Appendix 16. Model Archival Summary for Sulfate Concentration at Station 06892350; Kansas River at De Soto, Kansas

This model archival summary summarizes the Sulfate concentration (Sulf) model developed to compute 15-minute Sulf from July 19, 2012 onward. This model supersedes all previous models.

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 was installed with sensors for water temperature, specific conductance, dissolved oxygen, pH, turbidity, and chlorophyll. 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. Linear regression models were developed using the open-source software package "R." Explanatory variables selected as inputs to linear 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 regression model is based on 59 concurrent measurements of Sulf concentration and specific conductance (SC) collected from July 19, 2012 through June 29, 2015. Samples were collected throughout the range of continuously observed hydrologic conditions. No samples were below laboratory detection limits. Summary statistics and the complete model-calibration dataset are provided below. Studentized residuals from the final model were inspected for values greater than 3 or less than negative 3. Values outside of that range are considered potential outliers and are investigated. One sample, June 11, 2014, was found to have potential errors in collection and processing, and has been removed from the dataset. All other potential outliers were not found to have errors associated with collection, processing, or analysis, and were therefore considered valid.

Sulfate 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 Sulf concentration at the USGS National Water Quality Laboratory in Lakewood, Colorado.

Model Development

Regression analysis was done using R by examining SC, streamflow, and other continuously measured data as explanatory variables for estimating Sulf concentration. A variety of models that predict Sulf, $(Sulf)^2$, $\sqrt{Sulf}$ and models that predict $\log_{10}(Sulf)$ were evaluated. The distribution of residuals was examined for normality, and plots of residuals (the difference between the measured and computed values) as compared to computed Sulf were examined for homoscedasticity (meaning that their departures from zero did not

change substantially over the range of computed values). This comparison lead to the conclusion that the most appropriate and reliable model would be one that estimated $\log_{10}$(Sulf).

SC was selected as the best predictor of Sulf based on residual plots, relatively high adjusted coefficient of determination (adjusted R^2) and relatively low model standard percentage error (*MSPE*), prediction error sum of squares (PRESS), and Mallow's C_p. 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 regression analysis for Sulf concentration at site number 06892350.

Sulf concentration-based model:

$$\log_{10}(Sulf) = 1.29 \times \log_{10}(SC) - 1.73$$

where

Sulf = sulfate in milligrams per liter (mg/L); and,

SC = specific conductance in microsiemens per centimeter at 25 degrees Celsius (µs/cm)

SC makes physical and statistical sense as explanatory variables for Sulf.

The log-transformed model may be retransformed to the original units so that Sulf can be calculated directly. The retransformation introduces a bias in the calculated constituent. This bias may be corrected using Duan's Bias Correction Factor (BCF). For this model, the calculated BCF is 1.01. The retransformed model, accounting for BCF is:

$$Sulf = SC^{1.29} + 0.0188$$

Previous Models

Start year	End year	Model
2000	2005	$\log_{10}(Sulf) = 1.24 \times \log_{10}(SC) - 1.57$

