

Appendix 3. Model Archive Summary for Best-Fit Regression Developed to Estimate Fecal Coliform Concentration at Station 01481000; Brandywine Creek at Chadds Ford, Pennsylvania

This model archive summary describes the regression model developed to estimate continuous instantaneous (30-minute) fecal coliform concentrations for the period of March 1 through October 31 each year starting in 2007.

Site and Model Information

U.S. Geological Survey (USGS) station number: 01481000

Station name: Brandywine Creek at Chadds Ford, Pennsylvania

Location: Latitude 39°52'11", longitude 75°35'37" referenced to North American Datum of 1927, Delaware County, PA, Hydrologic Unit 02040205, on left bank 27 (feet) ft upstream from Penn Central Railroad bridge at Chadds Ford, 150 ft upstream from Harvey Run, and 1,200 ft downstream from highway bridge on U.S. Highway 1.

Equipment: A Yellow Spring Instrument (YSI) 6920V2 monitor equipped with sensors for temperature, pH, specific conductance, and dissolved oxygen, and an Optical Monitoring System 600 series (YSI 6136) turbidity sensor. On April 22, 2009 the YSI 6136 turbidity sensor replaced earlier Analite (2005 to October 2008) and In-situ Troll RDO instrumentation (October 2008 to April 2009). The monitor is housed in a 6-inch perforated plastic pipe placed in the stream about 5 ft from the left bank. Readings from the sensors were recorded every 30 minutes from 2007-12 and have been recorded every 15-minutes since 2013 and transmitted hourly by way of satellite.

Date regression model was created: October 2016.

Period of data for model calibration: March 1 – October 31, years 2007-15.

Model application date: October 2016 onward

Computed by: Lisa Senior, October 2016.

Reviewed by: Kirk White (Supervisory Hydrologist), Matt Gyves (Hydrologic Technician) April 6, 2017

Approved by: Joseph Duris (Pennsylvania Water Science Center Water Quality Specialist) May 26, 2017

Model Calibration Dataset

All data were collected using USGS protocols and are stored in the National Water Information System (NWIS) database. Linear regression models were developed using the SPlus and open-source software package "R." Explanatory variables selected as inputs to linear regression were: specific conductance, pH, water temperature, dissolved oxygen, turbidity, and streamflow. Seasonality components (sine and cosine terms calculated using Julian day as a fraction of the year) were also evaluated as explanatory variables in the models to determine if seasonal changes affected the model. All combinations of physicochemical properties and seasonality components were evaluated to determine which combinations produced the best models.

The final regression model is based on 101 concurrent measurements of fecal coliform and turbidity concentrations from March 1 – October 31 of each year for years 2007-15, plus computed seasonality variables. Fecal coliform concentrations were determined from analysis of discrete samples, and turbidity concentrations were determined from continuous record of 30-minute values, interpolated when necessary to correspond with collection time of the discrete sample for bacteria analysis. Samples were collected through a range of hydrologic conditions during the March-October sampling period each year. Studentized residuals for final

model were inspected and considered for potential removal as outliers if residual values were greater than 3 or less than -3; however, no samples met these criteria, and no samples were removed from the dataset.

Fecal Coliform Data

Discrete grab samples for bacteria analysis were collected from midpoint of the stream near the gaging station and chilled until processed in the laboratory at the USGS office in Exton, Pa. within 6 hours of sample collection. The number of fecal coliform colonies in a sample was determined by membrane filtration using a 0.7 micron filter and subsequent plating and incubation using standard methods. At the laboratory, a range of dilutions was plated for each stream sample to obtain optimal counts (20-60 colonies) on at least one plate.

Model Development

Regression analysis was done using S+ and R by examining turbidity (*Turb*), streamflow (*Q*), and other continuously measured data in addition to computed seasonality terms ($\sin 2\pi JD$ and $\cos 2\pi JD$) as explanatory variables for estimating fecal coliform (*FC*) concentration. A variety of linear regression models that predict *FC* and $\log_{10}(FC)$ were evaluated. The distribution of residuals was examined for normality, and plots of residuals (the difference between the measured and computed values) as compared to computed *FC* were examined for homoscedasticity (meaning that their departures from zero did not change substantially over the range of computed values). This comparison led to the conclusion that the most appropriate and reliable model would be one that estimated $\log_{10}(FC)$.

$\log_{10}(Turb)$ and seasonality explanatory variables were selected as the best predictors of $\log_{10}(FC)$ based on residual plots, relatively high adjusted coefficient of determination (adjusted R^2), and relatively low model residual standard error (or root mean square error, *RMSE*) and low standard percentage error (*MSPE*).

Model Summary

Final regression model for fecal coliform (*FC*) concentration at site number 01480617, *FC* concentration-based model:

$$\log_{10}(FC) = 1.1 \times \log_{10}(Turb) - 0.288 \times \sin(2\pi JD/365) - 0.774 \times \cos(2\pi JD/365) + 1.12$$

where

FC = fecal coliform in colony-forming units per 100 milliliter (cfu/mL) (parameter 31625);

Turb = turbidity in formazin nephelometric units (FNU) (parameter 63680);

Sin & *Cos* = sine and cosine functions used to compute seasonality variables; and,

JD = Julian day (day of year).

