Appendix 29. Model Archival Summary for Fecal Coliform Bacteria Density at U.S. Geological Survey Site 06887500, Kansas River at Wamego, Kansas, during October 2013 through September 2019

This model archival summary summarizes the fecal coliform bacteria (FCB; U.S. Geological Survey [USGS] parameter code 31625) density model developed to compute 15-minute FCB densities from October 2013 onward. This model supersedes all previous models.

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: 06887500
Site name: Kansas River at Wamego, Kansas
Location: Lat 39°11'54", long 96°18'19" referenced to North American Datum of 1927, in SW 1/4 NW 1/4 SE 1/4 sec.9, T.10 S., R.10 E., Pottawatomie County, Kans., hydrologic unit 10270102.

Equipment: A YSI 6600 water-quality monitor equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, and turbidity (TBY) was installed from August 2012 through June 2014. A Xylem YSI EXO2 water-quality monitor equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, TBY, and chlorophyll and phycocyanin fluorescence was installed during June 2014 through September 2019. The monitor was housed in a 4-inch-diameter galvanized steel pipe. Readings from the water-quality monitor were recorded every 15 minutes and transmitted by way of satellite, hourly.

Date model was created: May 5, 2020

Model calibration data period: October 21, 2013, through September 23, 2019

Model application date: October 21, 2013, onward

Model-Calibration Dataset

All data were collected using USGS protocols (Wagner and others, 2006; U.S. Geological Survey, variously dated) and are stored in the National Water Information System (U.S. Geological Survey, 2020) database and available to the public. Ordinary least squares analysis was used to develop regression models using R programming language (R Core Team, 2020). Potential explanatory variables that were evaluated individually and in combination included streamflow, water temperature, specific conductance, dissolved oxygen, pH, TBY, and chlorophyll and phycocyanin fluorescence. These potential explanatory variables were interpolated within the 15-minute continuous record based on sample time. The maximum time span between two continuous data points used for interpolation was 2 hours (in order 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 1 hour in the continuous data record resulted in missing interpolated data). Seasonal components (sine and cosine variables) were also evaluated as potential explanatory variables.

The final selected regression model was based on 77 concurrent measurements of FCB density, sensor-measured TBY, and seasonal components (sine and cosine variables) during October 21, 2013, through September 23, 2019. Samples were collected throughout the range of continuously observed hydrologic conditions. No samples had densities below laboratory detection limits. Thirty-three sample densities were qualified as “estimated.” Summary statistics and the complete model-calibration dataset are provided below. Potential outliers were identified using the methods described in Rasmussen and others (2009). Additionally, studentized residuals from the final model were inspected for values greater than three or less than negative three. Values outside of that range were considered potential outliers and were investigated. All potential outliers were not found to have errors associated with collection, processing, or analysis and were therefore considered valid.

This model is specific to the Kansas River at Wamego, Kans., during this study period and cannot be applied to data collected from other sites on the Kansas River or data collected from other waterbodies.

Fecal Coliform Bacteria Sampling Details

Indicator bacteria samples typically were collected either from the downstream side of the bridge or instream within 100 feet of the bridge. The grab sample collection method with weighted basket was used for all indicator bacteria samples (contrary to the equal-width-increment collection method used for all other analytes; U.S. Geological Survey, variously dated). During July 2012 through June 2017, grab samples were collected every 2 weeks during March through October, once a month during November through February, and during selected reservoir release and runoff events. During July 2017 through September 2019, grab samples were

collected on a monthly to bimonthly basis, depending on flow conditions. An open-mouth bottle with weighted-basket sampler was used. Additional detail on sample collection is available in Foster and Graham (2016) and Graham and others (2018). Samples were analyzed for FCB density at the USGS Kansas Water Science Center in Lawrence, Kans.

Model Development

Ordinary least squares regression analysis was done using R programming language (R Core Team, 2020) to relate discretely collected FCB density to sensor-measured TBY and seasonal components (sine and cosine variables). 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). Previously published explanatory variables were also strongly considered for continuity.

TBY and seasonal components (sine and cosine variables) were selected as good surrogates for FCB based on residual plots, coefficient of determination (R^2), and model standard percentage error. Values for all the aforementioned statistics were computed and are included below along with all relevant sample data and additional statistical information.

Model Summary

The following is a summary of final regression analysis for FCB density at USGS site 06887500:

FCB density-based model:

$$\log FCB = 0.968 \times \log TBY - 0.286 \times \sin(2\pi D) - 0.339 \times \cos(2\pi D) + 0.442$$

where

log = logarithm base 10;

FCB = fecal coliform bacteria density, in colonies per 100 milliliters;

TBY = turbidity, in formazin nephelometric units;

sin = sine;

D = date, in decimal years; and

cos = cosine.

TBY makes physical and statistical sense as an explanatory variable for FCB because of its positive correlation with suspended material to which fecal indicator bacteria can physically bind. Increases in turbidity during precipitation runoff events also makes seasonality (sine and cosine variables) a logical explanatory variable for FCB.

