Appendix 7. Model Archive Summary for Nitrate Plus Nitrite at U.S. Geological Survey Site 07144780, North Fork Ninnescah River above Cheney Reservoir, Kansas, during January 1, 1999, through December 31, 2019

This model archive summary summarizes the nitrate plus nitrite (NO_3NO_2) model developed to compute hourly or daily nitrate plus nitrite concentrations during January 1, 1999, onward. This model supersedes all prior models used during this period. The methods used follow U.S. Geological Survey (USGS) guidance as referenced in relevant Office of Surface Water/Office of Water Quality Technical Memoranda and USGS Techniques and Methods, book 3, chapter C4 (Rasmussen and others, 2009).

Site and Model Information

Site number: 07144780
Site name: North Fork Ninnescah River above Cheney Reservoir, Kansas
Location: Lat 37°51'45", long 98°00'49" referenced to North American Datum of 1927, in NE 1/4 SE 1/4 NE 1/4 sec.19, T.25 S., R.6 W., Reno County, Kans., Hydrologic Unit 11030014, on right bank at upstream side of county highway bridge, 10 miles south of Hutchinson, 18.1 miles upstream from Cheney Dam.

Equipment: A YSI 6600 Extended Deployment System water-quality monitor equipped with sensors for water temperature, specific conductance, pH, dissolved oxygen, and turbidity (a YSI Model 6026 turbidity sensor [November 9, 1998, to December 1, 2010] and a YSI Model 6136 turbidity sensor [October 17, 2009, to November 12, 2015; March 31, 2017, to June 7, 2017]) (YSI Incorporated, 2007, 2012a). The YSI 6600 water-quality monitor was in operation during November 9, 1998, through November 12, 2015.

A Xylem YSI EXO2 water-quality monitor equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, and turbidity (YSI Incorporated, 2012b). The YSI EXO2 water-quality monitor began operation on November 13, 2015. Monitors were housed in a 4-inch diameter polyvinyl chloride (PVC) pipe and placed in a location representative of the stream cross section. Monitor readings were recorded and satellite transmitted hourly.

Date model was developed: April 26, 2019

Model calibration data period: January 26, 1999, to September 28, 2017

Model Data

All data were collected using USGS protocols (U.S. Geological Survey, 2006; Wagner and others, 2006; Sauer and Turnipseed, 2010; Turnipseed and Sauer, 2010) and are stored in the National Water Information System (NWIS) database (https://doi.org/10.5066/F7P55KJN; U.S. Geological Survey, 2020). Explanatory variables were evaluated individually and in combination. Potential explanatory variables included streamflow, water temperature, specific conductance, pH, dissolved oxygen, and turbidity. Seasonal components (sine and cosine variables) were also evaluated as explanatory variables.

The regression model is based on 123 concomitant values of discretely collected nitrate plus nitrite samples and continuously measured specific conductance during January 26, 1999, through September 28, 2017. Discrete samples were collected over a range of streamflows. Two samples were less than the minimum reporting level (less than [<] 0.02 and <0.01 milligram per liter); therefore, a Tobit regression model was developed to compute estimates of nitrate plus nitrite using the absolute maximum likelihood estimation approach (Hald, 1949; Cohen, 1950; Tobin, 1958; Helsel and others, 2020). Summary statistics and the complete model-calibration data are provided below. Potential outliers were identified using the methods described in Rasmussen and others (2009). Additionally, outlier test criteria, including leverage and Cook's distance (Cook's D), were used to estimate potential outlier effect on the final Tobit regression model (Cook, 1977). The sample collected on March 22, 2006, had a large specific conductance value likely from previous road salt application and was removed from the model calibration dataset.

