Appendix 2. Model Archive Summary for Best-Fit Regression Developed to Estimate Fecal Coliform Concentration at Station 01480870; East Branch Brandywine Creek below Downingtown, Pennsylvania

This model archive summary describes the regression model developed to estimate continuous instantaneous (15 or 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: 01480870

Station name: East Branch Brandywine Creek below Downingtown, Pennsylvania

Location: Latitude 39°58'07", longitude 75°40'25" referenced to North American Datum of 1927, Chester County, PA, Hydrologic Unit 02040205, on left bank at downstream side of Sugars Bridge (U.S. Highway 322), 2,000 feet (ft) upstream from Valley Creek, 1.5 mile (mi) north of Marshallton, and 3.3 mi southeast of Downingtown.

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 August 7, 2008, the YSI 6136 turbidity sensor replaced earlier Analite instrumentation that had been in operation since 2005. The monitor is housed in a 6-inch perforated plastic pipe placed in the stream about 30 ft downstream of gage orifice and 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: January 2017

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

Model application date: March 2017 onward

Computed by: Lisa Senior, January 2017.

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 the seasonality components were evaluated to determine which combinations produced the best models.

The final regression model is based on 102 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*), water temperature (parameter 00010 or *Temp*), and other continuously measured data in addition to computed seasonality terms (*sin2piJD* and *cos2piJD)* 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*), *sin2piJD,* and *cos2piJD* 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 01480870,
FC concentration-based model:

$$\log_{10}(FC) = 0.879 \times \log10(Turb) - 0.444 \times \sin(2\pi JD/365) - 0.638 \times \cos(2\pi JD/365) + 1.68$$

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 seasonality variables sin2piJD and cos2piJD [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 1.45. The retransformed model, using the BCF, is:

$FC = 1.45 \times 10^{\log_{10}(FC)}$ or $FC = (Turb^{0.879} \times 69.4)/[10^{0.444\sin(2\pi JD/365)} \times 10^{0.638\cos(2\pi JD/365)}]$

