

Appendix 15. Model Archival Summary for Sulfate Concentration at Station 06887500; Kansas River at Wamego, 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: 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 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 55 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. None of the Sulf samples were deemed outliers.

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, $(\text{Sulf})^2$, $\sqrt{\text{Sulf}}$ and models that predict $\log_{10}(\text{Sulf})$ 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 led to the conclusion that the most appropriate and reliable model would be one that estimated $\log_{10}(\text{Sulf})$.

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 06887500.

Sulf concentration-based model:

$$\log_{10}(Sulf) = 1.05 \times \log_{10}(SC) - 1.06$$

where

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

SC = specific conductance in microsiemens per centimeter at 25 degrees Celsius ($\mu\text{s}/\text{cm}$)

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:

$$Sulf = SC^{1.05} + 0.088$$

Previous Models

Start year	End year	Model
2000	2005	$\log_{10}(Sulf) = 1.05 \times \log_{10}(SC) - 1.05$

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; 06887500; Kansas River at Wamego, KS

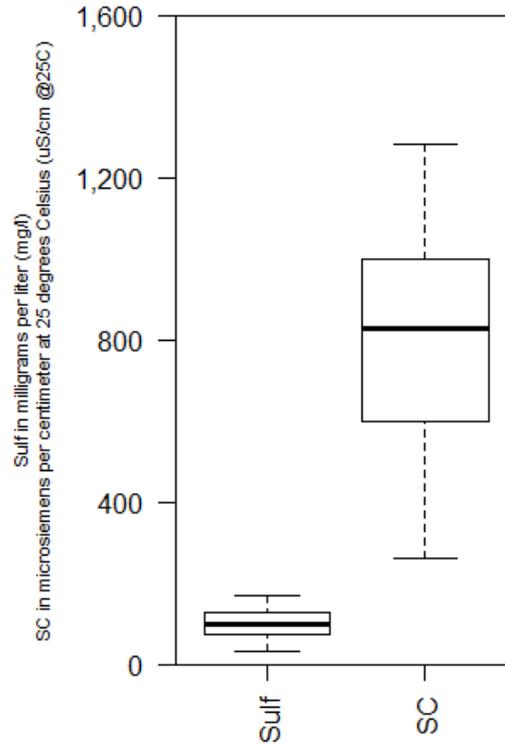
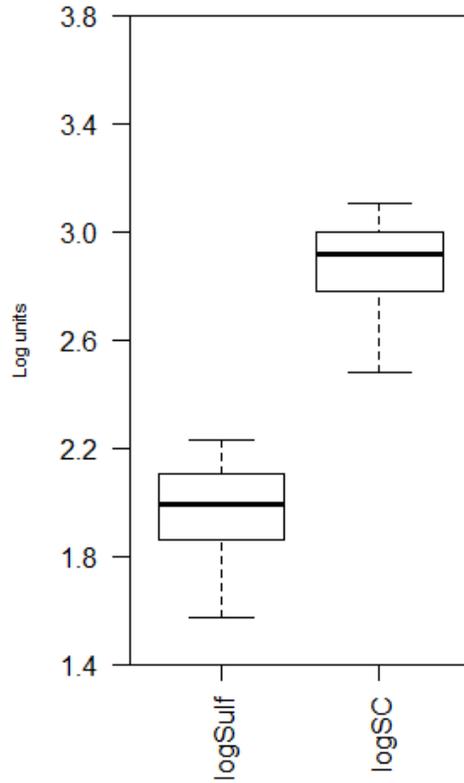
Model Form

