

Appendix 28. Model Archival Summary for Fecal Coliform Concentration at Station 06892350; Kansas River at De Soto, Kansas

This model archival summary summarizes the fecal coliform concentration (Fcoli) model developed to compute 15-minute Fcoli from July 19, 2012 onward. This model supersedes all previous models.

Site and Model Information

Site number: 06892350

Site name: Kansas River at De Soto, Kansas

Location: Lat 38°59'00", long 94°57'52" referenced to North American Datum of 1927, in NE 1/4 SE 1/4 sec.28, T.12 S., R.22 E., Leavenworth County, KS, Hydrologic Unit 10270104.

Equipment: An YSI 6600 water-quality monitor equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, turbidity, and chlorophyll was installed from August 2012 through May 2014. From June 2014 to the present (2015) a Xylem YSI EXO2 water-quality monitor equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, turbidity, and chlorophyll. The monitor is housed in a 4-inch diameter galvanized steel pipe. Readings from the water-quality monitor are recorded every 15 minutes and transmits data by way of satellite, hourly.

Date model was created: October 15, 2015

Model calibration data period: July 19, 2012 – June 29, 2015

Model application date: July 19, 2012 onward

Model-Calibration Dataset

All data were collected using U.S. Geological Survey (USGS) protocols and are stored in the National Water Information System (NWIS) database. Linear regression models were developed using the open-source software package “R.” Explanatory variables selected as inputs to linear regression were physicochemical properties: specific conductance, pH, water temperature, dissolved oxygen, turbidity, chlorophyll fluorescence, and streamflow. Seasonal components (sine and cosine variables) were also evaluated as explanatory variables in the models to determine if seasonal changes affected the model. All combinations of physicochemical properties and a seasonal component were evaluated to determine which combinations produced the best models.

The final selected regression model is based on 48 concurrent measurements of Fcoli concentration and turbidity (Turb) collected from July 19, 2012 through June 29, 2015. Samples were collected throughout the range of continuously observed hydrologic conditions. One sample was below the laboratory detection limits of 1.0. Twenty samples were estimated due to colony counts outside of the ideal range. Summary statistics and the complete model-calibration dataset are provided below. Studentized residuals from the final model were inspected for values greater than 3 or less than negative 3. Values outside of that range are considered potential outliers and are investigated. One sample, June 11, 2014, was found to have potential errors in collection and processing, and has been removed from the dataset. All other potential outliers were not found to have errors associated with collection, processing, or analysis, and were therefore considered valid.

Fecal Coliform Sampling Details

Cross-section samples are typically collected either from the downstream side of the bridge or instream within 100 feet of the bridge. The equal-width-increment (EWI) method is used, and samples typically are composited for analysis. Cross-section samples are collected every 2 weeks from March through October, once a month from November through February, and during selected runoff events. A FISP US DH-95, D-95, or D-96A1 depth integrating sampler is used from the bridge; and a DH-81 or DH-95 hand sampler is used for boat samples. Samples are analyzed for Fcoli concentration at the USGS Kansas Water Science Center (KSWSC) Lab in Lawrence, Kansas.

Model Development

Regression analysis was done using R by examining Turb, streamflow, and other continuously measured data as explanatory variables for estimating Fcoli concentration. A variety of models that predict Fcoli, $(Fcoli)^2$, \sqrt{Fcoli} and models that predict $\log_{10}(Fcoli)$ 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 Fcoli 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 $\log_{10}(\text{Fcoli})$.

Turb was and seasonality were selected as the best predictors of Fcoli based on residual plots, relatively high adjusted coefficient of determination (adjusted R^2) and relatively low model standard percentage error ($MSPE$), prediction error sum of squares (PRESS), and Mallows's C_p . Values for all of the afore mentioned statistics and metrics were computed for various models and are included below along with all relevant sample data and more in-depth statistical information.

Model Summary

Summary of final regression analysis for Fcoli concentration at site number 06892350.

Fcoli concentration-based model:

$$\log_{10}(\text{Fcoli}) = 0.877 \times \log_{10}(\text{Turb}) - 0.228 \times \sin\left(\frac{2\pi DY}{365}\right) - 0.59 \times \cos\left(\frac{2\pi DY}{365}\right) + 0.481$$

where

Fcoli = fecal coliform in colonies per 100 milliliter (col/mL);

Turb = turbidity in formazin nephelometric units (FNU);

Sin \& Cos = seasonality component; and,

DY = day of the year.