Model Statistics, Data, and Plots

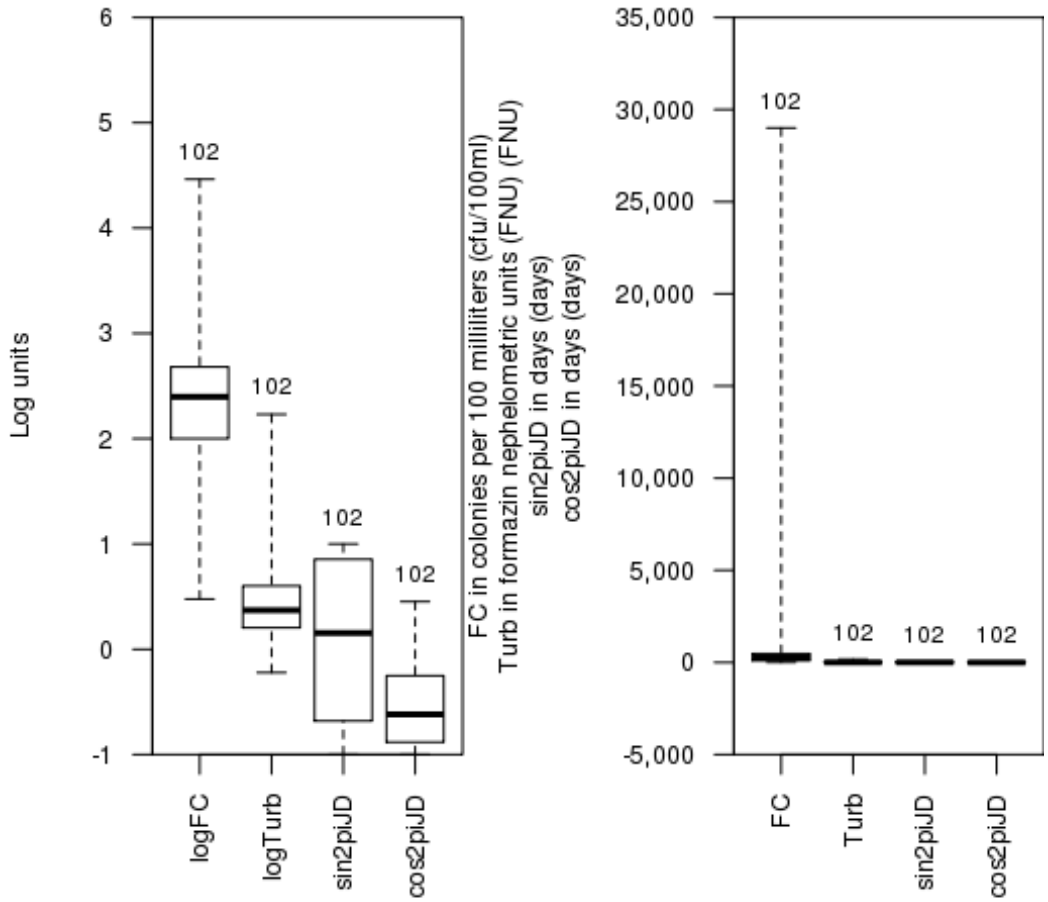
Model

logFC = + 0.879 * logTurb - 0.444 * sin2piJD - 0.638 * cos2piJD + 1.68

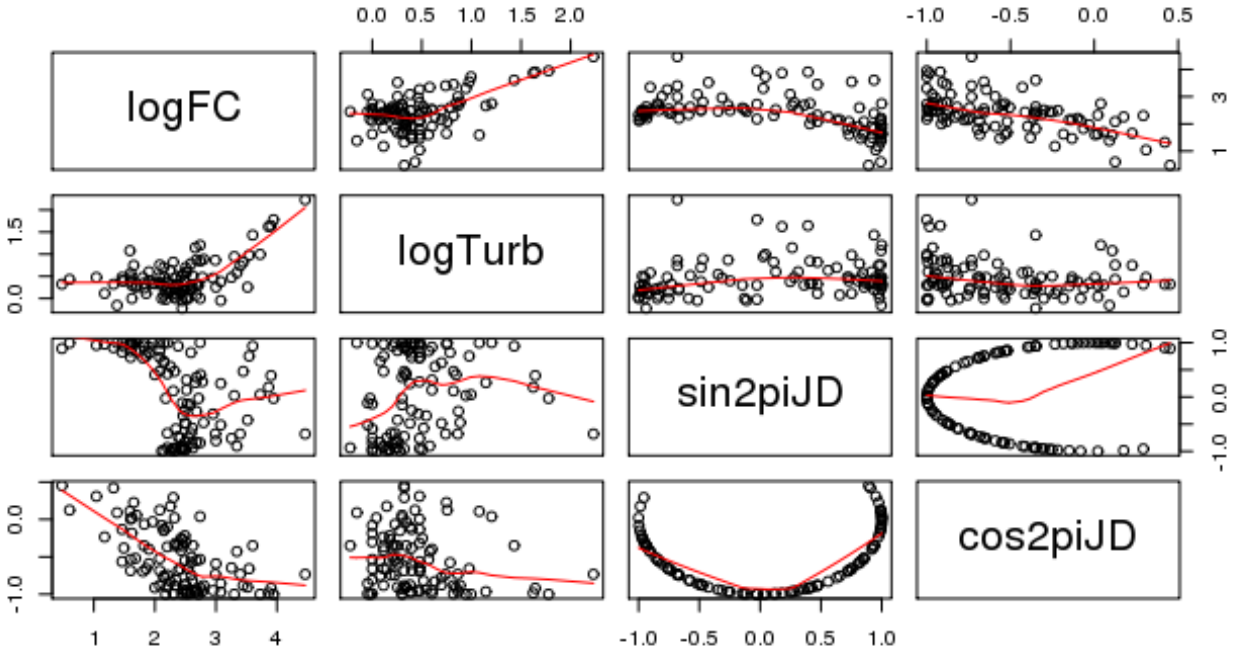
Variable Summary Statistics

	logFC	FC	logTurb	sin2piJD	cos2piJD	Turb
Minimum	0.477	3	-0.222	-1.0000	-1.000	0.60
1st Quartile	2.000	100	0.204	-0.6810	-0.882	1.60
Median	2.400	250	0.371	0.1540	-0.618	2.35
Mean	2.380	1020	0.457	0.0763	-0.530	6.41
3rd Quartile	2.680	480	0.602	0.8540	-0.251	4.00
Maximum	4.460	29000	2.230	1.0000	0.452	170.00

Box Plots



Exploratory Plots



Basic Model Statistics

```
Number of Observations                              102
Standard error (RMSE)                             0.379
Average Model standard percentage error (MSPE)     98.7
Coefficient of determination (R²)                 0.708
Adjusted Coefficient of Determination (Adj. R²)     0.7
Bias Correction Factor (BCF)                       1.45

Variance Inflation Factors (VIF)
 logTurb sin2piJD cos2piJD
    1.11     1.12     1.13
```

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	1.680	0.0705	23.80	1.76e-42
logTurb	0.879	0.0950	9.25	5.14e-15
sin2piJD	-0.444	0.0533	-8.33	5.00e-13
cos2piJD	-0.638	0.1000	-6.35	6.75e-09

Correlation Matrix

	Intercept	logTurb	sin2piJD	cos2piJD
Intercept	1.000	-0.406	-0.107	0.611
logTurb	-0.406	1.000	-0.251	0.259
sin2piJD	-0.107	-0.251	1.000	-0.270
cos2piJD	0.611	0.259	-0.270	1.000