Nitrate Plus Nitrite

Discrete samples were collected from the downstream side of the bridge or instream within 50 feet of the bridge using equal-width-increment, multiple vertical, single vertical, or grab methods following U.S. Geological Survey (2006) and Rasmussen and others (2014). Discrete samples were collected on a semifixed to event-based schedule ranging from 2 to 17 samples per year with a Federal Interagency Sedimentation Project U.S. DH–95 or D–95 with a Teflon bottle, cap, and nozzle depth-integrating sampler; a DH–81 with a Teflon bottle, cap, and nozzle hand sampler; or a grab sample with a Teflon bottle depending on sample location. Samples were analyzed for nitrate plus nitrite by the Wichita Municipal Water and Wastewater Laboratory in Wichita, Kans., according to standard methods (American Public Health Association and others, 1995).

Continuous Data

Specific conductance was measured with a YSI 6600 sensor during November 9, 1998, through November 12, 2015, and a YSI EXO2 sensor during November 13, 2015, through December 31, 2019. Concomitant specific conductance values were time interpolated. If continuous data was not available (2 or more hours of specific conductance values bracketing the sample collection time were missing) because of fouling, changes in equipment, or unsuitable site conditions, then the field monitor specific conductance value measured sampling was substituted. If no concomitant continuous data were available, the sample was not included in the dataset.

Model Development

Stepwise regression analysis was done using R programming language (R Core Team, 2019) to relate discretely collected nitrate plus nitrite to specific conductance and other continuously measured data. The distribution of residuals was examined for normality and plots of residuals (the difference between the measured and model calculated values) compared to model calculated nitrate plus nitrite were examined for homoscedasticity (departures from zero did not change substantially over the range of model calculated values).

A total of 1.6 percent of the model-calibration dataset consisted of censored results (less than minimum reporting level). Tobit regression models were developed using absolute maximum likelihood estimation methods using the *smwrQW* (v.0.7.9) package in R programming language (R Core Team, 2019).

Specific conductance and seasonality were selected as the best predictors of nitrate plus nitrite based on residual plots, a higher pseudocoefficient of determination (pseudo-R^2), and relatively low estimated standard residual error (RSE). Seasonality was included as an explanatory variable because nitrate plus nitrite seems to have a cyclical pattern potentially affected by groundwater during low seasonal flow.

Model Summary

Summary of final nitrate plus nitrite regression analysis at USGS site 07144780:

Nitrate plus nitrite-based model:

$$NO3NO2 = 0.000399 \times SPC + 0.0224 \times \sin(2\pi D) + 0.563 \times \cos(2\pi D) + 0.533\ ,$$

where,

$NO3NO2$ = nitrate plus nitrite, filtered, in milligrams per liter as nitrogen;
SPC = specific conductance in microsiemens per centimeter at 25 degrees Celsius; and
D = date in decimal years.

Model Statistics, Plots and Data

Definitions for terms used in this output can be found at the end of this document.

Model

$$NO3NO2 = 0.000399 \times SPC + 0.0224 \times \sin(2\pi D) + 0.563 \times \cos(2\pi D) + 0.533$$

Computation method: Absolute Maximum Likelihood Estimation (AMLE)

Explanatory Variables

```
Coefficients:
             Estimate Std. Error z-score p-value
(Intercept) 0.5333743  8.564e-02  6.2283  0.0000
SPC         0.0003986  8.143e-05  4.8953  0.0000
sin2piD     0.0223780  3.981e-02  0.5621  0.5661
cos2piD     0.5626100  5.281e-02 10.6530  0.0000
```

Basic Model Statistics

```
Estimated residual standard error (Unbiased) = 0.3099
Number of observations = 123, number censored = 2 (1.6 percent)

Log-likelihood (model) = -30.5
Log-likelihood (intercept only) = -86.01
  Chi-square = 111
  degrees of freedom = 3
  p-value = <0.0001

Computation method: AMLE

Pseudo-R-squared: 0.5962

  Akaike Information Criterion: 71
  Bayesian Information Criterion: 85.06


Variance inflation factors
    SPC 1.06
sin2piD 1.01
cos2piD 1.06
```

