

Appendix 5. Logistic Model Archival Summary for Geosmin Concentration > 2 nanograms per liter at Station 06887500; Kansas River at Wamego, Kansas

This model archival summary (MAS) summarizes the logistic model for the probability of geosmin (Geos) concentrations > 2 nanograms per liter (ng/L) model developed to compute 15-minute Geos 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 59 concurrent measurements of Geos concentration and chlorophyll (Chl) collected from July 19, 2012 through June 29, 2015, and models the probability of Geos presence at concentrations >2.0 ng/L or absence (≤ 2.0 ng/L). Samples were collected throughout the range of continuously observed hydrologic conditions. Twenty-eight samples were below the model threshold of 2.0 ng/L. Thirteen samples were below the laboratory detection limit of 1.0 ng/L. Summary statistics and the complete model-calibration dataset are provided below.

Geosmin 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 events. A FISP US DH-95, D-95 or D-96A1 depth integrating sampler is used from the bridge. Samples are analyzed for Geos concentration at the Engineering Performance Solutions Laboratory in Jacksonville, FL.

Model Development

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

Probability of Geos occurrence model:

$$\text{logit}(P) = 1.325 + 0.830\sin\left(\frac{2\pi D}{365}\right) + 1.219\cos\left(\frac{2\pi D}{365}\right) - 0.0527(Chl)$$

where

P = probability of geosmin presence (>2.0 ng/L);

Chl = chlorophyll in micrograms per liter ($\mu\text{g/L}$);

D = day of year; and,

Sin & Cos = seasonality component.

Chl and seasonality make physical and statistical sense as explanatory variables for Geos.

Previous Models

No previous models.

Probability of Geosmin > 2 ng/L Record

The Geos 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 Geosmin; 06887500; Kansas River at Wamego, KS

Model Form

Logit P = 1.325 + (0.830 * Sin) + (1.219 * Cos) - (0.0527 * Chl)

Variable Summary Statistics

	Geos	Geos Bin	Chl
Minimum	0.500	0.000	1.140
1st Quartile	1.050	0.000	5.390
Median	2.100	1.000	11.060
Mean	2.756	0.525	17.628
3rd Quartile	3.100	1.000	20.409
Maximum	16.200	1.000	154.650

Model Calibration

Multiple Logistic Regression

Logit P = 1.325 + (0.830 * Sin) + (1.219 * Cos) - (0.0527 * Chl)

N = 59

Estimation Criterion: Maximum likelihood

Dependent Variable: Binary Geos

Positive response (1): 1

Reference response (0): 0

Number of unique independent variable combinations: 59

Pearson Chi-square Statistic: 56.577 (P = 0.379)

Likelihood Ratio Test Statistic: 19.603 (P = <0.001)

-2*Log(Likelihood) = 62.036

Hosmer-Lemeshow Statistic: 9.979 (P = 0.266)

Threshold probability for positive classification: 0.530

Classification Table:

	Predicted Reference	Predicted Positive	Totals	Accuracy
Actual Reference Responses	21	7	28	0.75
Actual Positive Responses	9	22	31	0.71
Totals	30	29	59	0.73

Details of the Logistic Regression Equation

Ind. Variable	Coefficient	Standard Error	Wald Statistic	P value	VIF
Constant	1.325	0.515	6.621	0.010	
Sin	0.830	0.459	3.263	0.071	1.009
Cos	1.219	0.501	5.917	0.015	1.076
Chl	-0.0527	0.0255	4.287	0.038	1.084

Ind. Variable	Odds Ratio	5% Conf. Lower	95% Conf. Upper
Constant	3.763	1.371	10.327
Sin	2.292	0.932	5.639
Cos	3.384	1.267	9.039
Chl	0.949	0.902	0.997