Turbidity and seasonality makes physical and statistical sense as explanatory variables for Fcoli.

The log-transformed model may be retransformed to the original units so that Fcoli 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). For this model, the calculated BCF is 1.86. The retransformed model, accounting for BCF is:

$$\text{Fcoli} = \frac{\text{Turb}^{0.877} \times 5.63}{10^{0.228 \sin\left(\frac{2\pi DY}{365}\right)} \times 10^{0.59 \cos\left(\frac{2\pi DY}{365}\right)}}$$

Previous Models

Start year	End year	Model
2000	2005	$\log_{10}(\text{Fcoli}) = 1.53 \times \log_{10}(\text{Turb}) - 1.05$

Fecal Coliform Concentration Record

The Fcoli record is computed using this regression model and stored at the National Real-Time Water Quality (NRTWQ) Web site. Data are computed at 15-minute intervals. The complete water-quality record can be found at <http://nrtwq.usgs.gov/ks>.

Remarks

None

R Output for Fecal coliform; 06892350; Kansas River at De Soto, KS

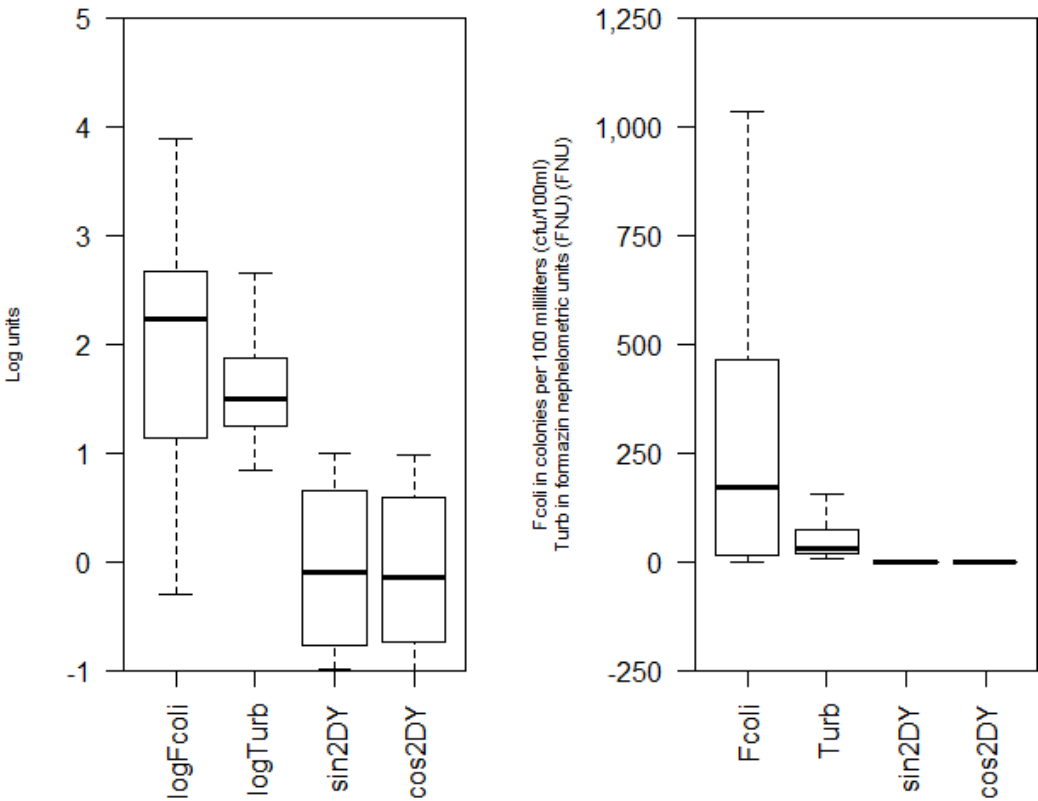
Model Form

$$\log F_{\text{coli}} = +0.877 * \log T_{\text{urb}} + -0.228 * \sin 2DY + -0.59 * \cos 2DY + 0.481$$