$$\log\text{Sulf} = + 1.05 * \log\text{SC} + -1.06$$

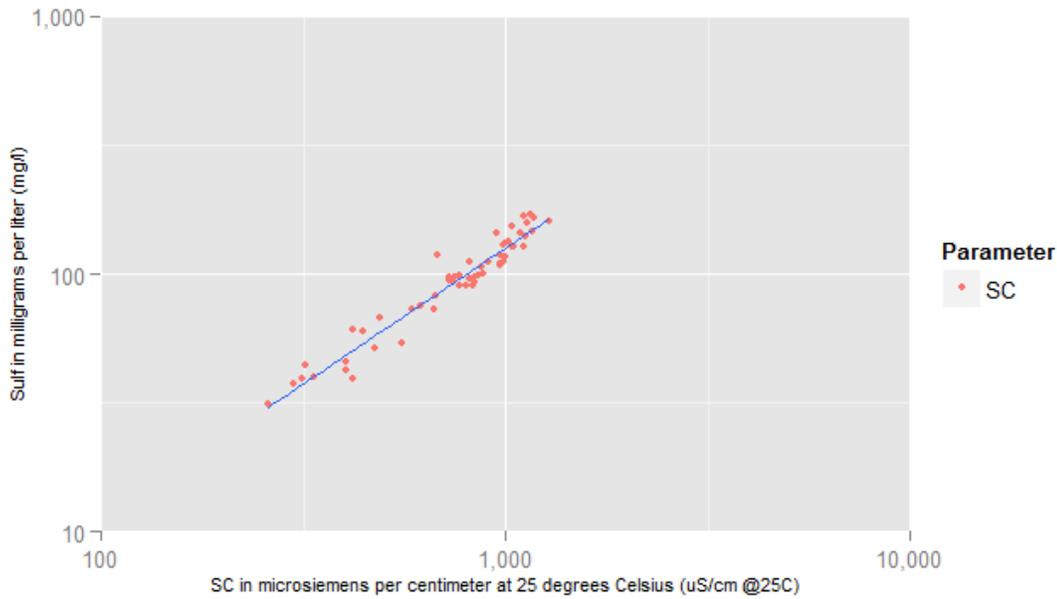
Variable Summary Statistics

	logSulf	logSC	Sulf	SC
Minimum	1.50	2.41	31.4	260
1st Quartile	1.86	2.77	72.5	585
Median	1.99	2.92	98.3	829
Mean	1.96	2.87	99.7	793
3rd Quartile	2.11	3.00	128.0	1000
Maximum	2.23	3.11	170.0	1280

Box Plot(s) of sample data



Exploratory Plot



Model Calibration

Basic Data

Number of Observations	55
Standard error (RMSE)	0.049
Upper Model standard percentage error (MSPE)	12
Lower Model standard percentage error (MSPE)	10.7
Coefficient of determination (R^2)	0.938
Adjusted Coefficient of Determination (Adj. R^2)	0.937
Bias Correction Factor (BCF)	1.01

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(> t)
(Intercept)	-1.06	0.1070	-9.9	1.21e-13
logSC	1.05	0.0372	28.3	1.20e-33

