

Appendix 1. Model Archive Summary for Best-Fit Regression Developed to Estimate Fecal Coliform Concentration at Station 01480617; West Branch Brandywine Creek at Modena, 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: 01480617

Station name: West Branch Brandywine Creek at Modena, Pennsylvania

Location: Latitude 39°57'42", longitude 75°48'06", referenced to North American Datum of 1927, Chester County, Pennsylvania, Hydrologic unit 02040205, on left bank at bridge on SR 15068 at Modena and 300 feet (ft) upstream from Denis Run.

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. The YSI 6136 turbidity sensor replaced earlier Analite instrumentation on February 28, 2007. 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 are recorded every 30 minutes 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 TIBCO Spotfire S+ 8.1 program 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 106 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 lead 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:

$$\log_{10}(FC) = 0.758 \times \log_{10}(Turb) - 0.293 \times \sin(2\pi JD/365) - 0.529 \times \cos(2\pi JD/365) + 2.13$$

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 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 1.38. The retransformed model, using the BCF, is:

$$FC = 1.38 \times 10^{\log_{10}(FC)} \text{ or } FC = (Turb^{0.8581} \times 187.9) / [10^{0.2931 \sin(2\pi JD/365)} \times 10^{0.5923 \cos(2\pi JD/365)}]$$

Model Statistics, Data, and Plots

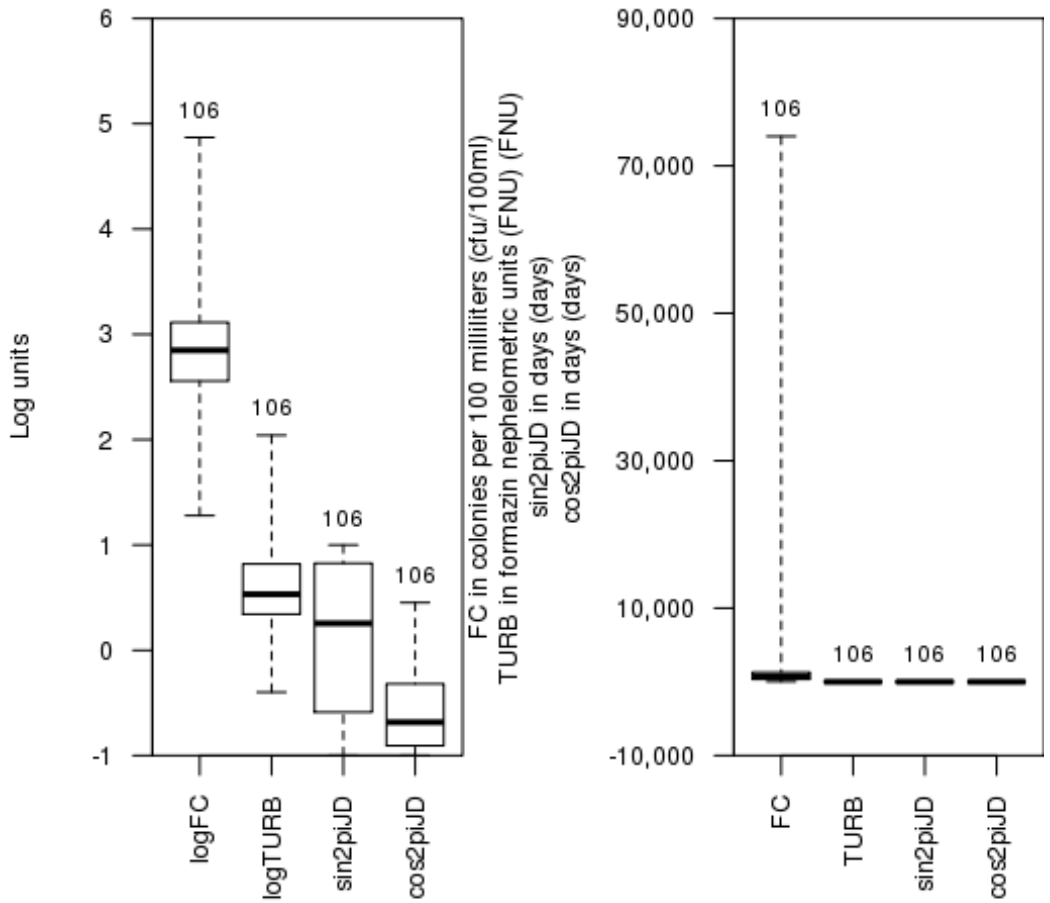
Model

$$\log FC = + 0.758 * \log TURB - 0.293 * \sin 2\pi JD - 0.529 * \cos 2\pi JD + 2.13$$