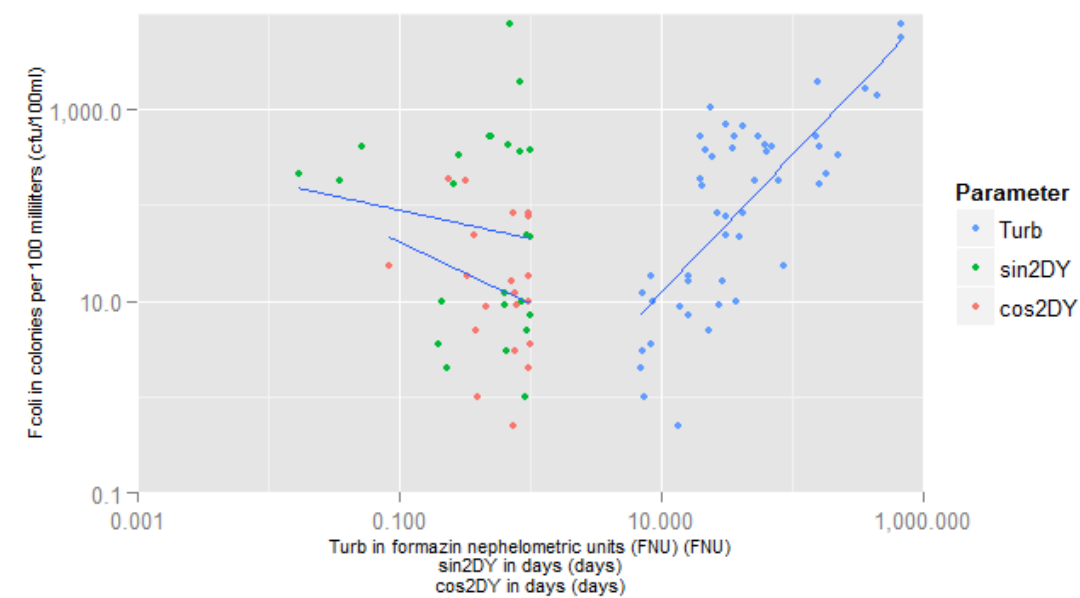
Variable Summary Statistics

	logFcoli	logTurb	sin2DY	cos2DY	Fcoli	Turb
Minimum	-0.301	0.845	-0.997	-1	0.5	7
1st Quartile	1.14	1.25	-0.768	-0.742	14	17.9
Median	2.24	1.49	-0.094	-0.14	172	31
Mean	1.95	1.6	-0.0708	-0.0953	569	90.1
3rd Quartile	2.67	1.87	0.647	0.59	465	73.5
Maximum	3.88	2.83	0.997	0.981	7670	674
Censored [< 1]	1	--	--	--	1	--

Box Plot(s) of sample data



Exploratory Plot



Model Calibration

Basic Data

Number of Observations	48
Standard error (RMSE)	0.573
Upper Model standard percentage error (MSPE)	274
Lower Model standard percentage error (MSPE)	73.2
Coefficient of determination (R²)	0.688
Adjusted Coefficient of Determination (Adj. R²)	0.666
Bias Correction Factor (BCF)	1.86

Variance Inflation Factors (VIF)

logTurb	sin2DY	cos2DY
1.896431	1.004927	1.891782

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(> t)
(Intercept)	0.481	0.350	1.38	0.176000
logTurb	0.877	0.220	4.00	0.000243
sin2DY	-0.228	0.116	-1.96	0.056600
cos2DY	-0.590	0.164	-3.59	0.000819

Correlation Matrix

	Intercept	logTurb	sin2DY	cos2DY
Intercept	1.0000	-0.971	-0.0444	-0.6430
logTurb	-0.9710	1.000	0.0700	0.6870
sin2DY	-0.0444	0.070	1.0000	0.0495
cos2DY	-0.6430	0.687	0.0495	1.0000