The logarithmically (log) transformed model may be retransformed to the original units so that FCB 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 1.78. The retransformed model, accounting for BCF is as follows:

$$FCB = 1.78 \times \left(TBY^{0.968} \times 10^{[-0.286 \times \sin(2\pi D)]} \times 10^{[-0.339 \times \cos(2\pi D)]} \times 10^{0.442}\right)$$

Previous Models

Start Year	End Year	Model Equation	Reference
1999	2003	logFCB = 1.19logTBY - 0.448	Rasmussen and others (2005)

Model Statistics, Data, and Plots

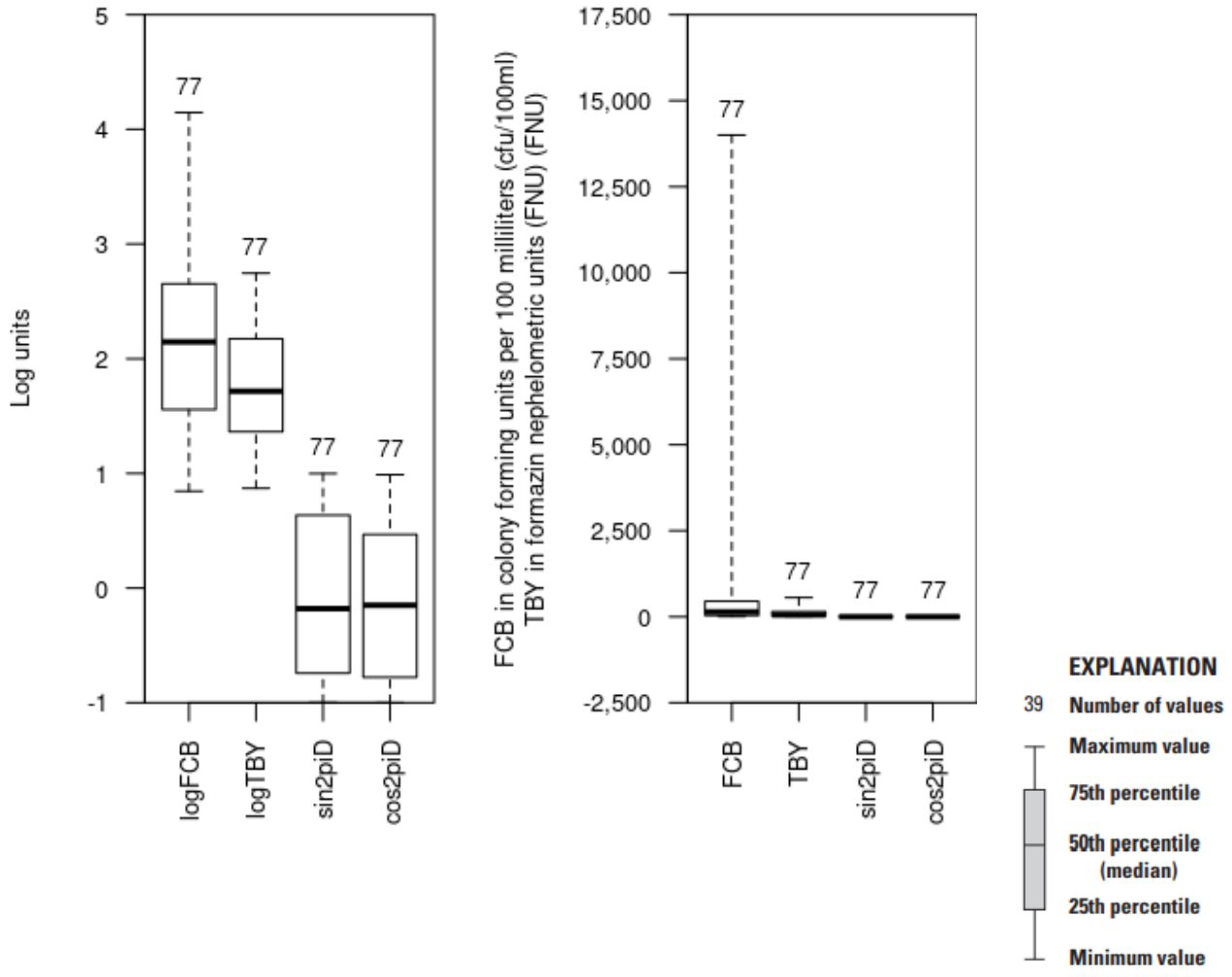
Model

logFCB = + 0.968 * logTBY - 0.286 * sin2piD - 0.339 * cos2piD + 0.442

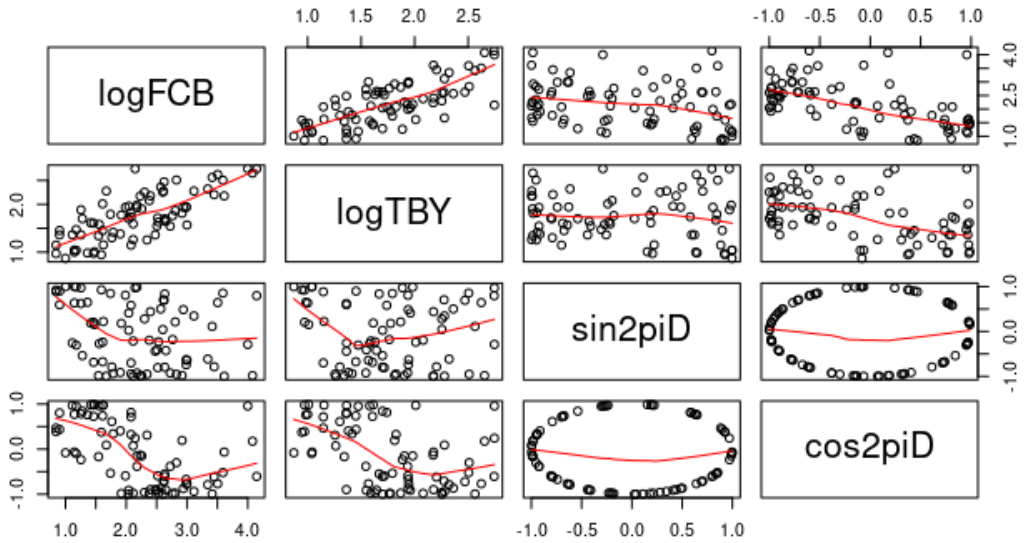
Variable Summary Statistics

	logFCB	FCB	logTBY	sin2piD	cos2piD	TBY
Minimum	0.845	7	0.87	-0.9980	-1.000	7.42
1st Quartile	1.560	36	1.36	-0.7410	-0.778	23.10
Median	2.150	140	1.72	-0.1800	-0.150	52.00
Mean	2.200	921	1.76	-0.0483	-0.116	108.00
3rd Quartile	2.650	450	2.17	0.6350	0.467	149.00
Maximum	4.150	14000	2.75	0.9980	0.988	557.00

Box Plots



Exploratory Plots