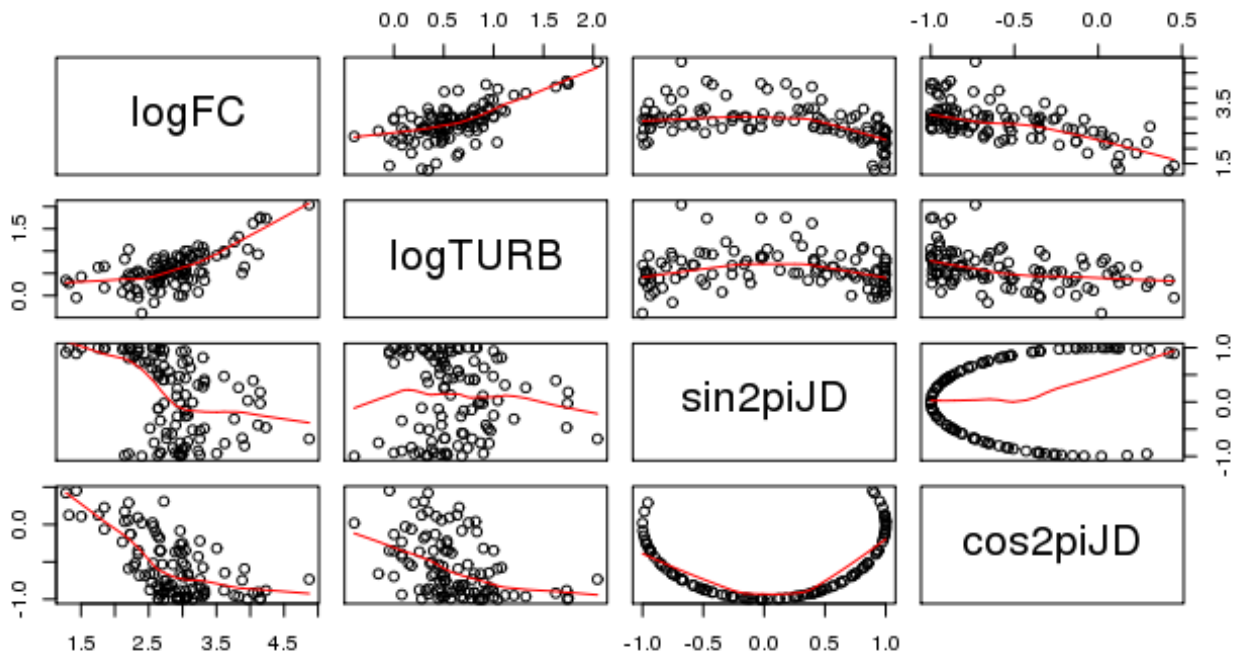
Variable Summary Statistics

	logFC	FC	logTURB	sin2piJD	cos2piJD	TURB
Minimum	1.28	19	-0.398	-1.000	-1.000	0.40
1st Quartile	2.56	360	0.342	-0.588	-0.906	2.20
Median	2.85	705	0.531	0.255	-0.683	3.40
Mean	2.86	2440	0.599	0.107	-0.566	7.29
3rd Quartile	3.11	1300	0.820	0.826	-0.319	6.60
Maximum	4.87	74000	2.040	1.000	0.453	110.00

Box Plots



Exploratory Plots



Basic Model Statistics

Number of Observations 106
 Standard error (RMSE) 0.348
 Average Model standard percentage error (MSPE) 89
 Coefficient of determination (R^2) 0.695
 Adjusted Coefficient of Determination (Adj. R^2) 0.686
 Bias Correction Factor (BCF) 1.38

Variance Inflation Factors (VIF)

	logTURB	sin2piJD	cos2piJD
	1.22	1.05	1.28

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(> t)
(Intercept)	2.130	0.0686	31.10	7.33e-54
logTURB	0.758	0.0911	8.32	4.12e-13
sin2piJD	-0.293	0.0485	-6.05	2.47e-08
cos2piJD	-0.529	0.0964	-5.49	2.98e-07

Correlation Matrix

	Intercept	logTURB	sin2piJD	cos2piJD
Intercept	1.000	-0.455	-0.211	0.476
logTURB	-0.455	1.000	-0.044	0.423
sin2piJD	-0.211	-0.044	1.000	-0.214
cos2piJD	0.476	0.423	-0.214	1.000