Sulfate Concentration Record

The Sulf 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

R Output for Sulfate; 06892350; Kansas River at De Soto, KS

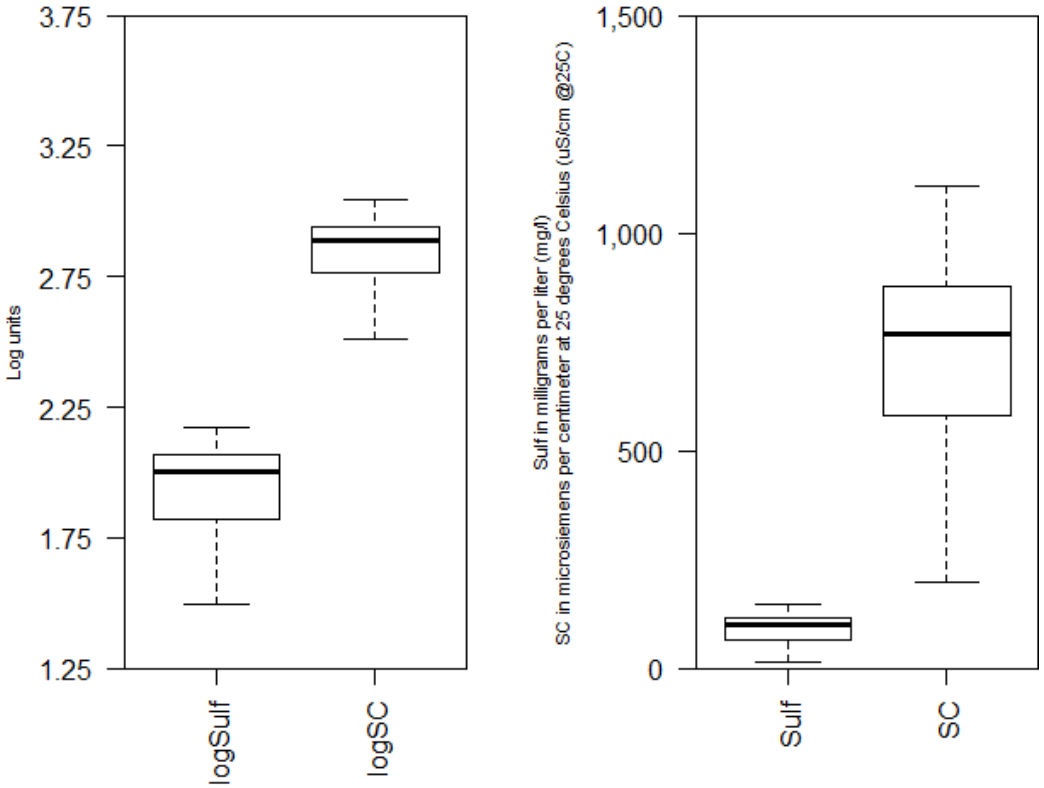
Model Form

logSulf = + 1.29 * logSC + -1.73

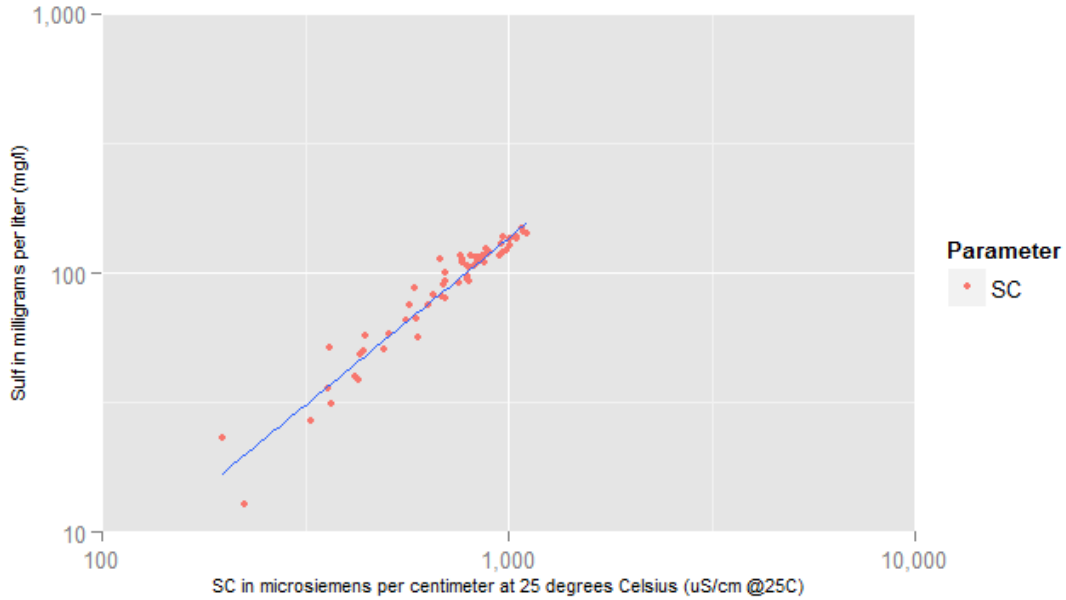
Variable Summary Statistics

	logSulf	logSC	Sulf	SC
Minimum	1.10	2.29	12.7	197
1st Quartile	1.82	2.76	65.7	569
Median	2.00	2.89	100.0	770
Mean	1.92	2.83	92.5	726
3rd Quartile	2.07	2.95	117.0	882
Maximum	2.17	3.05	149.0	1110

Box Plot(s) of sample data



Exploratory Plot



Model Calibration

Basic Data

Number of Observations	59