Red line shows the locally weighted scatterplot smoothing (LOWESS).

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

Basic Model Statistics

```
Number of Observations                              77
Standard error (RMSE)                            0.476
Average Model standard percentage error (MSPE)     133
Coefficient of determination (R²)                0.672
Adjusted Coefficient of Determination (Adj. R²) 0.658
Bias Correction Factor (BCF)                      1.78

Variance Inflation Factors (VIF)
 logTBY sin2piD cos2piD
   1.28    1.01    1.28
```

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.442	0.2200	2.01	4.79e-02
logTBY	0.968	0.1240	7.80	3.28e-11
sin2piD	-0.286	0.0775	-3.69	4.35e-04
cos2piD	-0.339	0.0874	-3.88	2.29e-04

Correlation Matrix

	Intercept	logTBY	sin2piD	cos2piD
Intercept	1.0000	-0.9680	-0.0704	-0.4130
logTBY	-0.9680	1.0000	0.0918	0.4640
sin2piD	-0.0704	0.0918	1.0000	0.0755
cos2piD	-0.4130	0.4640	0.0755	1.0000

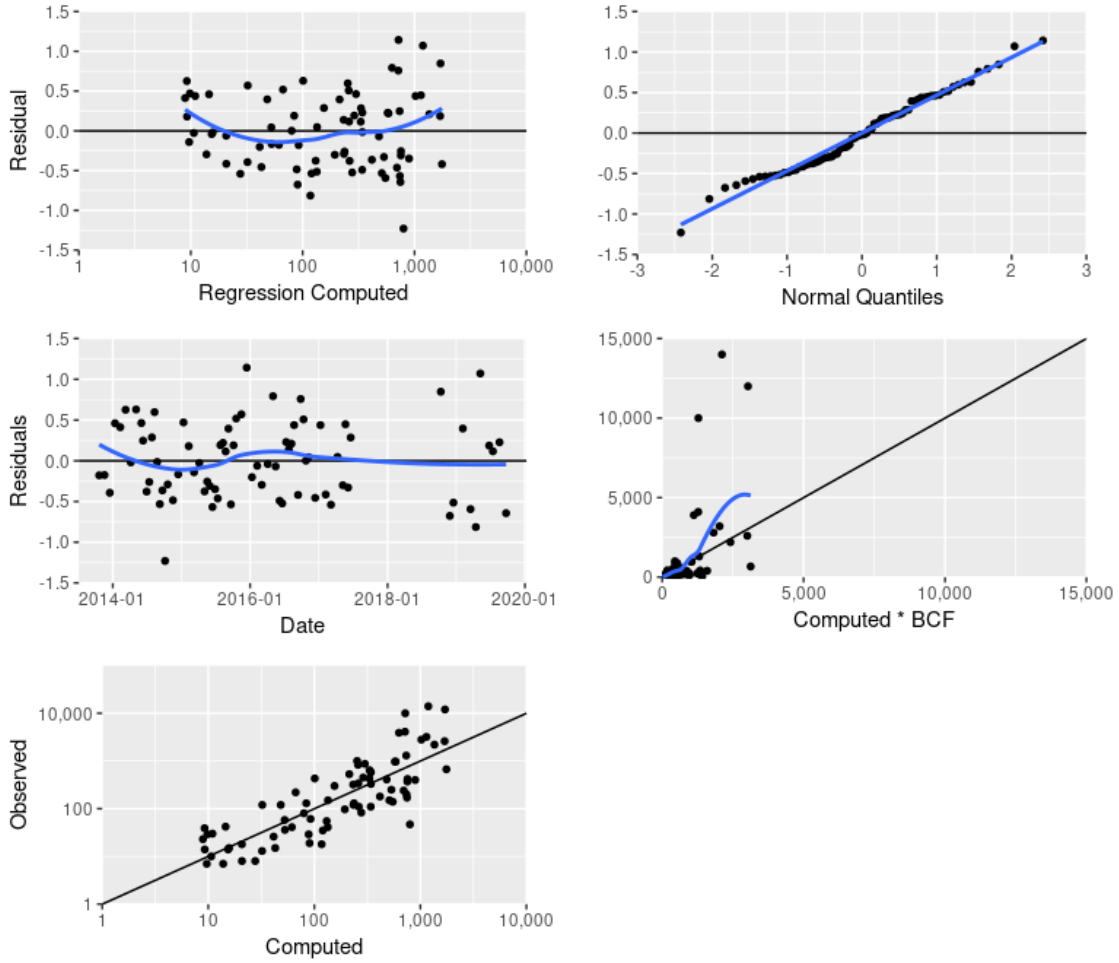
Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.156	0.319	0.456