Outlier Test Criteria

```
leverage   cooksD
 0.07317  0.84393
```

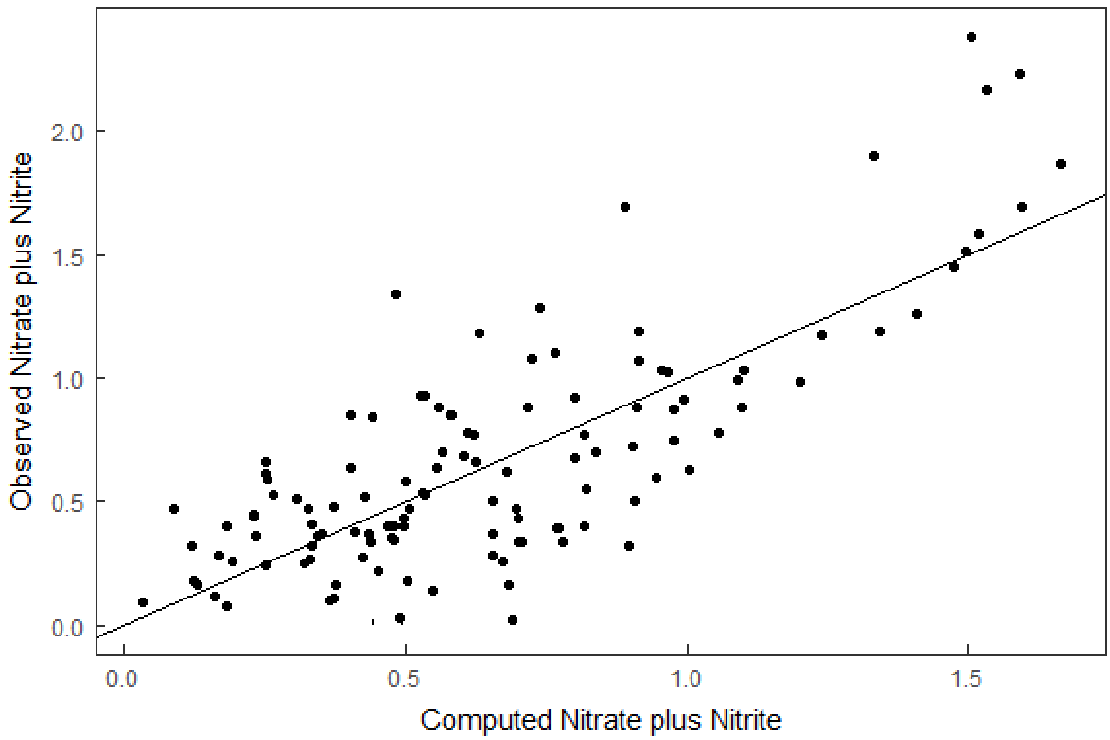
Flagged Observations

```
    Observations exceeding at least one test criterion
   nitratenitrite  ycen   yhat  resids leverage  cooksD
26           0.32 FALSE 0.8964 -0.5764  0.08009 0.08185
```

