Appendix 14. Model Archive Summary for Total Phosphorus at U.S. Geological Survey Site 07144780, North Fork Ninnescah River above Cheney Reservoir, Kansas, during October 17, 2009, through December 31, 2019

This model archive summary summarizes the total phosphorus (TP) model developed to compute hourly or daily TP during October 17, 2009, 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: April 23, 2010, to April 20, 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 30 concomitant values of discretely measured TP samples and continuously measured turbidity during April 23, 2010, through May 2, 2017. Discrete samples were collected over a range of streamflows and turbidity conditions. Only one sample was less than the minimum reporting level (less than [<] 0.03 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 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.

Total Phosphorus

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

Continuous Data

Turbidity was measured using a YSI model 6136 sensor installed during October 17, 2009, through November 12, 2015, and March 31, 2017, through June 7, 2017. Concomitant turbidity values were time interpolated. If the continuous data were not available (2 or more hours of turbidity values bracketing the sample collection time were missing) because of fouling, changes in equipment, or unsuitable site conditions, then the field monitor turbidity value measured during sampling was substituted. If no 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 TP concentration to turbidity 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 TP were examined for homoscedasticity (departures from zero did not change substantially over the range of model calculated values). Previously published explanatory variables were also strongly considered for continuity; however, the best explanatory variable(s) was ultimately selected.

A total of 3.3 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).

Turbidity was selected as the best predictor of logarithm base 10 ($\log_{10}$) (TP) based on residual plots, a higher pseudocoefficient of determination (pseudo-R^2), and relatively low estimated standard residual error (RSE).

Model Summary

Summary of final TP regression analysis at USGS site 07144780:

TP-based model:

$$\log_{10}(TP) = 0.7167 \times \log_{10}(TBY6136) - 1.938,$$

where,

TP = phosphorus, total, unfiltered, in milligrams per liter as phosphorus, and
TBY6136 = turbidity, YSI model 6136, in formazin nephelometric units.

The log-transformed model may be retransformed to original units so that TP 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; Duan, 1983). For this model, the calculated BCF is 1.07. The retransformed model, accounting for BCF, is as follows:

$TP = (TBY6136^{0.7167} \times 10^{-1.938}) \times 1.07$

Model Statistics, Plots and Data

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

Model

$\log_{10}(TP) = 0.7167 \times \log_{10}(TBY6136) - 1.938$

Computation method: Absolute Maximum Likelihood Estimation (AMLE)

Explanatory Variables

```
Coefficients:
            Estimate Std. Error z-score p-value
(Intercept)  -1.9379    0.13409 -14.453       0
logTBY6136    0.7167    0.07642   9.378       0
```

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) = 0.1682
Number of observations = 30, number censored = 1 (3.3 percent)

Log-likelihood (model) = 10.42
Log-likelihood (intercept only) = -42.95
  Chi-square = 106.7
  degrees of freedom = 1
  p-value = <0.0001

Computation method: AMLE

Pseudo-R-squared: 0.7592

  Akaike Information Criterion: -14.84
  Bayesian Information Criterion: -10.64
```

Outlier Test Criteria