Flagged Observations

	logFCB	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
201410060900	1.67	2.90	-1.230	-2.66	-2.78	0.0569	0.1070	-0.683
201512140910	4.00	2.86	1.140	2.63	2.74	0.1640	0.3390	1.220
201605020920	3.59	2.80	0.792	1.73	1.75	0.0684	0.0547	0.474
201810110940	4.08	3.23	0.847	1.88	1.91	0.0977	0.0952	0.628
201903180940	2.15	2.74	-0.594	-1.34	-1.35	0.1370	0.0714	-0.537
201905091520	4.15	3.07	1.070	2.36	2.44	0.0894	0.1370	0.764

Statistical Plots

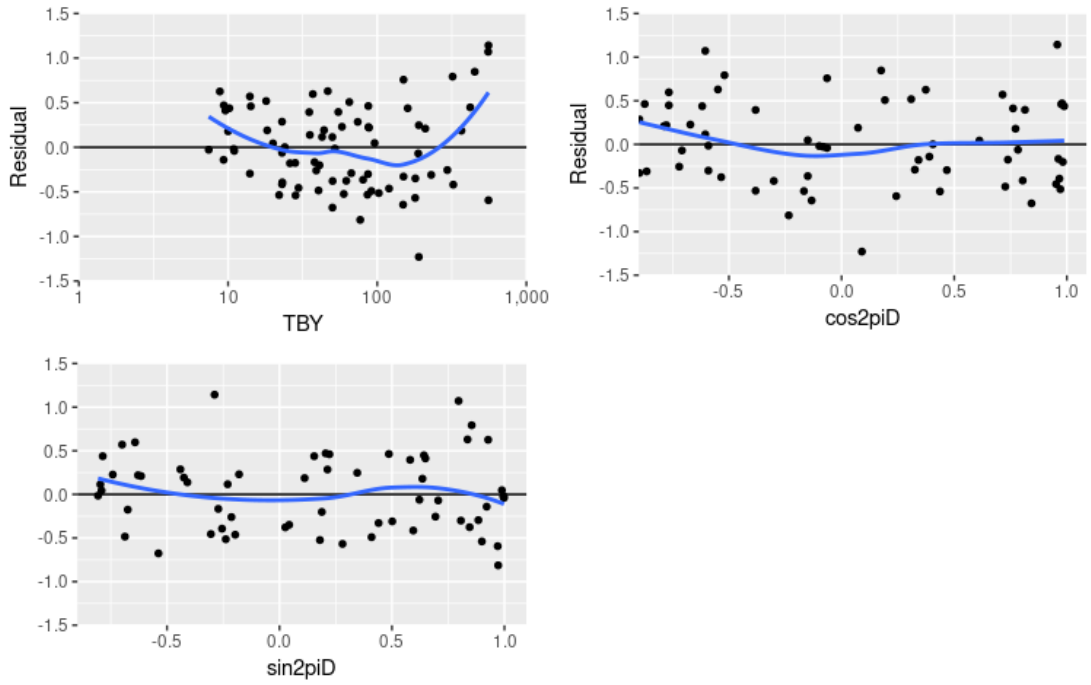


First row (left): Residual FCB related to regression computed FCB with local polynomial regression fitting, or locally estimated scatterplot smoothing (LOESS), indicated by the blue line.

First row (right): Residual FCB related to the corresponding normal quantile of the residual with simple linear regression, indicated by the blue line.