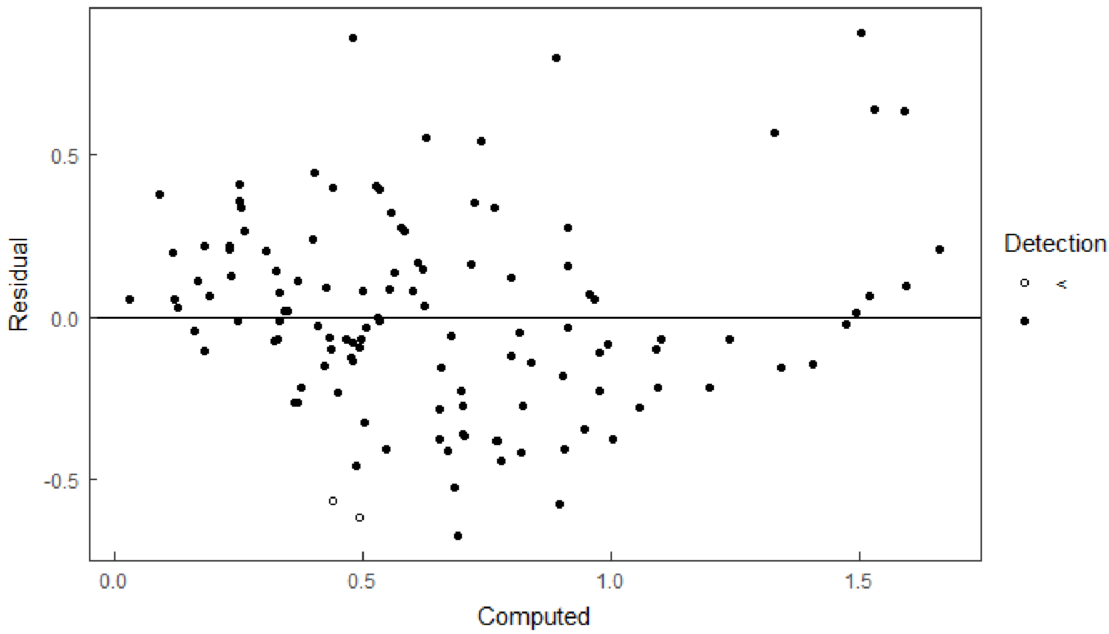
95% Confidence Intervals

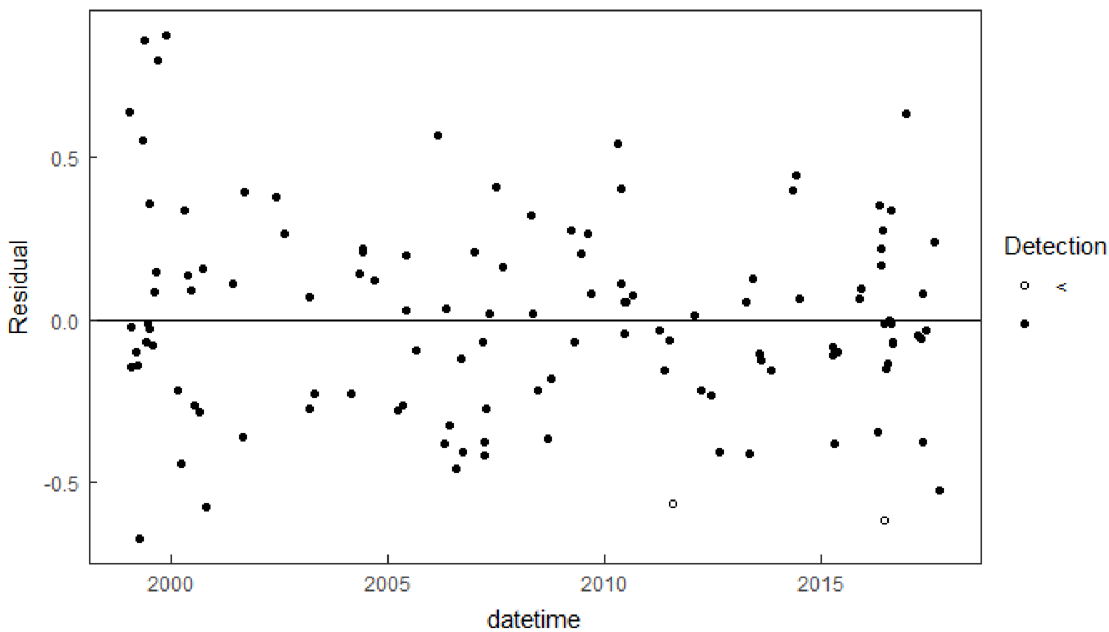
```
                    2.5 %       97.5 %
(Intercept)  0.3655296675 0.7012190277
SPC          0.0002390112 0.0005581986
sin2piD     -0.0556490377 0.1004050003
cos2piD      0.4590997621 0.6661202456
```

Plots



Observed Nitrate plus Nitrite
Computed Nitrate plus Nitrite
0.0
0.5
1.0
1.5
2.0





Variable Summary Statistics

Independent Variable (xvar) - SPC

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  156.5   579.0   974.3   889.4  1175.6  1461.7 
```

Standard Deviation

