

Appendix 6. Logistic Model Archival Summary for Geosmin Concentration > 2 nanograms per liter at Station 06892350; Kansas River at De Soto, Kansas

This model archival summary (MAS) summarizes the logistic model for the probability of geosmin (Geos) concentrations > 2 nanograms per liter (ng/L) developed to compute 15-minute Geos 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 61 concurrent measurements of Geos concentration and streamflow (Q) 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-seven samples were the model threshold of 2.0 ng/L. Seven 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 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 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 Geos concentration at the Engineering Performance Solutions Laboratory in Jacksonville, FL.

Model Development

Logistic regression analysis was done using SigmaPlot by examining seasonality, Q, and other continuously measured data as explanatory variables for estimating Geos concentration. Q and seasonality 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 06892350.

Probability of Geos occurrence model:

$$\text{logit}(P) = 0.236 + 0.585\sin\left(\frac{2\pi D}{365}\right) + 1.084\cos\left(\frac{2\pi D}{365}\right) + 0.0000473(Q)$$

where

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

Q = streamflow in cubic feet per second (ft³/s); and,

D = day of year; and,

Sin \& Cos = seasonality component.

Q 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; 06892350; Kansas River at De Soto, KS

Model Form

$$\text{Logit P} = 0.236 + (0.585 * \text{Sin}) + (1.084 * \text{Cos}) + (0.0000473 * \text{Q})$$

Variable Summary Statistics

	Geos	Geos Bin	Q
Minimum	0.500	0.000	705.980
1st Quartile	1.500	0.000	1269.300
Median	2.300	1.000	2037.680
Mean	3.836	0.557	6768.175
3rd Quartile	3.800	1.000	4138.710
Maximum	41.600	1.000	80628.890

Model Calibration

Multiple Logistic Regression

$$\text{Logit P} = 0.236 + (0.585 * \text{Sin}) + (1.084 * \text{Cos}) + (0.0000473 * \text{Q})$$

N = 61

Estimation Criterion: Maximum likelihood

Dependent Variable: Geos Bin

Positive response (1): 1

Reference response (0): 0

Number of unique independent variable combinations: 61

Pearson Chi-square Statistic: 60.455 (P = 0.318)

Likelihood Ratio Test Statistic: 10.502 (P = 0.015)

-2*Log(Likelihood) = 73.257

Hosmer-Lemeshow Statistic: 12.503 (P = 0.130)

Threshold probability for positive classification: 0.500

Classification Table:

	Predicted Reference	Predicted Positive	Totals	Accuracy
Actual Reference Responses	20	7	27	0.74
Actual Positive Responses	8	26	34	0.76
Totals	28	33	61	0.75

Details of the Logistic Regression Equation

Ind. Variable	Coefficient	Standard Error	Wald Statistic	P value	VIF
Constant	0.236	0.332	0.503	0.478	
Sin	0.585	0.413	2.007	0.157	1.027
Cos	1.084	0.442	6.004	0.014	1.138
Q	0.0000473	0.0000303	2.426	0.119	1.166

Ind. Variable	Odds Ratio	5% Conf. Lower	95% Conf. Upper
Constant	1.266	0.660	2.429
Sin	1.795	0.799	4.032
Cos	2.956	1.242	7.035
Q	1.000	1.000	1.000