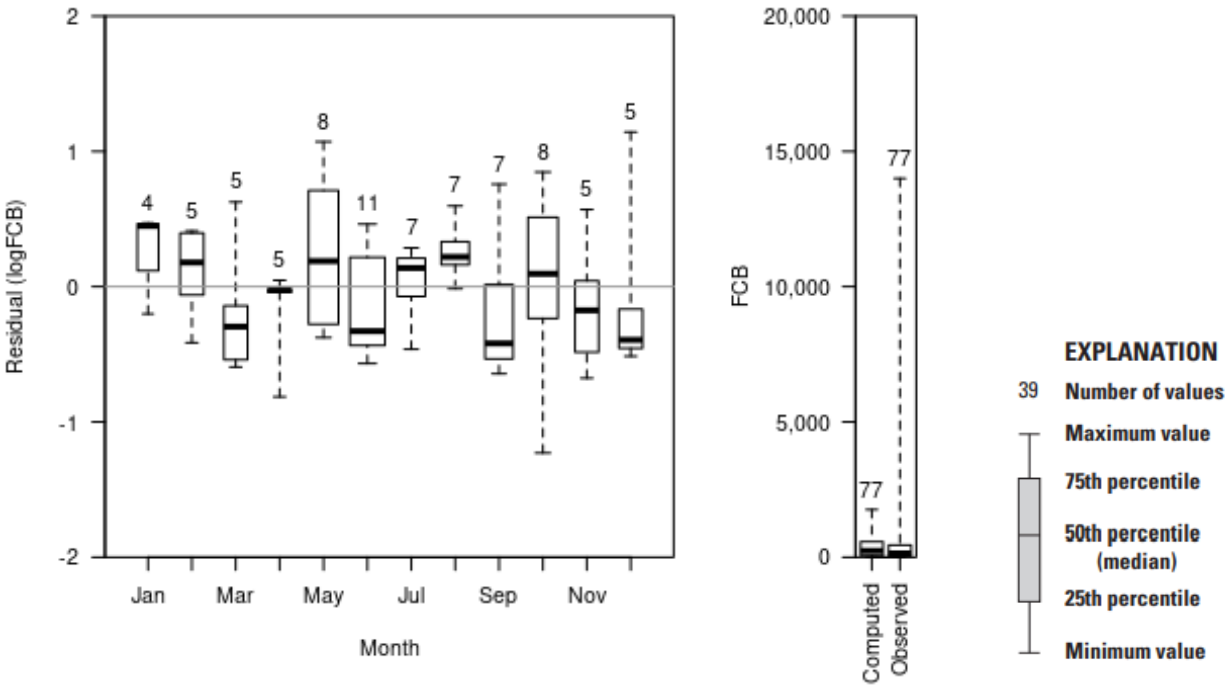
Second row: Residual FCB related to date (left) and regression computed FCB multiplied by the BCF (right) with LOESS, indicated by the blue line.

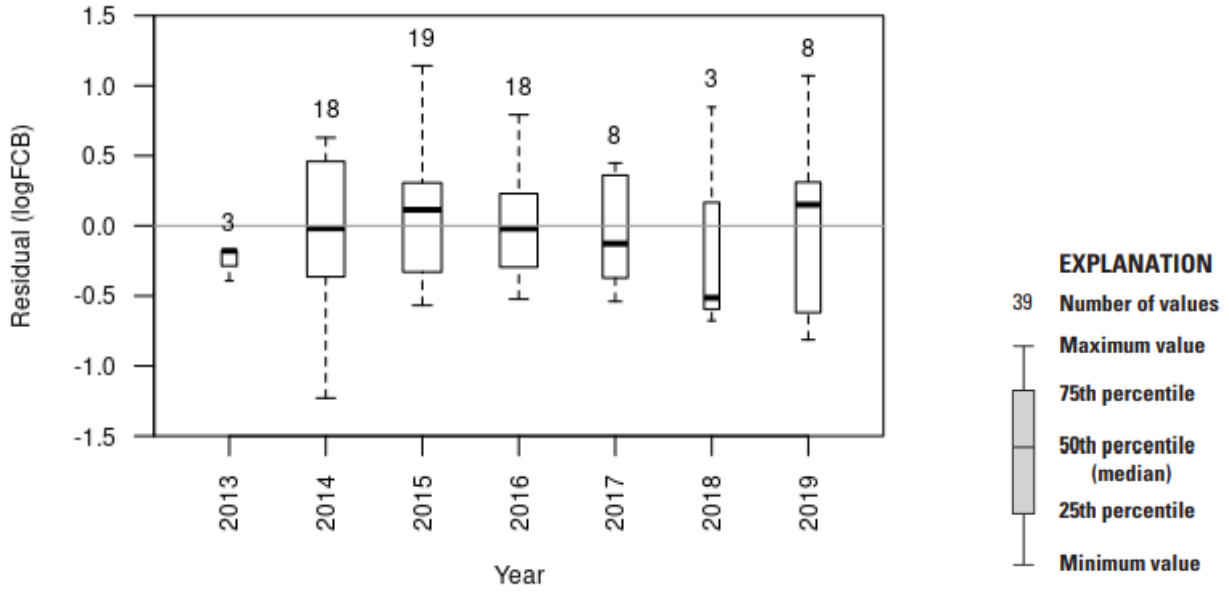
Third row: Observed FCB related to regression computed FCB.