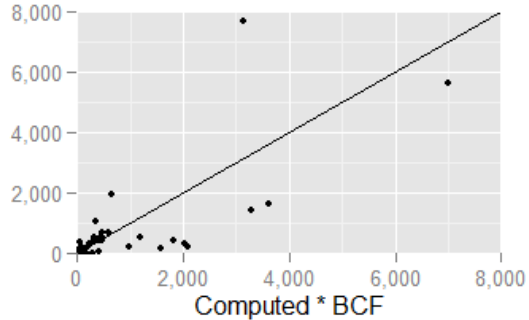
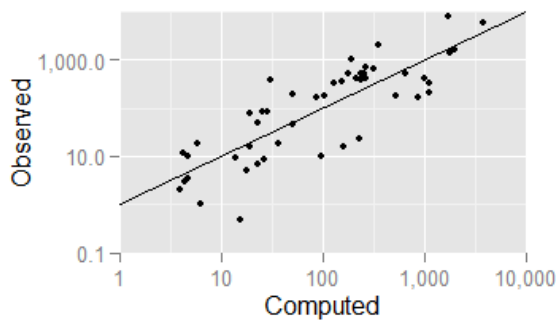
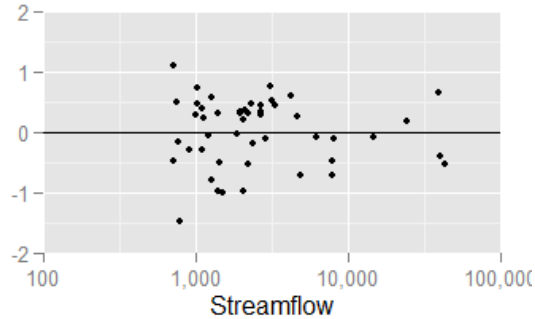
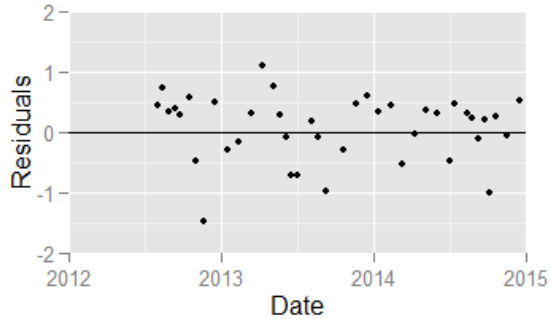
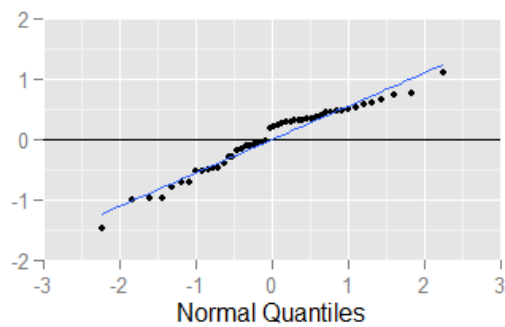
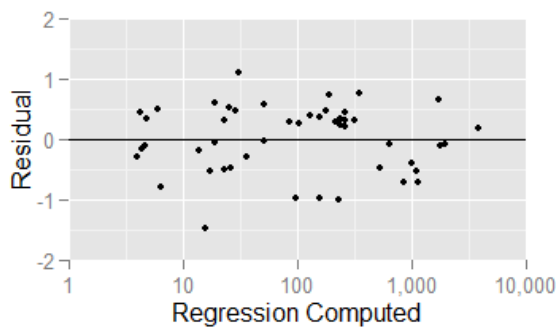
Test Criteria

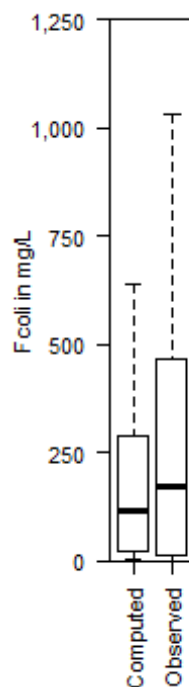
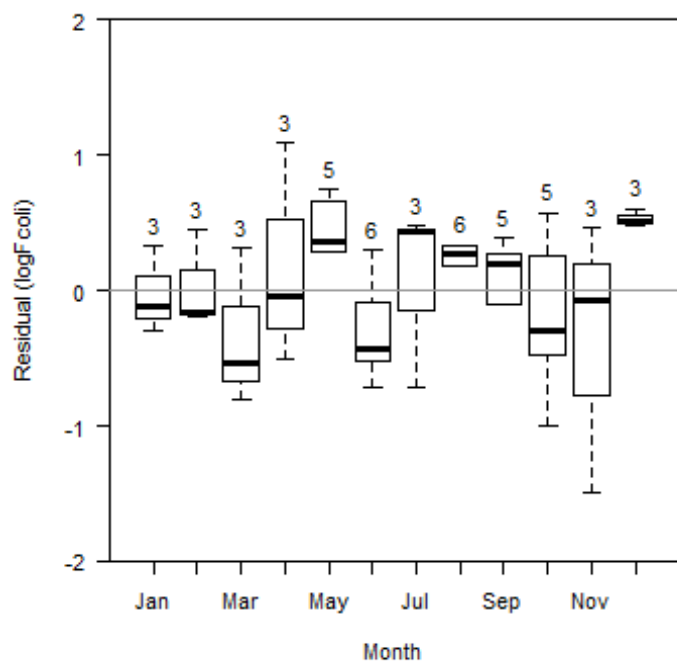
Leverage	Cook's D	DFFITS
0.18750	0.26323	0.50000