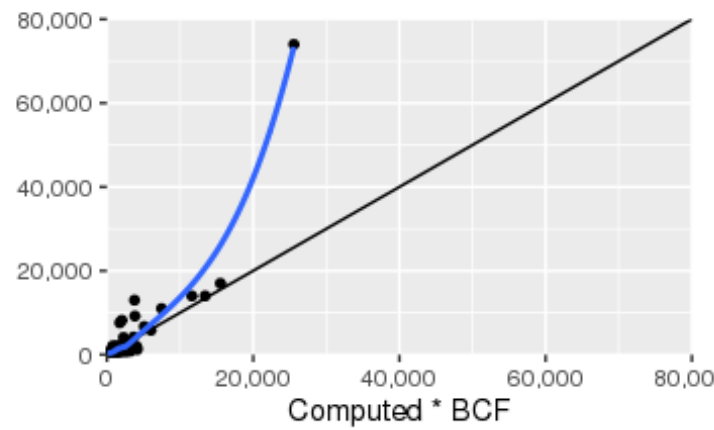
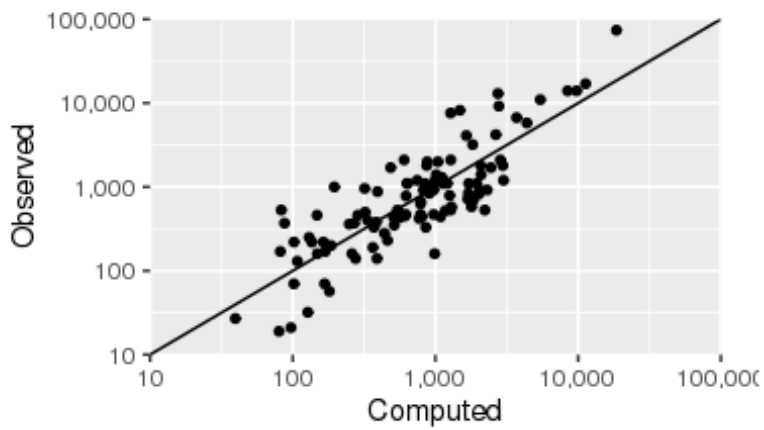
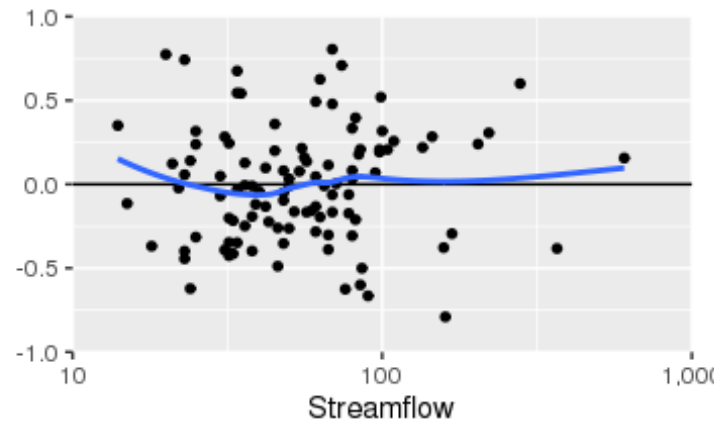
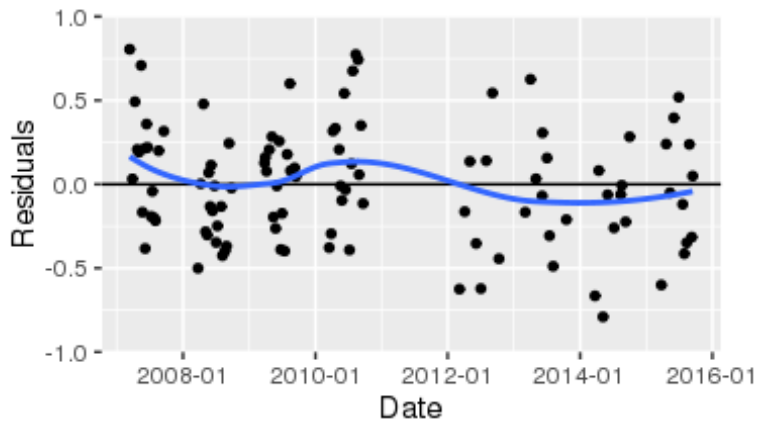
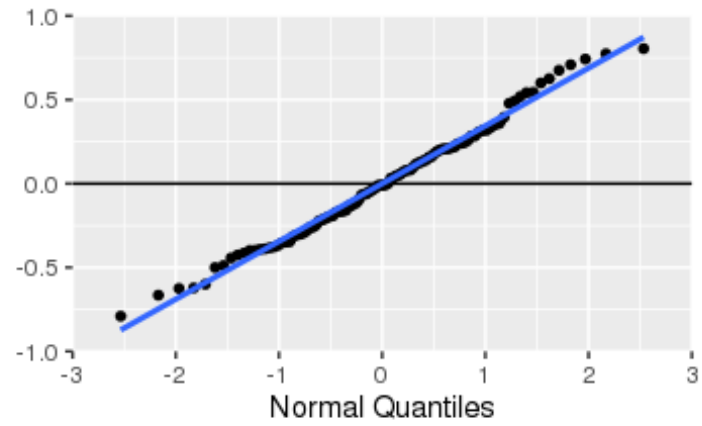
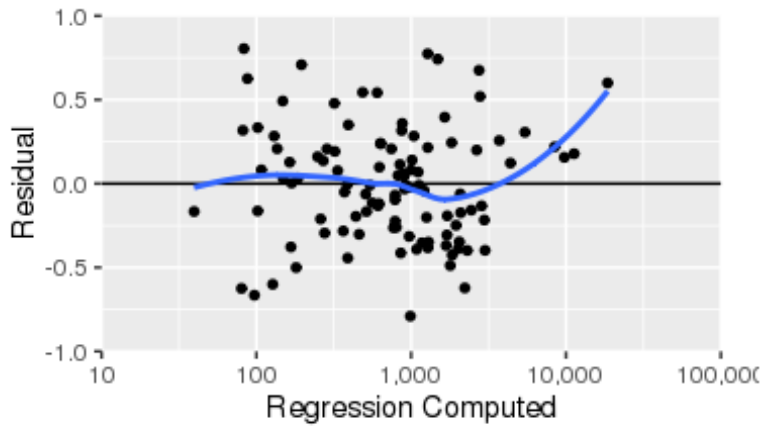
Outlier Test Criteria