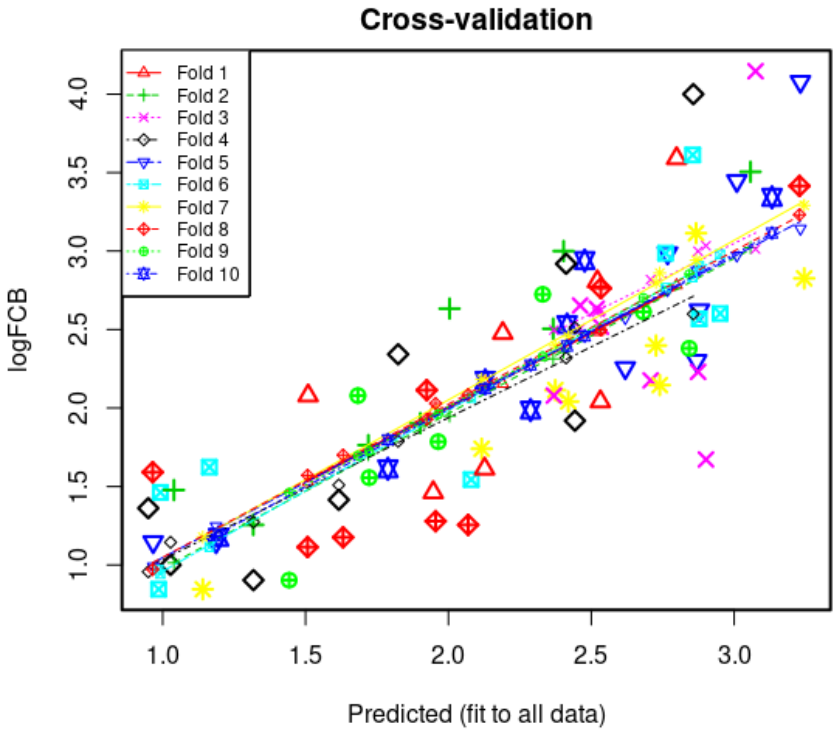
First row: Residual FCB related to TBY (left) and cos2piD (right) with LOESS, indicated by the blue line.

Second row: Residual FCB related to sin2piD (right) with LOESS, indicated by the blue line.





Cross-Validation



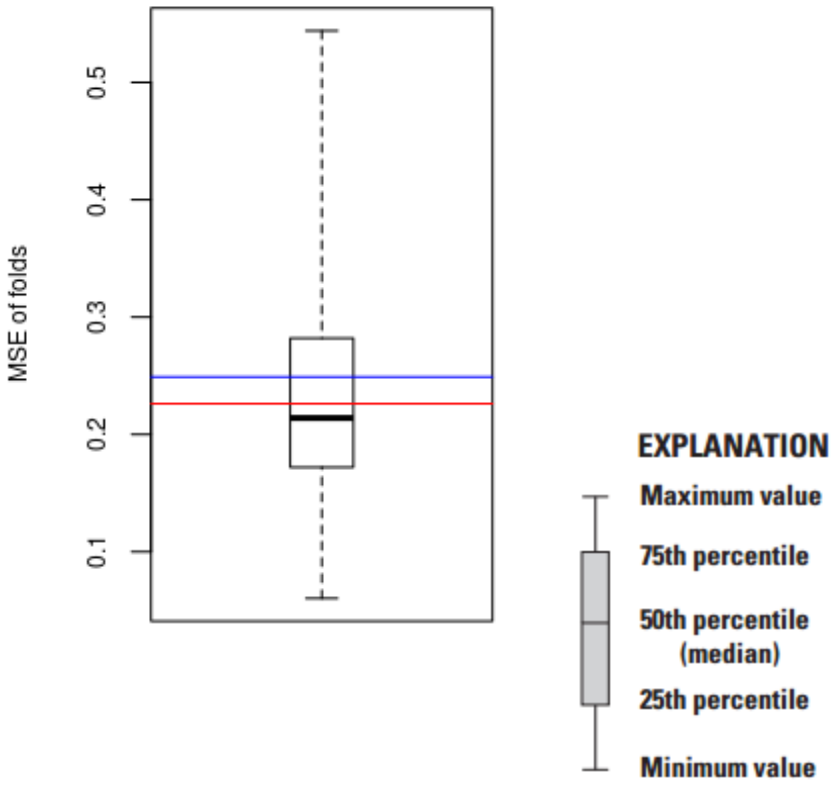

```
Fold - equal partition of the data (10 percent of the data).

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.


              Minimum MSE of folds:  0.0602
                 Mean MSE of folds:  0.2490
               Median MSE of folds:  0.2140
              Maximum MSE of folds:  0.5440
 (Mean MSE of folds) / (Model MSE):  1.1000
```