Turb and seasonality terms makes physical and statistical sense as explanatory variables for *FC* because previous studies showed bacteria concentrations were related to turbidity concentrations and were highest in summer months. The negative coefficients for seasonal variables $\sin 2\pi JD$ and $\cos 2\pi JD$ [computed as $\sin(2\pi JD/365)$ and $\cos(2\pi JD/365)$] have the effect to increase predicted *FC* the most during the peak of summer. The transformed model may be retransformed to the original units so that *FC* concentrations can be calculated directly. A potential bias that is introduced because of retransformation can be corrected using Duan's bias correction factor (BCF). For this model the BCF is 1.40. The retransformed model, using the BCF, is:

$$FC = 1.40 \times 10^{\log_{10}(FC)} \text{ or } FC = (Turb^{1.0957} \times 18.3) / [10^{0.2877 \sin(2\pi JD/365)} \times 10^{0.7742 \cos(2\pi JD/365)}]$$

Model Statistics, Data, and Plots

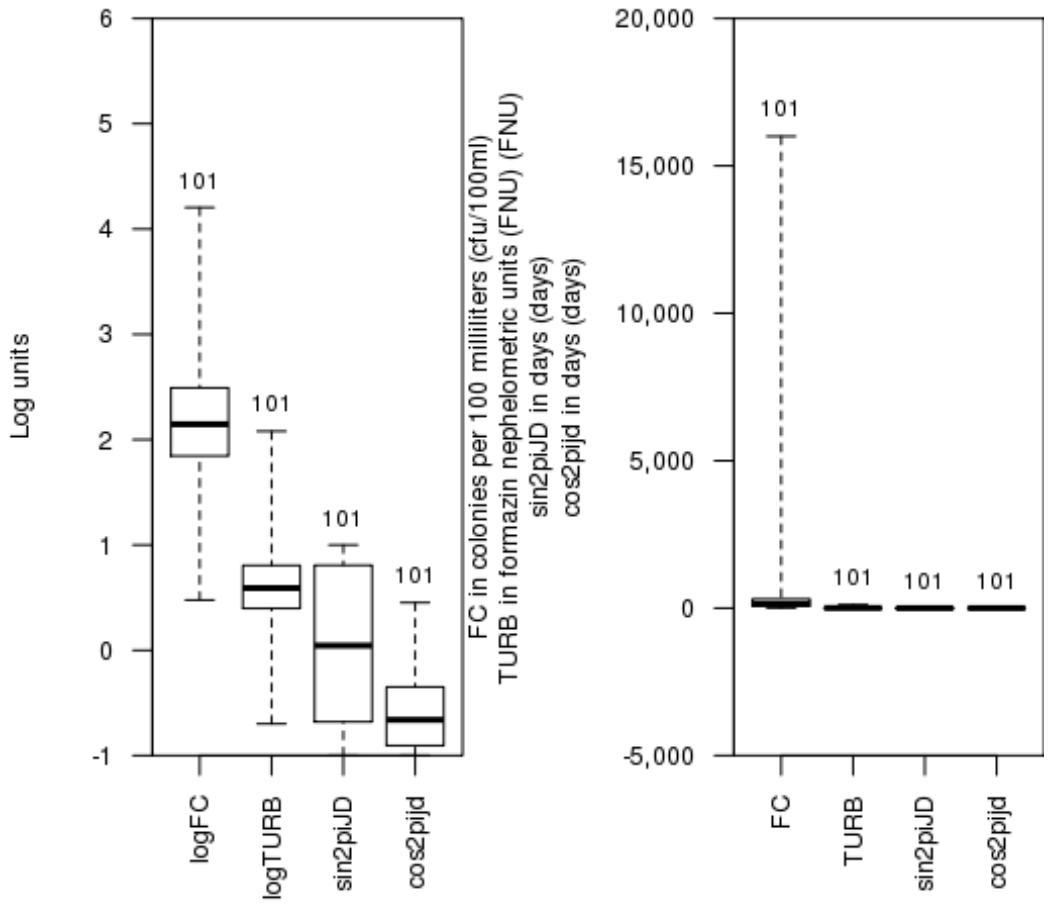
Model

$$\log FC = + 1.1 * \log TURB - 0.288 * \sin 2\pi jD - 0.774 * \cos 2\pi jD + 1.12$$