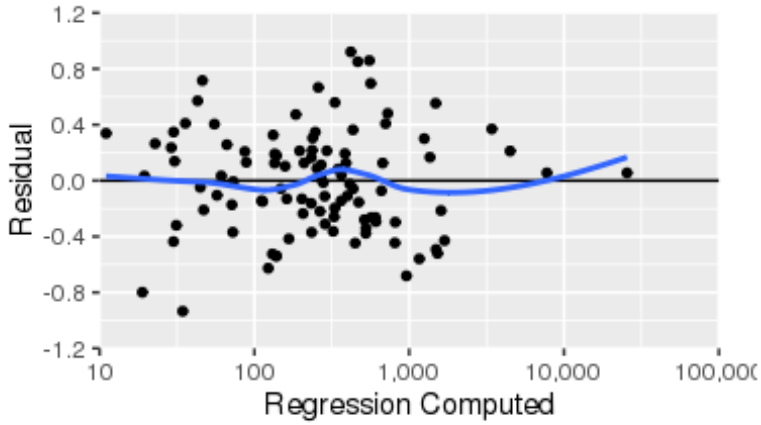
Outlier Test Criteria

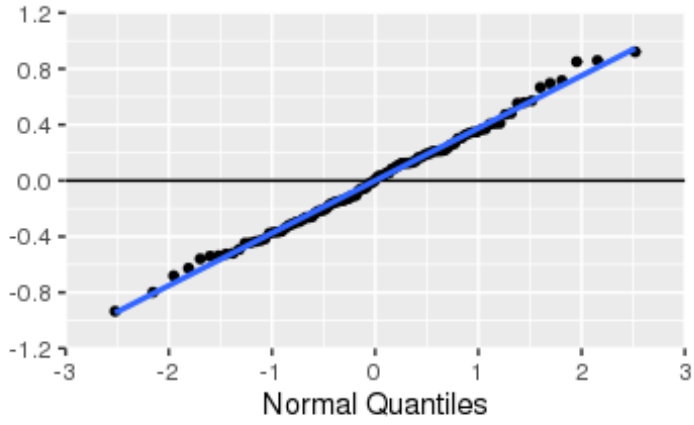
Leverage	Cook's D	DFFITS
0.0882	0.2647	0.3430

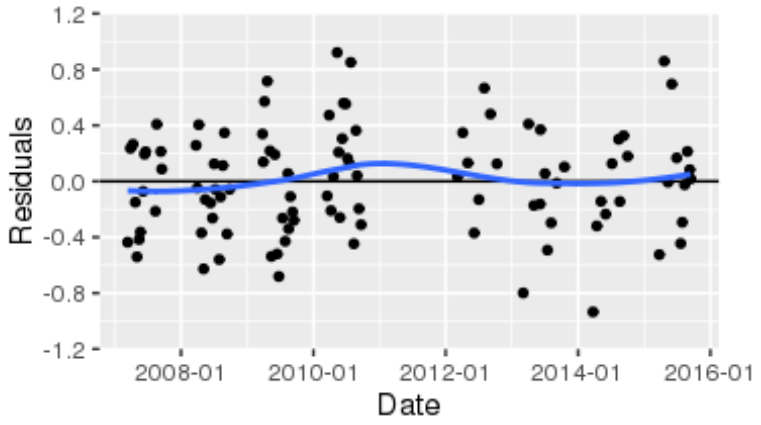
Flagged Observations

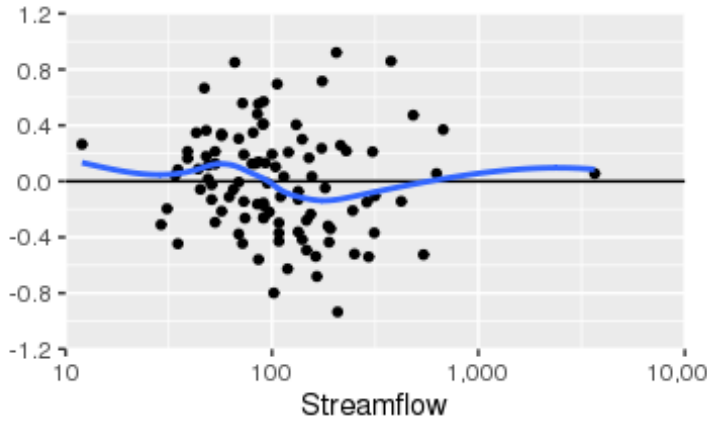
	logFC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
06/20/2007 08:30	3.860	3.65	0.2110	0.585	0.583	0.0906	0.008520	0.1840
08/13/2009 07:45	4.460	4.41	0.0554	0.167	0.166	0.2300	0.002070	0.0906
03/29/2010 09:45	2.740	2.27	0.4730	1.300	1.300	0.0763	0.034800	0.3750
05/12/2010 10:30	3.540	2.62	0.9210	2.470	2.540	0.0303	0.047500	0.4480
07/26/2010 09:30	3.520	2.67	0.8510	2.270	2.320	0.0243	0.032100	0.3660
08/01/2012 13:30	3.080	2.41	0.6670	1.790	1.810	0.0366	0.030500	0.3540
03/04/2013 12:00	0.477	1.28	-0.7990	-2.190	-2.230	0.0732	0.094700	-0.6280
06/07/2013 12:00	3.900	3.53	0.3690	1.020	1.020	0.0908	0.026100	0.3230
07/02/2013 12:45	3.940	3.89	0.0550	0.154	0.154	0.1150	0.000777	0.0555
10/17/2013 14:00	2.300	2.20	0.1030	0.285	0.284	0.0904	0.002020	0.0895
03/24/2014 13:45	0.602	1.54	-0.9350	-2.520	-2.600	0.0443	0.073800	-0.5590
03/25/2015 12:45	1.590	2.12	-0.5250	-1.440	-1.450	0.0711	0.039600	-0.4000
04/21/2015 12:15	3.600	2.74	0.8590	2.360	2.410	0.0745	0.112000	0.6850

