

# **Appendix 31. Model Archival Summary for Actinomyces**

## **Concentration at Station 06892350; Kansas River at De Soto, Kansas**

This model archival summary summarizes the Actinomyces concentration (Act) model developed to compute 15-minute Act 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 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 50 concurrent measurements of Act concentration and turbidity (Turb) collected from July 19, 2012 through June 29, 2015. Samples were collected throughout the range of continuously observed hydrologic conditions. No samples were below laboratory detection limits. Four 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.

### **Actinomyces 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 Act concentration at the Ohio Water Microbiology Laboratory in Columbus, Ohio.

### **Model Development**

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

Turb and seasonality were selected as the best predictors of Act 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 Act concentration at site number 06892350.

Act concentration-based model:

$$\log_{10}(\text{Act}) = 1.54 \times \log_{10}(\text{Turb}) + 0.246 \times \sin\left(\frac{2\pi DY}{365}\right) + 0.194 \times \cos\left(\frac{2\pi DY}{365}\right) - 0.171$$

where

$\text{Act}$  = actinomycetes in colonies per liter (col/mL);

$\text{Turb}$  = turbidity in formazin nephelometric units (FNU);

$\text{Sin \& Cos}$  = seasonality component; and,

$DY$  = day of the year.

Turbidity and the day of the year make physical and statistical sense as explanatory variables for Turb.

The log-transformed model may be retransformed to the original units so that Turb 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 1.24. The retransformed model, accounting for BCF is:

$$\text{Act} = \text{Turb}^{1.54} \times 10^{0.246 \sin\left(\frac{2\pi DY}{365}\right)} \times 10^{0.194 \cos\left(\frac{2\pi DY}{365}\right)} \times 0.836$$

## Previous Models

No previous models have been published.

## Actinomycetes Concentration Record

The Act 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 Actinomycetes; 06892350; Kansas River at De Soto, KS

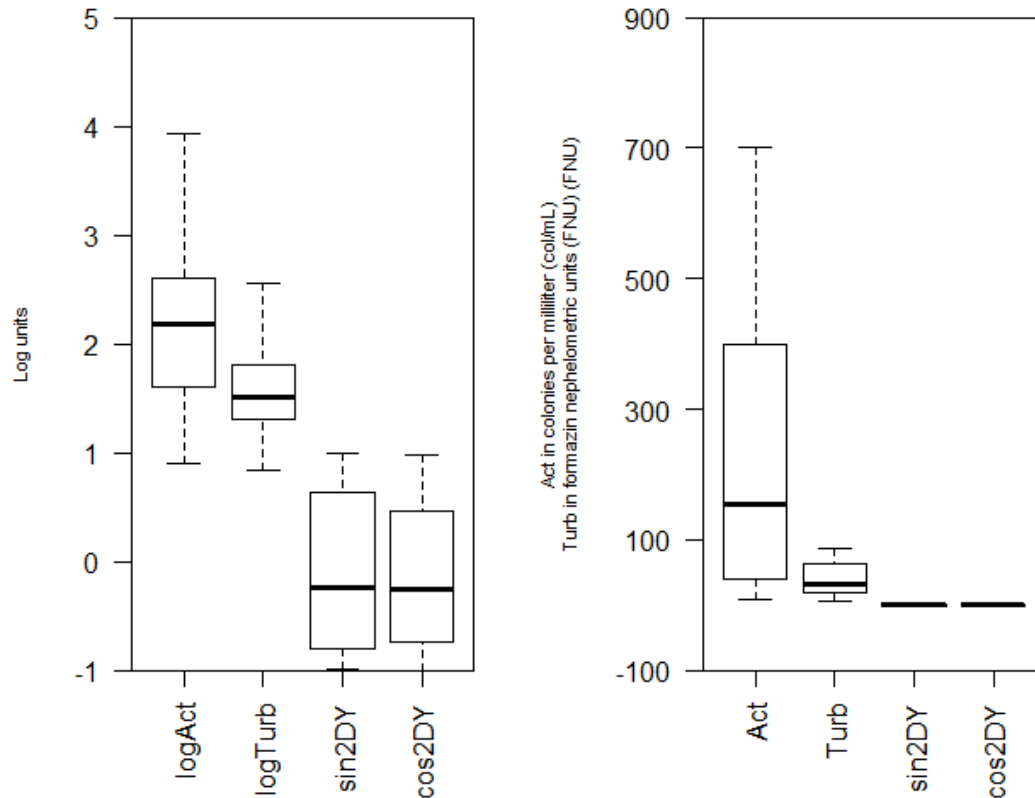
### Model Form

$\log\text{Act} = + 1.54 * \log\text{Turb} + 0.246 * \sin 2\text{DY} + 0.194 * \cos 2\text{DY} + -0.171$