Flagged Observations

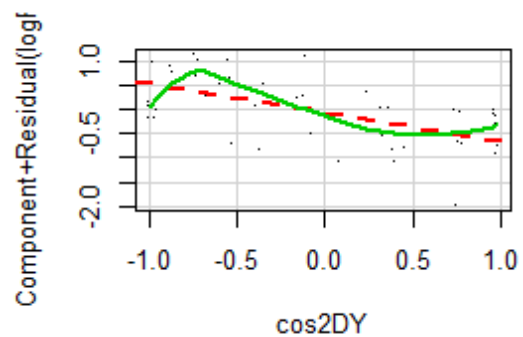
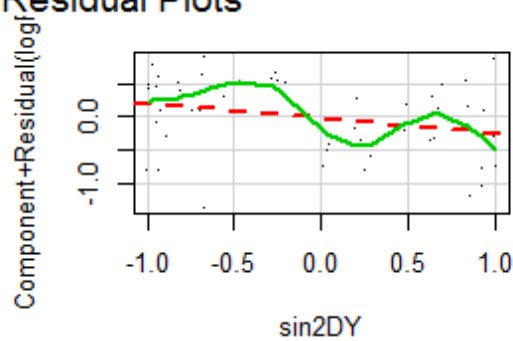
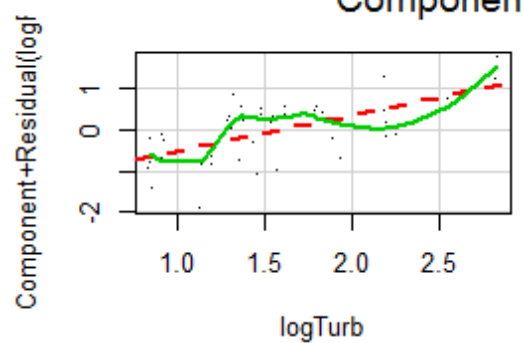
	logFcoli	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
11/19/2012 11:30	-0.301	1.185	-1.4860	-2.687	-2.905	0.06653	0.12860	-0.7756
4/8/2013 11:00	2.580	1.486	1.0940	1.987	2.059	0.07541	0.08051	0.5880
9/9/2013 10:30	1.204	2.193	-0.9893	-1.791	-1.839	0.06899	0.05940	-0.5005
10/6/2014 13:30	1.362	2.359	-0.9969	-1.816	-1.866	0.08045	0.07210	-0.5520
5/4/2015 13:30	1.000	1.979	-0.9791	-1.773	-1.819	0.07031	0.05947	-0.5004
5/18/2015 15:30	3.885	3.226	0.6587	1.280	1.290	0.19230	0.09748	0.6291

