

Appendix 1. Logistic Model Archival Summary for Cyanobacteria Concentration > 2,000 cells per milliliter at Station 06887500; Kansas River at Wamego, Kansas

This model archival summary (MAS) summarizes the logistic model for the probability of cyanobacteria (Cyano) concentrations > 2,000 cells per milliliter (cells/mL) developed to compute 15-minute Cyano from July 19, 2012 onward.

Site and Model Information

Site number: 06887500

Site name: Kansas River at Wamego, Kansas

Location: Lat 39°11'54", long 96°18'19" referenced to North American Datum of 1927, in SW 1/4 NW 1/4 SE 1/4 sec.9, T.10 S., R.10 E., Pottawatomie County, KS, Hydrologic Unit 10270102.

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 56 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 >2,000 cells/mL or absence (\leq 2,000 cells/mL). Samples were collected throughout the range of continuously observed hydrologic conditions. Forty samples were below the model threshold of 2,000 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 from the downstream side 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. 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, 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 06887500.

Probability of Cyano occurrence model:

$$\text{logit}(P) = 0.0887 - 0.875\sin\left(\frac{2\pi D}{365}\right) - 1.914\cos\left(\frac{2\pi D}{365}\right) - 0.0319(\text{Turb})$$

where

P = probability of cyanobacteria presence (>2,000 cells/mL);

$Turb$ = turbidity in formazin nephelometric units (FNU); and,

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 Cyanobacteria > 2,000 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; 06887500; Kansas River at Wamego, KS

Model Form

Logit P = 0.0887 – (0.875 * Sin) – (1.914 * Cos) - (0.0319 * Turb)

Variable Summary Statistics

	Cyano	Cyano Bin	Turb
Minimum	0.000	0.000	5.200
1st Quartile	0.000	0.000	15.125
Median	283.101	0.000	32.700
Mean	4459.842	0.286	68.170
3rd Quartile	3053.000	1.000	82.233
Maximum	68955.000	1.000	298.750

Model Calibration

Multiple Logistic Regression

Logit P = 0.0887 - (0.875 * Sin) - (1.914 * Cos) - (0.0319 * TBY)

N = 56

Estimation Criterion: Maximum likelihood

Dependent Variable: Cyano Binary

Positive response (1): 1

Reference response (0): 0

Number of unique independent variable combinations: 56

Pearson Chi-square Statistic: 57.812 (P = 0.238)

Likelihood Ratio Test Statistic: 19.962 (P = <0.001)

-2*Log(Likelihood) = 47.044

Hosmer-Lemeshow Statistic: 8.164 (P = 0.418)

Threshold probability for positive classification: 0.310

Classification Table:

	Predicted Reference	Predicted Positive	Totals	Accuracy
Actual Reference Responses	33	7	40	0.82
Actual Positive Responses	4	12	16	0.75
Totals	37	19	56	0.80

Details of the Logistic Regression Equation

Ind. Variable	Coefficient	Standard Error	Wald Statistic	P value	VIF
Constant	0.0887	0.581	0.0233	0.879	
Sin	-0.875	0.507	2.977	0.084	1.000
Cos	-1.914	0.662	8.356	0.004	1.344
TBY	-0.0319	0.0135	5.588	0.018	1.344

Ind. Variable	Odds Ratio	5% Conf. Lower	95% Conf. Upper
Constant	1.093	0.350	3.413
Sin	0.417	0.154	1.126
Cos	0.147	0.0403	0.540
TBY	0.969	0.943	0.995

