Appendix 2. 4-Year Logistic Regression Model Archival Summary for Microcystin Occurrence at Station 07144790, 2013–16

This model archival summary summarizes the logistic model for the probability of microcystin occurrence developed to compute hourly microcystin from January 1, 2013, onward.

Station and Model Information

Station number: 07144790
Station name: Cheney Re Nr Cheney, KS
Station location: Latitude 37°43'34", Longitude 97°47'38" referenced to the North American Datum of 1927, in SE¼NE¼NW¼ sec. 6, T. 27 S., R. 04 W., Sedgwick County, Kansas, Hydrologic Unit 11030014.

Equipment: From April 2001 through September 2014, a YSI 6600 water-quality monitor was installed and equipped with sensors for water temperature, specific conductance, dissolved oxygen (YSI Clark cell [from April 2001 through January 2007] or YSI model 6150 optical [from February 2007 through September 2014]), pH, turbidity (YSI model 6026 [from April 2001 through September 2006] or YSI 6136 [from October 2006 through September 2014]), and chlorophyll. From October 2014 to the present (December 2016), a Xylem YSI EXO2 water-quality monitor has been used and is equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, turbidity, and chlorophyll fluorescence (YSI model 6025 sensor). The Xylem monitor is housed in a 4-inch diameter galvanized steel pipe. Readings from the water-quality monitor are recorded hourly and data are transmitted hourly by satellite.

Date model was created: August 16, 2016

Model calibration data period: January 15, 2013, through June 15, 2016

Model application date: August 2016 onward

Model-Calibration Dataset

All data were collected using U.S. Geological Survey (USGS) protocols (U.S. Geological Survey, variously dated; https://water.usgs.gov/owq/FieldManual/) and are stored in the National Water Information System database (https://doi.org/10.5066/F7P55KJN). 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, chlorophyll fluorescence, and elevation of the reservoir surface. Seasonal components (sine and cosine variables) also were 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 was based on 45 concurrent measurements of microcystin occurrence collected from January 15, 2013, through June 15, 2016, and models the probability of the presence or absence of microcystin. Samples were collected throughout the range of continuously observed hydrologic conditions. In total, 16 samples were below the threshold for positive classification (0.1 microgram per liter [µg/L]). Summary statistics and the complete model-calibration dataset are provided below. Studentized residuals were inspected for values outside the 95-percent confidence interval and leverage values for independent variables were inspected for values greater than 2. Values outside of the specified ranges were considered potential outliers and were investigated. No outliers were identified in the model-calibration dataset.

Microcystin Sampling Details

Discrete water-quality samples were collected monthly to biweekly during January 2013 through June 2016. Samples were collected as integrated photic-zone (depth at which light is about 1 percent of that at the surface) samples using a double check-valve bailer; these samples were depth integrated. Total microcystin was analyzed by the USGS Organic Geochemistry Research Laborabory, Lawrence, Kans. All samples were lysed by three sequential freeze-thaw cycles and filtered using 0.7-micrometer glass-fiber filters before analysis for microcystin. Abraxis® enzyme-linked immunosorbent assays were used to measure microcystin (congener independent).

Model Development

Logistic regression analysis was done using SigmaPlot by examining seasonality and other continuously measured data as explanatory variables for estimating microcystin occurrence. Seasonality was selected as the best predictor of microcystin 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. A model classification table with a threshold probability for positive classification (TPPC) of 0.5 also was 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 aforementioned statistics and metrics were computed for various models and are included below along with all relevant sample data and more indepth statistical information.

Model Summary

Summary of final logistic regression analysis for microcystin occurrence at USGS station 07144790.

Probability of microcystin occurrence model:

$$\text{logit}(P) = -0.190 - 1.868\sin\left(\frac{2\pi D}{365}\right) - 1.109\cos\left(\frac{2\pi D}{365}\right) + 0.0910\left(Chl\right) \qquad (2\text{–}1)$$

where

logit(P) is the logistic probability of microcystin occurrence (concentrations greater than or equal to 0.1 microgram per liter);

D is the Julian day of the year;

Chl is fluorescence at wavelength of 650 to 700 nanometers, in micrograms per liter as chlorophyll.