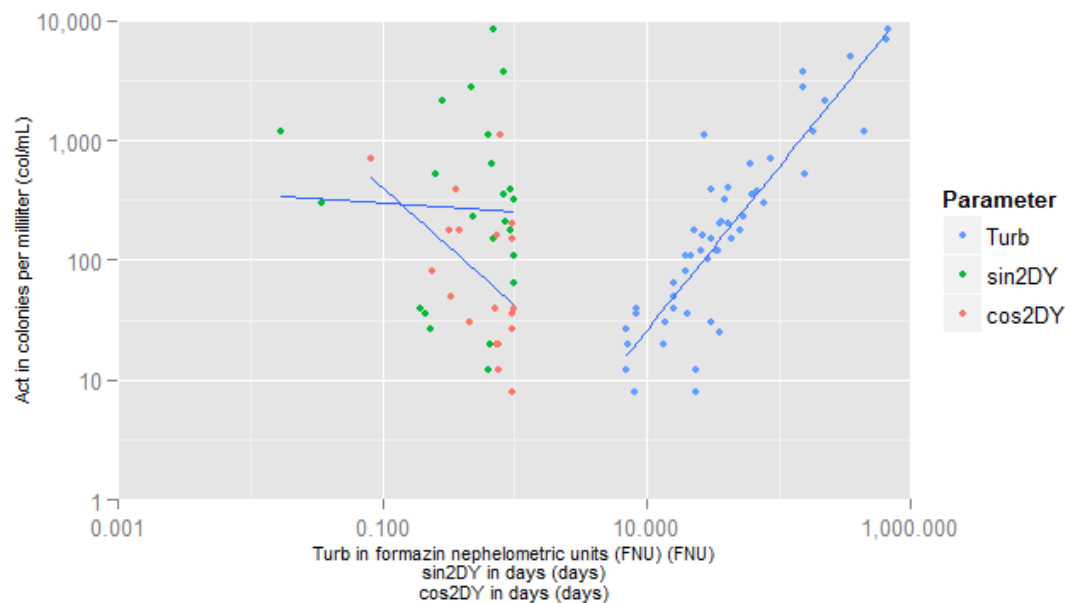
### Variable Summary Statistics

	logAct	logTurb	sin2DY	cos2DY	Act	Turb
Minimum	0.903	0.845	-0.997	-1.000	8	7.0
1st Quartile	1.600	1.300	-0.802	-0.746	40	20.0
Median	2.190	1.510	-0.233	-0.257	155	32.4
Mean	2.230	1.590	-0.106	-0.134	789	86.0
3rd Quartile	2.600	1.800	0.640	0.460	400	63.6
Maximum	3.930	2.830	0.997	0.981	8500	674.0

### Box Plot(s) of sample data



## Exploratory Plot



## Model Calibration

### Basic Data

Number of Observations	50
Standard error (RMSE)	0.306
Upper Model standard percentage error (MSPE)	102
Lower Model standard percentage error (MSPE)	50.6
Coefficient of determination ( $R^2$ )	0.85
Adjusted Coefficient of Determination (Adj. $R^2$ )	0.84
Bias Correction Factor (BCF)	1.24

### Variance Inflation Factors (VIF)

logTurb	sin2DY	cos2DY
1.676436	1.000965	1.677125

## Explanatory Variables

	Coefficients	Standard Error	t value	Pr(> t )
(Intercept)	-0.171	0.1810	-0.949	3.48e-01
logTurb	1.540	0.1140	13.500	1.45e-17
sin2DY	0.246	0.0615	4.000	2.31e-04
cos2DY	0.194	0.0814	2.380	2.14e-02

## Correlation Matrix

	Intercept	logTurb	sin2DY	cos2DY
Intercept	1.0000	-0.9690	0.0148	-0.5790
logTurb	-0.9690	1.0000	0.0231	0.6350
sin2DY	0.0148	0.0231	1.0000	0.0307
cos2DY	-0.5790	0.6350	0.0307	1.0000