```
[1] 359.6959
```

Dependent Variable (yvar) - Nitrate plus Nitrite

```
    Min.   1st Qu.    Median      Mean    3rd Qu.      Max. 
   <0.01      0.34   0.52296     0.659       0.88      2.38 
```

Standard Deviation

```
[1] 0.4829
```

Model Calibration Dataset

	datetime	nitratenitrite	SPC	Computed_NO3NO2	residuals
1	1999-01-26 11:50:00	2.17	1210	1.5326	0.637435
2	1999-01-31 14:25:00	1.26	949	1.4075	-0.147552
3	1999-02-03 10:45:00	1.45	1154	1.4745	-0.024494
4	1999-03-17 11:40:00	0.99	979	1.0912	-0.101224
5	1999-04-06 13:55:00	0.7	829	0.8401	-0.140103
6	1999-04-16 12:55:00	0.02	697	0.6915	-0.671474
7	1999-05-13 10:25:00	1.18	1130	0.6300	0.549982
8	1999-05-24 10:45:00	1.34	950	0.4823	0.857691
9	1999-06-10 12:00:00	0.4	1133	0.4686	-0.068573
10	1999-06-25 11:15:00	0.32	896	0.3340	-0.014031
11	1999-07-02 10:15:00	0.61	705	0.2515	0.358491
12	1999-07-14 11:20:00	0.38	1080	0.4094	-0.029454
13	1999-07-29 09:55:00	0.4	1147	0.4795	-0.079501
14	1999-08-12 10:35:00	0.64	1155	0.5542	0.085780
15	1999-08-26 10:50:00	0.77	1081	0.6212	0.148796
16	1999-09-22 11:20:00	1.69	1163	0.8905	0.799463
17	1999-12-02 10:35:00	2.38	1227	1.5056	0.874368
18	2000-02-25 10:40:00	0.986	822	1.2004	-0.214367
19	2000-03-24 13:50:00	0.34	390	0.7809	-0.440935
20	2000-04-27 10:45:00	1.1	1165	0.7678	0.334216
21	2000-05-25 10:20:00	0.7	1188	0.5649	0.135091
22	2000-06-21 12:00:00	0.52	1117	0.4272	0.092844
23	2000-07-26 11:50:00	0.11	896	0.3717	-0.261676
24	2000-08-29 11:00:00	0.37	1087	0.6551	-0.285127
25	2000-09-28 10:30:00	1.07	1054	0.9142	0.155810
26	2000-10-26 10:50:00	0.32	345	0.8964	-0.576425
27	2001-06-06 11:35:00	0.28	342	0.1698	0.110229
28	2001-09-04 11:05:00	0.34	1100	0.7018	-0.361781
29	2001-09-19 10:25:00	0.93	343	0.5351	0.394849
30	2002-06-12 11:10:00	0.47	204	0.0908	0.379215
31	2002-08-14 11:35:00	0.53	398	0.2647	0.265291
32	2003-03-18 12:00:00	1.03	667	0.9576	0.072418
33	2003-03-19 12:20:00	0.55	351	0.8223	-0.272298
34	2003-04-21 11:30:00	0.47	832	0.6987	-0.228697
35	2004-03-05 12:10:00	0.75	448	0.9777	-0.227679
36	2004-05-14 10:35:00	0.47	408	0.3277	0.142249
37	2004-06-14 09:45:00	0.44	582	0.2316	0.208434
38	2004-06-14 09:50:00	0.45	584	0.2324	0.217637
39	2004-09-08 10:25:00	0.92	1240	0.8011	0.118942
40	2005-03-24 10:15:00	0.78	1060	1.0577	-0.277680
41	2005-05-16 11:40:00	0.1	520	0.3651	-0.265054
42	2005-06-10 10:55:00	0.32	258	0.1198	0.200157
43	2005-06-13 09:25:00	0.16	310	0.1296	0.030405
44	2005-08-29 09:35:00	0.4	707	0.4956	-0.095630
45	2006-03-02 09:50:00	1.9	1250	1.3309	0.569052
46	2006-05-01 11:15:00	0.39	1242	0.7724	-0.382453
47	2006-05-12 10:30:00	0.66	1100	0.6258	0.034145
48	2006-06-05 10:15:00	0.18	1171	0.5050	-0.325019
49	2006-07-31 10:30:00	0.03	1150	0.4893	-0.459276
50	2006-09-07 10:50:00	0.68	1280	0.7994	-0.119393
51	2006-09-21 10:00:00	0.5	1230	0.9076	-0.407610