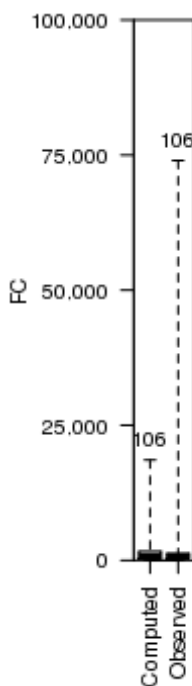
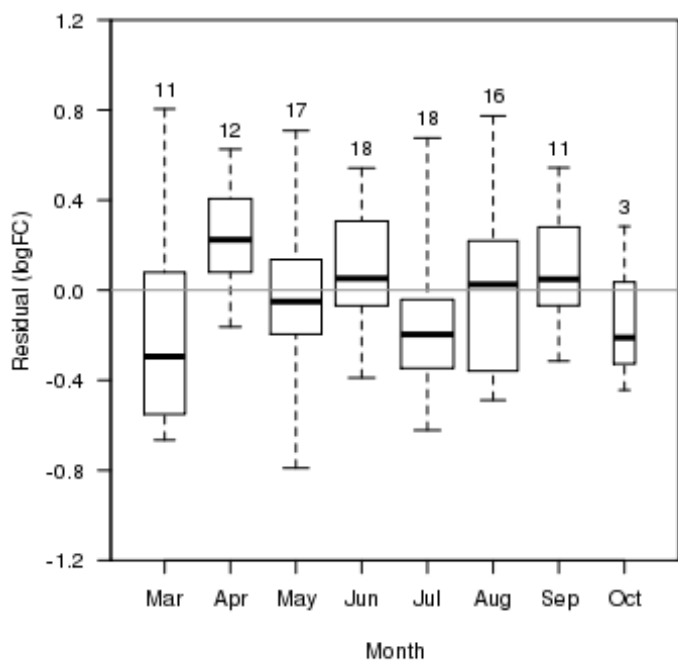
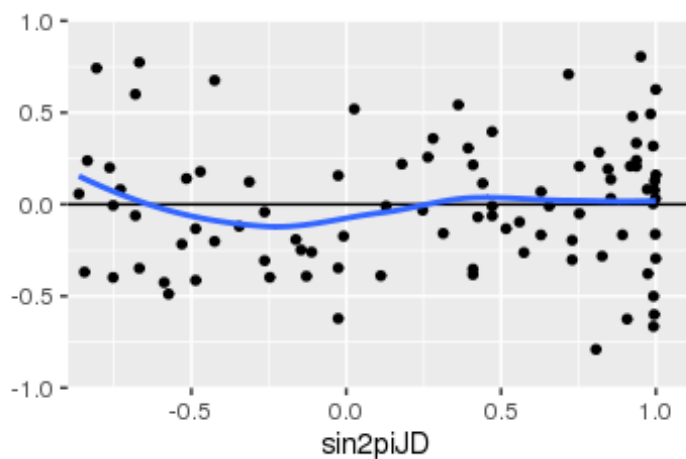
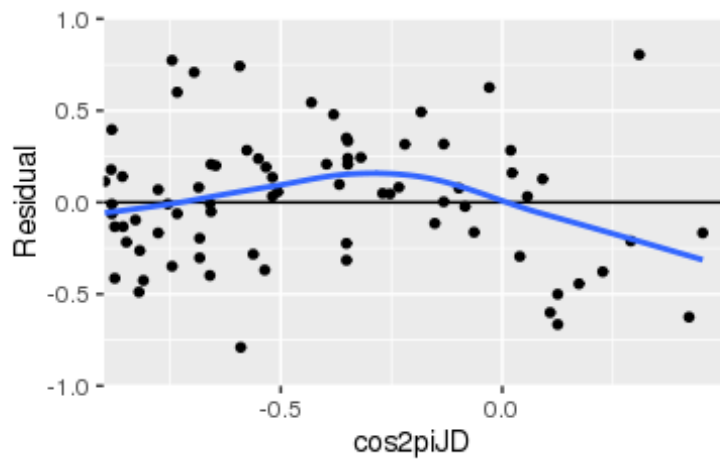
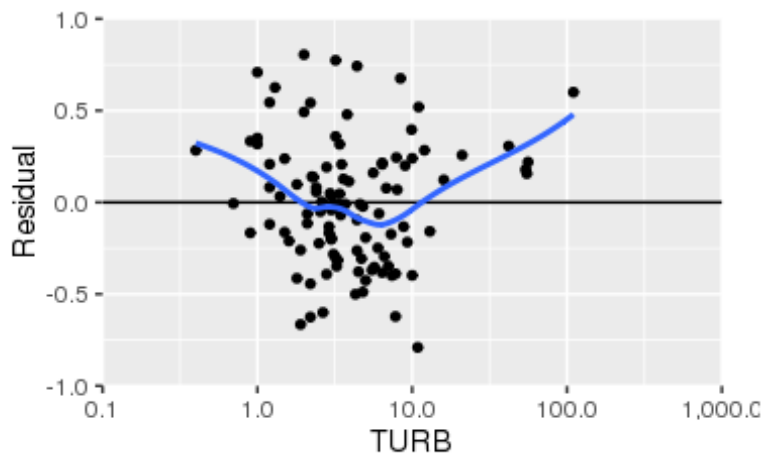
Leverage	Cook's D	DFFITS
0.0849	0.2647	0.3365