Data

Date	Julian Day	Sin	Cos	Chl	Geosmin	Geosmin Binary (>2)	Computed Probability	Correct Classification
7/19/2012	201	-0.3131	-0.9497	16.398	1	0	0.2774632	yes
7/30/2012	212	-0.4863	-0.8738	11.597	2.6	1	0.3197311	no
8/13/2012	226	-0.6808	-0.7325	20.487	1.9	0	0.2291886	yes
8/27/2012	240	-0.8359	-0.5488	18.28	2.8	1	0.268681	no
9/10/2012	254	-0.9428	-0.3335	14.59	3.1	1	0.3468789	no
9/24/2012	268	-0.9951	-0.0988	14.71	1.9	0	0.4021619	yes
10/15/2012	289	-0.9657	0.2595	11.06	2.5	1	0.5639586	yes
10/29/2012	303	-0.8759	0.4825	12.59	0.9	0	0.6278406	no
11/19/2012	324	-0.6486	0.7611	10.1793	2.3	1	0.7646388	yes
12/17/2012	352	-0.2219	0.9751	8.44	4	1	0.8681674	yes
1/14/2013	14	0.2387	0.9711	9.44	7.8	1	0.9010895	yes
2/11/2013	42	0.6616	0.7498	9.01	8	1	0.9099631	yes
3/11/2013	70	0.9338	0.3577	9.91	14.3	1	0.8822046	yes
4/8/2013	98	0.9933	-0.1159	50.2	16.2	1	0.3453885	no
5/6/2013	126	0.8264	-0.5632	18.1	4.7	1	0.5913969	yes
5/20/2013	140	0.6681	-0.7441	20.33	1.3	0	0.4750745	yes
6/3/2013	154	0.4712	-0.8820	1.54	<1.0	0	0.6363567	no
6/17/2013	168	0.2470	-0.9690	4.3	<1.0	0	0.5304803	no
7/1/2013	182	0.0086	-1.0000	1.45	<1.0	0	0.5092098	yes
7/15/2013	196	-0.2303	-0.9731	15.66	1.3	0	0.2935927	yes
8/5/2013	217	-0.5596	-0.8288	1.14	<1.0	0	0.4478231	yes
8/12/2013	224	-0.6552	-0.7555	9.17	2.5	1	0.3491266	no
8/19/2013	231	-0.7412	-0.6713	1.5	3.8	1	0.4533594	no
9/9/2013	252	-0.9307	-0.3657	22.72	2.1	1	0.2514561	no
9/23/2013	266	-0.9911	-0.1330	44.2	1.1	0	0.120256	yes
10/21/2013	294	-0.9399	0.3416	18.17	1.6	0	0.5009529	yes
11/18/2013	322	-0.6744	0.7383	8	2.4	1	0.7762635	yes
12/9/2013	343	-0.3697	0.9291	4.45	2.3	1	0.8717656	yes
12/16/2013	350	-0.2554	0.9668	3.34	3.3	1	0.8924533	yes
1/13/2014	13	0.2219	0.9751	1.34	2.1	1	0.9326095	yes
2/10/2014	41	0.6486	0.7611	6	2.8	1	0.9223619	yes
3/10/2014	69	0.9275	0.3737	4.66	3.1	1	0.909261	yes
4/7/2014	97	0.9951	-0.0988	15.73	7	1	0.7686695	yes

5/5/2014	125	0.8359	-0.5488	43.48	1.6	0	0.280235	yes
5/19/2014	139	0.6808	-0.7325	48.74	1.4	0	0.1717468	yes
5/28/2014	148	0.5596	-0.8288	154.65	1.5	0	0.0006255	yes
6/2/2014	153	0.4863	-0.8738	36.3	1.7	0	0.2225427	yes
6/11/2014	162	0.3456	-0.9384	5.9	4	1	0.5391278	yes
6/30/2014	181	0.0258	-0.9997	6.13	1.5	0	0.4513251	yes
7/14/2014	195	-0.2135	-0.9769	39.08	<1.0	0	0.1087287	yes
7/28/2014	209	-0.4405	-0.8977	21.3	1.4	0	0.2213315	yes
8/4/2014	216	-0.5452	-0.8383	31.3	<1.0	0	0.1418927	yes
8/11/2014	223	-0.6421	-0.7667	51.22	1.4	0	0.0550367	yes
8/25/2014	237	-0.8065	-0.5913	33.88	<1.0	0	0.1357223	yes
9/8/2014	251	-0.9243	-0.3817	7.55	5.3	1	0.4243368	no
9/22/2014	265	-0.9887	-0.1501	18.53	<1.0	0	0.3418534	yes
10/6/2014	279	-0.9959	0.0903	25.8	<1.0	0	0.3205018	yes
10/20/2014	293	-0.9456	0.3253	8.12	<1.0	0	0.6246316	no
11/17/2014	321	-0.6871	0.7266	4.2	2.2	1	0.8052867	yes
12/15/2014	349	-0.2720	0.9623	5.5	<1.0	0	0.8789827	no
1/12/2015	12	0.2051	0.9787	2.37	2.6	1	0.9284882	yes
2/9/2015	40	0.6354	0.7722	11.55	2.7	1	0.8988693	yes
3/9/2015	68	0.9210	0.3896	7.52	<1.0	0	0.8973176	no
4/6/2015	96	0.9967	-0.0817	17.8	4.2	1	0.7528446	yes
5/4/2015	124	0.8452	-0.5344	20.96	2.1	1	0.5670144	yes
5/18/2015	138	0.6933	-0.7207	4.58	4.1	1	0.6857427	yes
6/1/2015	152	0.5012	-0.8653	4.95	2.3	1	0.6047087	yes
6/15/2015	166	0.2802	-0.9599	4.68	<1.0	0	0.5351045	no
6/29/2015	180	0.0430	-0.9991	5.28	5.4	1	0.4661771	no

Definitions and National Water Information System (parameter code)

Geos: Geosmin in ng/L (51285)

Chl: Chlorophyll in ug/L (32318)