52	2006-09-21 10:05:00	0.5	1230	0.9076	-0.407610
53	2007-01-09 10:30:00	1.87	1430	1.6627	0.207296

54	2007-03-14 10:20:00	1.17	1283	1.2400	-0.070056
55	2007-03-22 10:00:00	1.03	1120	1.1006	-0.070598
56	2007-03-26 10:40:00	0.63	974	1.0044	-0.374411
57	2007-03-31 12:30:00	0.4	629	0.8186	-0.418575
58	2007-04-16 12:15:00	0.43	722	0.7013	-0.271306
59	2007-05-07 10:30:00	0.37	311	0.3508	0.019177
60	2007-06-29 10:25:00	0.66	704	0.2527	0.407333
61	2007-09-04 11:25:00	0.88	1140	0.7177	0.162275
62	2008-04-24 11:40:00	0.88	574	0.5589	0.321087
63	2008-05-09 11:35:00	0.36	353	0.3436	0.016441
64	2008-06-19 09:45:00	0.16	979	0.3763	-0.216288
65	2008-09-15 10:55:00	0.34	842	0.7062	-0.366216
66	2008-10-16 10:10:00	0.72	591	0.9026	-0.182644
67	2009-03-31 11:20:00	1.19	871	0.9149	0.275096
68	2009-04-27 12:15:00	0.43	466	0.4979	-0.067941
69	2009-06-17 10:40:00	0.51	787	0.3073	0.202703
70	2009-08-20 10:50:00	0.85	1098	0.5839	0.266091
71	2009-09-10 11:30:00	0.58	464	0.5006	0.079354
72	2010-04-23 10:00:00	1.28	980	0.7391	0.540924
73	2010-05-17 16:40:00	0.93	945	0.5270	0.402968
74	2010-05-27 10:00:00	0.48	716	0.3708	0.109171
75	2010-06-14 11:30:00	0.18	304	0.1237	0.056283
76	2010-06-16 10:15:00	0.12	417	0.1629	-0.042891
77	2010-07-06 10:30:00	0.09	156	0.0331	0.056899
78	2010-08-25 11:00:00	0.41	378	0.3333	0.076650
79	2011-04-13 10:00:00	0.88	1180	0.9126	-0.032627
80	2011-05-23 10:20:00	0.5	1373	0.6574	-0.157378
81	2011-06-28 10:00:00	0.37	1160	0.4355	-0.065514
82	2011-07-27 11:20:00	<0.02	1070	0.4411	-0.563860
83	2012-02-06 09:45:00	1.51	1242	1.4942	0.015752
84	2012-03-23 10:15:00	0.88	1158	1.0965	-0.216544
85	2012-06-20 09:15:00	0.22	1170	0.4503	-0.230320
86	2012-08-27 09:30:00	0.14	860	0.5487	-0.408684
87	2013-04-11 10:10:00	1.02	1270	0.9677	0.052319
88	2013-05-10 10:00:00	0.26	1180	0.6732	-0.413239
89	2013-05-31 10:00:00	0.36	431	0.2348	0.125183
90	2013-08-05 10:05:00	0.08	322	0.1827	-0.102729
91	2013-08-16 08:30:00	0.35	897	0.4764	-0.126453
92	2013-10-31 10:00:00	1.19	1380	1.3439	-0.153943
93	2014-05-13 10:00:00	0.84	659	0.4425	0.397525
94	2014-06-10 10:30:00	0.85	972	0.4042	0.445802
95	2014-07-02 09:10:00	0.26	559	0.1934	0.066587
96	2015-04-08 09:45:00	0.91	1260	0.9926	-0.082618
97	2015-04-14 09:55:00	0.87	1369	0.9785	-0.108479
98	2015-04-21 10:15:00	0.39	1010	0.7694	-0.379449
99	2015-05-26 10:45:00	0.34	868	0.4374	-0.097356
100	2015-11-25 11:00:00	1.59	1360	1.5205	0.065530
101	2015-12-01 12:10:00	1.69	1462	1.5942	0.096458
102	2016-04-19 10:25:00	0.6	1432	0.9471	-0.347143
103	2016-05-11 12:10:00	1.08	1350	0.7255	0.352604
104	2016-05-24 09:30:00	0.779	1290	0.6116	0.167464
105	2016-05-25 10:10:00	0.4	225	0.1812	0.218788
106	2016-06-06 11:20:00	0.853	1380	0.5793	0.273862
107	2016-06-15 09:00:00	0.523	1350	0.5347	-0.011730
108	2016-06-22 10:00:00	<0.01	1290	0.4942	-0.617342