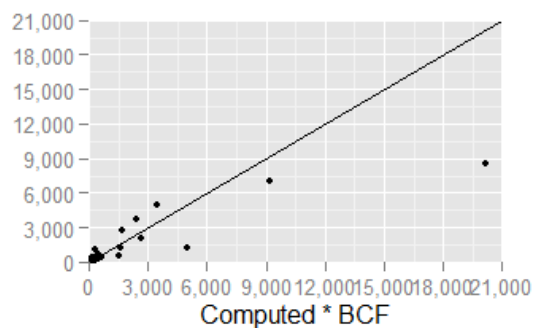
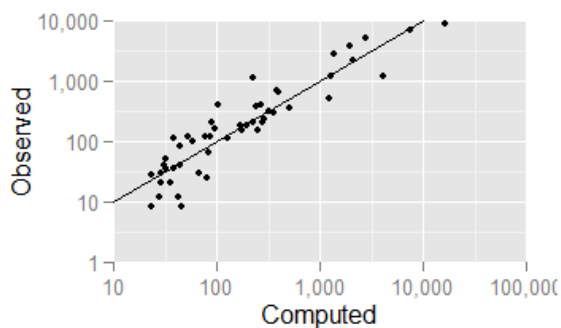
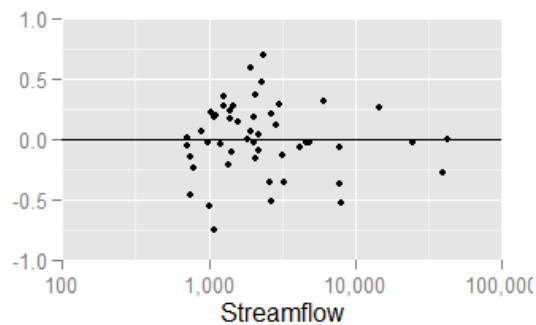
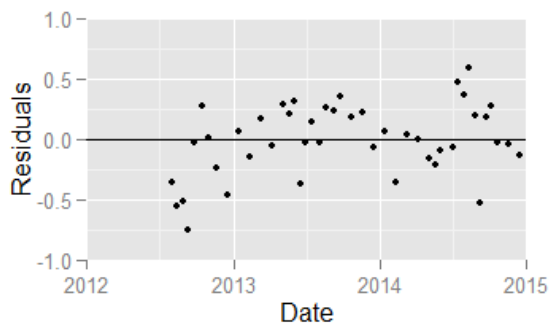
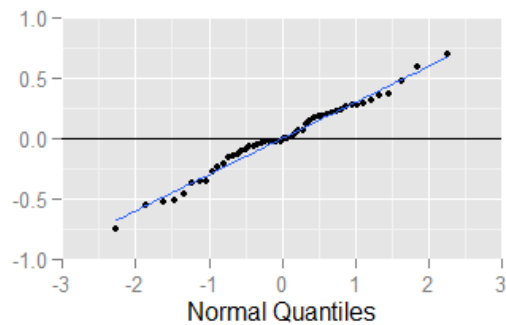
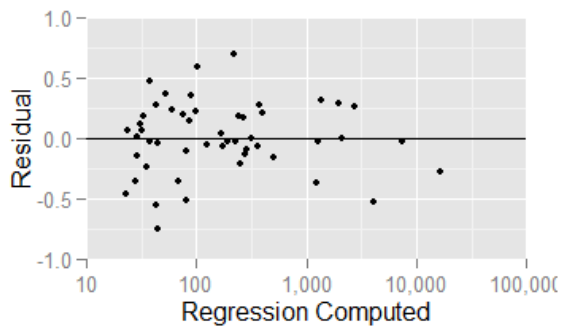
## Test Criteria