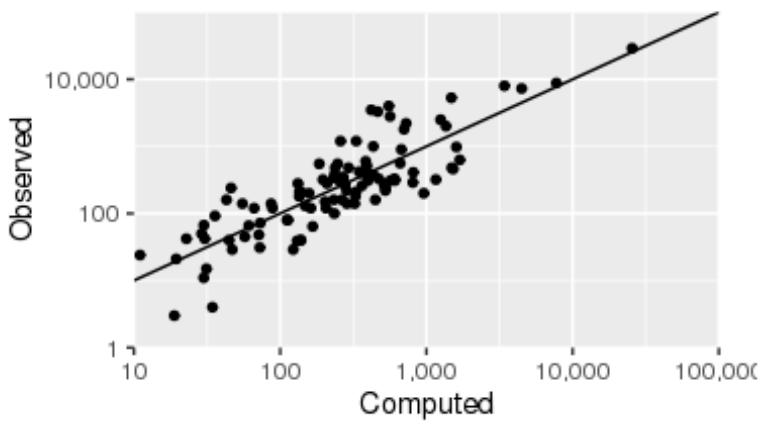
Statistical Plots

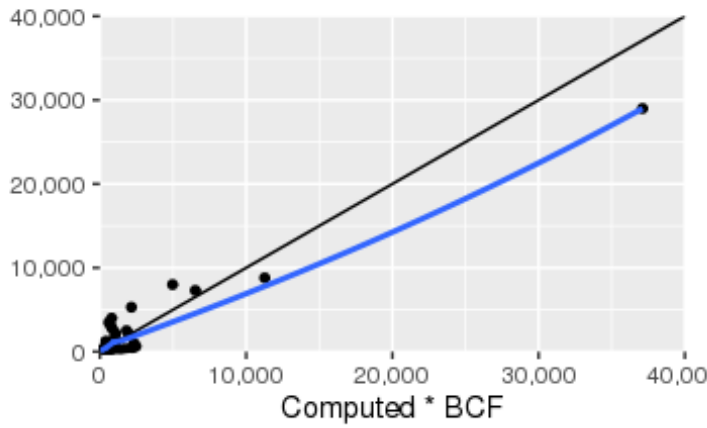


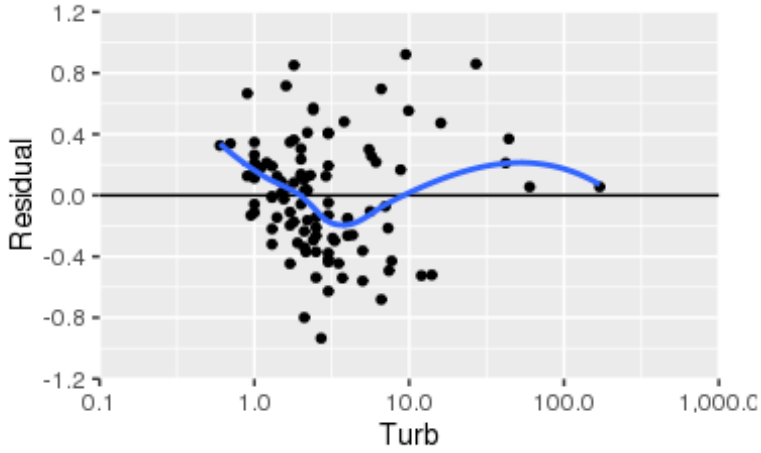


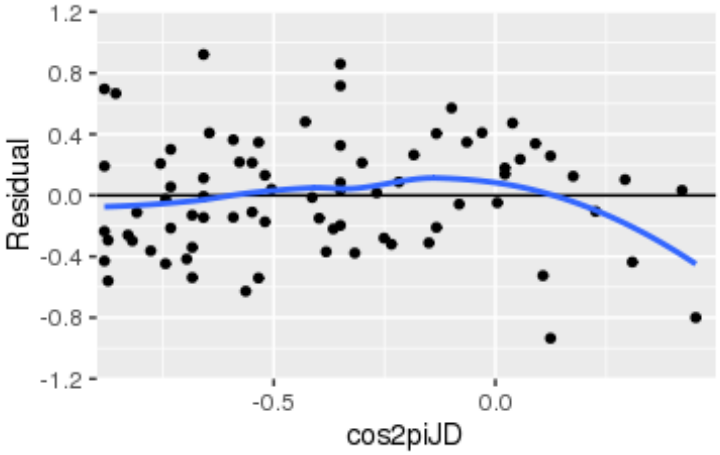


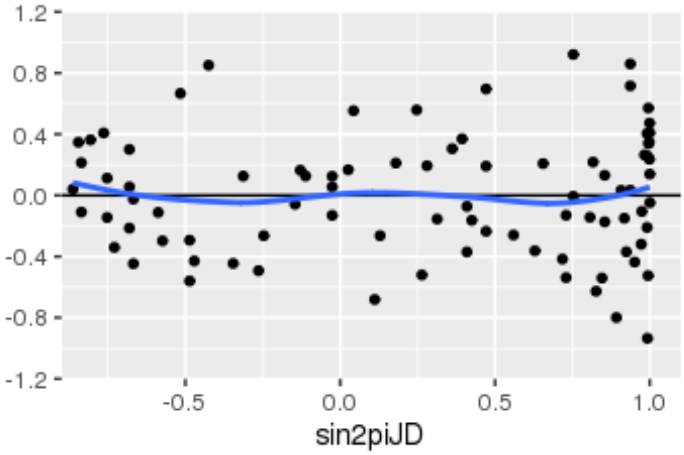


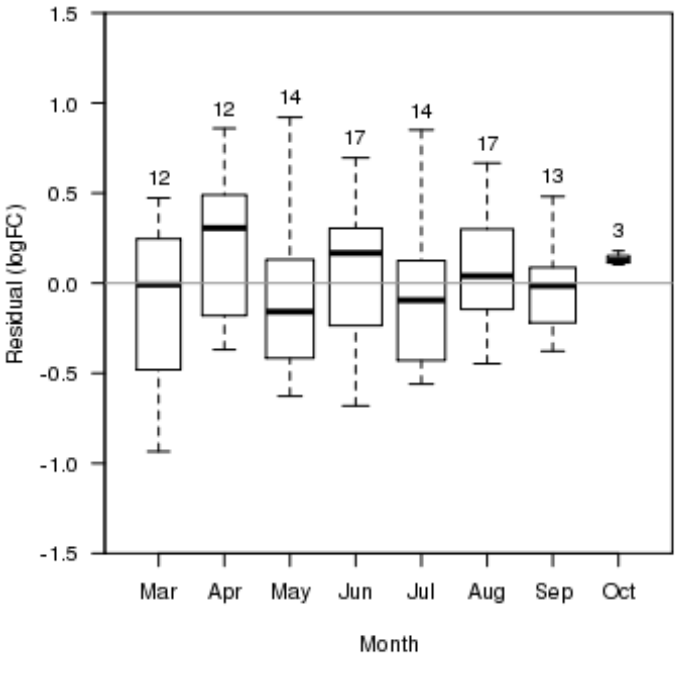


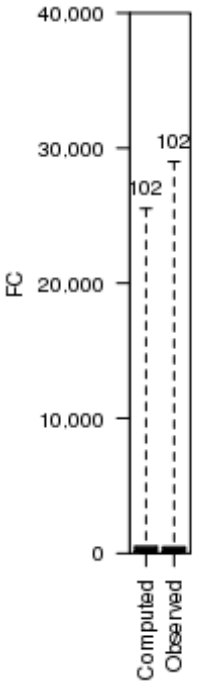


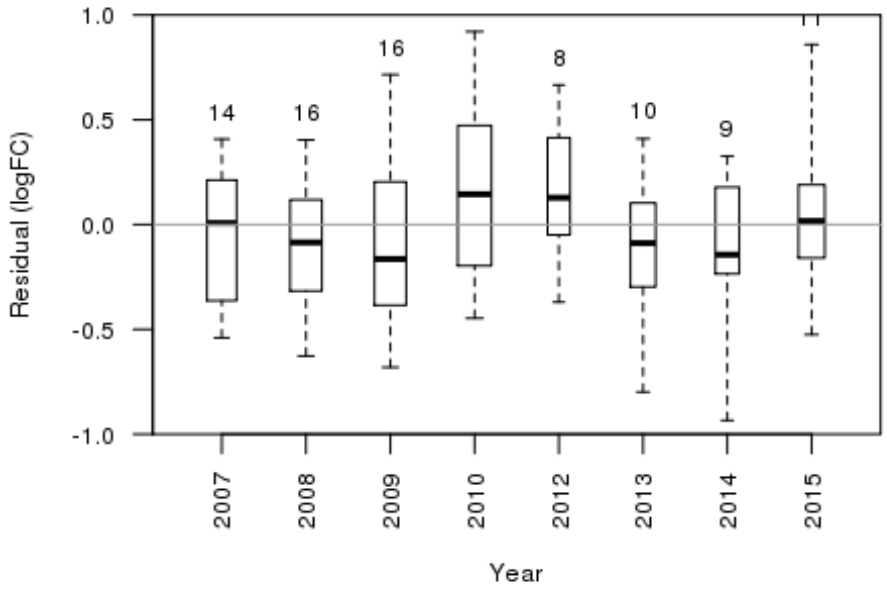

Residual
1.2
0.8
0.4
0.0
-0.4
-0.8
-1.2
0.1
1.0
10.0
100.0
1,000.0
Turb


Residual
1.2
0.8
0.4
0.0
-0.4
-0.8
-1.2
-0.5
0.0
cos2piJD

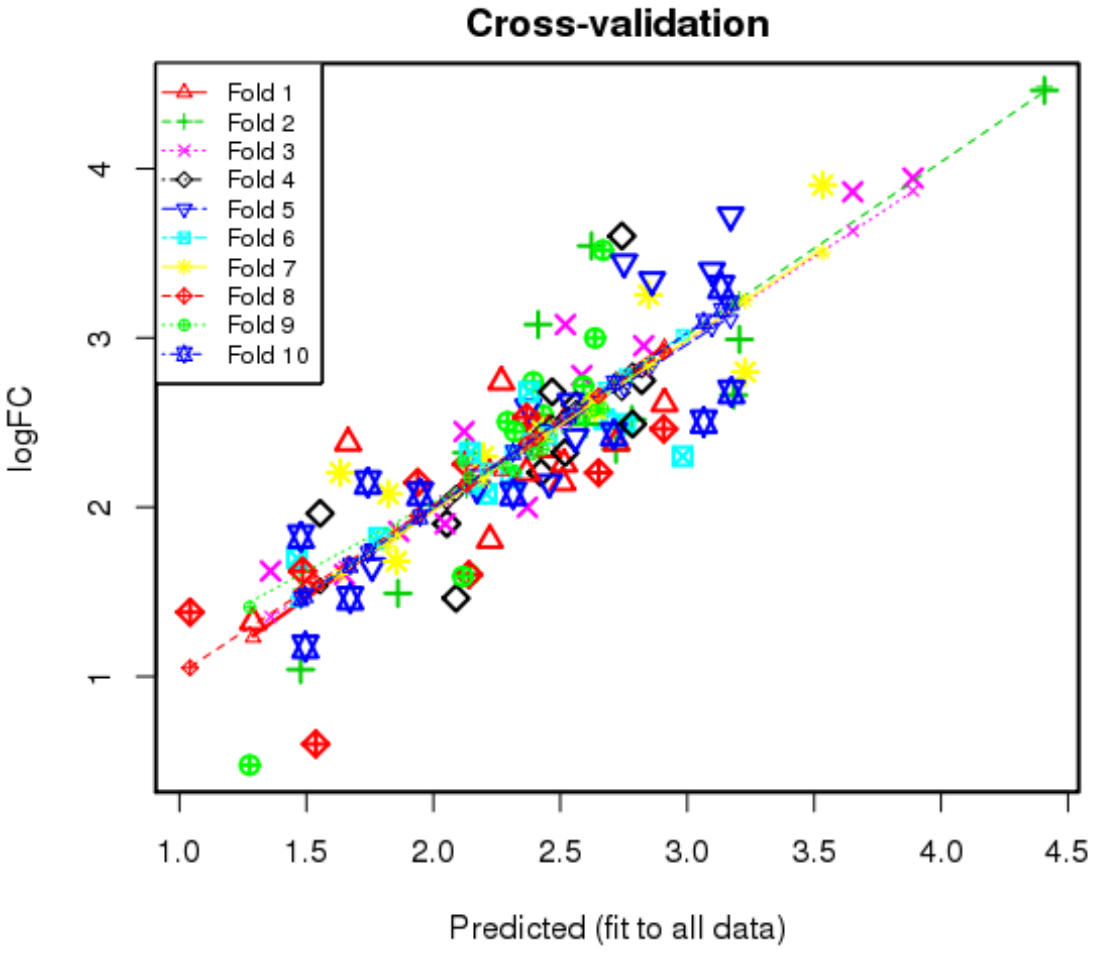

1.2
0.8
0.4
0.0
-0.4
-0.8
-1.2
-0.5
0.0
0.5
1.0
sin2piJD


Residual (logFC)
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
