Appendix 3. Logistic Model Archival Summary for Microcystin Occurrence at Station 06887500; Kansas River at Wamego, Kansas

This model archival summary (MAS) summarizes the logistic model for the probability of microcystin (MC) occurrence model developed to compute 15-minute MC from July 19, 2012 onward. This model supersedes all previous models.

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 59 concurrent measurements of MC concentration, streamflow (Q), and turbidity (Turb) collected from July 19, 2012 through June 29, 2015, and models the probability of MC presence (≥0.10 µg/L) or absence (<0.10 µg/L). Samples were collected throughout the range of continuously observed hydrologic conditions. Forty-six samples were below the laboratory detection limit of 0.10 µg/L. Summary statistics and the complete model-calibration dataset are provided below.

Microcystin Sampling Details

Cross-section samples are typically collected either 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 events. A FISP US DH-95, D-95 or D-96A1 depth integrating sampler is used from the bridge. Samples are analyzed for MC concentration at the Organic Geochemistry Research Laboratory in Lawrence, KS.

Model Development

Logistic regression analysis was done using SigmaPlot by examining Q, Turb, seasonality, and other continuously measured data as explanatory variables for estimating MC concentration. Q, Turb, and seasonality were selected as the best predictors of MC 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 MC concentration at site number 06887500.

Probability of MC occurrence model:

$$\text{logit}(P) = -0.872 - 1.716\sin\left(\frac{2\pi D}{365}\right) - 1.313\cos\left(\frac{2\pi D}{365}\right) + 0.000349(Q) - 0.0490(Turb)$$

where

P = probability of microcystin presence (≥0.10 µg/L);

Q = streamflow in cubic feet per second (ft^3/s);

Turb = turbidity in formazin nephelometric units (FNU);

D = day of year; and,

Sin & Cos = seasonality component.

Q, Turb, and seasonality make physical and statistical sense as explanatory variables for MC.

Previous Models

No previous models.

Probability of Microcystin Occurrence Record

The MC 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 Microcystin; 06887500; Kansas River at Wamego, KS

Model Form

Logit P = -0.872 – (1.716 * Sin) – (1.313 * Cos) + (0.000349 * Q) – (0.0490 * Turb)

Variable Summary Statistics

	MC	MC Bin	Q	Turb
Minimum	0.050	0.000	572.980	5.200
1st Quartile	0.050	0.000	1011.465	16.500
Median	0.050	0.000	1277.050	33.400
Mean	0.115	0.220	4052.866	67.137
3rd Quartile	0.050	0.000	3658.495	81.085
Maximum	1.700	1.000	30851.830	298.750

Model Calibration

Multiple Logistic Regression

Logit P = -0.872 - (1.716 * Sin) - (1.313 * Cos) + (0.000349 * Q) - (0.0490 * Turb)

N = 59
Estimation Criterion: Maximum likelihood
Dependent Variable: MC Binary
Positive response (1): 1
Reference response (0): 0
Number of unique independent variable combinations: 59

Pearson Chi-square Statistic: 46.475 (P = 0.725)
Likelihood Ratio Test Statistic: 20.411 (P = <0.001)
-2*Log(Likelihood) = 41.814
Hosmer-Lemeshow Statistic: 6.534 (P = 0.588)

Threshold probability for positive classification: 0.400
Classification Table:

	Predicted Reference	Predicted Positive	Totals	Accuracy
Actual Reference Responses	40	6	46	0.87
Actual Positive Responses	3	10	13	0.77
Totals	43	16	59	0.85

Details of the Logistic Regression Equation

Ind. Variable	Coefficient	Standard Error	Wald Statistic	P value	VIF
Constant	-0.872	0.688	1.609	0.205	
Sin	-1.716	0.736	5.436	0.020	1.016
Cos	-1.313	0.636	4.260	0.039	1.319
Q	0.000349	0.000178	3.855	0.050	2.415
Turb	-0.0490	0.0243	4.085	0.043	2.683

Ind. Variable	Odds Ratio	5% Conf. Lower	95% Conf. Upper
Constant	0.418	0.109	1.609
Sin	0.180	0.0425	0.761
Cos	0.269	0.0774	0.936
Q	1.000	1.000	1.001
Turb	0.952	0.908	0.999