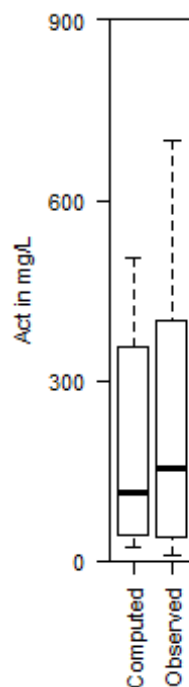
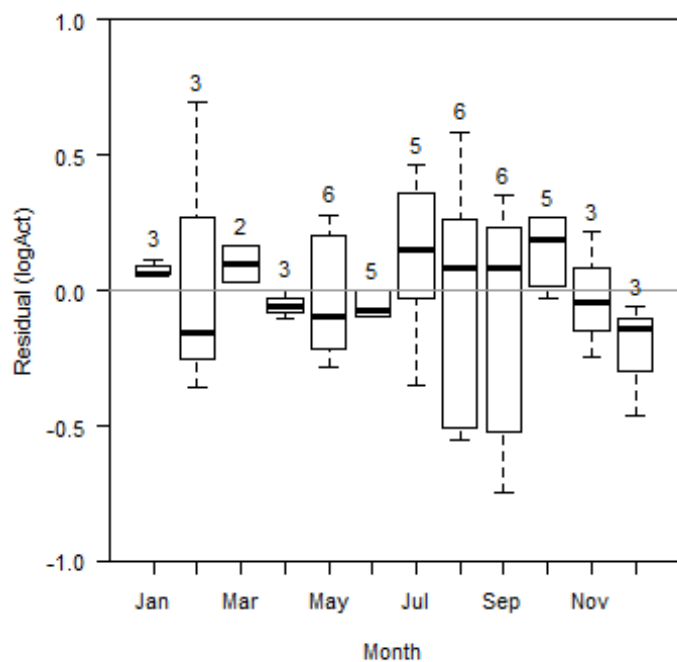
Leverage	Cook's D	DFFITS
0.1800000	0.2633403	0.4898979

## Flagged Observations

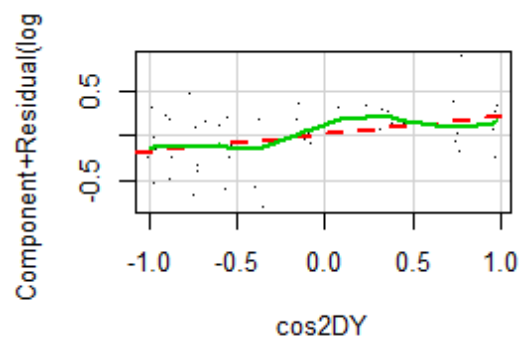
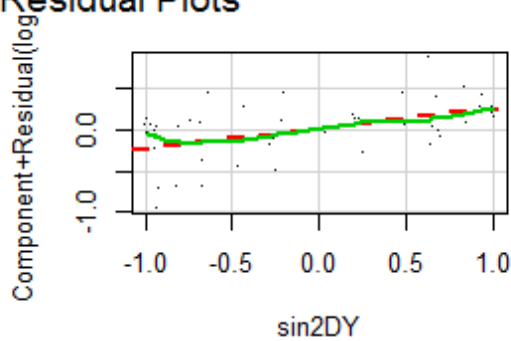
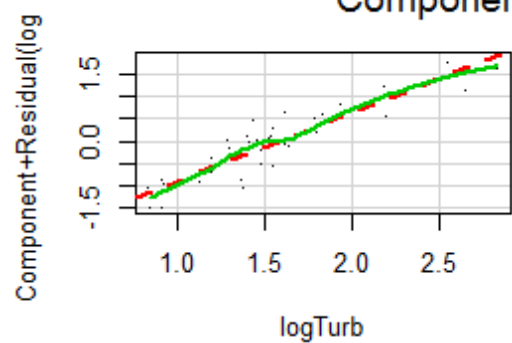
	logAct	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
8/13/2012 11:10	1.0790	1.628	-0.5490	-1.875	-1.930	0.08488	0.08155	-0.5878
9/10/2012 12:10	0.9031	1.652	-0.7490	-2.531	-2.698	0.06468	0.11070	-0.7095
7/14/2014 12:50	2.0410	1.578	0.4638	1.612	1.641	0.11580	0.08509	0.5940
8/11/2014 15:10	2.6020	2.015	0.5871	1.978	2.045	0.05905	0.06135	0.5122
9/8/2014 12:40	3.0790	3.604	-0.5249	-1.884	-1.939	0.17070	0.18260	-0.8798
2/9/2015 13:50	3.0410	2.347	0.6948	2.377	2.511	0.08786	0.13610	0.7792
5/18/2015 15:30	3.9290	4.210	-0.2810	-1.023	-1.023	0.19360	0.06275	-0.5013