Red line - Model MSE

Blue line - Mean MSE of folds

Model-Calibration Dataset

	Date	logFCB	logTBY	sin2piD	cos2piD	FCB	TBY	Computed logFCB	Computed FCB	Residual	Normal Quantiles	Censored Values
1	2013-10-21	1.79	1.41	-0.94	0.342	61	26	1.96	164	-0.179	-0.263	--
2	2013-11-18	1.61	1.45	-0.674	0.738	41	28.2	1.79	109	-0.175	-0.229	--
3	2013-12-16	1.11	1.36	-0.255	0.967	13	23.1	1.51	57.1	-0.393	-0.746	--
4	2014-01-13	1.62	1.15	0.222	0.975	42	14.2	1.16	25.9	0.46	0.93	--
5	2014-02-10	1.36	0.982	0.649	0.761	23	9.6	0.95	15.8	0.412	0.746	--
6	2014-03-10	1.59	0.945	0.928	0.374	39	8.81	0.965	16.4	0.626	1.37	--
7	2014-04-07	1.18	1.04	0.995	-0.0988	15	10.9	1.2	27.9	-0.0197	0	--
8	2014-05-05	2.63	1.67	0.836	-0.549	430	46.6	2	179	0.629	1.46	--
9	2014-06-02	2.94	1.94	0.486	-0.874	870	87.1	2.48	533	0.463	0.981	--
10	2014-06-11	3.11	2.28	0.346	-0.938	1300	190	2.87	1310	0.247	0.545	--
11	2014-06-30	2.04	1.7	0.0258	-1	110	50.2	2.42	467	-0.378	-0.703	--
12	2014-07-14	2.11	1.59	-0.214	-0.977	130	39	2.37	421	-0.26	-0.365	--
13	2014-07-28	2.48	1.36	-0.441	-0.898	300	23	2.19	275	0.287	0.622	--
14	2014-08-11	3	1.57	-0.642	-0.767	1000	37	2.4	450	0.597	1.29	--
15	2014-08-25	2.52	1.72	-0.806	-0.591	330	52	2.53	607	-0.0152	0.0325	--
16	2014-09-08	2.18	1.93	-0.924	-0.382	150	86	2.71	907	-0.532	-1.22	--
17	2014-09-22	2.26	1.9	-0.989	-0.15	180	80.3	2.62	739	-0.364	-0.622	--
18	2014-10-06	1.67	2.28	-0.996	0.0903	47	190	2.9	1420	-1.23	-2.42	--
19	2014-10-20	2.08	1.83	-0.946	0.325	120	67	2.37	416	-0.29	-0.4	--
20	2014-11-17	1.46	1.61	-0.687	0.727	29	40.4	1.95	157	-0.484	-0.981	--
21	2014-12-15	1.56	1.58	-0.272	0.962	36	38	1.72	93.8	-0.166	-0.196	--
22	2015-01-12	1.46	0.971	0.205	0.979	29	9.36	0.992	17.4	0.471	1.04	--
23	2015-02-09	1.15	1	0.635	0.772	14	10	0.967	16.5	0.179	0.263	--
24	2015-03-09	0.845	0.97	0.921	0.39	7	9.33	0.986	17.2	-0.141	-0.163	--

25	2015-04-06	1	0.87	0.997	-0.0817	10	7.42	1.03	18.9	-0.0276	-0.0325	--
26	2015-05-04	1.74	1.79	0.845	-0.534	55	62	2.12	232	-0.376	-0.662	--
27	2015-05-18	2.62	2.47	0.693	-0.721	420	295	2.88	1340	-0.255	-0.331	--
28	2015-06-01	2.57	2.36	0.501	-0.865	370	230	2.88	1340	-0.31	-0.508	--
29	2015-06-15	2.3	2.25	0.28	-0.96	200	180	2.87	1310	-0.568	-1.46	--
30	2015-06-29	2.6	2.26	0.043	-0.999	400	180	2.95	1590	-0.349	-0.583	--
31	2015-07-13	2.38	2.08	-0.197	-0.98	240	120	2.84	1240	-0.463	-0.93	--
32	2015-07-27	2.65	1.64	-0.425	-0.905	450	44	2.46	514	0.192	0.365	--
33	2015-08-10	2.99	1.94	-0.629	-0.778	970	88	2.77	1040	0.22	0.435	--
34	2015-08-24	2.63	1.7	-0.796	-0.605	430	50	2.52	587	0.115	0.163	--
35	2015-09-08	2.72	1.54	-0.924	-0.382	530	35	2.33	380	0.394	0.662	--
36	2015-09-21	1.54	1.34	-0.986	-0.167	35	22	2.08	214	-0.536	-1.29	--
37	2015-10-05	2.11	1.26	-0.997	0.0731	130	18.3	1.92	149	0.19	0.331	--
38	2015-10-19	2.34	1.26	-0.951	0.309	220	18	1.82	119	0.518	1.15	--
39	2015-11-16	2.08	1.15	-0.699	0.715	120	14	1.51	57.4	0.57	1.22	--
40	2015-12-14	4	2.75	-0.288	0.957	10000	557	2.86	1280	1.14	2.42	--
41	2016-01-11	1.41	1.61	0.188	0.982	26	41	1.62	73.5	-0.201	-0.297	--
42	2016-02-08	1.26	1.36	0.622	0.783	18	23	1.32	36.9	-0.0617	-0.0976	--
43	2016-03-03	0.845	1.15	0.884	0.467	7	14	1.14	24.6	-0.295	-0.435	--
44	2016-04-04	1.15	1.04	0.998	-0.0645	14	11	1.19	27.3	-0.0407	-0.065	--
45	2016-05-02	3.59	2.51	0.854	-0.52	3900	320	2.8	1120	0.792	1.68	--
46	2016-05-16	2.61	2.27	0.706	-0.709	410	188	2.68	855	-0.0696	-0.13	--
47	2016-06-06	2.04	1.96	0.409	-0.912	110	91.3	2.53	605	-0.49	-1.04	--
48	2016-06-20	1.92	1.78	0.18	-0.984	83	59.7	2.44	493	-0.524	-1.15	--
49	2016-07-11	2.76	1.76	-0.18	-0.984	580	58	2.53	607	0.23	0.508	--
50	2016-07-25	2.51	1.55	-0.409	-0.912	320	35.3	2.37	414	0.138	0.229	--
51	2016-08-08	3.34	2.32	-0.615	-0.788	2200	210	3.13	2410	0.21	0.4	--
52	2016-08-22	3.45	2.2	-0.786	-0.619	2800	160	3.01	1820	0.438	0.834	--
53	2016-09-12	2.83	2.51	-0.954	-0.301	670	323	3.25	3130	-0.419	-0.834	--
54	2016-09-26	3.61	2.18	-0.998	-0.0645	4100	150	2.86	1270	0.758	1.56	--
55	2016-10-11	2.92	1.81	-0.981	0.192	830	65	2.41	459	0.507	1.09	--
56	2016-10-24	1.9	1.38	-0.914	0.405	80	24	1.9	142	0.00139	0.065	--
57	2016-11-07	1.76	1.3	-0.791	0.612	58	20	1.72	93.3	0.0435	0.0976	--
58	2016-12-12	1.18	1.47	-0.305	0.952	15	29.7	1.63	76.1	-0.455	-0.881	--
59	2017-01-09	1.48	1.01	0.154	0.988	30	10.2	1.04	19.5	0.438	0.789	--
60	2017-02-06	0.903	1.36	0.595	0.804	8	23	1.32	36.9	-0.415	-0.789	--
61	2017-03-06	0.903	1.45	0.9	0.437	8	28.3	1.44	49.3	-0.54	-1.37	--
62	2017-04-10	2.18	1.98	0.989	-0.15	150	96	2.13	239	0.0472	0.13	--
63	2017-05-08	1.99	1.94	0.806	-0.591	97	86.7	2.29	345	-0.301	-0.471	--
64	2017-05-22	3.51	2.62	0.642	-0.767	3200	420	3.06	2030	0.448	0.881	--
65	2017-06-05	2.4	2.18	0.441	-0.898	250	150	2.73	947	-0.329	-0.545	--
66	2017-06-19	2.81	1.87	0.214	-0.977	640	74	2.52	590	0.285	0.583	--
67	2018-10-11	4.08	2.65	-0.984	0.176	12000	450	3.23	3030	0.847	1.83	--
68	2018-11-28	1.28	1.7	-0.538	0.843	19	50.1	1.96	160	-0.676	-1.83	--
69	2018-12-17	1.61	2.01	-0.239	0.971	41	102	2.13	238	-0.514	-1.09	--
70	2019-02-05	2.08	1.74	0.581	0.814	120	54.8	1.68	85.7	0.396	0.703	--
71	2019-03-18	2.15	2.75	0.97	0.243	140	557	2.74	976	-0.594	-1.56	--
72	2019-04-15	1.26	1.89	0.972	-0.234	18	76.9	2.07	208	-0.814	-2.04	--
73	2019-05-09	4.15	2.74	0.796	-0.605	14000	554	3.07	2110	1.07	2.04	--
74	2019-06-25	3.41	2.56	0.112	-0.994	2600	367	3.23	3010	0.186	0.297	--
75	2019-07-15	2.53	1.63	-0.23	-0.973	340	42.7	2.42	463	0.116	0.196	--
76	2019-08-19	2.99	1.94	-0.741	-0.671	970	87.3	2.76	1020	0.227	0.471	--
77	2019-09-23	2.23	2.17	-0.991	-0.133	170	149	2.87	1330	-0.643	-1.68	--

