Appendix 10. Model Archive Summary for Nitrate at Station 07144780; North Fork Ninnescah River above Cheney Reservoir, Kansas, during January 1, 1999, through December 31, 2019

This model archive summary summarizes the nitrate (NO_3) model developed to compute hourly or daily nitrate concentrations during January 1, 1999, through December 31, 2019. This model is used concomitantly with other models during this period to calculate concentrations when other explanatory variables are not available for the purposes of load and concentration model calculations. 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), and other standard USGS methods (Sauer and Turnipseed, 2010; Turnipseed and Sauer, 2010).

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 Sutron Satlink 2 High Data Rate Collection Platform and a Design Analysis Water Log H350/355 nonsubmersible pressure transducer transfers real-time stage, precipitation, and water quality data via satellite. The primary reference gage is a Type-A wire-weight gage located on the downstream bridge guardrail. Check-bar elevation is 21.804 feet. The orifice is enclosed in 1 1/4-inch pipe, which runs from the gage house, under the bridge, and along an I-beam where it is attached to the concrete pier closest to the left edge of water.

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 100 concomitant values of discretely collected nitrate samples and continuously measured streamflow during January 26, 1999, through September 28, 2017. Discrete samples were collected over a range of streamflows. One sample was less than the minimum reporting level (less than [<] 0.01 milligram per liter); therefore, a Tobit regression model was developed to compute estimates of nitrate 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 influence on the final Tobit regression model (Cook, 1977). None of the samples in this dataset were deemed outliers or removed from the model calibration dataset. Other data deemed outliers and removed in previously published versions of this model (Stone and others, 2013) were examined and retained in the dataset if there were no clear issues, explanations, or conditions that would cause a result to be invalid for model calibration.

Nitrate

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 1 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 by the Wichita Municipal Water and Wastewater Laboratory in Wichita, Kans., according to standard methods (American Public Health Association and others, 1995).

Continuous Data

The streamflow data were measured using a nonsubmersible pressure transducer during January 1, 1999, through December 31, 2019. The continuous streamflow data used were time interpolated values from the continuous time series. If the continuous data were not 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 to streamflow 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 were examined for homoscedasticity (departures from zero did not change substantially over the range of model calculated values).

A total of 1 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).

Streamflow and seasonality were selected as good predictors of nitrate based on residual plots, a higher pseudocoefficient of determination (pseudo-R^2), and relatively low estimated standard residual error (RSE). This model was developed with the sole purpose to fill in gaps of missing data of the primary model for concentration and load estimations. Seasonality was included as an explanatory variable because nitrate seems to have a cyclical pattern potentially influenced by groundwater during low seasonal flow.

Model Summary

Summary of final nitrate regression analysis at USGS site 07144780:

Nitrate-based model:

$$NO3 = -0.199 \times \log_{10}(Q) + 0.0929 \times \sin(2\pi D) + 0.604 \times \cos(2\pi D) + 1.339 ,$$

where,

$NO3$ = nitrate, in milligrams per liter as nitrogen;
Q = streamflow, in cubic feet per second; and
D = date, in decimal years.

Previous Models

Version	Model Equation	Reference
1.0	$NO3 = -0.278 \times \log_{10}(Q) + 0.0903 \times \sin(2\pi D) + 0.652 \times \cos(2\pi D) + 1.55$	Stone and others (2013)

Model Statistics, Plots and Data

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

Model

$$NO3 = -0.199 \times \log_{10}(Q) + 0.0929 \times \sin(2\pi D) + 0.604 \times \cos(2\pi D) + 1.339$$

Computation method: Absolute Maximum Likelihood Estimation (AMLE)

Explanatory Variables

```
Coefficients:
            Estimate Std. Error z-score p-value
(Intercept)  1.33885    0.11739  11.405  0.0000
logQ        -0.19872    0.04922  -4.037  0.0001
sin2piD      0.09286    0.04413   2.104  0.0327
cos2piD      0.60412    0.05698  10.602  0.0000
```

Basic Model Statistics

For a detailed definition and explanation of the terms used below, refer to Helsel and others (2020).

```
Estimated residual standard error (Unbiased) = 0.299
Number of observations = 100, number censored = 1 (1 percent)

Log-likelihood (model) = -20.42
Log-likelihood (intercept only) = -68.16
  Chi-square = 95.47
  degrees of freedom = 3
  p-value = <0.0001

Computation method: AMLE

Pseudo-R-squared: 0.594

  Akaike Information Criterion: 50.84
  Bayesian Information Criterion: 63.86


Variance inflation factors
   logQ 1.11
sin2piD 1.13
cos2piD 1.02
```