12
12
14
17
14
17
13
3
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Month

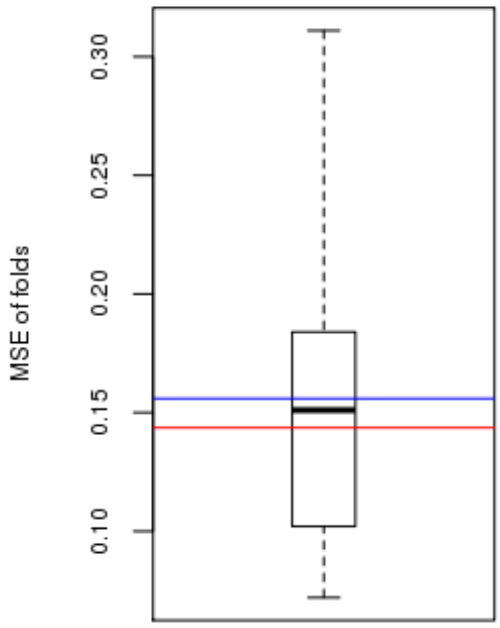

FC
40,000
30,000
20,000
10,000
0
102
102
Computed
Observed



Cross Validation



```
             Minimum MSE of folds:  0.072
                Mean MSE of folds:  0.156
              Median MSE of folds:  0.151
             Maximum MSE of folds:  0.311
(Mean MSE of folds) / (Model MSE):  1.090
```



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
1	2007-03-13	1.04	0.477	0.951	0.309	11	3	1.48	43.6	-0.436	-1.16	--
2	2007-03-28	1.7	0.301	0.998	0.0559	50	2	1.46	42.2	0.236	0.704	--
3	2007-04-11	1.62	0	0.983	-0.184	42	1	1.36	33.2	0.265	0.768	--
4	2007-04-24	1.9	0.602	0.918	-0.398	80	4	2.05	164	-0.15	-0.363	--
5	2007-05-03	1.6	0.568	0.845	-0.534	40	3.7	2.14	202	-0.541	-1.6	--
6	2007-05-15	1.81	0.477	0.718	-0.696	64	3	2.22	243	-0.416	-1.07	--
7	2007-05-22	2.15	0.699	0.629	-0.778	140	5	2.51	469	-0.363	-0.908	--
8	2007-06-06	2.75	0.845	0.409	-0.912	560	7	2.82	962	-0.0725	-0.16	--
9	2007-06-14	2.78	0.477	0.28	-0.96	600	3	2.58	559	0.193	0.526	--
10	2007-06-20	3.86	1.62	0.18	-0.984	7300	42	3.65	6520	0.211	0.583	--
11	2007-08-13	2.99	0.863	-0.681	-0.732	980	7.3	3.21	2330	-0.214	-0.526	--
12	2007-08-20	3.26	0.477	-0.764	-0.645	1800	3	2.85	1020	0.408	1.16	--