Standard error (RMSE)	0.0572
Upper Model standard percentage error (MSPE)	14.1
Lower Model standard percentage error (MSPE)	12.3
Coefficient of determination (R^2)	0.935
Adjusted Coefficient of Determination (Adj. R^2)	0.934
Bias Correction Factor (BCF)	1.01

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	-1.73	0.1280	-13.5	1.85e-19
logSC	1.29	0.0449	28.7	1.54e-35

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.998
E.vars	-0.998	1.000

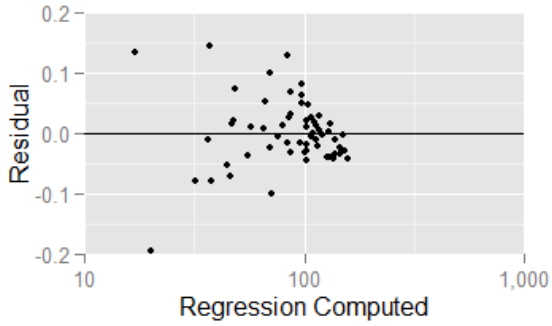
Test Criteria

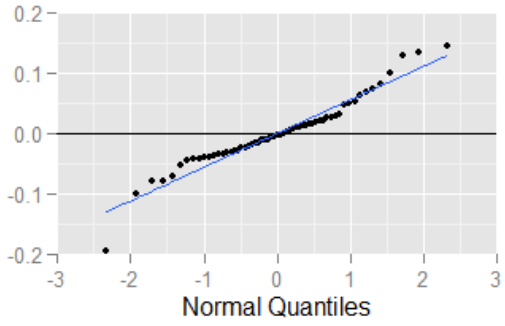
Leverage	Cook's D	DFFITS
0.05084746	0.10555214	0.26037782