## 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
logAct ~ logTurb	1	0.3618	78.07	77.61	21.05	6.73	1 ± 93	
logAct ~ logQ	1	0.5202	54.66	53.72	92.62	13.94	1 ± 150	
logAct ~ Turb	1	0.5372	51.65	50.64	101.8	15.31	1 ± 160	
logAct ~ logTurb + sin2DY	2	0.3209	83.11	82.39	7.652	5.402	1 ± 81	
logAct ~ logTurb + cos2DY	2	0.3514	79.74	78.88	17.94	6.626	1.676 ± 90	
logAct ~ Turb + logTurb	2	0.3642	78.24	77.31	22.53	6.811	3.374 ± 94	
logAct ~ logTurb + sin2DY + cos2DY	3	0.306	84.96	83.98	3.98	5.149	1.676 ± 76	
logAct ~ Turb + logTurb + sin2DY	3	0.3226	83.28	82.19	9.109	5.459	3.374 ± 81	
logAct ~ Q + logTurb + sin2DY	3	0.3236	83.18	82.09	9.417	5.469	1.846 ± 82	
logAct ~ Turb + logTurb + sin2DY + cos2DY	4	0.2992	85.94	84.69	3.001	4.965	3.893 ± 74	
logAct ~ Q + logTurb + sin2DY + cos2DY	4	0.3068	85.22	83.9	5.195	5.308	1.897 ± 77	
logAct ~ logQ + logTurb + sin2DY + cos2DY	4	0.3084	85.06	83.73	5.672	5.268	3.138 ± 77	
logAct ~ Q + Turb + logTurb + sin2DY + cos2DY	5	0.3026	85.94	84.34	5	5.141	2.607 ± 75	
logAct ~ logQ + Turb + logTurb + sin2DY + cos2DY	5	0.3026	85.94	84.34	5.001	5.105	3.496 ± 75	
logAct ~ Q + logQ + logTurb + sin2DY + cos2DY	5	0.3102	85.22	83.54	7.186	5.493	3.722 ± 78	
logAct ~ Q + logQ + Turb + logTurb + sin2DY + cos2DY	6	0.3061	85.94	83.97	7	5.333	4.618 ± 76	

## Data

	Date	logAct	logTurb	sin2DY	cos2DY	Act	Turb	Computed	Computed	Residual	Normal
0								logAct	Act		Quantiles
1	2012-07-30	1.477	1.488	-0.4705	-0.8824	30	30.79	1.83	84.17	-0.353	-1.03
2	2012-08-13	1.079	1.371	-0.6656	-0.7463	12	23.5	1.628	52.87	-0.549	-1.85
3	2012-08-27	1.398	1.556	-0.823	-0.568	25	36	1.909	100.9	-0.511	-1.46
4	2012-09-10	0.9031	1.38	-0.9355	-0.3533	8	24	1.652	55.86	-0.749	-2.26
5	2012-09-24	1.556	1.314	-0.9928	-0.1195	36	20.59	1.581	47.43	-0.0247	-0.075
6	2012-10-15	1.903	1.301	-0.9712	0.2384	80	20	1.636	53.86	0.267	0.949
7	2012-10-29	1.477	1.146	-0.888	0.4598	30	14	1.462	36.02	0.0156	0.125
8	2012-11-19	1.301	1.13	-0.6668	0.7452	20	13.5	1.547	43.84	-0.246	-0.873
9	2012-12-17	0.9031	0.9165	-0.2467	0.9691	8	8.25	1.365	28.82	-0.462	-1.33
10	2013-01-14	1.431	0.8451	0.2296	0.9733	27	7	1.373	29.36	0.0586	0.278
11	2013-02-11	1.301	0.8579	0.6544	0.7562	20	7.21	1.455	35.47	-0.154	-0.671
12	2013-03-11	2.591	1.495	0.9298	0.3681	390	31.28	2.427	332.7	0.164	0.438
13	2013-04-08	2.041	1.332	0.9944	-0.106	110	21.5	2.101	157	-0.0594	-0.33
14	2013-05-06	3.568	2.188	0.8314	-0.5557	3700	154	3.288	2416	0.28	1.12
15	2013-05-20	2.806	1.788	0.675	-0.7378	640	61.4	2.6	496	0.206	0.671
16	2013-06-03	3.447	2.184	0.4786	-0.878	2800	152.9	3.134	1694	0.313	1.22
17	2013-06-17	2.716	2.204	0.2557	-0.9668	520	160	3.093	1540	-0.377	-1.22
18	2013-07-01	3.079	2.258	0.01697	-0.9999	1200	181	3.11	1602	-0.0306	-0.125
19	2013-07-15	2.079	1.526	-0.2203	-0.9754	120	33.6	1.932	106.4	0.147	0.384
20	2013-08-05	3.845	2.822	-0.5515	-0.8341	7000	663.3	3.869	9212	-0.0243	-0.025
21	2013-08-19	3.699	2.551	-0.734	-0.6792	5000	356	3.439	3420	0.26	0.873
22	2013-09-09	2	1.459	-0.9269	-0.3754	100	28.75	1.771	73.37	0.229	0.802
23	2013-09-23	2.301	1.554	-0.9896	-0.1439	200	35.84	1.947	110.2	0.354	1.33
24	2013-10-21	1.699	1.204	-0.9434	0.3318	50	16	1.512	40.48	0.187	0.551
25	2013-11-18	2.204	1.422	-0.6797	0.7335	160	26.4	1.989	121.4	0.215	0.735
26	2013-12-16	2.176	1.489	-0.2664	0.9639	150	30.8	2.238	215.4	-0.0622	-0.384
27	2014-01-13	1.556	0.9294	0.2107	0.9775	36	8.5	1.499	39.23	0.0577	0.227
28	2014-02-10	1.079	0.8482	0.6397	0.7687	12	7.05	1.439	34.17	-0.359	-1.12
29	2014-03-10	2.255	1.362	0.9245	0.3812	180	23	2.223	208	0.0322	0.176
30	2014-04-07	2.505	1.591	0.996	-0.0893	320	39	2.502	395.4	0.00316	0.075
31	2014-05-05	2.544	1.803	0.8402	-0.5422	350	63.57	2.702	626.8	-0.158	-0.735
32	2014-05-19	2.176	1.648	0.6874	-0.7263	150	44.5	2.391	306.1	-0.215	-0.802
33	2014-06-02	2.362	1.739	0.4938	-0.8696	230	54.83	2.455	354.7	-0.0931	-0.494