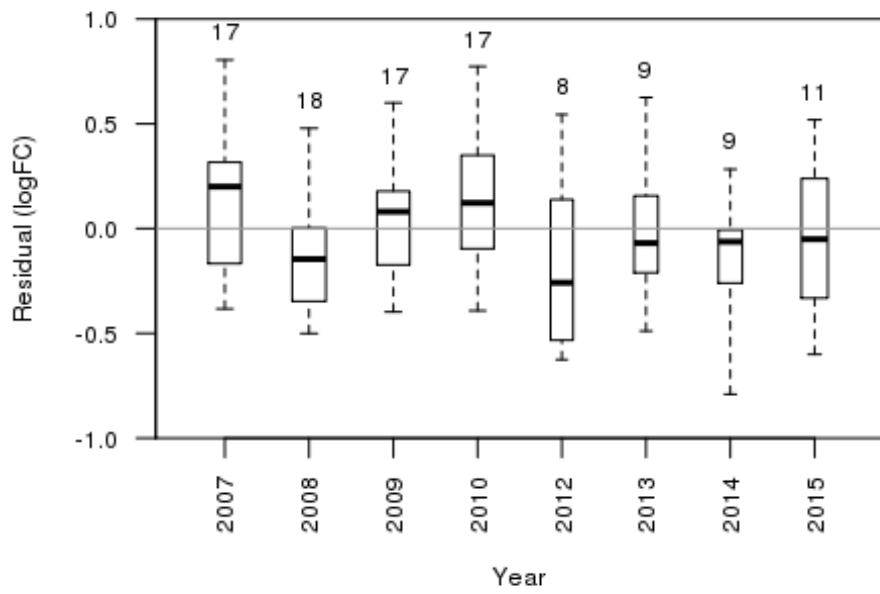
Flagged Observations

	logFC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS	
03/13/2007 07:30	2.72	1.92	0.805		2.390	2.440	0.0608	0.0921	0.622
05/15/2007 10:00	3.00	2.29	0.709		2.090	2.120	0.0498	0.0571	0.486
03/24/2008 08:27	1.76	2.26	-0.501		-1.480	-1.490	0.0527	0.0304	-0.351
07/29/2009 12:18	4.23	4.05	0.178		0.536	0.535	0.0887	0.0070	0.167
08/13/2009 08:09	4.87	4.27	0.600		1.870	1.900	0.1530	0.1580	0.806
08/12/2010 08:09	3.88	3.11	0.774		2.250	2.300	0.0226	0.0293	0.349
08/24/2010 08:18	3.91	3.17	0.742		2.160	2.200	0.0255	0.0305	0.356
03/06/2012 08:27	1.28	1.90	-0.625		-1.870	-1.890	0.0734	0.0689	-0.532
09/04/2012 10:00	3.23	2.69	0.544		1.600	1.610	0.0454	0.0304	0.352
10/10/2012 09:18	2.15	2.59	-0.444		-1.330	-1.330	0.0798	0.0383	-0.393
04/02/2013 08:18	2.57	1.94	0.626		1.840	1.860	0.0406	0.0357	0.382
10/17/2013 11:27	2.20	2.41	-0.211		-0.635	-0.633	0.0913	0.0101	-0.201
03/24/2014 11:27	1.32	1.99	-0.665		-1.960	-1.980	0.0456	0.0457	-0.434
05/07/2014 09:09	2.20	2.99	-0.790		-2.310	-2.360	0.0307	0.0421	-0.420
10/01/2014 09:27	2.40	2.11	0.283		0.857	0.856	0.0989	0.0202	0.284
03/25/2015 10:27	1.51	2.11	-0.600		-1.770	-1.780	0.0451	0.0368	-0.388