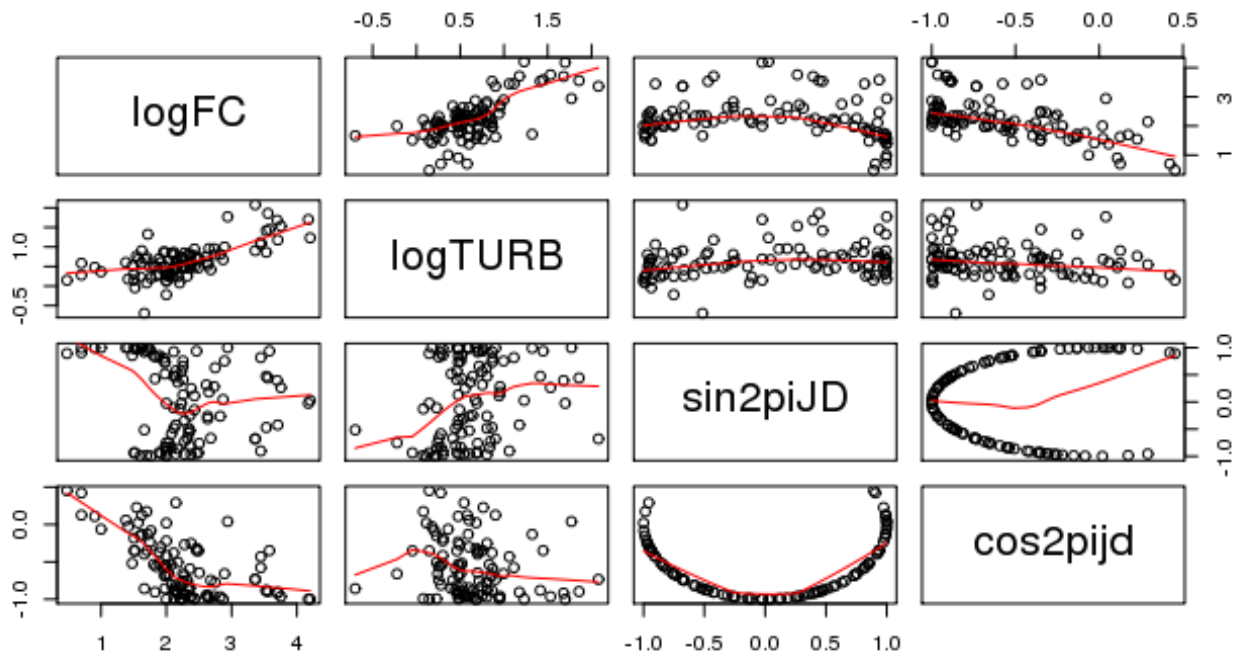
Variable Summary Statistics

	logFC	FC	logTURB	sin2piJD	cos2pijd	TURB
Minimum	0.477	3	-0.699	-1.0000	-1.000	0.2
1st Quartile	1.850	70	0.398	-0.6790	-0.905	2.5
Median	2.150	140	0.591	0.0446	-0.660	3.9
Mean	2.250	861	0.639	0.0376	-0.568	8.5
3rd Quartile	2.490	310	0.806	0.8070	-0.349	6.4
Maximum	4.200	16000	2.080	1.0000	0.453	120.0

Box Plots



Exploratory Plots



Basic Model Statistics

Number of Observations 101
 Standard error (RMSE) 0.349
 Average Model standard percentage error (MSPE) 89.3
 Coefficient of determination (R^2) 0.766
 Adjusted Coefficient of Determination (Adj. R^2) 0.759
 Bias Correction Factor (BCF) 1.4

Variance Inflation Factors (VIF)

logTURB	sin2piJD	cos2pijd
1.13	1.12	1.09

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(> t)
(Intercept)	1.120	0.0751	14.9	8.71e-27
logTURB	1.100	0.0845	13.0	6.71e-23
sin2piJD	-0.288	0.0504	-5.7	1.27e-07
cos2pijd	-0.774	0.0944	-8.2	1.00e-12

Correlation Matrix

	Intercept	logTURB	sin2piJD	cos2pijd
Intercept	1.0000	-0.546	0.0246	0.552
logTURB	-0.5460	1.000	-0.2830	0.233
sin2piJD	0.0246	-0.283	1.0000	-0.216
cos2pijd	0.5520	0.233	-0.2160	1.000

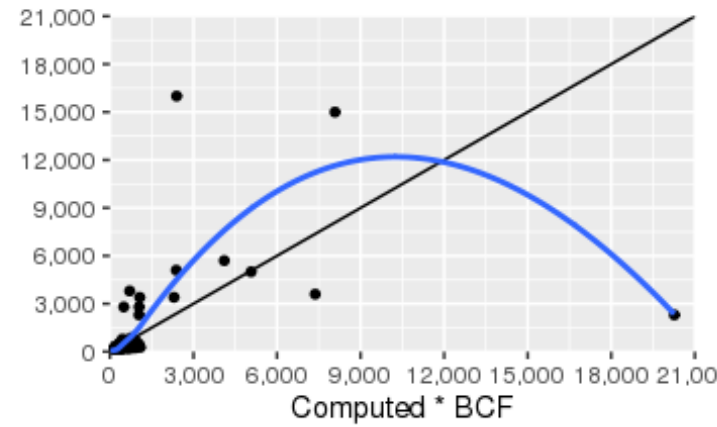
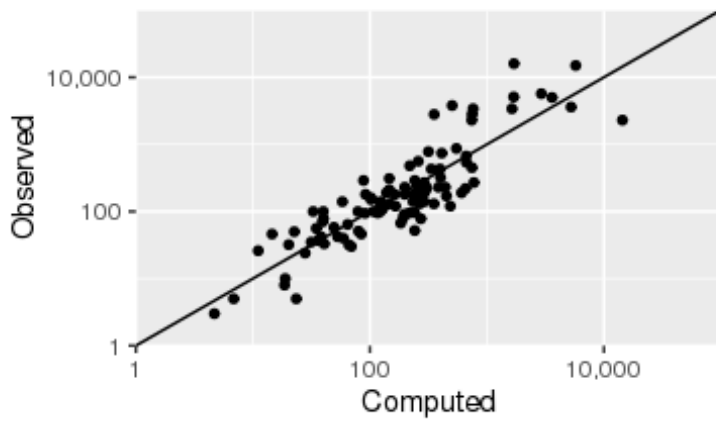
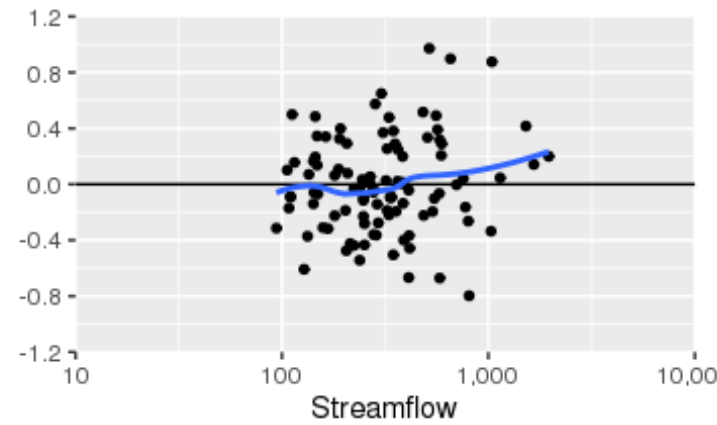
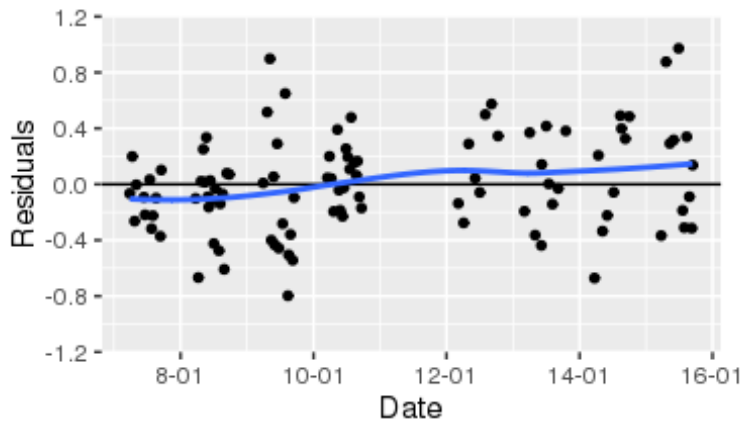
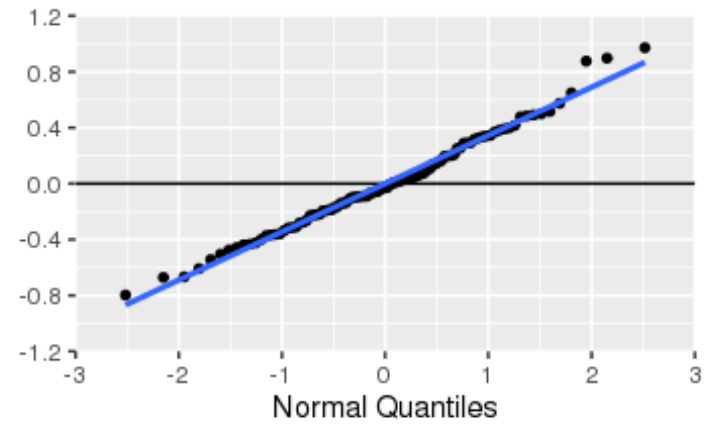
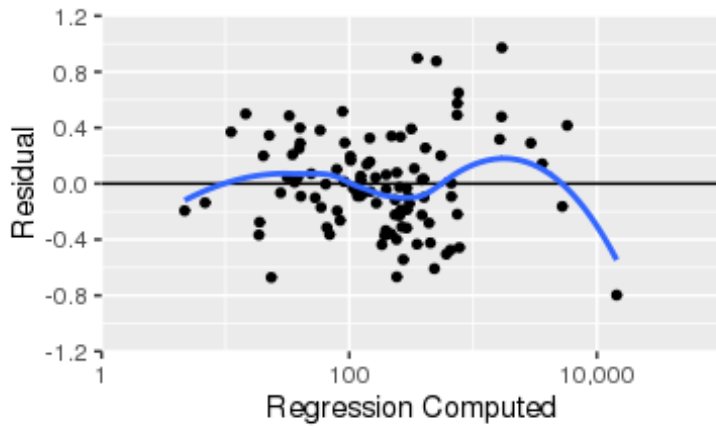
Outlier Test Criteria