Seasonality (the information contained in the sine [sin] and cosine [cos] component of the equation; Helsel and Hirsch, 2002) and *Chl* make physical and statistical sense as explanatory variables for microcystin.

Previously Published Model

$$\text{logit}(P) = -1.305 - 1.990\sin\left(\frac{2\pi D}{365}\right) - 1.340\cos\left(\frac{2\pi D}{365}\right) + 0.0511\left(Chl\right) \qquad (2\text{–}2)$$

Model author: Stone and others (2013)

Model data period: June 2005 through December 2009

Probability of Microcystin Occurrence Record

The microcystin occurrence record is computed using this regression model, and the complete water-quality record is stored at the National Real-Time Water Quality website: http://nrtwq.usgs.gov/ks. Data are computed at 60-minute intervals.

SigmaPlot® Output for Microcystin at Station 07144790

4-Year Model Form

$$\text{logit}(P) = -0.190 - 1.868\sin\left(\frac{2\pi D}{365}\right) - 1.109\cos\left(\frac{2\pi D}{365}\right) + 0.0910\left(Chl\right) \qquad (2\text{–}3)$$

Variable Summary Statistics

[µg/L, microgram per liter; *Chl*, fluorescence at wavelength of 650 to 700 nanometers, in micrograms per liter as chlorophyll; <, less than; --, not measured]

Summary statistic	Microcystin (µg/L)	Microcystin binary	*Chl* (µg/L)
Minimum	<0.1	0	2
1st quartile	<0.1	0	7.16
Median	0.1	1	9
Mean	--	1	11
3d quartile	0.4	1	12
Maximum	7.0	1	45

Model Calibration Using Multiple Logistic Regression

See the 4-year model form in equation 2–3 above.

Number of samples=45
Missing observations=186
Estimation criterion: Maximum likelihood
Dependent Variable: Microcystin binary (abbr) plus or minus (±)
Positive response=1
Reference response=0
Number of unique independent variable combinations=45

Pearson Chi-square Statistic=48.443 (probability value [*p*-value]=0.169)
Likelihood Ratio Test Statistic=14.971 (*p*-value=0.002)
-2*Log(Likelihood)=43.603
Hosmer-Lemeshow Statistic=9.465 (*p*-value=0.305)

TPPC=0.56

Classification table	Predicted reference responses	Predicted positive responses	Total actual responses	Percent correctly classified responses
Actual reference responses	12	4	16	75
Actual positive responses	4	25	29	86
Total	16	29	45	81

Details of the logistic regression equation:

[*p*-value, probability value; VIF, Variance Inflation Factor; --, not measured; sin, sine of the seasonality component; cos, cosine of the seasonality component; <, less than; *Chl*, fluorescence at wavelength of 650 to 700 nanometers, in micrograms per liter as chlorophyll]

Independent variable	Coefficient	Standard error	Wald statistic	*p*-value	VIF
Constant	–0.190	0.647	0.0866	0.769	--
Sin	–1.868	0.631	8.753	0.003	1.099
Cos	–1.109	0.542	4.180	0.041	1.043
Chl	0.0910	0.0525	3.006	0.083	1.143

Independent variable	Odds ratio	Lower 5-percent confidence interval	Upper 95-percent confidence interval
Constant	0.827	0.232	2.939
sin	0.154	0.0448	0.532
cos	0.330	0.114	0.955
Chl	1.095	0.988	1.214

Data Used in Model Development

[sin, sine of the seasonality component; cos, cosine of the seasonality component; µg/L, microgram per liter; ≥, greater than or equal to; *Chl*, fluorescence at wavelength of 650 to 700 nanometers, in µg/L as chlorophyll; <, less than]