```
109 2016-07-08 10:35:00         0.277 1136          0.4253 -0.148234
110 2016-07-22 13:30:00         0.347 1205          0.4811 -0.134554
111 2016-08-03 11:20:00         0.532 1208          0.5311  0.000767
112 2016-08-07 12:30:00          0.59  464          0.2555  0.334537
113 2016-08-13 11:30:00          0.24  361          0.2499 -0.009928
114 2016-08-27 09:20:00          0.25  292          0.3223 -0.072276
115 2016-08-31 10:45:00         0.264  232          0.3308 -0.066898
116 2016-12-13 12:00:00          2.23 1323          1.5930  0.634226
117 2017-03-29 10:45:00          0.77  576          0.8168 -0.046773
118 2017-04-20 12:00:00          0.62  762          0.6799 -0.059879
119 2017-05-02 09:50:00          0.28  972          0.6558 -0.375838
120 2017-05-13 14:30:00         0.681 1060          0.6023  0.079065
121 2017-06-07 11:30:00         0.472 1200          0.5074 -0.035060
122 2017-08-11 11:00:00          0.64  789          0.4021  0.237848
123 2017-09-28 10:30:00          0.16  502          0.6846 -0.524638
```

Definitions

```
NO3NO2: nitrate plus nitrite, filtered, in milligrams per liter as nitrogen (00631)
SPC: specific conductance in microsiemens per centimeter at 25 degrees Celsius (00095)
D: date, in decimal years
Leverage: an outlier’s measure in the x-direction (Helsel and others, 2020).
p-value: the probability that the independent variable has no effect on the dependent
      variable (Helsel and others, 2020).
Pseudo-R-squared: pseudocoefficient of determination. An estimation of the proportion of
      variance in the response variable explained by the model (McKelvey and Zavoina,
      1975).
z-score: the estimated coefficient divided by its associated standard error (Helsel and
      others, 2020).
```

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References Cited

American Public Health Association, American Water Works Association, and Water Environment Federation, 1995, Standard methods for the examination of water and wastewater (19th ed.): Washington, D.C., American Public Health Association, 905 p.

Christensen, V.G., Graham, J.L., Milligan, C.R., Pope, L.M., and Ziegler, A.C., 2006, Water quality and relation to taste-and-odor compounds in the North Fork Ninnescah River and Cheney Reservoir, south-central Kansas, 1997–2003: U.S. Geological Survey Scientific Investigations Report 2006–5095, 43 p. [Also available at https://doi.org/10.3133/sir20065095.]

Cohen, A.C., Jr., 1950, Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples: Annals of Mathematical Statistics, v. 21, no. 4, p. 557–569, accessed October 2019 at https://doi.org/10.1214/aoms/1177729751.

Cook, D.R., 1977, Detection of influential observations in linear regression: Technometrics, v. 19, no. 1, p. 15–18. [Also available at https://www.jstor.org/stable/1268249.]

Hald, A., 1949, Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point: Scandinavian Actuarial Journal, v. 1949, no. 1, p. 119–134. [Also available at https://doi.org/10.1080/03461238.1949.10419767.]

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J., 2020, Statistical methods in water resources—Supporting materials: U.S. Geological Survey data release, accessed August 2020 at https://doi.org/10.5066/P9JWL6XR.