Outlier Test Criteria

```
leverage   cooksD
  0.0900   0.8451
```

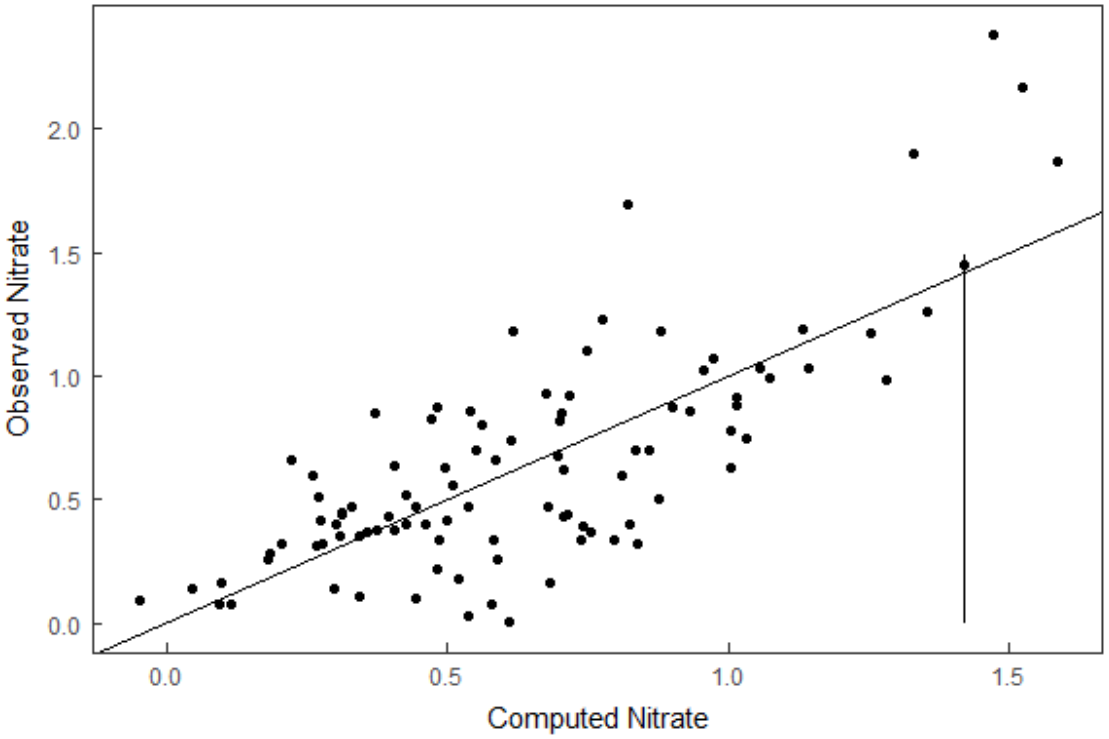
Flagged Observations

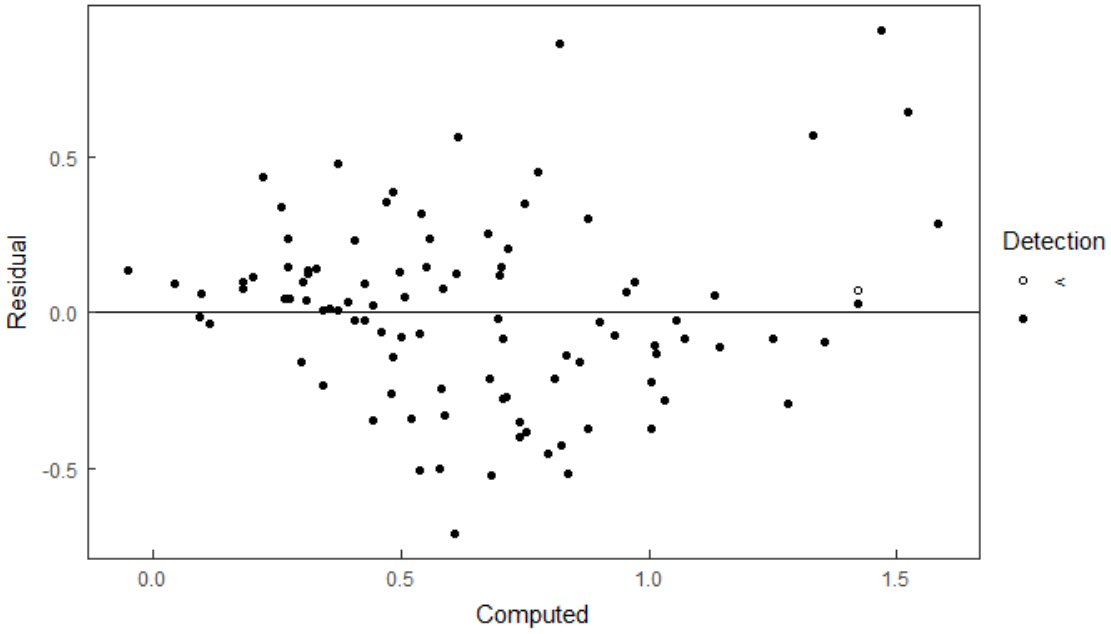
```
    Observations exceeding at least one test criterion
   Nitrate  ycen     yhat  resids leverage   cooksD
26    0.32 FALSE  0.83755 -0.5175  0.11855 0.114300
76    0.09 FALSE -0.04856  0.1386  0.09219 0.006006
```

