Appendix 2. Logistic Model Archival Summary for Cyanobacteria Concentration > 10,000 cells per milliliter at Station 06892350; Kansas River at De Soto, Kansas

This model archival summary (MAS) summarizes the logistic model for the probability of cyanobacteria (Cyano) concentrations > 10,000 cells per milliliter (cells/mL) developed to compute 15-minute Cyano from July 19, 2012 onward.

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 has been used. 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. Logistic model equations were developed using the multiple logistic regression routine in SigmaPlot® version 11.0 (Systat Software, Inc., 2008). Explanatory variables were evaluated individually and in selected combinations. Explanatory variables selected as inputs to logistic 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 logistic regression model is based on 57 concurrent measurements of Cyano concentration and turbidity (Turb) collected from July 19, 2012 through June 29, 2015, and models the probability of Cyano presence at concentrations >10,000 cells/mL or absence (≤10,000 cells/mL). Samples were collected throughout the range of continuously observed hydrologic conditions. Forty-four samples were below the model threshold of 10,000 cells/mL. Nineteen samples had no detection of Cyano (0 cells/mL. Summary statistics and the complete model-calibration dataset are provided below.

Cyanobacteria 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 and harmful algal bloom 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 Cyano concentration at the BSA Environmental Services, Inc., Laboratory in Beachwood, OH.

Model Development

Logistic regression analysis was done using SigmaPlot by examining Turb, seasonality, streamflow, and other continuously measured data as explanatory variables for estimating Cyano concentration. Turb and seasonality were selected as the best predictors of Cyano based on a relatively low Pearson Chi-square Statistic, relatively high Likelihood Ratio Test Statistic, relatively low -2 Log Likelihood Statistic, relatively high Hosmer-Lemeshow Statistic, significant Wald Statistic, and relatively low Variance Inflation Factor (VIF). A model classification table with a threshold probability for positive classification (TPPC) of 0.5 was also used in final model selection. After the best model was selected, the TPPC for the model was adjusted based on the fraction of data classified as positive to make the model more conservative (more likely to overestimate a positive response) by guarding more strongly against false negatives. 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 logistic regression analysis for Cyano concentration at site number 06892350.

Probability of Cyano occurrence model:

$$\text{logit}(P) = 0.724 - 1.254\sin\left(\frac{2\pi D}{365}\right) - 3.725\cos\left(\frac{2\pi D}{365}\right) - 0.0928(Turb)$$

where

P = probability of cyanobacteria presence (>10,000 cells/mL); and,

Turb = turbidity in formazin nephelometric units (FNU);

D = day of year; and,

Sin & Cos = seasonality component.

Turb and seasonality make physical and statistical sense as explanatory variables for Cyano.

Previous Models

No previous models.

Probability of > 10,000 cells/mL Concentration Record

The Cyano 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

SigmaPlot Output for Cyanobacteria; 06892350; Kansas River at De Soto, KS

Model Form

Logit P = 0.724 – (1.254 * Sin) – (3.725 * Cos) – (0.0928 * Turb)

Variable Summary Statistics

	Cyano	Cyano Bin	Turb
Minimum	0.000	0.000	7.000
1st Quartile	0.000	0.000	20.590
Median	342.233	0.000	33.600
Mean	21251.668	0.228	111.686
3rd Quartile	7862.935	0.000	69.000
Maximum	295202.128	1.000	1529.000

Model Calibration

Multiple Logistic Regression

Logit P = 0.724 - (1.254 * Sin) - (3.725 * Cos) - (0.0928 * Turb)

N = 57
Estimation Criterion: Maximum likelihood
Dependent Variable: Cyano Binary
Positive response (1): 1
Reference response (0): 0
Number of unique independent variable combinations: 57

Pearson Chi-square Statistic: 37.949 (P = 0.928)
Likelihood Ratio Test Statistic: 33.185 (P = <0.001)
-2*Log(Likelihood) = 28.025
Hosmer-Lemeshow Statistic: 2.415 (P = 0.966)

Threshold probability for positive classification: 0.480
Classification Table:

	Predicted Reference	Predicted Positive	Totals	Accuracy
Actual Reference Responses	41	3	44	0.93
Actual Positive Responses	2	11	13	0.85
Totals	43	14	57	0.91

Details of the Logistic Regression Equation

Ind. Variable	Coefficient	Standard Error	Wald Statistic	P value	VIF
Constant	0.724	1.393	0.270	0.603	
Sin	-1.254	0.740	2.868	0.090	1.006
Cos	-3.725	1.247	8.923	0.003	1.127
Turb	-0.0928	0.0488	3.620	0.057	1.130

Ind. Variable	Odds Ratio	5% Conf. Lower	95% Conf. Upper
Constant	2.063	0.134	31.642
Sin	0.285	0.0669	1.218
Cos	0.0241	0.00209	0.278
Turb	0.911	0.828	1.003