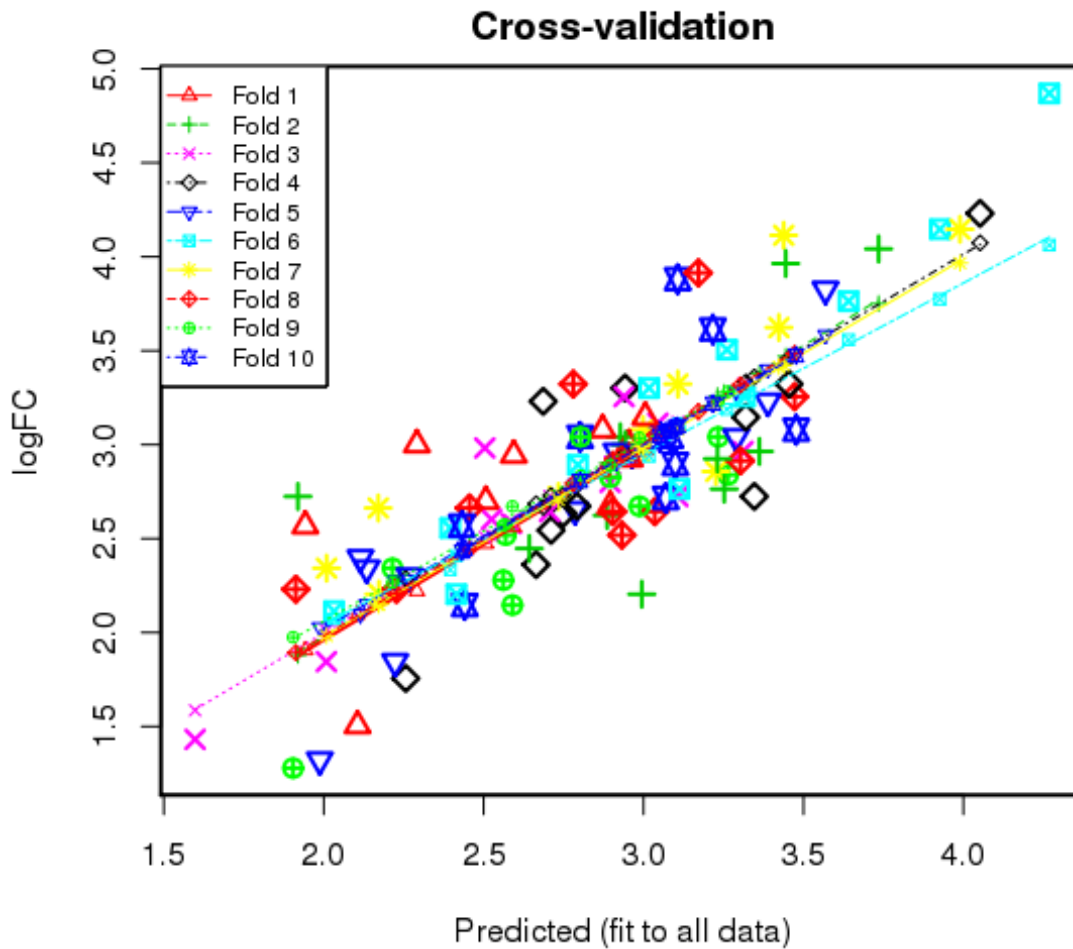
Statistical Plots



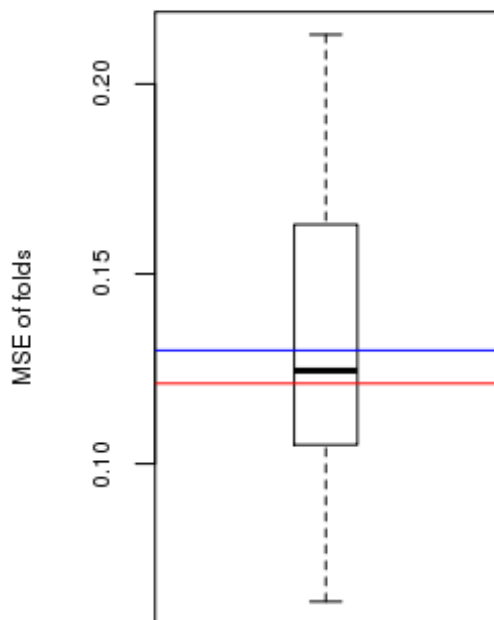




Cross Validation



Minimum MSE of folds: 0.0637
 Mean MSE of folds: 0.1300
 Median MSE of folds: 0.1240
 Maximum MSE of folds: 0.2130
 (Mean MSE of folds) / (Model MSE): 1.0700



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	2007-03-13	2.72	0.301	0.951	0.309	530	2	1.92	114	0.805	2.53	--
2	2007-03-28	2.2	0.477	0.998	0.0567	160	3	2.17	205	0.0317	0.0827	--
3	2007-04-11	2.66	0.301	0.983	-0.183	460	2	2.17	204	0.492	1.28	--
4	2007-04-24	2.34	0.0792	0.918	-0.397	220	1.2	2.13	188	0.208	0.644	--
5	2007-05-03	2.7	0.447	0.845	-0.533	500	2.8	2.51	442	0.192	0.531	--
6	2007-05-15	3	0	0.718	-0.696	1000	1	2.29	269	0.709	1.83	--
7	2007-05-22	2.54	0.462	0.629	-0.777	350	2.9	2.71	707	-0.167	-0.426	--
8	2007-06-06	3.32	0.806	0.409	-0.912	2100	6.4	3.11	1760	0.215	0.673	--
9	2007-06-06	2.72	0.806	0.409	-0.912	530	6.4	3.11	1760	-0.383	-1.09	--
10	2007-06-14	3.3	0.505	0.28	-0.96	2000	3.2	2.94	1200	0.359	1.14	--
11	2007-06-20	4.15	1.75	0.18	-0.983	14000	56	3.93	11600	0.22	0.703	--