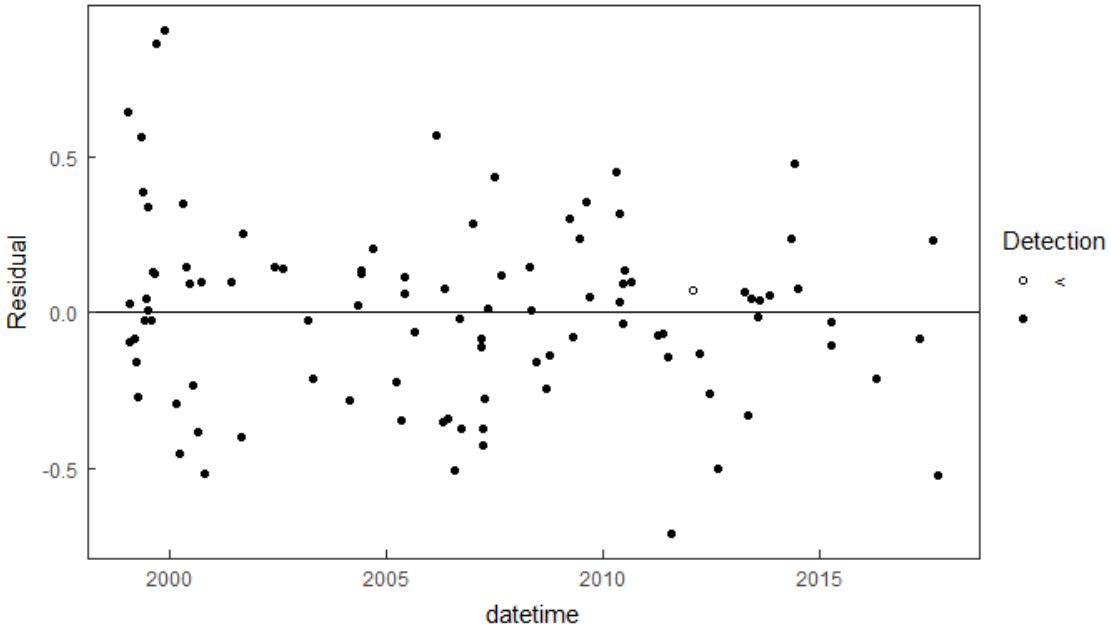
95% Confidence Intervals

```
                   2.5 %     97.5 %
(Intercept)  1.108772804  1.5689360
logQ        -0.295195025 -0.1022408
sin2piD      0.006369252  0.1793589
cos2piD      0.492431945  0.7158018
```

Plots







Variable Summary Statistics

Independent Variable (xvar) - Q

```
    Min.  1st Qu.   Median     Mean  3d Qu.       Max.
   5.291   75.868  214.301  648.064  506.462 8,216.570
```

Standard Deviation

```
[1] 1236.16
```

Dependent Variable (yvar) – Nitrate

```
    Min.  1st Qu.   Median     Mean    3d Qu.      Max.
   <0.01     0.45    0.515   0.6522       0.87     2.38
```