Data

Date	Julian Day	Sin	Cos	Turb	Cyano Abundance	Cyano Binary (>2,000)	Computed Probability	Correct Classification
7/19/2012	201	-0.3131	-0.9497	33.4	20901	1	0.7531343	yes
8/13/2012	226	-0.6808	-0.7325	17.1	6830	1	0.8236205	yes
8/27/2012	240	-0.8359	-0.5488	154	7642	1	0.0455816	no
9/10/2012	254	-0.9428	-0.3335	23.71	68955	1	0.6890076	yes
9/24/2012	268	-0.9951	-0.0988	31.33	340	0	0.5371789	no
10/15/2012	289	-0.9657	0.2595	17	4736	1	0.4736236	yes
10/29/2012	303	-0.8759	0.4825	22	4725	1	0.3163507	yes
11/19/2012	324	-0.6486	0.7611	11.69	0	0	0.2361562	yes
12/17/2012	352	-0.2219	0.9751	5.2	0	0	0.1480837	yes
1/14/2013	14	0.2387	0.9711	7.88	1149	0	0.0970455	yes
2/11/2013	42	0.6616	0.7498	6.2	2305	1	0.1068567	no
3/11/2013	70	0.9338	0.3577	9.65	1056	0	0.151781	yes
4/8/2013	98	0.9933	-0.1159	19	4725	1	0.2379084	no
5/6/2013	126	0.8264	-0.5632	110	0	0	0.0445864	yes
5/20/2013	140	0.6681	-0.7441	49	5561	1	0.3465869	yes
6/3/2013	154	0.4712	-0.8820	210	0	0	0.004805	yes
6/17/2013	168	0.2470	-0.9690	116.67	0	0	0.1198619	yes
7/1/2013	182	0.0086	-1.0000	98.5	0	0	0.2411553	yes
7/15/2013	196	-0.2303	-0.9731	43	17891	1	0.6860147	yes
8/5/2013	217	-0.5596	-0.8288	294	471	0	0.0007363	yes
8/12/2013	224	-0.6552	-0.7555	81	236	0	0.3832992	no
8/19/2013	231	-0.7412	-0.6713	260	299	0	0.0018863	yes
9/9/2013	252	-0.9307	-0.3657	16	3040	1	0.7488773	yes
9/23/2013	266	-0.9911	-0.1330	11.9	4220	1	0.6965103	yes
10/21/2013	294	-0.9399	0.3416	21	0	0	0.3982094	no
11/18/2013	322	-0.6744	0.7383	25.33	8	0	0.1761332	yes
12/16/2013	350	-0.2554	0.9668	24.33	299	0	0.0898931	yes
1/13/2014	13	0.2219	0.9751	12	620	0	0.0866812	yes
2/10/2014	41	0.6486	0.7611	9.6	0	0	0.0960482	yes
3/10/2014	69	0.9275	0.3737	9.07	0	0	0.150925	yes
4/7/2014	97	0.9951	-0.0988	12.5	908	0	0.2706698	yes
5/5/2014	125	0.8359	-0.5488	53.99	3850	1	0.2118368	no
5/19/2014	139	0.6808	-0.7325	32	0	0	0.4687512	no

6/2/2014	153	0.4863	-0.8738	70	1768	0	0.2897485	yes
6/11/2014	162	0.3456	-0.9384	192.5	0	0	0.0103802	yes
6/30/2014	181	0.0258	-0.9997	47.79	41	0	0.611959	no
7/14/2014	195	-0.2135	-0.9769	39	678	0	0.7113142	no
7/28/2014	209	-0.4405	-0.8977	23	3092	1	0.8113922	yes
8/4/2014	216	-0.5452	-0.8383	24	33295	1	0.8029675	yes
8/11/2014	223	-0.6421	-0.7667	32	47853	1	0.7497604	yes
8/25/2014	237	-0.8065	-0.5913	52.5	0	0	0.5625381	no
9/8/2014	251	-0.9243	-0.3817	85.42	0	0	0.2503466	yes
9/22/2014	265	-0.9887	-0.1501	81.17	1176	0	0.206144	yes
10/6/2014	279	-0.9959	0.0903	190	145	0	0.0050996	yes
10/20/2014	293	-0.9456	0.3253	67.8	218	0	0.1335817	yes
11/17/2014	321	-0.6871	0.7266	39.5	0	0	0.123327	yes
12/15/2014	349	-0.2720	0.9623	38	0	0	0.0613545	yes
1/12/2015	12	0.2051	0.9787	9.36	308.711525	0	0.094237	yes
2/9/2015	40	0.6354	0.7722	10	0	0	0.0941231	yes
3/9/2015	68	0.9210	0.3896	9.46	0	0	0.1462254	yes
4/6/2015	96	0.9967	-0.0817	7.42	0	0	0.2966365	yes
5/4/2015	124	0.8452	-0.5344	65	0	0	0.1543424	yes
5/18/2015	138	0.6933	-0.7207	298.75	19.000189	0	0.0001721	yes
6/1/2015	152	0.5012	-0.8653	230	97.2253385	0	0.0024007	yes
6/15/2015	166	0.2802	-0.9599	181.79	26.0002586	0	0.0160267	yes
6/29/2015	180	0.0430	-0.9991	175	267.201594	0	0.0261264	yes

Definitions and National Water Information System (parameter code)

Cyano: Cyanophyta in cells/mL (98062)

Turb: Turbidity in FNU (63680)