Statistical Plots





Component + Residual Plots



EXPLANATION

- Non-parametric smoothed fit
- Linear best fit

Models considered

Model Formula	Number of Variables	Standard Error	R2	Adjusted R2	Cp	PRESS	VIF	MSPE
logFcoli ~ logTurb	1	0.6544	57.35	56.42	13.22	21.04	1 ± 210	
logFcoli ~ cos2DY	1	0.6806	53.86	52.86	17.9	23.13	1 ± 230	
logFcoli ~ Turb	1	0.8037	35.67	34.27	42.31	31.71	1 ± 310	
logFcoli ~ logTurb + cos2DY	2	0.5903	66.05	64.54	3.553	17.67	1.887 ± 180	
logFcoli ~ Turb + cos2DY	2	0.6116	63.56	61.94	6.889	18.46	1.228 ± 190	
logFcoli ~ logTurb + sin2DY	2	0.6439	59.6	57.8	12.2	20.96	1.002 ± 210	
logFcoli ~ logTurb + sin2DY + cos2DY	3	0.5726	68.77	66.64	1.903	17.11	1.896 ± 170	
logFcoli ~ Turb + sin2DY + cos2DY	3	0.5902	66.82	64.55	4.52	17.84	1.229 ± 180	
logFcoli ~ logQ + logTurb + cos2DY	3	0.5922	66.58	64.31	4.829	18.13	2.904 ± 180	
logFcoli ~ Turb + logTurb + sin2DY + cos2DY	4	0.5766	69.04	66.16	3.53	17.68	3.755 ± 180	
logFcoli ~ Q + logTurb + sin2DY + cos2DY	4	0.5779	68.91	66.02	3.707	17.86	1.799 ± 180	
logFcoli ~ logQ + logTurb + sin2DY + cos2DY	4	0.5781	68.88	65.99	3.748	17.93	3.091 ± 180	
logFcoli ~ Q + Turb + logTurb + sin2DY + cos2DY	5	0.5797	69.44	65.8	5	17.99	2.098 ± 180	
logFcoli ~ logQ + Turb + logTurb + sin2DY + cos2DY	5	0.5813	69.27	65.62	5.222	18.37	3.275 ± 180	
logFcoli ~ Q + logQ + logTurb + sin2DY + cos2DY	5	0.5846	68.92	65.22	5.699	19.03	4.341 ± 180	
logFcoli ~ Q + logQ + Turb + logTurb + sin2DY + cos2DY	6	0.5867	69.44	64.97	7	19.15	4.813 ± 180	