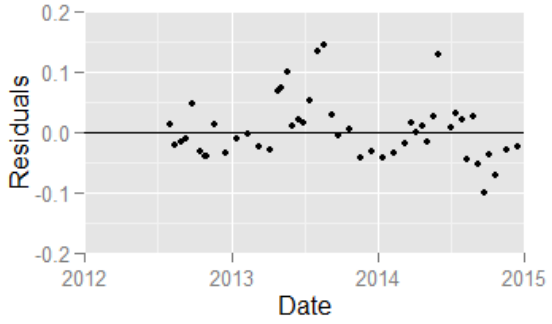
Flagged Observations

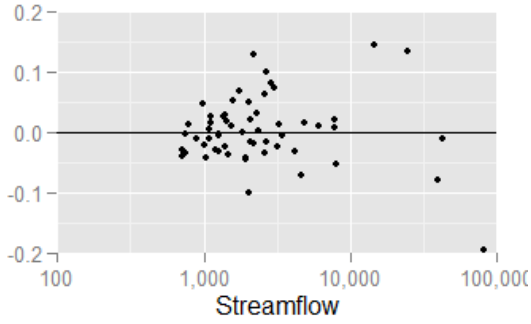
	logSulf	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
8/5/2013 11:30	1.361	1.227	0.13360	2.6040	2.7500	0.19630	0.828100	1.35900
8/19/2013 10:00	1.714	1.570	0.14340	2.5870	2.7290	0.06289	0.224600	0.70710
6/2/2014 13:00	2.051	1.922	0.12920	2.2770	2.3670	0.01695	0.044700	0.31080
10/20/2014 14:00	1.589	1.661	-0.07241	-1.2930	-1.3000	0.04226	0.036860	-0.27320
5/18/2015 15:30	1.429	1.507	-0.07844	-1.4290	-1.4430	0.08080	0.089790	-0.42780
6/6/2015 19:50	1.104	1.300	-0.19630	-3.7430	-4.2720	0.16050	1.340000	-1.86800
6/15/2015 14:50	1.551	1.563	-0.01221	-0.2205	-0.2186	0.06486	0.001686	-0.05758
6/29/2015 12:30	1.495	1.575	-0.07986	-1.4400	-1.4540	0.06167	0.068160	-0.37280