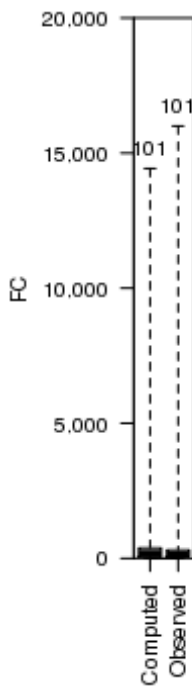
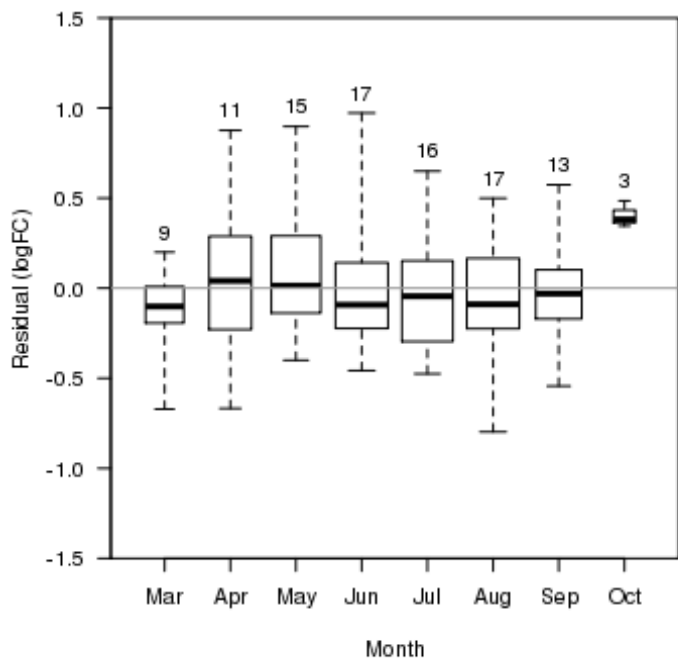
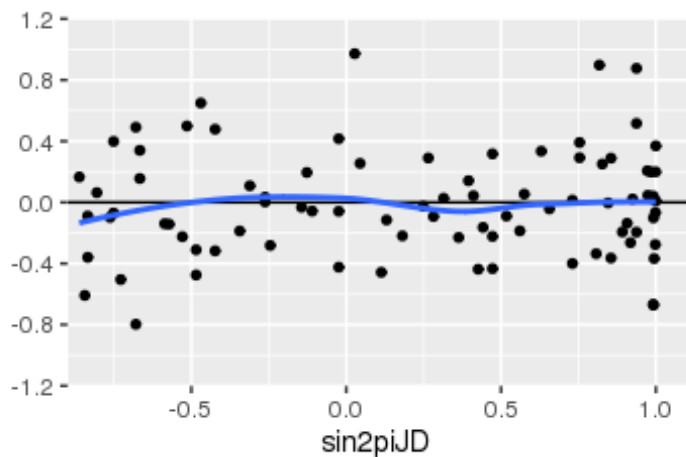
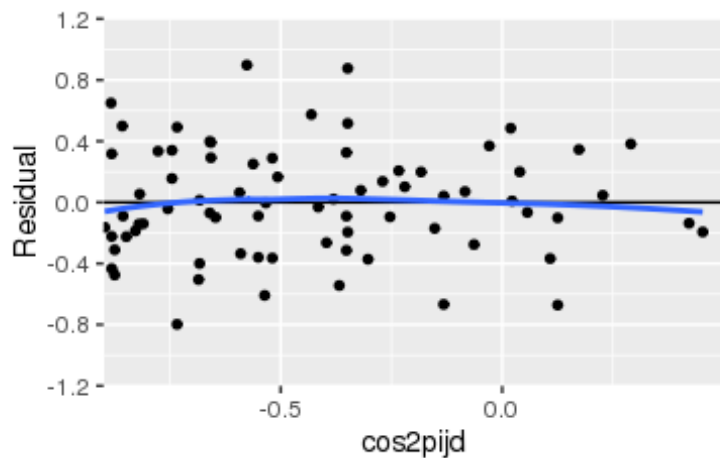
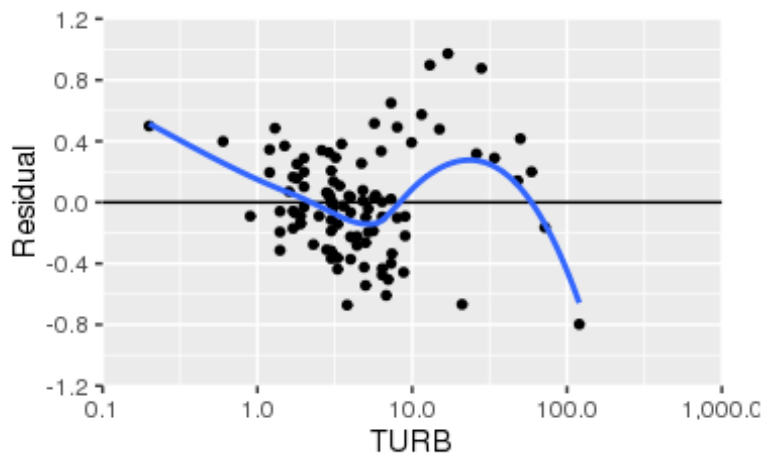
Leverage	Cook's D	DFFITS
0.0891	0.2647	0.3447