Data

Date	Julian Day	Sin	Cos	Q	Turb	Total MC	MC Binary (>0.1)	Computed Probability	Correct Classification
7/19/2012	201	-0.3131	-0.9497	1792.93	33.4	0.11	1	0.474879421	yes
7/30/2012	212	-0.4863	-0.8738	2942.07	30.1	0.48	1	0.659216101	yes
8/13/2012	226	-0.6808	-0.7325	999.49	17.1	0.82	1	0.683042236	yes
8/27/2012	240	-0.8359	-0.5488	2165.58	154	<0.10	0	0.004012231	yes
9/10/2012	254	-0.9428	-0.3335	1070.06	23.71	0.1	1	0.597243071	yes
9/24/2012	268	-0.9951	-0.0988	1035.1	31.33	<0.10	0	0.447618714	no
10/15/2012	289	-0.9657	0.2595	1046.08	17	0.12	1	0.493872302	yes
10/29/2012	303	-0.8759	0.4825	647.03	22	<0.10	0	0.298212697	yes
11/19/2012	324	-0.6486	0.7611	615.02	11.69	<0.10	0	0.246541181	yes
12/17/2012	352	-0.2219	0.9751	588.06	5.2	<0.10	0	0.139265925	yes
1/14/2013	14	0.2387	0.9711	767.23	7.88	<0.10	0	0.064438829	yes
2/11/2013	42	0.6616	0.7498	727.24	6.2	<0.10	0	0.045558362	yes
3/11/2013	70	0.9338	0.3577	615.02	9.65	<0.10	0	0.03906029	yes
4/8/2013	98	0.9933	-0.1159	572.98	19	<0.10	0	0.040841045	yes
5/6/2013	126	0.8264	-0.5632	2391.67	110	<0.10	0	0.00221321	yes
5/20/2013	140	0.6681	-0.7441	1632.97	49	<0.10	0	0.053400332	yes
6/3/2013	154	0.4712	-0.882	9016.16	210	<0.10	0	0.000463653	yes
6/17/2013	168	0.247	-0.969	6940.33	116.67	<0.10	0	0.034754157	yes
7/1/2013	182	8.61E-03	-1	3892.15	98.5	<0.10	0	0.045355731	yes
7/15/2013	196	-0.2303	-0.9731	1418.81	43	<0.10	0	0.307191004	yes
8/5/2013	217	-0.5596	-0.8288	20227.17	294	<0.10	0	0.002056401	yes
8/12/2013	224	-0.6552	-0.7555	8001.42	81	0.16	1	0.515837581	yes
8/19/2013	231	-0.7412	-0.6713	11039.03	260	<0.10	0	0.000491065	yes
9/9/2013	252	-0.9307	-0.3657	1129.19	16	<0.10	0	0.693016085	no
9/23/2013	266	-0.9911	-0.133	1054.38	11.9	<0.10	0	0.687268087	no
10/21/2013	294	-0.9399	0.3416	1123.35	21	<0.10	0	0.414406454	no
11/18/2013	322	-0.6744	0.7383	1043.08	25.33	0.1	1	0.173283979	no
12/9/2013	343	-0.3697	0.9291	4193.91	35	<0.10	0	0.153073615	yes
12/16/2013	350	-0.2554	0.9668	4130.66	24.33	<0.10	0	0.18921819	yes
1/13/2014	13	0.2219	0.9751	2227.22	12	<0.10	0	0.087505744	yes
2/10/2014	41	0.6486	0.7611	3424.84	9.6	<0.10	0	0.094479005	yes
3/10/2014	69	0.9275	0.3737	1035.58	9.07	<0.10	0	0.045746724	yes
4/7/2014	97	0.9951	-0.0988	1054.38	12.5	<0.10	0	0.063260119	yes

5/5/2014	125	0.8359	-0.5488	1184.72	53.99	<0.10	0	0.021446258	yes
5/19/2014	139	0.6808	-0.7325	1023.44	32	<0.10	0	0.09186881	yes
5/28/2014	148	0.5596	-0.8288	1380.18	78.5	<0.10	0	0.016100596	yes
6/2/2014	153	0.4863	-0.8738	1827.35	70	<0.10	0	0.033734697	yes
6/11/2014	162	0.3456	-0.9384	17325.69	192.5	<0.10	0	0.025865544	yes
6/30/2014	181	0.0258	-0.9997	6117.01	47.79	<0.10	0	0.546517912	no
7/14/2014	195	-0.2135	-0.9769	1500.19	39	<0.10	0	0.351476745	yes
7/28/2014	209	-0.4405	-0.8977	1198.29	23	1.7	1	0.587184763	yes
8/4/2014	216	-0.5452	-0.8383	1123.35	24	0.48	1	0.593589406	yes
8/11/2014	223	-0.6421	-0.7667	963.36	32	<0.10	0	0.500731857	no
8/25/2014	237	-0.8065	-0.5913	909.13	52.5	<0.10	0	0.274969351	yes
9/8/2014	251	-0.9243	-0.3817	7454.33	85.42	0.11	1	0.407750034	yes
9/22/2014	265	-0.9887	-0.1501	2063.19	81.17	<0.10	0	0.096257375	yes
10/6/2014	279	-0.9959	0.0903	1097.83	190	<0.10	0	0.000269951	yes
10/20/2014	293	-0.9456	0.3253	985.2	67.8	<0.10	0	0.065495677	yes
11/17/2014	321	-0.6871	0.7266	976.43	39.5	0.1	1	0.095904911	no
12/15/2014	349	-0.272	0.9623	2063.22	38	<0.10	0	0.056651299	yes
1/12/2015	12	0.2051	0.9787	2195.54	9.36	<0.10	0	0.09957503	yes
2/9/2015	40	0.6354	0.7722	864.61	10	<0.10	0	0.040503732	yes
3/9/2015	68	0.921	0.3896	922.42	9.46	<0.10	0	0.042854643	yes
4/6/2015	96	0.9967	-0.0817	983.01	7.42	0.1	1	0.076131441	no
5/4/2015	124	0.8452	-0.5344	1277.05	65	<0.10	0	0.012578113	yes
5/18/2015	138	0.6933	-0.7207	22236.84	298.75	<0.10	0	0.000332585	yes
6/1/2015	152	0.5012	-0.8653	7555.13	230	<0.10	0	9.70703E-05	yes
6/15/2015	166	0.2802	-0.9599	22479.56	181.79	<0.10	0	0.237833758	yes
6/29/2015	180	0.043	-0.9991	30851.83	175	0.11	1	0.927454902	yes

Definitions and National Water Information System (parameter code)

MC: Microcystin (65210)
Turb: Turbidity in FNU (63680)
Q: Streamflow in cubic feet per second (00060)