12	2007-07-10	3.04	0.699	-0.163	-0.987	1100	5	3.23	2360	-0.192	-0.478	--
13	2007-07-16	3.04	0.477	-0.264	-0.965	1100	3	3.08	1670	-0.0418	-0.106	--
14	2007-07-26	2.9	0.477	-0.425	-0.906	790	3	3.1	1730	-0.202	-0.531	--
15	2007-08-02	3.26	0.968	-0.531	-0.849	1800	9.3	3.47	4080	-0.217	-0.587	--
16	2007-08-20	3.62	0.954	-0.764	-0.647	4200	9	3.42	3650	0.2	0.559	--
17	2007-09-17	3.26	0.531	-0.976	-0.22	1800	3.4	2.94	1200	0.316	0.971	--
18	2008-03-24	1.76	0.633	0.992	0.125	57	4.3	2.26	248	-0.501	-1.62	--
19	2008-04-08	2.23	0.415	0.991	-0.132	170	2.6	2.23	232	0.00301	0.059	--
20	2008-04-23	2.98	0.58	0.924	-0.381	960	3.8	2.5	439	0.479	1.23	--
21	2008-05-05	2.28	0.491	0.826	-0.562	190	3.1	2.56	501	-0.283	-0.733	--
22	2008-05-14	2.36	0.505	0.73	-0.683	230	3.2	2.66	635	-0.302	-0.797	--
23	2008-05-22	3.11	0.903	0.629	-0.777	1300	8	3.04	1530	0.069	0.202	--
24	2008-05-30	2.65	0.462	0.516	-0.856	450	2.9	2.79	840	-0.132	-0.299	--
25	2008-06-04	3.04	0.591	0.441	-0.897	1100	3.9	2.93	1160	0.114	0.324	--
26	2008-06-12	3.23	1.11	0.313	-0.949	1700	13	3.39	3370	-0.158	-0.349	--
27	2008-06-23	3.04	0.568	0.129	-0.991	1100	3.7	3.05	1550	-0.00983	-0.0118	--
28	2008-07-02	2.96	0.845	-0.0258	-1	920	7	3.31	2810	-0.347	-0.898	--
29	2008-07-09	3.04	0.778	-0.146	-0.99	1100	6	3.29	2680	-0.248	-0.644	--
30	2008-07-30	3.32	0.944	-0.486	-0.875	2100	8.8	3.45	3920	-0.133	-0.324	--
31	2008-08-06	2.84	0.699	-0.588	-0.81	690	5	3.26	2530	-0.426	-1.4	--
32	2008-08-19	2.96	0.869	-0.753	-0.66	920	7.4	3.36	3170	-0.399	-1.28	--
33	2008-08-28	2.86	0.74	-0.845	-0.536	720	5.5	3.23	2320	-0.369	-1.01	--
34	2008-09-11	3.51	0.898	-0.948	-0.319	3200	7.9	3.26	2510	0.244	0.797	--
35	2008-09-25	2.96	0.681	-0.997	-0.084	920	4.8	2.99	1330	-0.0227	-0.059	--
36	2009-03-26	2.34	0.556	0.996	0.091	220	3.6	2.22	226	0.127	0.374	--
37	2009-03-30	2.56	0.748	1	0.0223	360	5.6	2.4	342	0.161	0.478	--
38	2009-04-06	2.6	0.833	0.995	-0.098	400	6.8	2.52	461	0.0773	0.226	--
39	2009-04-21	2.66	0.544	0.937	-0.349	460	3.5	2.46	393	0.207	0.587	--
40	2009-05-06	3.3	1.08	0.817	-0.576	2000	12	3.02	1430	0.284	0.898	--
41	2009-05-14	2.45	0.477	0.73	-0.683	280	3	2.64	605	-0.196	-0.504	--
42	2009-05-26	2.62	0.643	0.574	-0.818	420	4.4	2.89	1060	-0.263	-0.703	--
43	2009-06-02	2.95	0.663	0.471	-0.881	900	4.6	2.96	1270	-0.01	-0.0354	--
44	2009-06-15	3.83	1.32	0.264	-0.964	6700	21	3.57	5100	0.257	0.83	--
45	2009-06-24	2.91	0.892	0.112	-0.994	820	7.8	3.3	2760	-0.389	-1.14	--
46	2009-07-01	3.15	0.863	-0.0086	-1	1400	7.3	3.32	2870	-0.174	-0.452	--
47	2009-07-15	3.08	1	-0.247	-0.969	1200	10	3.48	4130	-0.398	-1.23	--
48	2009-07-29	4.23	1.73	-0.471	-0.883	17000	54	4.05	15500	0.178	0.504	--
49	2009-08-13	4.87	2.04	-0.681	-0.734	74000	110	4.27	25600	0.6	1.54	--
50	2009-08-17	3.08	0.38	-0.73	-0.685	1200	2.4	3	1370	0.0809	0.25	--
51	2009-09-08	2.89	0.255	-0.931	-0.368	780	1.8	2.79	857	0.0977	0.299	--
52	2009-09-15	3	0.531	-0.968	-0.253	1000	3.4	2.95	1240	0.0458	0.13	--
53	2010-03-18	1.85	0.653	0.974	0.227	70	4.5	2.22	230	-0.378	-1.05	--
54	2010-03-29	2.15	0.82	0.999	0.0395	140	6.6	2.44	380	-0.295	-0.765	--
55	2010-04-08	2.23	0	0.991	-0.132	170	1	1.91	113	0.318	1.01	--
56	2010-04-21	2.34	-0.0458	0.937	-0.349	220	0.9	2.01	140	0.334	1.05	--
57	2010-05-12	3.08	0.806	0.753	-0.658	1200	6.4	2.87	1020	0.207	0.615	--
58	2010-05-20	2.72	0.519	0.655	-0.755	530	3.3	2.73	746	-0.00964	0.0118	--
59	2010-05-27	2.8	0.643	0.56	-0.828	630	4.4	2.9	1080	-0.0961	-0.226	--
60	2010-06-09	3.32	0.342	0.362	-0.932	2100	2.2	2.78	829	0.542	1.4	--
61	2010-06-16	2.92	0.505	0.247	-0.969	840	3.2	2.96	1250	-0.0324	-0.0827	--
62	2010-07-08	2.64	0.447	-0.129	-0.992	440	2.8	3.04	1490	-0.392	-1.18	--
63	2010-07-19	3.76	1.2	-0.313	-0.95	5800	16	3.64	6020	0.122	0.349	--
64	2010-07-26	4.11	0.924	-0.425	-0.906	13000	8.4	3.44	3770	0.676	1.71	--
65	2010-08-12	3.88	0.505	-0.668	-0.745	7600	3.2	3.11	1760	0.774	2.17	--
66	2010-08-24	3.91	0.643	-0.806	-0.593	8200	4.4	3.17	2040	0.742	1.97	--
67	2010-08-30	3	0.38	-0.863	-0.507	1000	2.4	2.94	1210	0.0571	0.178	--
68	2010-09-09	2.94	0	-0.937	-0.352	880	1	2.59	541	0.35	1.09	--
69	2010-09-21	2.63	0.322	-0.989	-0.152	430	2.1	2.75	770	-0.115	-0.25	--
70	2012-03-06	1.28	0.342	0.907	0.422	19	2.2	1.9	110	-0.625	-1.97	--
71	2012-04-04	1.85	0.176	0.998	-0.0637	70	1.5	2.01	140	-0.163	-0.374	--
72	2012-05-02	2.57	0.362	0.854	-0.519	370	2.3	2.43	372	0.136	0.4	--
73	2012-06-06	2.72	0.756	0.409	-0.912	520	5.7	3.07	1610	-0.353	-0.971	--
74	2012-07-02	2.72	0.892	-0.0258	-1	530	7.8	3.35	3060	-0.622	-1.83	--
75	2012-08-01	3.15	0.352	-0.516	-0.858	1400	2.25	3.01	1390	0.141	0.426	--
76	2012-09-04	3.23	0.0792	-0.903	-0.431	1700	1.2	2.69	668	0.544	1.46	--
77	2012-10-10	2.15	0.342	-0.984	0.173	140	2.2	2.59	535	-0.444	-1.46	--
78	2013-03-04	1.43	-0.0458	0.892	0.453	27	0.9	1.6	54.5	-0.166	-0.4	--
79	2013-04-02	2.57	0.114	1	-0.0293	370	1.3	1.94	120	0.626	1.62	--