Correlation Matrix

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

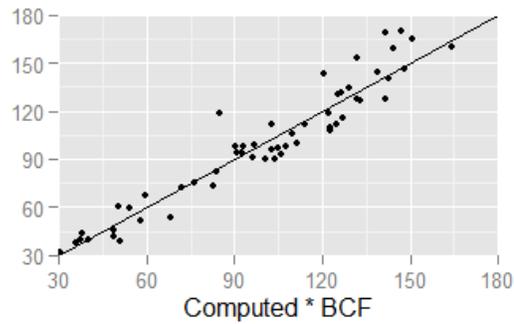
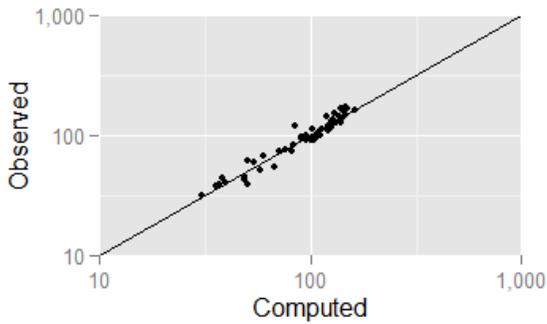
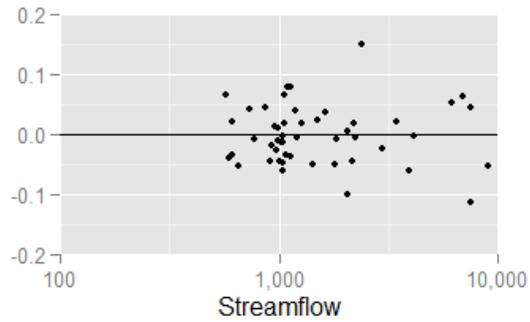
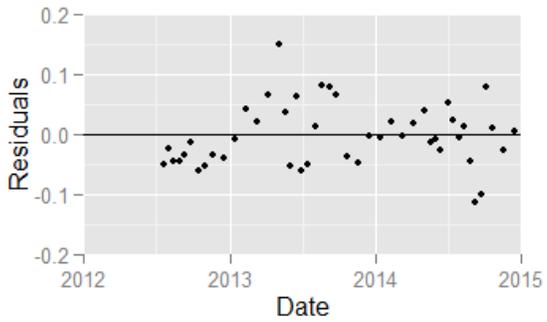
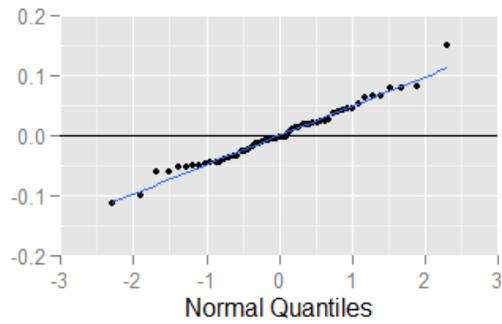
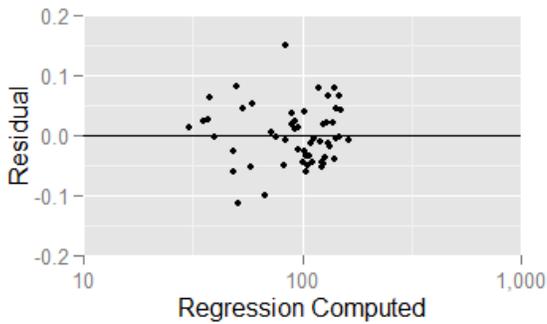
Test Criteria

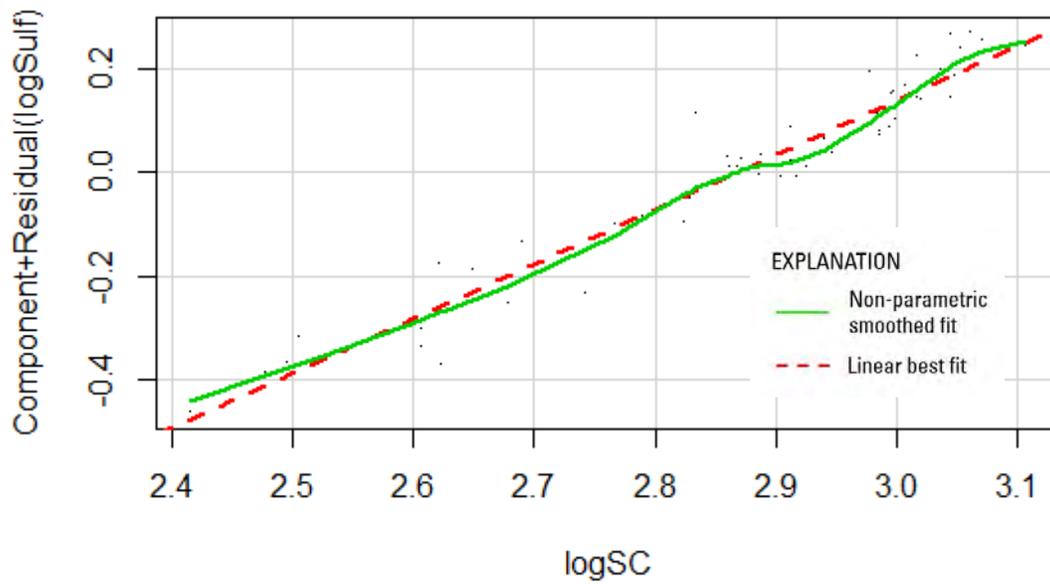
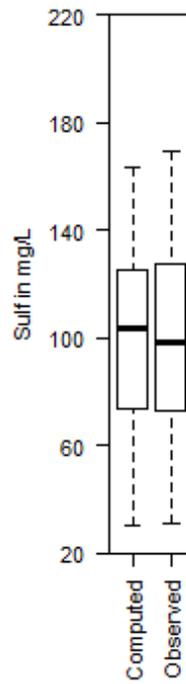
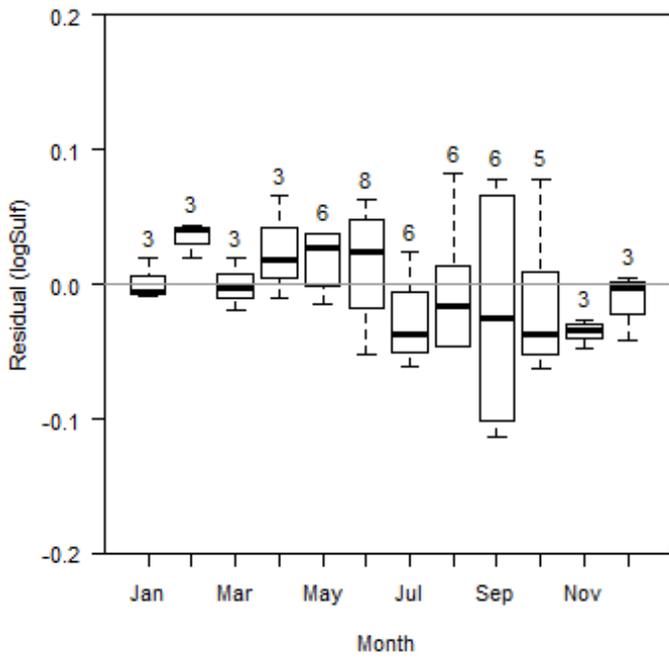
Leverage	Cook's D	DFFITS
0.05454545	0.10556635	0.26967994