13	2007-09-12	2.51	0	-0.954	-0.301	320	1	2.29	285	0.213	0.613	--
14	2007-09-17	2.49	0.176	-0.976	-0.218	310	1.5	2.4	369	0.0874	0.185	--
15	2008-03-24	2.08	0.756	0.992	0.124	120	5.7	1.82	96.5	0.257	0.736	--
16	2008-03-31	1.6	0.477	1	0.0043	40	3	1.65	65	-0.0482	-0.0859	--
17	2008-04-08	2.15	0.477	0.991	-0.133	140	3	1.74	80.2	0.404	1.11	--
18	2008-04-23	1.49	0.398	0.924	-0.382	31	2.5	1.86	105	-0.369	-0.945	--
19	2008-05-05	1.46	0.477	0.826	-0.563	29	3	2.09	179	-0.627	-1.81	--
20	2008-05-14	2.08	0.477	0.73	-0.684	120	3	2.21	235	-0.13	-0.26	--
21	2008-06-12	2.52	0.602	0.313	-0.95	330	4	2.67	686	-0.155	-0.39	--
22	2008-06-23	2.52	0.602	0.129	-0.992	330	4	2.78	880	-0.264	-0.673	--
23	2008-07-02	2.72	0.301	-0.0258	-1	520	2	2.59	567	0.125	0.26	--
24	2008-07-09	2.58	0.301	-0.146	-0.989	380	2	2.64	631	-0.0581	-0.135	--
25	2008-07-30	2.51	0.699	-0.486	-0.874	320	5	3.06	1690	-0.56	-1.7	--
26	2008-08-06	2.34	0	-0.588	-0.809	220	1	2.45	414	-0.112	-0.235	--
27	2008-08-19	2.54	0	-0.753	-0.658	350	1	2.43	392	0.113	0.235	--
28	2008-08-28	2.74	0	-0.845	-0.534	550	1	2.39	360	0.347	0.945	--
29	2008-09-11	2.34	0.477	-0.948	-0.317	220	3	2.72	762	-0.377	-1.03	--
30	2008-09-25	2.11	0	-0.997	-0.0817	130	1	2.17	216	-0.0576	-0.111	--
31	2009-03-26	1.38	-0.155	0.996	0.0903	24	0.7	1.04	16	0.338	0.908	--
32	2009-03-30	1.62	0.301	1	0.0215	42	2	1.48	44.4	0.139	0.39	--
33	2009-04-06	2.2	0.38	0.995	-0.0988	160	2.4	1.63	62.4	0.571	1.52	--
34	2009-04-21	2.38	0.204	0.937	-0.35	240	1.6	1.66	67.1	0.716	1.81	--
35	2009-05-06	2.59	0.785	0.817	-0.577	390	6.1	2.37	344	0.217	0.673	--
36	2009-05-14	1.6	0.398	0.73	-0.684	40	2.5	2.14	201	-0.538	-1.52	--
37	2009-06-02	2.32	0.114	0.471	-0.882	210	1.3	2.13	197	0.191	0.498	--
38	2009-06-15	2.66	1.15	0.264	-0.965	460	14	3.18	2220	-0.52	-1.38	--
39	2009-06-24	2.3	0.82	0.112	-0.994	200	6.6	2.98	1400	-0.681	-1.95	--
40	2009-07-15	2.49	0.398	-0.247	-0.969	310	2.5	2.75	827	-0.264	-0.643	--
41	2009-07-29	2.8	0.886	-0.471	-0.882	630	7.7	3.23	2460	-0.429	-1.11	--
42	2009-08-13	4.46	2.23	-0.681	-0.732	29000	170	4.41	37100	0.0554	0.135	--
43	2009-08-17	2.38	0.322	-0.73	-0.684	240	2.1	2.72	764	-0.34	-0.871	--
44	2009-08-27	2.49	0.23	-0.836	-0.549	310	1.7	2.6	580	-0.11	-0.21	--
45	2009-09-08	2.2	0.114	-0.931	-0.366	160	1.3	2.42	386	-0.22	-0.554	--
46	2009-09-15	2.43	0.505	-0.968	-0.251	270	3.2	2.71	747	-0.28	-0.704	--
47	2010-03-18	1.65	0.748	0.974	0.226	45	5.6	1.76	83.3	-0.105	-0.185	--
48	2010-03-29	2.74	1.2	0.999	0.0387	550	16	2.27	269	0.473	1.26	--
49	2010-04-08	1.46	0.398	0.991	-0.133	29	2.5	1.67	68.3	-0.21	-0.498	--
50	2010-04-21	1.82	0.342	0.937	-0.35	66	2.2	1.79	88.8	0.0338	0.0614	--
51	2010-05-12	3.54	0.978	0.753	-0.658	3500	9.5	2.62	610	0.921	2.52	--
52	2010-05-20	2.15	0.0792	0.655	-0.755	140	1.2	1.94	126	0.208	0.554	--
53	2010-05-27	2.26	0.633	0.56	-0.829	180	4.3	2.51	475	-0.259	-0.613	--
54	2010-06-09	2.68	0.301	0.362	-0.932	480	2	2.38	346	0.305	0.836	--
55	2010-06-16	3.08	0.38	0.247	-0.969	1200	2.4	2.52	482	0.559	1.44	--
56	2010-06-28	3.72	0.996	0.043	-0.999	5300	9.9	3.17	2150	0.553	1.38	--
57	2010-07-08	2.53	0	-0.129	-0.992	340	1	2.37	339	0.164	0.416	--
58	2010-07-19	2.95	0.462	-0.313	-0.95	900	2.9	2.83	980	0.126	0.312	--
59	2010-07-26	3.52	0.255	-0.425	-0.905	3300	1.8	2.67	677	0.851	1.95	--
60	2010-08-12	2.2	0.23	-0.668	-0.744	160	1.7	2.65	651	-0.447	-1.26	--
61	2010-08-24	3	0.255	-0.806	-0.591	1000	1.8	2.64	630	0.363	1.03	--
62	2010-08-30	2.6	0.204	-0.863	-0.505	400	1.6	2.56	530	0.0402	0.0859	--
63	2010-09-09	2.32	0.23	-0.937	-0.35	210	1.7	2.52	480	-0.196	-0.471	--
64	2010-09-21	2.15	0.279	-0.989	-0.15	140	1.9	2.46	416	-0.311	-0.802	--
65	2012-03-06	1.32	0.322	0.907	0.421	21	2.1	1.29	28.3	0.0329	0.0368	--
66	2012-04-04	1.83	0.23	0.998	-0.0645	67	1.7	1.48	43.7	0.348	0.985	--
67	2012-05-02	2.08	0.362	0.854	-0.52	120	2.3	1.95	129	0.131	0.363	--
68	2012-06-06	2	0.332	0.409	-0.912	100	2.15	2.37	341	-0.37	-0.985	--
69	2012-07-02	2.18	-0.0223	-0.0258	-1	150	0.95	2.31	295	-0.131	-0.286	--
70	2012-08-01	3.08	-0.0458	-0.516	-0.857	1200	0.9	2.41	376	0.667	1.6	--
71	2012-09-04	3.34	0.58	-0.903	-0.429	2200	3.8	2.86	1060	0.481	1.32	--
72	2012-10-10	2.26	0.146	-0.984	0.176	180	1.4	2.13	196	0.125	0.286	--
73	2013-03-04	0.477	0.322	0.892	0.452	3	2.1	1.28	27.5	-0.799	-2.15	--
74	2013-04-02	1.96	0.342	1	-0.0301	92	2.2	1.55	52.1	0.41	1.21	--
75	2013-05-02	1.68	0.255	0.854	-0.52	48	1.8	1.85	104	-0.173	-0.443	--
76	2013-06-05	2.2	0.342	0.425	-0.905	160	2.2	2.37	339	-0.163	-0.416	--
77	2013-06-07	3.9	1.64	0.394	-0.919	8000	44	3.53	4970	0.369	1.07	--
78	2013-07-02	3.94	1.78	-0.0258	-1	8800	60	3.89	11300	0.055	0.111	--
79	2013-07-16	2.68	0.869	-0.264	-0.965	480	7.4	3.17	2170	-0.493	-1.32	--
80	2013-08-05	2.61	0.519	-0.574	-0.819	410	3.3	2.91	1180	-0.297	-0.768	--