Definitions

D: Date, in decimal years.

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

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

E.vars: Explanatory variables.

FCB: Fecal coliforms, in colonies per 100 milliliters (31625).

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

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

LOWESS: Locally weighted scatterplot smoothing (Cleveland, 1979; Helsel and others, 2020).

MSE: Model standard 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).

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

TBY: Turbidity, in formazin nephelometric units (63680).

References Cited

Cleveland, W.S., 1979, Robust locally weighted regression and smoothing scatterplots: Journal of the American Statistical Association, v. 74, no. 368, p. 829-836.

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.]

Foster, G.M., and Graham, J.L., 2016, Logistic and linear regression model documentation for statistical relations between continuous real-time and discrete water-quality constituents in the Kansas River, Kansas, July 2012 through June 2015: U.S. Geological Survey Open-File Report 2016–1040, 27 p., accessed July 2020 at https://doi.org/10.3133/ofr20161040.

Graham, J.L., Foster, G.M., Williams, T.J., Mahoney, M.D., May, M.R., and Loftin, K.A., 2018, Water-quality conditions with an emphasis on cyanobacteria and associated toxins and taste-and-odor compounds in the Kansas River, Kansas, July 2012 through September 2016: U.S. Geological Survey Scientific Investigations Report 2018–5089, 55 p. [Also available at https://doi.org/10.3133/sir20185089.]

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, ver. 1.1.]

R Core Team, 2020, R—A language and environment for statistical computing, version 4.0.3: Vienna, Austria, R Foundation for Statistical Computing, accessed December 2020 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, 53 p. [Also available at https://doi.org/10.3133/tm3C4.]

Rasmussen, T.J., Ziegler, A.C., and Rasmussen, P.P., 2005, Estimation of constituent concentrations, densities, loads, and yields in lower Kansas River, northeast Kansas, using regression models and continuous water-quality monitoring, January 2000 through December 2003: U.S. Geological Survey Scientific Investigations Report 2005–5165, 117 p. [Also available at https://doi.org/10.3133/sir20055165.]

U.S. Geological Survey, 2020, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 2020 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], accessed July 2020 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, 51 p. plus 8 attachments. [Also available at https://doi.org/10.3133/tm1D3.] [Supersedes USGS Water-Resources Investigations Report 2000–4252.]