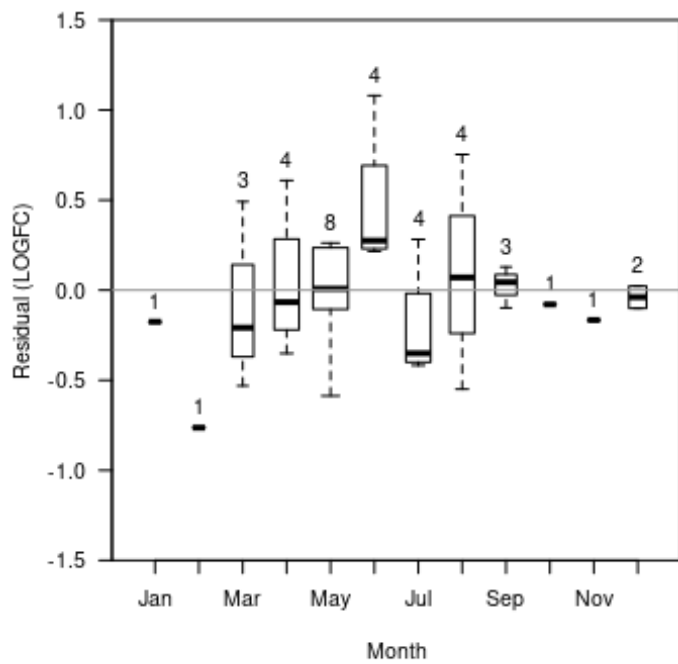

Boxplot of MSE of folds with EXPLANATION legend: Maximum value, 75th percentile, 50th percentile (median), 25th percentile, Minimum value.
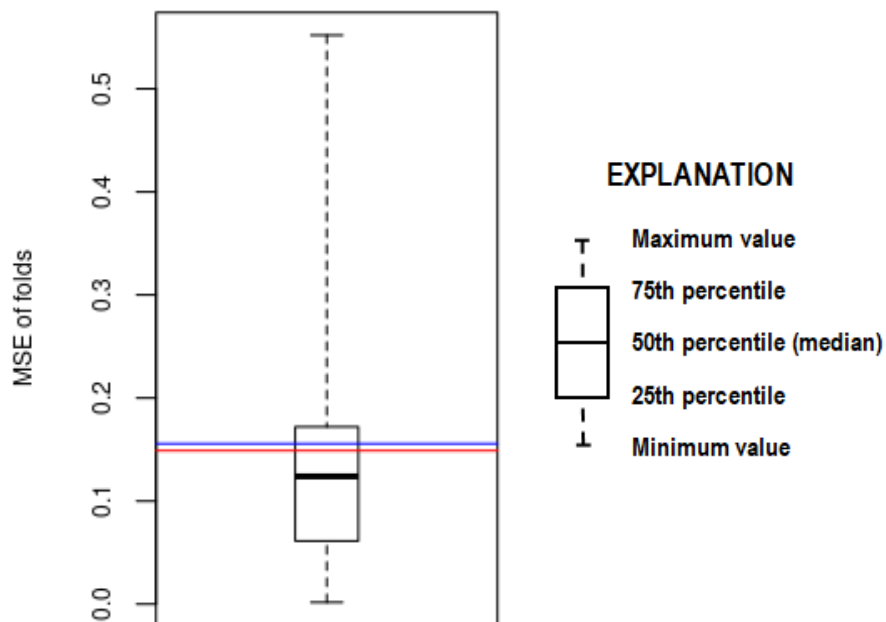
Red line - Model MSE

Blue line - Mean MSE of folds

Model-Calibration Dataset

	Date	LOGFC	LOGTBY	FC	TBY	Computed LOGFC	Computed FC	Residual	Normal Quantiles
1	12/9/2014	1.56	0.751	36	5.64	1.66	70.4	-0.101	-0.391
2	4/6/2015	1.71	1.08	51	12	2.06	177	-0.351	-0.907
3	4/22/2015	3.32	2.15	2100	140	3.36	3580	-0.0416	-0.104
4	5/20/2015	3.85	2.36	7000	228	3.62	6490	0.223	0.627
5	5/27/2015	2.92	2.26	833	183	3.51	4980	-0.586	-1.7
6	6/17/2015	3.58	2.15	3800	140	3.36	3580	0.216	0.545
7	7/13/2015	3.34	1.9	2200	79	3.06	1780	0.283	1.02
8	8/27/2015	3.18	2.44	1500	277	3.73	8250	-0.55	-1.46
9	4/21/2016	3.78	1.99	6000	97	3.17	2290	0.609	1.46
10	5/26/2016	3.41	2.34	2600	217	3.6	6110	-0.181	-0.627
11	6/17/2016	4.4	2.11	25000	128	3.32	3220	1.08	2.13
12	7/6/2016	2.88	2.09	750	123	3.29	3040	-0.418	-1.14
13	8/11/2016	3.26	2	1800	100	3.19	2370	0.0703	0.317
14	9/13/2016	3.43	2.17	2700	146	3.39	3780	0.0444	0.245
15	3/30/2017	4.22	2.45	16700	280	3.73	8370	0.492	1.28
16	5/1/2017	3.66	2.18	4600	150	3.4	3900	0.262	0.907
17	5/31/2017	2.76	1.45	571	28	2.51	500	0.248	0.807
18	6/28/2017	2.64	1.36	440	22.8	2.4	388	0.245	0.714
19	7/13/2017	2.23	1.48	170	30.2	2.55	550	-0.319	-0.807
20	8/2/2017	2.27	1.19	184	15.5	2.19	243	0.0709	0.391
21	8/16/2017	3.24	1.43	1730	26.8	2.49	474	0.754	1.7
22	9/6/2017	2.18	1.26	152	18.2	2.28	296	-0.0983	-0.317
23	11/15/2017	1.95	1.13	90	13.5	2.12	204	-0.165	-0.466
24	1/31/2018	1.24	0.556	17.5	3.6	1.42	40.7	-0.176	-0.545
25	3/22/2018	1.81	1.05	65	11.2	2.02	163	-0.208	-0.714
26	5/2/2018	2.08	1.12	120	13.2	2.11	199	-0.0302	-0.0346
27	5/23/2018	2.47	1.4	293	25.3	2.45	441	0.0133	0.104
28	7/19/2018	3.6	2.65	4000	450	3.98	14900	-0.382	-1.02
29	9/6/2018	4.01	2.57	10300	373	3.89	11900	0.129	0.466
30	12/4/2018	3.35	2.12	2250	131	3.33	3310	0.0232	0.174
31	2/27/2019	2.26	1.87	184	74.5	3.03	1660	-0.764	-2.13
32	3/14/2019	3.31	2.53	2030	341	3.84	10700	-0.531	-1.28
33	4/11/2019	3	1.92	1000	83.7	3.09	1910	-0.0902	-0.245
34	5/1/2019	3.78	2.48	6000	304	3.78	9260	0.00202	0.0346
35	6/12/2019	2.78	1.42	600	26.3	2.47	462	0.304	1.14
36	10/9/2019	2.41	1.44	260	27.2	2.49	483	-0.0789	-0.174

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.]

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.]