34	2014-06-30	2.477	1.892	0.0349	-0.9994	300	78	2.552	443.9	-0.0751	-0.438
35	2014-07-14	2.041	1.294	-0.205	-0.9788	110	19.67	1.578	47.05	0.464	1.63
36	2014-07-28	2.079	1.411	-0.4333	-0.9012	120	25.75	1.716	64.76	0.363	1.46
37	2014-08-11	2.602	1.621	-0.6366	-0.7712	400	41.8	2.015	128.8	0.587	1.85
38	2014-08-25	2.079	1.539	-0.8019	-0.5974	120	34.6	1.882	94.81	0.197	0.61
39	2014-09-08	3.079	2.652	-0.9209	-0.3899	1200	449	3.604	5001	-0.525	-1.63
40	2014-09-22	2.568	1.839	-0.987	-0.1607	370	69	2.382	299.8	0.186	0.494
41	2014-10-06	2.845	1.936	-0.9966	0.08203	700	86.33	2.576	469	0.269	1.03
42	2014-10-20	2.255	1.711	-0.9481	0.3179	180	51.4	2.288	241.3	-0.0323	-0.176
43	2014-11-17	1.602	1.204	-0.6933	0.7207	40	16	1.649	55.47	-0.047	-0.278
44	2014-12-15	2.301	1.623	-0.2816	0.9595	200	41.99	2.441	343.3	-0.14	-0.61
45	2015-01-12	1.602	0.9227	0.1963	0.9805	40	8.37	1.485	38.05	0.117	0.33
46	2015-02-09	3.041	1.439	0.6298	0.7768	1100	27.5	2.347	276.4	0.695	2.26
47	2015-04-06	1.806	1.206	0.9972	-0.07441	64	16.07	1.913	101.9	-0.107	-0.551
48	2015-05-04	2.322	1.574	0.8496	-0.5274	210	37.49	2.355	281.7	-0.0325	-0.227
49	2015-05-18	3.929	2.829	0.6982	-0.7159	8500	674.5	4.21	20210	-0.281	-0.949
50	2015-06-15	3.322	2.348	0.2872	-0.9579	2100	223	3.324	2621	-0.00131	0.025

Definitions and National Water Information System (parameter code)

Act: Actinomycetes in col/mL (63688)

Turb: Turbidity in FNU (63680)

DY: Date in decimal years