Flagged Observations

	logSulf	Estimate	Residual	Standard	Residual	Studentized	Residual	Leverage	Cook's D	DFFITS
4/8/2013 7:40	2.230	2.164	0.065590		1.36500		1.37700	0.03977	3.859e-02	0.28020
5/6/2013 8:00	2.075	1.925	0.149400		3.07600		3.36200	0.01884	9.086e-02	0.46590
6/17/2013 8:30	1.642	1.580	0.062900		1.34700		1.35800	0.09374	9.390e-02	0.43680
7/1/2013 8:10	1.624	1.686	-0.061380		-1.28900		-1.29800	0.05752	5.074e-02	-0.32060
8/5/2013 7:30	1.497	1.484	0.013130		0.28820		0.28570	0.13650	6.562e-03	0.11360
8/19/2013 9:30	1.783	1.701	0.081940		1.71700		1.75100	0.05319	8.286e-02	0.41490
6/11/2014 9:00	1.659	1.686	-0.026600		-0.55870		-0.55510	0.05752	9.526e-03	-0.13710
9/8/2014 8:00	1.591	1.704	-0.113100		-2.37000		-2.48300	0.05253	1.557e-01	-0.58460
9/22/2014 9:20	1.728	1.830	-0.101400		-2.09600		-2.16800	0.02713	6.125e-02	-0.36200
10/6/2014 9:00	2.226	2.148	0.078430		1.62900		1.65600	0.03640	5.015e-02	0.32190
5/18/2015 10:30	1.600	1.601	-0.001914		-0.04081		-0.04042	0.08533	7.767e-05	-0.01235
6/15/2015 7:50	1.595	1.569	0.025710		0.55210		0.54840	0.09781	1.652e-02	0.18060
6/29/2015 8:20	1.573	1.550	0.023770		0.51270		0.50910	0.10610	1.560e-02	0.17540