Data

	Date	logFcoli	logTurb	sin2DY	cos2DY	Fcoli	Turb	Computed logFcoli	Computed Fcoli	Residual	Normal Quantiles	Censored Values
0												
1	2012-07-30	2.845	1.488	-0.4705	-0.8824	700	30.79	2.414	483.9	0.431	0.704	--
2	2012-08-13	3.014	1.371	-0.6656	-0.7463	1033	23.5	2.276	351.6	0.738	1.61	--
3	2012-08-27	2.708	1.556	-0.823	-0.568	510	36	2.369	435.9	0.339	0.516	--
4	2012-09-10	2.505	1.38	-0.9355	-0.3533	320	24	2.114	242.1	0.392	0.639	--
5	2012-09-24	2.204	1.314	-0.9928	-0.1195	160	20.59	1.93	158.8	0.274	0.183	--
6	2012-10-15	2.279	1.301	-0.9712	0.2384	190	20	1.703	94.13	0.575	1.19	--
7	2012-10-29	0.9445	1.146	-0.888	0.4598	8.8	14	1.418	48.79	-0.474	-0.772	--
8	2012-11-19	-0.301	1.13	-0.6668	0.7452	0.5	13.5	1.185	28.56	-1.49	-2.24	< 1
9	2012-12-17	1.255	0.9165	-0.2467	0.9691	18	8.25	0.77	10.97	0.485	1	--
10	2013-01-14	0.301	0.8451	0.2296	0.9733	2	7	0.5963	7.356	-0.295	-0.516	--
11	2013-02-11	0.4771	0.8579	0.6544	0.7562	3	7.21	0.6387	8.11	-0.162	-0.4	--
12	2013-03-11	1.681	1.495	0.9298	0.3681	48	31.28	1.364	43.07	0.317	0.4	--
13	2013-04-08	2.58	1.332	0.9944	-0.106	380	21.5	1.486	57.03	1.09	2.24	--
14	2013-05-06	3.284	2.188	0.8314	-0.5557	1925	154	2.538	643.6	0.746	1.84	--
15	2013-05-20	2.623	1.788	0.675	-0.7378	420	61.4	2.331	399.3	0.292	0.236	--
16	2013-06-03	2.724	2.184	0.4786	-0.878	530	152.9	2.806	1192	-0.0817	-0.236	--
17	2013-06-17	2.215	2.204	0.2557	-0.9668	164	160	2.927	1573	-0.712	-1.09	--
18	2013-07-01	2.326	2.258	0.01697	-0.9999	212	181	3.047	2078	-0.721	-1.19	--
19	2013-08-05	3.751	2.822	-0.5515	-0.8341	5642	663.3	3.574	6991	0.177	-0.026	--
20	2013-08-19	3.212	2.551	-0.734	-0.6792	1629	356	3.287	3611	-0.0755	-0.183	--
21	2013-09-09	1.204	1.459	-0.9269	-0.3754	16	28.75	2.193	290.9	-0.989	-1.61	--
22	2013-10-21	1.255	1.204	-0.9434	0.3318	18	16	1.557	67.19	-0.302	-0.576	--
23	2013-11-18	1.919	1.422	-0.6797	0.7335	83	26.4	1.451	52.62	0.468	0.845	--
24	2013-12-16	1.886	1.489	-0.2664	0.9639	77	30.8	1.28	35.47	0.607	1.31	--
25	2014-01-13	1	0.9294	0.2107	0.9775	10	8.5	0.6721	8.758	0.328	0.458	--
26	2014-02-10	1.079	0.8482	0.6397	0.7687	12	7.05	0.6262	7.879	0.453	0.772	--
27	2014-03-10	0.699	1.362	0.9245	0.3812	5	23	1.24	32.39	-0.541	-1	--
28	2014-04-07	1.663	1.591	0.996	-0.0893	46	39	1.702	93.93	-0.0397	-0.0781	--
29	2014-05-05	2.556	1.803	0.8402	-0.5422	360	63.57	2.191	289.4	0.365	0.576	--
30	2014-06-02	2.708	1.739	0.4938	-0.8696	510	54.83	2.407	475.5	0.301	0.29	--
31	2014-06-30	2.255	1.892	0.0349	-0.9994	180	78	2.722	983.2	-0.467	-0.704	--
32	2014-07-14	2.716	1.294	-0.205	-0.9788	520	19.67	2.24	323.8	0.476	0.921	--
33	2014-08-11	2.82	1.621	-0.6366	-0.7712	661	41.8	2.503	593.6	0.317	0.345	--
34	2014-08-25	2.591	1.539	-0.8019	-0.5974	390	34.6	2.366	433.3	0.225	0.0781	--
35	2014-09-08	3.146	2.652	-0.9209	-0.3899	1400	449	3.248	3297	-0.102	-0.29	--
36	2014-09-22	2.613	1.839	-0.987	-0.1607	410	69	2.414	483.6	0.199	0.026	--
37	2014-10-06	1.362	1.936	-0.9966	0.08203	23	86.33	2.359	425.5	-0.997	-1.84	--

38	2014-10-20	2.262	1.711	-0.9481	0.3179	183	51.4	2.011	191.1	0.251	0.13	--
39	2014-11-17	1.204	1.204	-0.6933	0.7207	16	16	1.271	34.75	-0.0666	-0.13	--
40	2014-12-15	1.919	1.623	-0.2816	0.9595	83	41.99	1.404	47.2	0.515	1.09	--
41	2015-01-12	0.5441	0.9227	0.1963	0.9805	3.5	8.37	0.6677	8.67	-0.124	-0.345	--
42	2015-02-09	0.9542	1.439	0.6298	0.7768	9	27.5	1.142	25.86	-0.188	-0.458	--
43	2015-03-09	0	0.8692	0.9174	0.398	1	7.4	0.7998	11.75	-0.8	-1.31	--
44	2015-04-06	0.8451	1.206	0.9972	-0.07441	7	16.07	1.356	42.26	-0.511	-0.845	--
45	2015-05-04	1	1.574	0.8496	-0.5274	10	37.49	1.979	177.6	-0.979	-1.44	--
46	2015-05-18	3.885	2.829	0.6982	-0.7159	7667	674.5	3.226	3135	0.659	1.44	--
47	2015-06-15	2.519	2.348	0.2872	-0.9579	330	223	3.041	2046	-0.522	-0.921	--
48	2015-06-29	2.602	2.204	0.05198	-0.9986	400	160	2.992	1828	-0.39	-0.639	--

Definitions and National Water Information System (parameter code)

Fcoli: Fecal coliforms in cfu/100mL (31625)

Turb: Turbidity in FNU (63680)

DY: Date in decimal years