McKelvey, R.D., and Zavoina, W., 1975, A statistical model for the analysis of ordinal level dependent variables: The Journal of Mathematical Sociology, v.4, no. 1, p. 103–120. [Also available at https://doi.org/10.1080/0022250X.1975.9989847.]

R Core Team, 2019, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, accessed August 2019 at https://www.R-project.org/.

Rasmussen, T.J., Bennett, T.J., Stone, M.L., Foster, G.M., Graham, J.L., and Putnam, J.E., 2014, Quality-assurance and data-management plan for water-quality activities in the Kansas Water Science Center, 2014: U.S. Geological Survey Open-File Report 2014–1233, 41 p., accessed April 2020 at https://doi.org/10.3133/ofr20141233.

Rasmussen, P.P., Gray, J.R., Glysson, G.D., and Ziegler, A.C., 2009, Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data: U.S. Geological Survey Techniques and Methods, book 3, chap. C4, 52 p. [Also available at https://doi.org/10.3133/tm3C4.]

Sauer, V.B., and Turnipseed, D.P., 2010, Stage measurement at gaging stations: U.S. Geological Survey Techniques and Methods, book 3, chap. A7, 45 p., accessed April 2020 at https://doi.org/10.3133/tm3A7.

Stone, M.L., Graham, J.L., and Gatotho, J.W., 2013, Continuous real-time water-quality monitoring and regression analysis to compute constituent concentrations and loads in the North Fork Ninnescah River upstream from Cheney Reservoir, south-central Kansas, 1999–2012: U.S. Geological Survey Scientific Investigations Report 2013–5071, 44 p., accessed July 2020 at https://doi.org/10.3133/sir20135071.

Tobin, J., 1958, Estimation of relationships for limited dependent variables: Econometrica, v. 26, no. 1, p. 24–36. [Also available at https://doi.org/10.2307/1907382.]

Turnipseed, D.P., and Sauer, V.B., 2010, Discharge measurements at gaging stations: U.S. Geological Survey Techniques and Methods, book 3, chap. A8, 87 p., accessed April 2020 at https://doi.org/10.3133/tm3A8.

U.S. Geological Survey, 2006, Collection of water samples (ver. 2.0, September 2006): U.S. Geological Survey Techniques of Water Resources Investigations, book 9, chap. A4 [variously paged]. [Also available at https://doi.org/10.3133/twri09A4.]

U.S. Geological Survey, 2020, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 20, 2020, at https://doi.org/10.5066/F7P55KJN.

Wagner, R.J., Boulger, R.W., Jr., Oblinger, C.J., and Smith, B.A., 2006, Guidelines and standard procedures for continuous water-quality monitors—Station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods, book 1, chap D3, 51 p. plus 8 attachments. [Also available at https://doi.org/10.3133/tm1D3.]

YSI Incorporated, 2007, YSI 6136 turbidity sensor: YSI Incorporated, 2 p., accessed November 2019 at https://www.ysi.com/File%20Library/Documents/Specification%20Sheets/E56-6136-Turbidity-Sensor.pdf.

YSI Incorporated, 2012a, 6-series multiparameter water quality sondes—User manual, revision J: YSI Incorporated, 379 p., accessed November 2019 at https://www.ysi.com/File%20Library/Documents/Manuals/069300-YSI-6-Series-Manual-RevJ.pdf.

YSI Incorporated, 2012b, EXO water quality field sensors—Features, specifications, and comparability to YSI 6-series sensors, revision B: YSI Incorporated, 14 p., accessed November 2019 at https://www.exowater.com/media/pdfs/EXO-6Series-Sensor-Comparison.pdf.