Statistical Plots





Models considered

Model Formula	Number of Variables	Standard Error	R2	Adjusted R2	Cp	PRESS	VIF	MSPE
logSulf ~ logSC	1	0.04903	93.78	93.67	5.62	0.1369	1 ± 11	
logSulf ~ SC	1	0.05518	92.13	91.98	20.7	0.1751	1 ± 13	
logSulf ~ logQ	1	0.1044	71.82	71.29	205.7	0.6126	1 ± 24	
logSulf ~ logSC + sin2DY	2	0.0463	94.56	94.35	0.5407	0.1261	1.001 ± 11	
logSulf ~ SC + logSC	2	0.04854	94.02	93.79	5.434	0.1353	27.19 ± 11	
logSulf ~ logQ + logSC	2	0.04866	93.99	93.76	5.721	0.1358	5.074 ± 11	
logSulf ~ SC + logSC + sin2DY	3	0.04644	94.63	94.32	1.876	0.1289	29.16 ± 11	
logSulf ~ logQ + logSC + sin2DY	3	0.04651	94.62	94.3	2.024	0.1288	5.432 ± 11	
logSulf ~ Q + logSC + sin2DY	3	0.04659	94.6	94.28	2.199	0.1298	3.022 ± 11	
logSulf ~ SC + logSC + sin2DY + cos2DY	4	0.04664	94.7	94.27	3.324	0.1302	31.9 ± 11	
logSulf ~ logQ + SC + logSC + sin2DY	4	0.04679	94.66	94.23	3.648	0.1324	6.084 ± 11	
logSulf ~ logQ + logSC + sin2DY + cos2DY	4	0.04683	94.65	94.22	3.714	0.1307	5.505 ± 11	
logSulf ~ logQ + SC + logSC + sin2DY + cos2DY	5	0.04699	94.72	94.18	5.081	0.1338	6.087 ± 11	
logSulf ~ Q + SC + logSC + sin2DY + cos2DY	5	0.04711	94.7	94.15	5.32	0.1364	4.919 ± 11	
logSulf ~ Q + logQ + SC + logSC + sin2DY	5	0.04724	94.67	94.12	5.597	0.1385	7.528 ± 11	
logSulf ~ Q + logQ + SC + logSC + sin2DY + cos2DY	6	0.04744	94.73	94.07	7	0.1396	7.574 ± 11	

Data

	Date	logSulf	logSC	Sulf	SC	Computed logSulf	Computed Sulf	Residual	Normal Quantiles
0									
1	2012-07-19	1.97	2.925	93.39	842	2.022	105.7	-0.0512	-1.18
2	2012-07-30	1.957	2.886	90.53	770	1.981	96.24	-0.0238	-0.421
3	2012-08-13	2.04	2.987	109.7	970	2.086	122.7	-0.0459	-0.87
4	2012-08-27	1.999	2.946	99.8	884	2.044	111.3	-0.0446	-0.806
5	2012-09-10	1.993	2.931	98.32	854	2.028	107.3	-0.0354	-0.63
6	2012-09-24	2.024	2.941	105.7	872.3	2.038	109.7	-0.0137	-0.276
7	2012-10-15	1.952	2.918	89.51	828.7	2.014	104	-0.0623	-1.67
8	2012-10-29	2.034	2.987	108.2	970.5	2.086	122.8	-0.0522	-1.38
9	2012-11-19	1.985	2.923	96.52	837	2.019	105.1	-0.0342	-0.575
10	2012-12-17	2.107	3.045	127.8	1110	2.148	141.4	-0.0413	-0.745
11	2013-01-14	2.204	3.107	159.8	1280	2.213	164.3	-0.00938	-0.183
12	2013-02-11	2.216	3.072	164.5	1180	2.176	150.8	0.0404	0.87
13	2013-03-11	2.159	3.037	144.2	1090	2.14	138.8	0.0193	0.471
14	2013-04-08	2.23	3.061	169.8	1151	2.164	146.9	0.0656	1.27
15	2013-05-06	2.075	2.834	118.7	682	1.925	84.69	0.149	2.29
16	2013-05-20	1.989	2.861	97.57	726	1.954	90.46	0.0356	0.745
17	2013-06-03	1.709	2.678	51.2	476.6	1.761	58.07	-0.052	-1.27
18	2013-06-17	1.642	2.506	43.9	320.3	1.58	38.22	0.0629	1.18
19	2013-07-01	1.624	2.606	42.1	404	1.686	48.8	-0.0614	-1.51
20	2013-07-15	1.864	2.824	73.16	667.5	1.915	82.8	-0.0511	-1.09
21	2013-08-05	1.497	2.415	31.4	259.8	1.484	30.66	0.0131	0.276
22	2013-08-19	1.783	2.621	60.7	418	1.701	50.58	0.0819	1.9
23	2013-09-09	2.156	2.979	143.3	952	2.078	120.3	0.0785	1.67
24	2013-09-23	2.184	3.017	152.8	1040	2.118	132.1	0.0662	1.38
25	2013-10-21	2.063	3	115.6	1000	2.1	126.7	-0.0374	-0.686
26	2013-11-18	2.047	2.994	111.5	987.2	2.094	125	-0.0472	-1.01
27	2013-12-16	1.877	2.79	75.3	616.3	1.879	76.13	-0.00205	0.0454
28	2014-01-13	2.147	3.049	140.2	1120	2.152	142.8	-0.00526	-0.0454
29	2014-02-10	2.127	3.007	134.1	1017	2.108	129	0.0194	0.522
30	2014-03-10	2.165	3.064	146.3	1160	2.168	148.2	-0.00275	0
31	2014-04-07	2.115	2.996	130.3	991.4	2.096	125.6	0.0187	0.323