81	2013-09-05	2.43	0.114	-0.911	-0.413	270	1.3	2.45	405	-0.0138	-0.0368	--
82	2013-10-17	2.3	0.322	-0.956	0.293	200	2.1	2.2	229	0.103	0.21	--
83	2014-03-24	0.602	0.431	0.992	0.124	4	2.7	1.54	50	-0.935	-2.52	--
84	2014-04-14	1.18	0.114	0.972	-0.234	15	1.3	1.5	45.5	-0.32	-0.836	--
85	2014-05-07	1.9	0.398	0.806	-0.591	80	2.5	2.05	162	-0.143	-0.312	--
86	2014-06-02	2.08	0.322	0.471	-0.882	120	2.1	2.31	300	-0.235	-0.583	--
87	2014-07-07	2.45	-0.0458	-0.112	-0.994	280	0.9	2.32	304	0.126	0.337	--
88	2014-08-13	3.4	0.74	-0.681	-0.732	2500	5.5	3.1	1820	0.301	0.802	--
89	2014-08-19	2.41	0.146	-0.753	-0.658	260	1.4	2.56	528	-0.145	-0.337	--
90	2014-09-09	2.45	-0.222	-0.937	-0.35	280	0.6	2.12	192	0.326	0.871	--
91	2014-10-01	2.32	0.0414	-1	0.0215	210	1.1	2.14	202	0.179	0.471	--
92	2015-03-25	1.59	1.08	0.994	0.107	39	12	2.12	190	-0.525	-1.44	--
93	2015-04-21	3.6	1.43	0.937	-0.35	4000	27	2.74	804	0.859	2.15	--
94	2015-05-12	1.86	0.114	0.753	-0.658	72	1.3	1.86	106	-0.00636	-0.0123	--
95	2015-06-02	3.45	0.82	0.471	-0.882	2800	6.6	2.75	820	0.696	1.7	--
96	2015-06-29	3.3	0.944	0.0258	-1	2000	8.8	3.13	1980	0.167	0.443	--
97	2015-07-21	2.46	0.544	-0.346	-0.938	290	3.5	2.91	1180	-0.445	-1.21	--
98	2015-07-30	2.49	0.38	-0.486	-0.874	310	2.4	2.78	886	-0.293	-0.736	--
99	2015-08-12	2.59	0.19	-0.668	-0.744	390	1.55	2.62	600	-0.0247	-0.0614	--
100	2015-08-27	2.68	0.0792	-0.836	-0.549	480	1.2	2.47	427	0.213	0.643	--
101	2015-09-09	2.62	0.255	-0.937	-0.35	420	1.8	2.54	504	0.0829	0.16	--
102	2015-09-14	2.45	0.176	-0.963	-0.268	280	1.5	2.43	392	0.0168	0.0123	--

Definitions

FC: Fecal coliforms in cfu/100ml (31625)
Turb: Turbidity in FNU (63680)
sin2piJD: sin(2πJD/365) in day/days
cos2piJD: cos(2πJD/365) in day/days

App Version 1.0