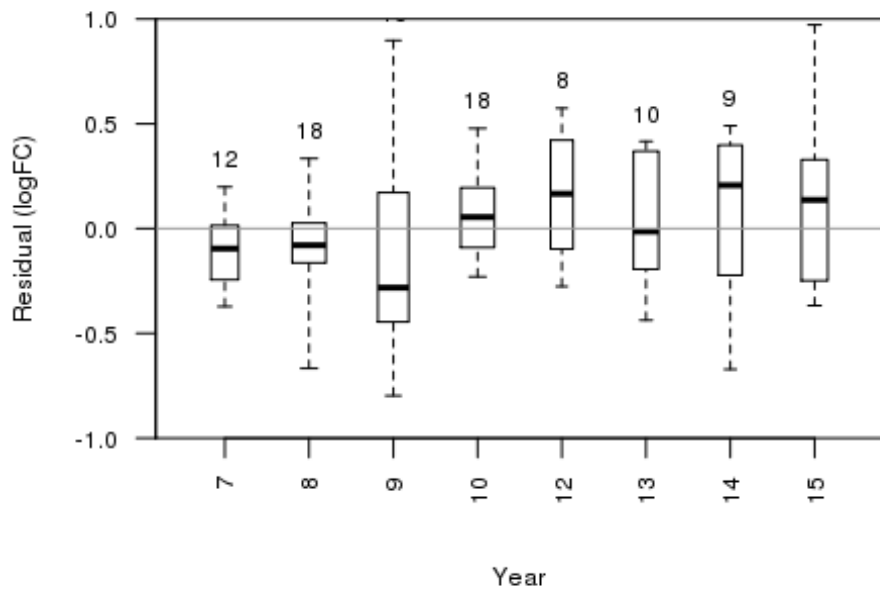
Flagged Observations

	logFC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
4/8/08 9:00	1.720	2.380	-0.667	-1.970	-2.000	0.0592	0.06110	-0.502
5/6/09 11:09	3.450	2.550	0.898	2.610	2.690	0.0285	0.04990	0.461
8/13/09 7:09	3.360	4.160	-0.798	-2.490	-2.560	0.1550	0.28400	-1.100
3/29/10 9:09	2.940	2.740	0.200	0.610	0.608	0.1210	0.01280	0.225
8/1/12 13:00	1.660	1.160	0.499	1.530	1.540	0.1220	0.08070	0.572
9/4/12 11:27	3.450	2.870	0.574	1.690	1.710	0.0519	0.03910	0.399
3/4/13 11:27	0.477	0.671	-0.194	-0.582	-0.580	0.0938	0.00878	-0.187
7/2/13 12:00	4.180	3.760	0.416	1.240	1.240	0.0764	0.03180	0.358
10/17/13 13:09	2.150	1.760	0.382	1.150	1.150	0.0949	0.03460	0.373
3/24/14 13:09	0.699	1.370	-0.671	-1.980	-2.010	0.0529	0.05460	-0.474
10/1/14 11:09	2.000	1.520	0.485	1.430	1.440	0.0638	0.03510	0.377
4/21/15 11:27	3.580	2.700	0.876	2.580	2.660	0.0562	0.09940	0.650
6/29/15 10:18	4.200	3.230	0.972	2.840	2.950	0.0363	0.07580	0.572