Data

Date	Julian Day	Sin	Cos	Q	Geosmin	Geosmin Binary (>2)	Computed Probability	Correct Classification
7/30/2012	212	-0.4863	-0.8738	3274.82	1	0	0.3013143	yes
8/9/2012	222	-0.6288	-0.7776	1176.39	12.7	1	0.28510325	no
8/13/2012	226	-0.6808	-0.7325	1016.7	3.3	1	0.2873335	no
8/27/2012	240	-0.8359	-0.5488	2647.49	2.1	1	0.32673377	no
9/10/2012	254	-0.9428	-0.3335	1093.04	1.5	0	0.3485297	yes
9/24/2012	268	-0.9951	-0.0988	985.65	1.2	0	0.39966359	yes
10/15/2012	289	-0.9657	0.2595	1253.89	2.8	1	0.5028456	yes
10/29/2012	303	-0.8759	0.4825	705.98	1.5	0	0.56948479	no
11/19/2012	324	-0.6486	0.7611	781.35	1.6	0	0.67221446	no
12/17/2012	352	-0.2219	0.9751	744.97	2.8	1	0.76817079	yes
1/14/2013	14	0.2387	0.9711	894.42	5.5	1	0.81308782	yes
2/11/2013	42	0.6616	0.7498	756.96	5.7	1	0.81325124	yes
3/11/2013	70	0.9338	0.3577	1403.56	21.1	1	0.77487249	yes
4/8/2013	98	0.9933	-0.1159	713.66	41.6	1	0.67366628	yes
5/6/2013	126	0.8264	-0.5632	3058.22	6.3	1	0.56297701	yes
5/20/2013	140	0.6681	-0.7441	2657.64	1.5	0	0.48640217	yes
6/3/2013	154	0.4712	-0.8820	6130.35	<1.0	0	0.46133248	yes
6/17/2013	168	0.2470	-0.9690	7780.09	<1.0	0	0.42497613	yes
7/1/2013	182	0.0086	-1.0000	4838.19	<1.0	0	0.35105076	yes
7/15/2013	196	-0.2303	-0.9731	1568.47	1.4	0	0.29325948	yes
7/25/2013	206	-0.3936	-0.9193	2009.52	2.3	1	0.2898885	no
8/5/2013	217	-0.5596	-0.8288	24344.62	7.2	1	0.5400753	yes
8/12/2013	224	-0.6552	-0.7555	8234.28	1	0	0.35957489	yes
8/19/2013	231	-0.7412	-0.6713	14567.8	3.2	1	0.44104315	no
9/9/2013	252	-0.9307	-0.3657	1392.34	1.9	0	0.34540886	yes
9/23/2013	266	-0.9911	-0.1330	1280.95	2	0	0.39469006	yes
10/21/2013	294	-0.9399	0.3416	1103.48	2.7	1	0.52706311	yes
11/18/2013	322	-0.6744	0.7383	1031.64	4.1	1	0.66602429	yes
12/9/2013	343	-0.3697	0.9291	4138.71	9.3	1	0.77244471	yes
12/16/2013	350	-0.2554	0.9668	4225.79	3	1	0.79151036	yes
12/23/2013	357	-0.1373	0.9905	2674.51	2.9	1	0.7950239	yes
1/13/2014	13	0.2219	0.9751	1940.3	1.9	0	0.81967206	no
2/10/2014	41	0.6486	0.7611	2632.32	2.2	1	0.82700137	yes
3/10/2014	69	0.9275	0.3737	2197.24	2.6	1	0.78367697	yes

4/7/2014	97	0.9951	-0.0988	1846.1	5	1	0.68953907	yes
5/5/2014	125	0.8359	-0.5488	2105.12	3.8	1	0.55707912	yes
5/19/2014	139	0.6808	-0.7325	1358.53	<1.0	0	0.47607262	yes
5/28/2014	148	0.5596	-0.8288	1967.79	1.8	0	0.43973723	yes
6/2/2014	153	0.4863	-0.8738	2180.67	1.7	0	0.41973299	yes
6/11/2014	162	0.3456	-0.9384	38493.89	5.9	1	0.77557013	yes
6/30/2014	181	0.0258	-0.9997	7769.93	1.5	0	0.38568662	yes
7/14/2014	195	-0.2135	-0.9769	2320.99	<1.0	0	0.30187947	yes
7/28/2014	209	-0.4405	-0.8977	2101.97	5.1	1	0.28994377	no
8/4/2014	216	-0.5452	-0.8383	1632.04	1.6	0	0.28604513	yes
8/11/2014	223	-0.6421	-0.7667	1943.05	2.3	1	0.2933906	no
8/25/2014	237	-0.8065	-0.5913	1131.06	3	1	0.30503918	no
9/8/2014	251	-0.9243	-0.3817	7997.12	1.4	0	0.41564388	yes
9/22/2014	265	-0.9887	-0.1501	2023.56	<1.0	0	0.39901066	yes
10/6/2014	279	-0.9959	0.0903	1484.34	1.5	0	0.45540872	yes
10/20/2014	293	-0.9456	0.3253	4648.23	3	1	0.56339105	yes
11/17/2014	321	-0.6871	0.7266	1212.24	1.5	0	0.66345192	no
12/15/2014	349	-0.2720	0.9623	3170.49	1.3	0	0.78066526	no
1/12/2015	12	0.2051	0.9787	2883.4	2.4	1	0.82532429	yes
2/9/2015	40	0.6354	0.7722	2349.39	2.8	1	0.82569391	yes
3/9/2015	68	0.9210	0.3896	1269.3	2.2	1	0.77846988	yes
4/6/2015	96	0.9967	-0.0817	1435.78	4.6	1	0.68956027	yes
5/4/2015	124	0.8452	-0.5344	2037.68	2	0	0.56150373	no
5/18/2015	138	0.6933	-0.7207	39000.54	3.7	1	0.84596825	yes
6/6/2015	157	0.4250	-0.9052	80628.89	4.2	1	0.96489755	yes
6/15/2015	166	0.2802	-0.9599	42779.19	<1.0	0	0.79914368	no
6/29/2015	180	0.0430	-0.9991	39832.05	8.3	1	0.74278394	yes

Definitions and National Water Information System (parameter code)

Geos: Geosmin in ng/L (51285)

Q: Streamflow in cubic feet per second (00060)