Standard Deviation

```
[1] 0.4623
```

Model-Calibration Data Set

	datetime	Nitrate	logQ	Q	Computed_Nitrate	residuals
1	1999-01-26 11:50:00	2.17	2.017	104.00	1.5229	0.64716
2	1999-01-31 14:25:00	1.26	2.779	601.50	1.3540	-0.09393
3	1999-02-03 10:45:00	1.45	2.382	241.00	1.4204	0.02963
4	1999-03-17 11:40:00	0.99	2.584	384.00	1.0718	-0.08176
5	1999-04-06 13:55:00	0.7	2.632	428.25	0.8591	-0.15910
6	1999-04-16 12:55:00	0.44	2.843	696.50	0.7121	-0.27205
7	1999-05-13 10:25:00	1.18	1.990	97.75	0.6156	0.56447
8	1999-05-24 10:45:00	0.87	2.199	158.00	0.4829	0.38715
9	1999-06-10 12:00:00	0.4	1.924	84.00	0.4269	-0.02684
10	1999-06-25 11:15:00	0.32	2.375	237.25	0.2769	0.04311
11	1999-07-02 10:15:00	0.6	2.386	243.25	0.2598	0.34019
12	1999-07-14 11:20:00	0.38	1.788	61.33	0.3736	0.00641
13	1999-07-29 09:55:00	0.38	1.778	60.00	0.4055	-0.02549
14	1999-08-12 10:35:00	0.63	1.633	43.00	0.4970	0.13299
15	1999-08-26 10:50:00	0.74	1.521	33.17	0.6121	0.12792
16	1999-09-22 11:20:00	1.69	1.690	49.00	0.8205	0.86948
17	1999-12-02 10:35:00	2.38	1.787	61.25	1.4697	0.91034
18	2000-02-25 10:40:00	0.986	2.414	259.33	1.2800	-0.29396
19	2000-03-24 13:50:00	0.34	3.578	3786.67	0.7952	-0.45513
20	2000-04-27 10:45:00	1.1	2.040	109.75	0.7482	0.35388
21	2000-05-25 10:20:00	0.7	1.778	60.00	0.5514	0.14865
22	2000-06-21 12:00:00	0.52	1.672	47.00	0.4257	0.09436
23	2000-07-26 11:50:00	0.11	2.068	117.00	0.3416	-0.23156
24	2000-08-29 11:00:00	0.37	0.968	9.30	0.7531	-0.38308
25	2000-09-28 10:30:00	1.07	1.290	19.50	0.9715	0.09852
26	2000-10-26 10:50:00	0.32	3.430	2690.00	0.8376	-0.51755
27	2001-06-06 11:35:00	0.28	3.261	1824.17	0.1835	0.09655
28	2001-09-04 11:05:00	0.34	1.255	18.00	0.7378	-0.39776
29	2001-09-19 10:25:00	0.93	2.273	187.33	0.6750	0.25509
30	2002-06-12 11:10:00	0.42	2.650	446.33	0.2725	0.14749
31	2002-08-14 11:35:00	0.47	2.539	346.08	0.3286	0.14146
32	2003-03-18 12:00:00	1.03	2.615	412.00	1.0561	-0.02602
33	2003-04-21 11:30:00	0.47	2.741	551.00	0.6803	-0.21023
34	2004-03-05 12:10:00	0.75	3.290	1951.67	1.0324	-0.28232
35	2004-05-14 10:35:00	0.47	2.760	575.83	0.4450	0.02508
36	2004-06-14 09:45:00	0.44	2.377	238.00	0.3127	0.12730
37	2004-06-14 09:50:00	0.45	2.377	238.00	0.3127	0.13730
38	2004-09-08 10:25:00	0.92	1.591	39.00	0.7153	0.20469
39	2005-03-24 10:15:00	0.78	2.576	376.75	1.0044	-0.22438
40	2005-05-16 11:40:00	0.1	2.721	526.00	0.4441	-0.34410
41	2005-06-10 10:55:00	0.32	3.048	1116.67	0.2036	0.11644
42	2005-06-13 09:25:00	0.16	3.498	3150.00	0.0991	0.06099
43	2005-08-29 09:35:00	0.4	2.407	255.17	0.4593	-0.05926
44	2006-03-02 09:50:00	1.9	1.967	92.67	1.3291	0.57090
45	2006-05-01 11:15:00	0.39	1.928	84.75	0.7406	-0.35061
46	2006-05-12 10:30:00	0.66	2.192	155.50	0.5844	0.07561
47	2006-06-05 10:15:00	0.18	1.591	39.00	0.5213	-0.34125
48	2006-07-31 10:30:00	0.03	1.154	14.25	0.5365	-0.50653
49	2006-09-07 10:50:00	0.68	1.591	39.00	0.6973	-0.01731
50	2006-09-21 10:00:00	0.5	1.362	23.00	0.8758	-0.37578
51	2006-09-21 10:05:00	0.5	1.362	23.00	0.8758	-0.37578
52	2007-01-09 10:30:00	1.87	1.845	70.00	1.5834	0.28659
53	2007-03-14 10:20:00	1.17	1.824	66.67	1.2514	-0.08141