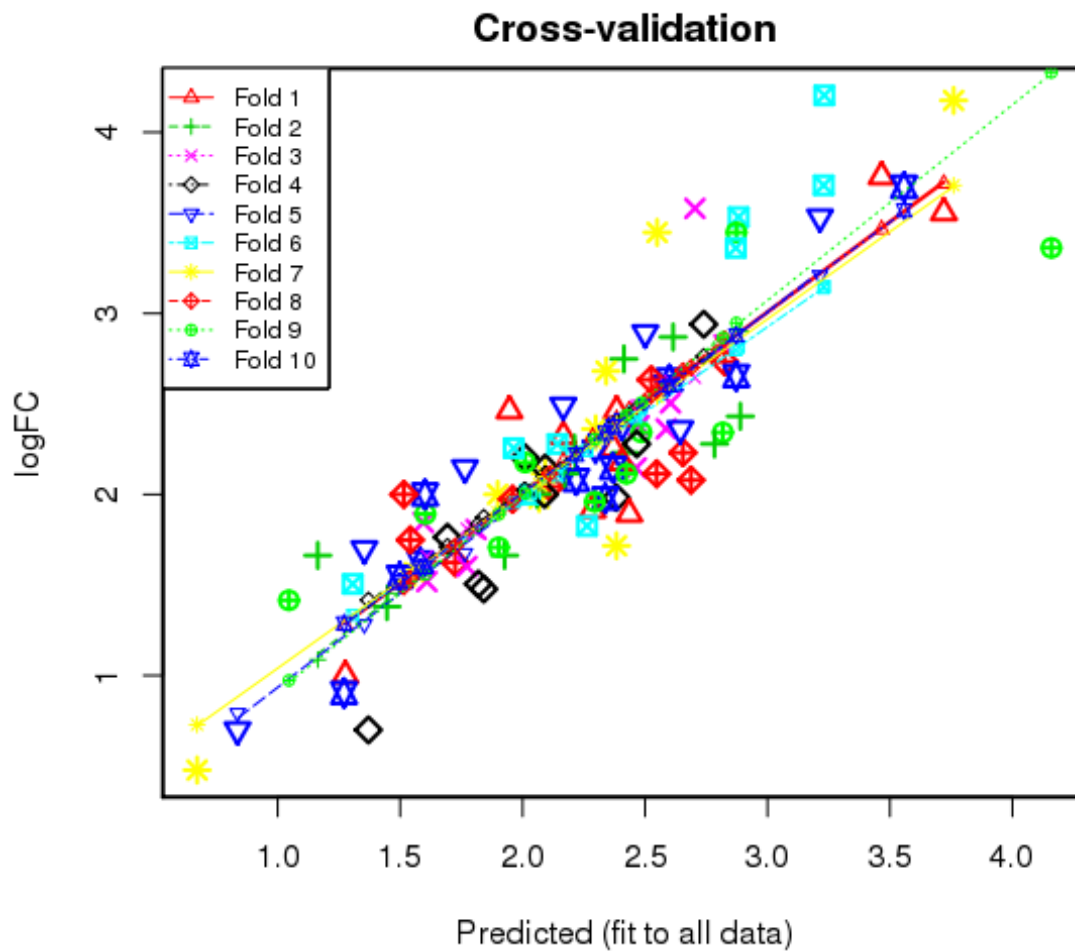
Statistical Plots



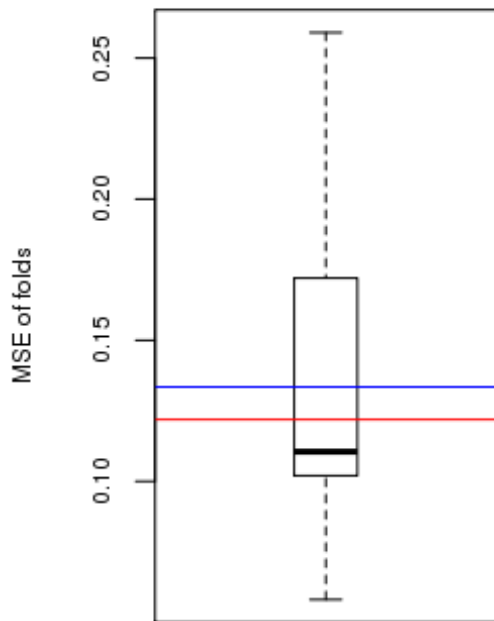




Cross Validation



Minimum MSE of folds: 0.0581
 Mean MSE of folds: 0.1330
 Median MSE of folds: 0.1100
 Maximum MSE of folds: 0.2590
 (Mean MSE of folds) / (Model MSE): 1.0900



Red line - Model MSE

Blue line - Mean MSE of folds

Model-Calibration Data Set

	Date	logFC	logTURB	sin2piJD	cos2pijd	FC	TURB	Computed logFC	Computed FC	Residual	Normal Quantiles	Censored Values
0												
1	7-03-28	1.38	0.602	0.998	0.0567	24	4	1.45	39.2	-0.0659	-0.124	--
2	7-04-11	1.51	0.301	0.983	-0.183	32	2	1.31	28.4	0.199	0.605	--
3	7-04-24	1.66	0.699	0.918	-0.397	46	5	1.93	119	-0.264	-0.761	--
4	7-05-03	1.81	0.477	0.846	-0.533	64	3	1.81	90.7	-0.00377	0.0496	--
5	7-06-14	2.73	0.954	0.282	-0.96	540	9	2.82	938	-0.0925	-0.25	--
6	7-06-20	2.65	0.954	0.181	-0.983	450	9	2.87	1050	-0.219	-0.635	--
7	7-07-16	2.63	0.602	-0.262	-0.965	430	4	2.6	559	0.0338	0.199	--
8	7-07-26	2.15	0.477	-0.423	-0.906	140	3	2.46	408	-0.317	-0.939	--
9	7-08-02	2.36	0.602	-0.529	-0.849	230	4	2.59	542	-0.225	-0.696	--
10	7-08-20	2.51	0.699	-0.763	-0.647	320	5	2.6	564	-0.0984	-0.302	--
11	7-09-12	1.91	0.602	-0.953	-0.303	82	4	2.29	271	-0.372	-1.15	--