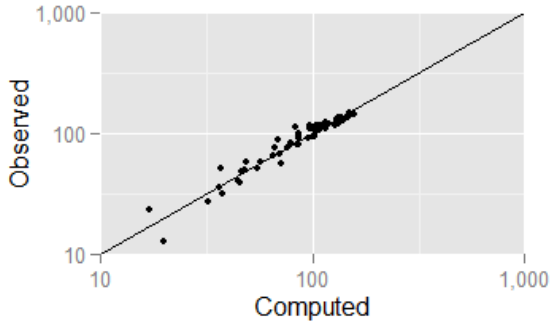
Statistical Plots

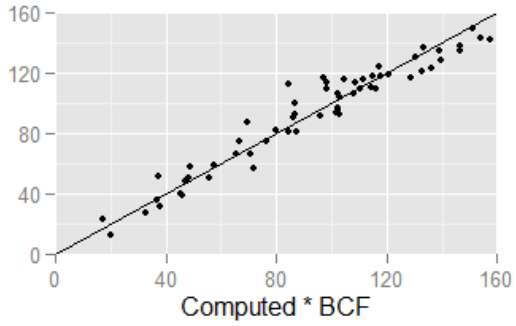


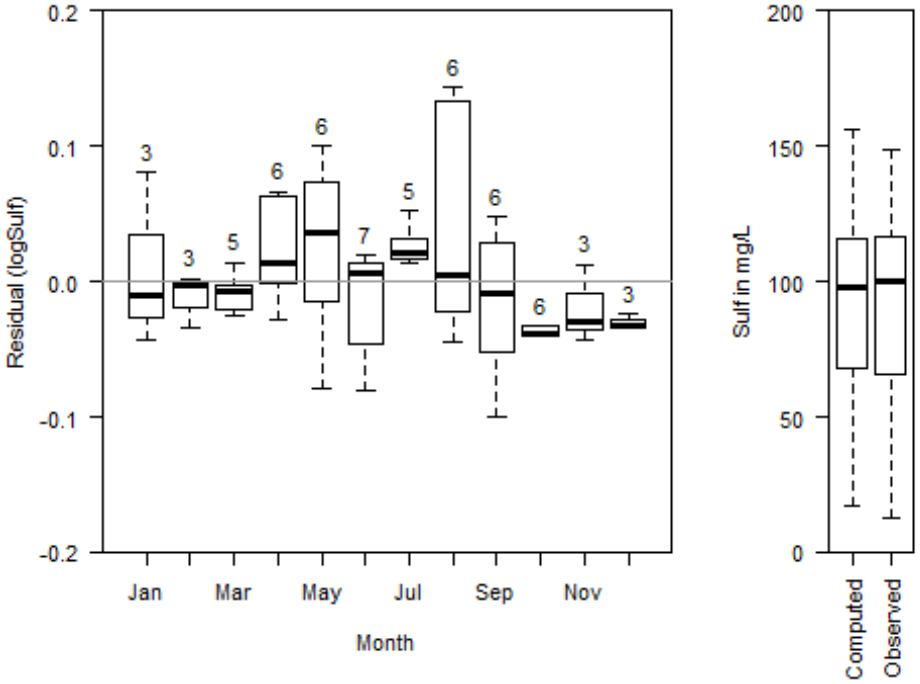


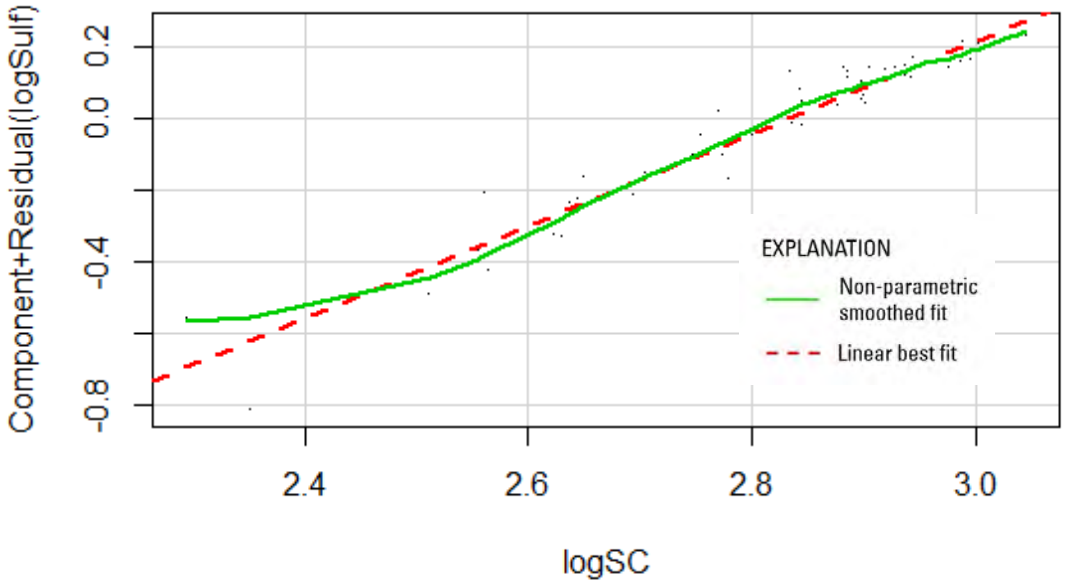









Residual (logSulf)
0.2
0.1
0.0
-0.1
-0.2
3
3
5
6
6
7
5
6
6
6
3
3
Jan
Mar
May
Jul
Sep
Nov
Month
Sulf in mg/L
200
150
100
50
0
Computed
Observed


Component+Residual(logSulf)
0.2
0.0
-0.4
-0.8
2.4
2.6
2.8
3.0
logSC
EXPLANATION
Non-parametric smoothed fit
Linear best fit