32	2014-05-05	2.047	2.913	111.4	818.7	2.009	102.7	0.0382	0.806
33	2014-05-19	2.104	3.017	127.1	1040	2.118	132.1	-0.014	-0.323
34	2014-06-02	1.912	2.829	81.65	675.1	1.921	83.8	-0.00855	-0.137
35	2014-06-11	1.659	2.606	45.61	404	1.686	48.8	-0.0266	-0.522
36	2014-06-30	1.827	2.69	67.16	490.3	1.774	59.83	0.0529	1.09
37	2014-07-14	1.989	2.872	97.43	744.4	1.965	92.87	0.0235	0.575
38	2014-07-28	2.049	2.957	111.8	905	2.055	114.1	-0.00591	-0.0909
39	2014-08-11	1.995	2.888	98.85	773	1.982	96.63	0.0125	0.229
40	2014-08-25	1.954	2.905	89.88	803	2	100.6	-0.0461	-0.939
41	2014-09-08	1.591	2.624	38.96	420.2	1.704	50.87	-0.113	-2.29
42	2014-09-22	1.728	2.743	53.48	553.3	1.83	67.96	-0.101	-1.9
43	2014-10-06	2.226	3.045	168.4	1110	2.148	141.4	0.0784	1.51
44	2014-10-20	1.971	2.869	93.58	740	1.962	92.29	0.00871	0.183
45	2014-11-17	1.983	2.913	96.07	818.5	2.009	102.6	-0.026	-0.471
46	2014-12-15	1.86	2.767	72.5	585.2	1.855	72.09	0.00516	0.137
47	2015-01-12	2.118	2.999	131.1	996.7	2.099	126.3	0.019	0.421
48	2015-02-09	2.2	3.053	158.6	1130	2.156	144.1	0.0442	1.01
49	2015-03-09	2.102	3.02	126.6	1048	2.122	133.1	-0.0192	-0.372
50	2015-04-06	2.074	2.985	118.7	965.5	2.084	122.1	-0.00981	-0.229
51	2015-05-04	1.974	2.862	94.14	728	1.955	90.72	0.0188	0.372
52	2015-05-18	1.6	2.526	39.77	336	1.601	40.19	-0.00191	0.0909
53	2015-06-01	1.774	2.649	59.44	445.5	1.73	54.09	0.0437	0.939
54	2015-06-15	1.595	2.496	39.37	313.3	1.569	37.34	0.0257	0.686
55	2015-06-29	1.573	2.477	37.44	300	1.55	35.67	0.0238	0.63

Definitions and National Water Information System (parameter code)

Sulf: Sulfate in mg/L (00945)

SC: Specific conductance in uS/cm @25C (00095)