12	7-09-17	2	0.301	-0.976	-0.22	100	2	1.9	111	0.102	0.407	--
13	8-03-24	1.62	0.903	0.992	0.125	42	8	1.72	74.5	-0.101	-0.328	--
14	8-04-08	1.72	1.32	0.991	-0.132	52	21	2.38	340	-0.667	-1.95	--
15	8-04-23	2.11	0.863	0.925	-0.381	130	7.3	2.09	174	0.0217	0.149	--
16	8-05-05	1.85	0.255	0.827	-0.562	70	1.8	1.59	55.2	0.251	0.696	--
17	8-05-14	1.97	0.477	0.73	-0.683	94	3	1.96	128	0.0141	0.124	--
18	8-05-22	2.75	0.799	0.63	-0.777	560	6.3	2.41	364	0.335	0.939	--
19	8-05-30	1.98	0.398	0.517	-0.856	95	2.5	2.07	164	-0.0896	-0.199	--
20	8-06-04	3.56	1.86	0.442	-0.897	3600	72	3.72	7370	-0.164	-0.462	--
21	8-06-12	2.61	0.752	0.314	-0.949	410	5.65	2.59	541	0.0268	0.174	--
22	8-06-23	2.26	0.477	0.13	-0.991	180	3	2.37	330	-0.115	-0.354	--
23	8-07-02	2.23	0.69	-0.0242	-1	170	4.9	2.65	634	-0.424	-1.25	--
24	8-07-09	2.43	0.491	-0.144	-0.99	270	3.1	2.46	408	-0.0321	-0.0248	--
25	8-07-30	2.34	0.806	-0.485	-0.875	220	6.4	2.82	923	-0.475	-1.51	--
26	8-08-06	2.08	0.279	-0.586	-0.81	120	1.9	2.22	232	-0.14	-0.407	--
27	8-08-19	2.04	0.243	-0.751	-0.66	110	1.75	2.11	181	-0.0695	-0.149	--
28	8-08-28	2.08	0.833	-0.844	-0.536	120	6.8	2.69	684	-0.608	-1.8	--
29	8-09-11	2.46	0.681	-0.948	-0.319	290	4.8	2.38	340	0.0786	0.381	--
30	8-09-25	1.76	0.204	-0.996	-0.084	58	1.6	1.69	69.2	0.0705	0.354	--
31	9-03-30	1.57	0.681	1	0.0223	37	4.8	1.56	50.9	0.00915	0.0992	--
32	9-04-21	2.46	0.756	0.937	-0.349	290	5.7	1.95	124	0.516	1.6	--
33	9-05-06	3.45	1.11	0.817	-0.576	2800	13	2.55	497	0.898	2.15	--
34	9-05-14	1.98	0.863	0.73	-0.683	96	7.3	2.38	339	-0.4	-1.2	--
35	9-05-26	2.15	0.462	0.575	-0.818	140	2.9	2.09	174	0.0538	0.302	--
36	9-06-02	2.11	0.806	0.472	-0.881	130	6.4	2.55	495	-0.433	-1.31	--
37	9-06-15	3.76	1.53	0.265	-0.964	5700	34	3.47	4100	0.29	0.795	--
38	9-06-24	2.43	0.944	0.113	-0.994	270	8.8	2.89	1090	-0.458	-1.44	--
39	9-07-15	2.36	0.643	-0.245	-0.969	230	4.4	2.64	618	-0.282	-0.829	--
40	9-07-29	3.53	0.863	-0.47	-0.883	3400	7.3	2.88	1070	0.649	1.8	--
41	9-08-13	3.36	2.08	-0.679	-0.734	2300	120	4.16	20300	-0.798	-2.52	--
42	9-08-17	2.28	0.845	-0.728	-0.685	190	7	2.78	853	-0.505	-1.6	--
43	9-08-27	1.98	0.505	-0.835	-0.551	95	3.2	2.34	305	-0.36	-1.02	--
44	9-09-08	1.89	0.699	-0.93	-0.368	78	5	2.44	383	-0.544	-1.69	--
45	9-09-15	2.38	0.806	-0.967	-0.253	240	6.4	2.48	420	-0.0951	-0.276	--
46	10-03-18	1.54	0.763	0.974	0.227	35	5.8	1.5	44.2	0.0458	0.276	--
47	10-03-29	2.94	1.77	0.999	0.0395	870	59	2.74	771	0.2	0.635	--
48	10-04-08	1.62	0.591	0.991	-0.132	42	3.9	1.58	53.7	0.041	0.225	--
49	10-04-21	1.71	0.716	0.937	-0.349	51	5.2	1.9	112	-0.195	-0.605	--
50	10-05-12	2.89	0.996	0.753	-0.658	780	9.9	2.5	445	0.391	1.15	--
51	10-05-20	2.26	0.716	0.656	-0.755	180	5.2	2.3	279	-0.0423	-0.0496	--
52	10-05-27	2.23	0.748	0.561	-0.828	170	5.6	2.42	367	-0.187	-0.518	--
53	10-06-09	2.18	0.613	0.363	-0.932	150	4.1	2.41	358	-0.23	-0.728	--
54	10-06-16	2.38	0.556	0.248	-0.969	240	3.6	2.41	357	-0.0253	0.0248	--
55	10-06-28	2.87	0.672	0.0446	-0.999	740	4.7	2.61	578	0.255	0.728	--
56	10-07-08	2.2	0.0792	-0.127	-0.992	160	1.2	2.01	143	0.195	0.575	--
57	10-07-19	2.63	0.531	-0.311	-0.95	430	3.4	2.53	471	0.108	0.434	--
58	10-07-26	3.71	1.18	-0.423	-0.906	5100	15	3.23	2380	0.478	1.31	--
59	10-08-12	2.32	0.255	-0.667	-0.745	210	1.8	2.17	206	0.156	0.518	--
60	10-08-24	2.36	0.447	-0.805	-0.593	230	2.8	2.3	279	0.0635	0.328	--
61	10-08-30	2.18	0.23	-0.862	-0.507	150	1.7	2.01	144	0.166	0.546	--
62	10-09-09	1.52	-0.0458	-0.936	-0.352	33	0.9	1.61	57.1	-0.0905	-0.225	--
63	10-09-21	1.6	0.23	-0.988	-0.152	40	1.7	1.77	83.1	-0.17	-0.489	--
64	12-03-06	0.699	0.279	0.907	0.422	5	1.9	0.836	9.62	-0.137	-0.381	--
65	12-04-04	1	0.362	0.998	-0.0637	10	2.3	1.28	26.5	-0.276	-0.795	--
66	12-05-02	1.89	0.301	0.855	-0.519	78	2	1.6	56.3	0.289	0.761	--
67	12-06-06	2.26	0.462	0.411	-0.912	180	2.9	2.21	229	0.0433	0.25	--
68	12-07-02	2	0.146	-0.0242	-1	100	1.4	2.06	161	-0.0586	-0.0992	--
69	12-08-01	1.66	-0.699	-0.514	-0.858	46	0.2	1.16	20.5	0.499	1.51	--
70	12-09-04	3.45	1.06	-0.902	-0.431	2800	11.5	2.87	1050	0.574	1.69	--
71	12-10-10	1.7	0.0792	-0.985	0.173	50	1.2	1.35	31.7	0.345	1.02	--
72	13-03-04	0.477	0.146	0.892	0.453	3	1.4	0.671	6.58	-0.194	-0.575	--
73	13-04-02	1.41	0.176	1	-0.0293	26	1.5	1.05	15.6	0.369	1.06	--
74	13-05-02	1.48	0.519	0.855	-0.519	30	3.3	1.84	97.5	-0.364	-1.06	--
75	13-06-05	1.83	0.519	0.426	-0.905	67	3.3	2.26	258	-0.437	-1.37	--
76	13-06-07	3.7	1.68	0.395	-0.919	5000	48	3.56	5070	0.142	0.489	--
77	13-07-02	4.18	1.7	-0.0242	-1	15000	50	3.76	8080	0.416	1.25	--
78	13-07-16	2.83	0.806	-0.262	-0.965	670	6.4	2.82	935	0.00271	0.0744	--
79	13-08-05	2.34	0.519	-0.572	-0.82	220	3.3	2.49	429	-0.143	-0.434	--