Models considered

Model Formula	Number of Variables	Standard Error	R2	Adjusted R2	Cp	PRESS	VIF	MSPE
logSulf ~ logSC	1	0.05724	93.51	93.4	40.49	0.2224	1	± 13
logSulf ~ SC	1	0.08259	86.49	86.25	143.8	0.4414	1	± 19
logSulf ~ logQ	1	0.1055	77.95	77.56	269.4	0.692	1	± 25
logSulf ~ Q + logSC	2	0.04798	95.52	95.36	12.91	0.154	2.089	± 11
logSulf ~ Q + SC	2	0.05415	94.29	94.09	30.94	0.1881	1.627	± 12
logSulf ~ logQ + logSC	2	0.05552	94	93.79	35.24	0.2116	4.375	± 13
logSulf ~ Q + logSC + sin2DY	3	0.04574	96	95.78	7.824	0.1417	2.193	± 11
logSulf ~ Q + logSC + cos2DY	3	0.04659	95.85	95.63	10.03	0.1501	2.106	± 11
logSulf ~ Q + logQ + logSC	3	0.04808	95.58	95.34	14	0.1555	3.472	± 11
logSulf ~ Q + logSC + sin2DY + cos2DY	4	0.04469	96.25	95.97	6.146	0.1392	2.222	± 10
logSulf ~ Q + SC + logSC + sin2DY	4	0.04578	96.07	95.78	8.86	0.1784	3.04	± 11
logSulf ~ Q + SC + logSC + cos2DY	4	0.04581	96.06	95.77	8.933	0.168	2.957	± 11
logSulf ~ Q + SC + logSC + sin2DY + cos2DY	5	0.0439	96.45	96.12	5.225	0.1606	3.06	± 10
logSulf ~ Q + logQ + logSC + sin2DY + cos2DY	5	0.04511	96.25	95.9	8.131	0.1477	3.5	± 10
logSulf ~ Q + logQ + SC + logSC + cos2DY	5	0.04562	96.17	95.8	9.396	0.1645	4.84	± 11
logSulf ~ Q + logQ + SC + logSC + sin2DY + cos2DY	6	0.04423	96.47	96.06	7	0.1653	4.86	± 10

Data

	Date	logSulf	logSC	Sulf	SC	Computed logSulf	Computed Sulf	Residual	Normal Quantiles
1	2012-07-30	1.913	2.816	81.82	655	1.899	79.97	0.0136	0.391
2	2012-08-13	2.039	2.942	109.3	875.5	2.062	116.2	-0.0229	-0.391
3	2012-08-27	1.961	2.877	91.31	753.2	1.977	95.73	-0.0169	-0.301
4	2012-09-10	2.043	2.938	110.5	866	2.055	114.6	-0.0122	-0.17
5	2012-09-24	2.063	2.906	115.6	805.3	2.015	104.3	0.048	0.919
6	2012-10-15	1.972	2.898	93.81	791	2.005	102	-0.0325	-0.633
7	2012-10-26	2.081	2.987	120.4	971	2.119	132.8	-0.0388	-0.919
8	2012-10-29	2.065	2.977	116.3	948	2.106	128.7	-0.0406	-0.986
9	2012-11-19	2.069	2.938	117.2	867	2.056	114.7	0.0127	0.345
10	2012-12-17	2.107	3.004	127.9	1010	2.141	139.7	-0.0346	-0.797
11	2013-01-14	2.13	3.003	134.8	1008	2.14	139.3	-0.0105	-0.127
12	2013-02-11	2.173	3.031	148.9	1075	2.176	151.4	-0.00343	0
13	2013-03-11	2.139	3.021	137.6	1050	2.163	146.8	-0.0245	-0.484
14	2013-04-08	2.156	3.037	143.1	1090	2.184	154.1	-0.0285	-0.532
15	2013-04-25	2.001	2.843	100.2	697	1.934	86.63	0.0667	1.22
16	2013-05-06	1.757	2.649	57.17	446	1.684	48.75	0.0728	1.31
17	2013-05-20	1.94	2.77	87.06	589.2	1.84	69.78	0.0997	1.55
18	2013-06-03	1.766	2.706	58.38	508	1.757	57.65	0.00912	0.257
19	2013-06-17	1.698	2.644	49.84	441	1.678	48.05	0.0196	0.582
20	2013-07-01	1.684	2.637	48.35	433.5	1.668	47	0.016	0.484
21	2013-07-15	1.873	2.755	74.7	569.4	1.821	66.77	0.0524	1.06
22	2013-08-05	1.361	2.294	22.96	197	1.227	17.02	0.134	1.93
23	2013-08-19	1.714	2.561	51.73	363.8	1.57	37.5	0.143	2.32
24	2013-09-09	2.094	2.945	124.1	882	2.066	117.3	0.028	0.797
25	2013-09-23	2.024	2.917	105.7	826	2.029	107.8	-0.00484	-0.0424
26	2013-10-21	2.072	2.947	118	886	2.068	118	0.00378	0.17
27	2013-11-18	2.088	2.996	122.4	990.6	2.131	136.2	-0.0428	-1.06
28	2013-12-16	1.905	2.846	80.29	701	1.937	87.27	-0.0326	-0.685
29	2014-01-13	2.151	3.045	141.5	1110	2.194	157.7	-0.0435	-1.13
30	2014-02-10	2.13	3.021	134.8	1050	2.163	146.8	-0.0336	-0.74
31	2014-03-10	1.986	2.899	96.86	793.3	2.006	102.3	-0.0202	-0.345