80	2013-05-02	2.3	0.146	0.854	-0.519	200	1.4	2.27	255	0.0326	0.106	--
81	2013-06-05	2.83	0.538	0.425	-0.905	670	3.45	2.9	1080	-0.0693	-0.202	--
82	2013-06-07	4.04	1.62	0.394	-0.919	11000	42	3.73	7470	0.306	0.934	--
83	2013-07-02	4.15	1.74	-0.0258	-1	14000	55	3.99	13400	0.157	0.452	--
84	2013-07-16	2.92	0.672	-0.264	-0.965	840	4.7	3.23	2340	-0.307	-0.83	--
85	2013-08-05	2.76	0.681	-0.574	-0.82	580	4.8	3.25	2460	-0.489	-1.54	--
86	2013-10-17	2.2	0.204	-0.956	0.29	160	1.6	2.41	358	-0.211	-0.559	--
87	2014-03-24	1.32	0.279	0.992	0.125	21	1.9	1.99	134	-0.665	-2.17	--
88	2014-04-14	2.11	0.0792	0.972	-0.234	130	1.2	2.03	148	0.0818	0.275	--
89	2014-05-07	2.2	1.04	0.806	-0.59	160	10.8	2.99	1360	-0.79	-2.53	--
90	2014-06-02	2.64	0.322	0.471	-0.881	440	2.1	2.71	699	-0.0627	-0.178	--
91	2014-07-07	2.64	0.279	-0.112	-0.994	440	1.9	2.9	1100	-0.26	-0.673	--
92	2014-08-13	3.26	0.785	-0.681	-0.734	1800	6.1	3.32	2850	-0.0615	-0.154	--
93	2014-08-19	2.58	-0.155	-0.753	-0.66	380	0.7	2.59	530	-0.00611	0.0354	--
94	2014-09-09	2.67	0.398	-0.937	-0.352	470	2.5	2.9	1080	-0.224	-0.615	--
95	2014-10-01	2.4	-0.398	-1	0.0191	250	0.4	2.11	179	0.283	0.863	--
96	2015-03-25	1.51	0.423	0.994	0.108	32	2.65	2.11	175	-0.6	-1.71	--
97	2015-04-21	3.04	1	0.937	-0.349	1100	10	2.8	871	0.24	0.765	--
98	2015-05-12	2.52	0.407	0.753	-0.658	330	2.55	2.57	510	-0.0505	-0.13	--
99	2015-06-02	3.61	0.996	0.471	-0.881	4100	9.9	3.22	2270	0.396	1.18	--
100	2015-06-29	3.96	1.04	0.0258	-1	9200	11	3.44	3830	0.519	1.34	--
101	2015-07-21	2.67	0.0792	-0.346	-0.939	470	1.2	2.79	852	-0.12	-0.275	--
102	2015-07-30	2.52	0.255	-0.486	-0.875	330	1.8	2.93	1180	-0.414	-1.34	--
103	2015-08-12	2.76	0.512	-0.668	-0.745	580	3.25	3.11	1780	-0.348	-0.934	--
104	2015-08-27	3.04	0.176	-0.836	-0.551	1100	1.5	2.8	875	0.238	0.733	--
105	2015-09-09	2.67	0.519	-0.937	-0.352	470	3.3	2.99	1340	-0.315	-0.863	--
106	2015-09-14	2.96	0.47	-0.963	-0.27	920	2.95	2.91	1130	0.0489	0.154	--

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