80	13-09-05	2	0.301	-0.91	-0.415	100	2	2.03	151	-0.0306	0	--
81	13-10-17	2.15	0.544	-0.957	0.29	140	3.5	1.76	81.6	0.382	1.11	--
82	14-03-24	0.699	0.58	0.992	0.125	5	3.8	1.37	33	-0.671	-2.15	--
83	14-04-14	1.75	0.477	0.972	-0.234	56	3	1.54	48.9	0.207	0.665	--
84	14-05-07	1.96	0.869	0.807	-0.59	91	7.4	2.29	277	-0.336	-0.978	--
85	14-06-02	2.15	0.643	0.472	-0.881	140	4.4	2.37	329	-0.223	-0.665	--
86	14-07-07	2.11	0.23	-0.11	-0.994	130	1.7	2.17	208	-0.0572	-0.0744	--
87	14-08-13	3.36	0.903	-0.679	-0.734	2300	8	2.87	1040	0.491	1.44	--
88	14-08-19	2	-0.222	-0.751	-0.66	100	0.6	1.6	56.1	0.398	1.2	--
89	14-09-09	2.49	0.462	-0.936	-0.352	310	2.9	2.17	206	0.326	0.901	--
90	14-10-01	2	0.114	-1	0.0191	100	1.3	1.52	46	0.485	1.37	--
91	15-03-25	0.903	0.477	0.994	0.108	8	3	1.27	26.2	-0.367	-1.11	--
92	15-04-21	3.58	1.45	0.937	-0.349	3800	28	2.7	710	0.876	1.95	--
93	15-05-12	2.26	0.505	0.753	-0.658	180	3.2	1.96	129	0.292	0.829	--
94	15-06-02	3.53	1.41	0.472	-0.881	3400	26	3.21	2300	0.317	0.864	--
95	15-06-29	4.2	1.23	0.0274	-1	16000	17	3.23	2390	0.972	2.52	--
96	15-07-21	2.28	0.477	-0.344	-0.939	190	3	2.47	411	-0.187	-0.546	--
97	15-07-30	2.11	0.447	-0.485	-0.875	130	2.8	2.42	373	-0.31	-0.864	--
98	15-08-12	2.68	0.415	-0.667	-0.745	480	2.6	2.34	308	0.34	0.978	--
99	15-08-27	2	0.279	-0.835	-0.551	100	1.9	2.09	173	-0.0894	-0.174	--
100	15-09-09	1.51	0.146	-0.936	-0.352	32	1.4	1.82	92.6	-0.314	-0.901	--
101	15-09-14	2.28	0.491	-0.963	-0.27	190	3.1	2.14	195	0.137	0.462	--

Definitions

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

TURB: Turbidity in FNU (63680)

sin2piJD: $\sin(2\pi JD/365)$ in day/days

cos2pijd: $\cos(2\pi JD/365)$ in day/days