54	2007-03-22 10:00:00	1.03	1.987	97.00	1.1415	-0.11151
55	2007-03-26 10:40:00	0.63	2.471	295.67	1.0051	-0.37506
56	2007-03-31 12:30:00	0.4	3.122	1325.00	0.8243	-0.42425
57	2007-04-16 12:15:00	0.43	2.871	743.50	0.7065	-0.27642
58	2007-05-07 10:30:00	0.37	3.571	3725.00	0.3563	0.01372
59	2007-06-29 10:25:00	0.66	2.603	401.00	0.2220	0.43800
60	2007-09-04 11:25:00	0.82	1.447	28.00	0.6997	0.12037
61	2008-04-24 11:40:00	0.85	2.426	266.60	0.7019	0.14816
62	2008-05-09 11:35:00	0.35	3.493	3110.31	0.3440	0.00602
63	2008-06-19 09:45:00	0.14	2.349	223.23	0.2981	-0.15806
64	2008-09-15 10:55:00	0.34	2.591	389.55	0.5825	-0.24243
65	2008-10-16 10:10:00	0.7	2.928	848.02	0.8345	-0.13447
66	2009-03-31 11:20:00	1.18	2.853	713.40	0.8777	0.30231
67	2009-04-27 12:15:00	0.42	3.338	2178.89	0.5003	-0.08028
68	2009-06-17 10:40:00	0.51	2.543	349.14	0.2711	0.23894
69	2009-08-20 10:50:00	0.83	2.014	103.19	0.4711	0.35894
70	2009-09-10 11:30:00	0.56	2.682	481.29	0.5076	0.05242
71	2010-04-23 10:00:00	1.23	2.159	144.33	0.7753	0.45475
72	2010-05-17 16:40:00	0.86	2.197	157.26	0.5398	0.32022
73	2010-05-27 10:00:00	0.43	2.533	341.23	0.3940	0.03600
74	2010-06-14 11:30:00	0.14	3.750	5625.32	0.0443	0.09572
75	2010-06-16 10:15:00	0.08	3.357	2277.10	0.1135	-0.03343
76	2010-07-06 10:30:00	0.09	3.915	8216.57	-0.0485	0.13856
77	2010-08-25 11:00:00	0.4	3.047	1114.21	0.3013	0.09872
78	2011-04-13 10:00:00	0.86	1.904	80.15	0.9301	-0.07012
79	2011-05-23 10:20:00	0.47	1.965	92.25	0.5370	-0.06701
80	2011-06-28 10:00:00	0.34	1.294	19.69	0.4842	-0.14423
81	2011-07-27 11:20:00	<0.01	0.724	5.29	0.6088	-0.71026
82	2012-02-06 09:45:00	1.49	2.313	205.38	1.4202	0.06980
83	2012-03-23 10:15:00	0.88	2.529	337.76	1.0138	-0.13381
84	2012-06-20 09:15:00	0.22	1.414	25.93	0.4803	-0.26032
85	2012-08-27 09:30:00	0.08	1.762	57.83	0.5795	-0.49949
86	2013-04-11 10:10:00	1.02	1.891	77.82	0.9537	0.06630
87	2013-05-10 10:00:00	0.26	2.258	180.98	0.5895	-0.32943
88	2013-05-31 10:00:00	0.31	3.033	1078.66	0.2667	0.04337
89	2013-08-05 10:05:00	0.08	3.478	3005.69	0.0951	-0.01508
90	2013-08-16 08:30:00	0.35	2.699	499.95	0.3090	0.04109
91	2013-10-31 10:00:00	1.19	2.153	142.12	1.1311	0.05892
92	2014-05-13 10:00:00	0.8	2.272	186.89	0.5596	0.24041
93	2014-06-10 10:30:00	0.85	2.199	157.96	0.3724	0.47766
94	2014-07-02 09:10:00	0.26	2.776	597.26	0.1823	0.07771
95	2015-04-08 09:45:00	0.91	1.754	56.75	1.0125	-0.10250
96	2015-04-14 09:55:00	0.87	1.999	99.67	0.9008	-0.03081
97	2016-04-19 10:25:00	0.6	2.138	137.55	0.8104	-0.21035
98	2017-04-20 12:00:00	0.62	2.666	463.31	0.7056	-0.08555
99	2017-08-11 11:00:00	0.64	2.066	116.31	0.4056	0.23439
100	2017-09-28 10:30:00	0.16	2.686	484.79	0.6839	-0.52384

Definitions

NO3: nitrate, filtered, in milligrams per liter as nitrogen (00618)
Q: streamflow, instantaneous, in cubic feet per second (00061)
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.]