32	2014-03-24	2.135	2.988	136.5	973.2	2.121	133.2	0.0143	0.437
33	2014-04-07	2.038	2.925	109.2	841.2	2.039	110.4	-0.000834	0.0848
34	2014-04-21	2.019	2.902	104.5	797.8	2.01	103.1	0.00935	0.301
35	2014-05-05	1.909	2.836	81.09	684.9	1.924	84.69	-0.0152	-0.257
36	2014-05-19	2.056	2.919	113.9	830.2	2.032	108.5	0.0245	0.685
37	2014-06-02	2.051	2.834	112.6	682.3	1.922	84.29	0.129	1.71
38	2014-06-30	1.817	2.748	65.69	559.7	1.811	65.32	0.00611	0.213
39	2014-07-14	1.968	2.845	92.8	699.3	1.936	87	0.0317	0.856
40	2014-07-28	2.027	2.899	106.5	792.8	2.006	102.3	0.0214	0.633
41	2014-08-11	1.965	2.902	92.23	798.2	2.01	103.2	-0.045	-1.22
42	2014-08-25	1.957	2.841	90.56	693.4	1.931	86.06	0.0258	0.74
43	2014-09-08	1.598	2.623	39.66	420.2	1.651	45.15	-0.0526	-1.31
44	2014-09-22	1.752	2.779	56.46	601.5	1.852	71.66	-0.0999	-1.93
45	2014-10-06	1.705	2.695	50.7	495.5	1.743	55.83	-0.0382	-0.856
46	2014-10-20	1.589	2.631	38.78	427.8	1.661	46.2	-0.0724	-1.42
47	2014-11-17	1.977	2.899	94.82	792.8	2.006	102.3	-0.0292	-0.582
48	2014-12-15	1.822	2.775	66.3	595.1	1.846	70.67	-0.0241	-0.437
49	2015-01-12	2.066	2.883	116.4	763	1.985	97.34	0.0812	1.42
50	2015-02-09	2.114	2.981	130.1	958	2.112	130.5	0.00223	0.127
51	2015-03-09	2.076	2.955	119.2	902	2.078	120.7	-0.00208	0.0424
52	2015-03-23	1.873	2.801	74.67	633	1.88	76.53	-0.00701	-0.0848
53	2015-04-06	2.062	2.928	115.4	848	2.044	111.5	0.0187	0.532
54	2015-04-20	2.053	2.886	113	769.5	1.989	98.4	0.0637	1.13
55	2015-05-04	2.038	2.886	109	769.3	1.989	98.36	0.0484	0.986
56	2015-05-18	1.429	2.512	26.86	325.1	1.507	32.45	-0.0784	-1.55
57	2015-06-07	1.104	2.351	12.71	224.5	1.3	20.14	-0.196	-2.32
58	2015-06-15	1.551	2.555	35.54	359	1.563	36.87	-0.0122	-0.213
59	2015-06-29	1.495	2.565	31.28	366.9	1.575	37.91	-0.0799	-1.71

Definitions and National Water Information System (parameter code)

Sulf: Sulfate in mg/L (00945)
SC: Specific conductance in uS/cm @25C (00095)