Data

Date	Julian Day	Sin	Cos	Turb	Cyano Abundance	Cyano Binary (>10000)	Computed Probability	Correct Classification
7/30/2012	212	-0.4863	-0.8738	30.79	58128.43533	1	0.8497322	yes
8/9/2012	222	-0.6288	-0.7776	26.3	72782.89542	1	0.8775304	yes
8/13/2012	226	-0.6808	-0.7325	23.5	263982.4222	1	0.8934216	yes
8/27/2012	240	-0.8359	-0.5488	36	16302.43919	1	0.6170352	yes
9/10/2012	254	-0.9428	-0.3335	24	295202.1279	1	0.7154081	yes
9/24/2012	268	-0.9951	-0.0988	20.59	4990.900709	0	0.605789	no
10/15/2012	289	-0.9657	0.2595	20	0	0	0.2916726	yes
10/29/2012	303	-0.8759	0.4825	14	10694.78723	1	0.2185718	no
11/19/2012	324	-0.6486	0.7611	13.5	2444.522796	0	0.0724039	yes
12/17/2012	352	-0.2219	0.9751	8.25	0	0	0.0324435	yes
1/14/2013	14	0.2387	0.9711	7	254.2330202	0	0.0210011	yes
2/11/2013	42	0.6616	0.7498	7.21	2300.192317	0	0.0274545	yes
3/11/2013	70	0.9338	0.3577	31.28	308.7115246	0	0.0091869	yes
4/8/2013	98	0.9933	-0.1159	21.5	23820.86125	1	0.1106971	no
5/6/2013	126	0.8264	-0.5632	154	0	0	3.727E-06	yes
5/20/2013	140	0.6681	-0.7441	61.4	0	0	0.0457693	yes
6/3/2013	154	0.4712	-0.8820	152.86	0	0	2.121E-05	yes
6/17/2013	168	0.2470	-0.9690	160	0	0	2.003E-05	yes
7/1/2013	182	0.0086	-1.0000	181	0	0	4.32E-06	yes
7/15/2013	196	-0.2303	-0.9731	33.6	101711.3676	1	0.8206708	yes
7/25/2013	206	-0.3936	-0.9193	50.5	174392.0326	1	0.489333	yes
8/5/2013	217	-0.5596	-0.8288	663.33	0	0	1.722E-25	yes
8/12/2013	224	-0.6552	-0.7555	270	342.2331915	0	1.037E-09	yes
8/19/2013	231	-0.7412	-0.6713	356	417.6685332	0	2.895E-13	yes
9/9/2013	252	-0.9307	-0.3657	28.75	10746.78775	1	0.6424977	yes
9/23/2013	266	-0.9911	-0.1330	35.84	7862.935335	0	0.296786	yes
10/21/2013	294	-0.9399	0.3416	16	199.7545159	0	0.2985227	yes
11/18/2013	322	-0.6744	0.7383	26.4	0	0	0.0258363	yes
12/16/2013	350	-0.2554	0.9668	30.8	117.6426596	0	0.0044313	yes
1/13/2014	13	0.2219	0.9751	8.5	537.5366977	0	0.0184355	yes
2/10/2014	41	0.6486	0.7611	7.05	0	0	0.0271659	yes
3/10/2014	69	0.9275	0.3737	23	0	0	0.0186252	yes
4/7/2014	97	0.9951	-0.0988	39	9473.927816	0	0.0224644	yes

5/5/2014	125	0.8359	-0.5488	63.57	1451.43541	0	0.0151215	yes
5/19/2014	139	0.6808	-0.7325	44.5	0	0	0.1781784	yes
6/2/2014	153	0.4863	-0.8738	54.83	7346.097893	0	0.1522883	yes
6/11/2014	162	0.3456	-0.9384	1529	0	0	1.109E-60	yes
6/30/2014	181	0.0258	-0.9997	78	0	0	0.0562495	yes
7/14/2014	195	-0.2135	-0.9769	19.67	1944.225772	0	0.9430158	no
7/28/2014	209	-0.4405	-0.8977	25.75	45790.74677	1	0.9030733	yes
8/4/2014	216	-0.5452	-0.8383	24.5	56035.18857	1	0.9053019	yes
8/11/2014	223	-0.6421	-0.7667	41.8	5347.393617	0	0.6241738	no
8/25/2014	237	-0.8065	-0.5913	34.6	29971.40451	1	0.6743567	yes
9/8/2014	251	-0.9243	-0.3817	448.95	0	0	2.229E-17	yes
9/22/2014	265	-0.9887	-0.1501	69	508.4660405	0	0.0202662	yes
10/6/2014	279	-0.9959	0.0903	86.33	145.2760116	0	0.0017057	yes
10/20/2014	293	-0.9456	0.3253	51.4	2553.947132	0	0.016781	yes
11/17/2014	321	-0.6871	0.7266	16	0	0	0.0687804	yes
12/15/2014	349	-0.2720	0.9623	41.99	290.5520231	0	0.0016343	yes
1/12/2015	12	0.2051	0.9787	8.37	290.5520231	0	0.0187908	yes
2/9/2015	40	0.6354	0.7722	27.5	0	0	0.0040704	yes
3/9/2015	68	0.9210	0.3896	7.4	817.697985	0	0.0712044	yes
4/6/2015	96	0.9967	-0.0817	16.07	120.0011937	0	0.1529297	yes
5/4/2015	124	0.8452	-0.5344	37.49	47.00046752	0	0.1391136	yes
5/18/2015	138	0.6933	-0.7207	674.5	0	0	8.492E-27	yes
6/15/2015	166	0.2802	-0.9599	222.96	0	0	5.4E-08	yes
6/29/2015	180	0.0430	-0.9991	160	1670.674133	0	2.893E-05	yes

Definitions and National Water Information System (parameter code)

Cyano: Cyanophyta in cells/mL (98062)
Turb: Turbidity in FNU (63680)