Date	Julian date	sin	cos	Microcystin (µg/L)	Microcystin binary (≥ 0.1 µg/L)	*Chl* (µg/L)	Computed probability	Correct classification
01/15/2013	015	0.262	0.965	<0.1	0	8.1	0.2724	Yes
01/23/2013	023	0.392	0.920	<0.1	0	9	0.2525	Yes
02/12/2013	043	0.679	0.734	<0.1	0	18.2	0.2274	Yes
03/19/2013	078	0.976	0.220	0.1	1	29.1	0.2413	No
04/09/2013	099	0.990	–0.139	<0.1	0	24.1	0.2798	Yes
05/07/2013	127	0.813	–0.583	<0.1	0	3.1	0.3634	Yes
06/06/2013	157	0.419	–0.908	<0.1	0	6.7	0.5218	Yes
07/08/2013	189	–0.119	–0.993	3	1	9	0.6731	Yes
07/23/2013	204	–0.368	–0.930	0.89	1	4.6	0.7156	Yes
08/06/2013	218	–0.579	–0.815	7.3	1	4.5	0.7445	Yes
08/19/2013	231	–0.746	–0.666	1.1	1	6	0.7602	Yes
09/09/2013	252	–0.933	–0.359	0.43	1	16.1	0.7676	Yes
09/25/2013	268	–0.996	–0.092	0.47	1	9.2	0.7412	Yes
10/21/2013	294	–0.937	0.348	0.3	1	8.2	0.6668	Yes
10/28/2013	301	–0.888	0.459	0.23	1	8.6	0.6388	Yes
11/06/2013	310	–0.808	0.590	0.14	1	9.9	0.5998	Yes
11/13/2013	318	–0.729	0.685	0.13	1	8.4	0.5624	Yes
11/18/2013	322	–0.669	0.743	0.12	1	8.4	0.5373	No
12/12/2013	346	–0.314	0.949	<0.1	0	7.2	0.4092	Yes
01/15/2014	015	0.262	0.965	0.11	1	7.2	0.2716	No
02/19/2014	050	0.763	0.646	0.11	1	9.5	0.2152	No
03/19/2014	078	0.976	0.220	0.18	1	45.2	0.2551	No
04/16/2014	106	0.966	–0.257	0.19	1	7.7	0.2855	No
05/20/2014	140	0.663	–0.748	0.13	1	10.03	0.4351	No
06/25/2014	176	0.105	–0.995	0.41	1	10.3	0.6209	Yes
07/10/2014	191	–0.153	–0.988	0.27	1	7.3	0.6786	Yes
07/22/2014	203	–0.352	–0.936	0.45	1	9.8	0.7180	Yes
08/05/2014	217	–0.565	–0.825	0.37	1	8.1	0.7461	Yes
09/16/2014	259	–0.970	–0.245	0.43	1	11	0.7565	Yes
10/28/2014	301	–0.889	0.458	0.19	1	6.9	0.6372	Yes
11/20/2014	324	–0.643	0.766	0.2	1	11.5	0.5303	No

Date	Julian date	sin	cos	Microcystin (µg/L)	Microcystin binary (≥ 0.1 µg/L)	*Chl* (µg/L)	Computed probability	Correct classification
12/16/2014	350	–0.255	0.967	<0.1	0	7.5	0.3918	Yes
01/13/2015	013	0.222	0.975	<0.1	0	12.4	0.2836	Yes
02/10/2015	041	0.649	0.761	0.12	1	17.2	0.2291	No
03/10/2015	069	0.928	0.374	<0.1	0	25.6	0.2292	Yes
04/15/2015	105	0.972	–0.234	<0.1	0	4.54	0.2782	Yes
05/06/2015	126	0.826	–0.563	<0.1	0	2.44	0.3564	Yes
06/09/2015	160	0.378	–0.926	0.19	1	1.57	0.5298	No
07/07/2015	188	–0.095	–0.996	0.4	1	7.16	0.6659	Yes
08/04/2015	216	–0.545	–0.838	<0.1	0	11.6	0.7468	No
09/08/2015	251	–0.924	–0.382	0.13	1	13.7	0.7666	Yes
11/09/2015	313	–0.780	0.625	<0.1	0	9.46	0.5866	No
02/17/2016	048	0.735	0.678	<0.1	0	6.83	0.2146	Yes
05/17/2016	138	0.693	–0.721	<0.1	0	3.57	0.4156	Yes
06/15/2016	167	0.264	–0.965	0.63	1	16	0.5821	Yes

References Cited

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources—Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 522 p. [Also available at https://pubs.usgs.gov/twri/twri4a3/.]

Systat Software, Inc., 2008, SigmaPlot® 11.0 statistics user's guide: Systat Software, Inc., 564 p.

U.S. Geological Survey, variously dated, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1–A10, accessed September 2016 at https://pubs.water.usgs.gov/twri9A.