```
Leverage Cook's D
    0.10     0.71
```

Flagged Observations

```
    Observations exceeding at least one test criterion
     logTP  ycen    yhat    resids leverage    cooksD
6  -0.4202 FALSE -0.2967 -0.123564   0.1031 0.0345773
7  -0.2840 FALSE -0.2453 -0.038727   0.1213 0.0041638
8  -1.1549 FALSE -1.1645  0.009575   0.1142 0.0002356
10 -1.3010 FALSE -1.1916 -0.109470   0.1242 0.0342754
16 -1.5229  TRUE -1.3553 -0.256050   0.1973 0.3546315
17 -1.2218 FALSE -1.1916 -0.030289   0.1242 0.0026239
```

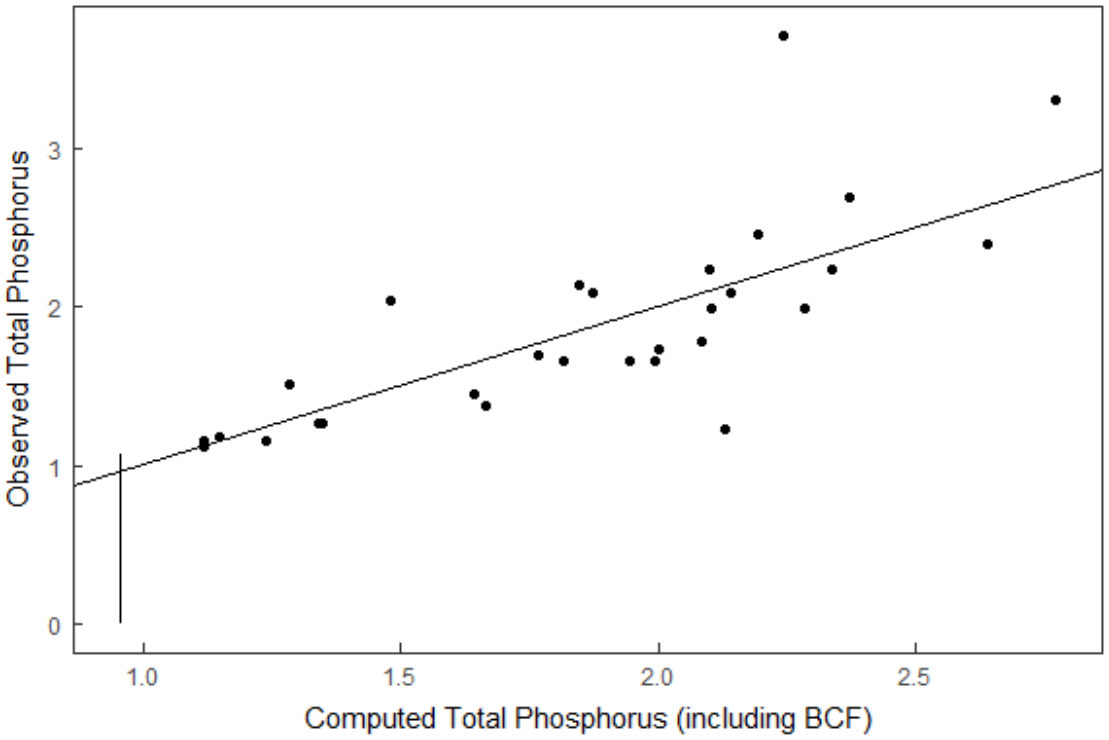
Bias correction factor

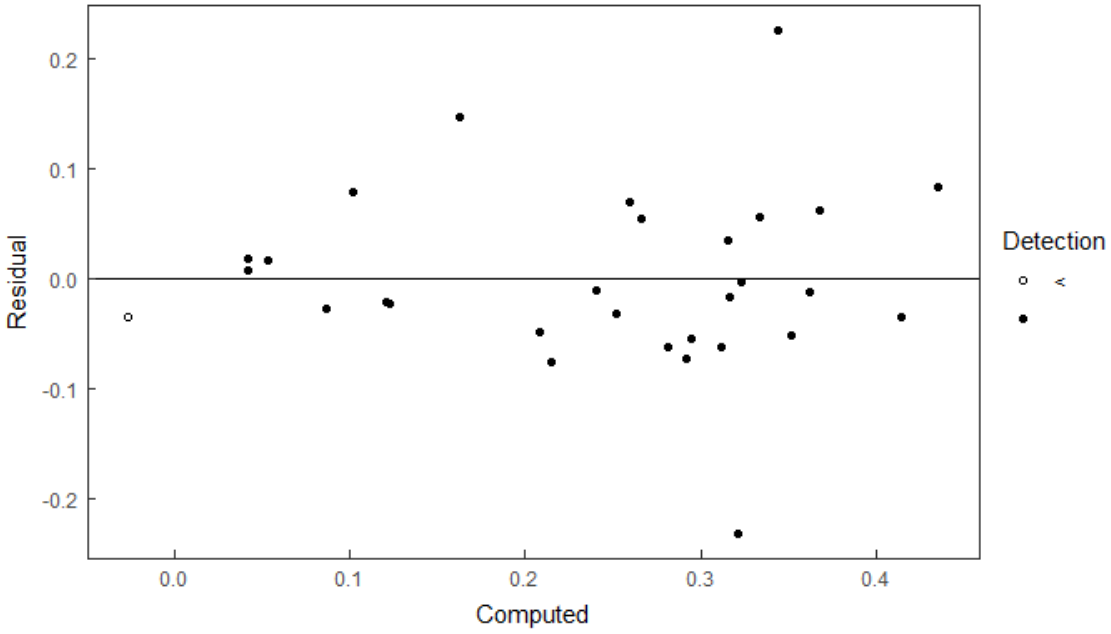
```
[1] 1.069753
```

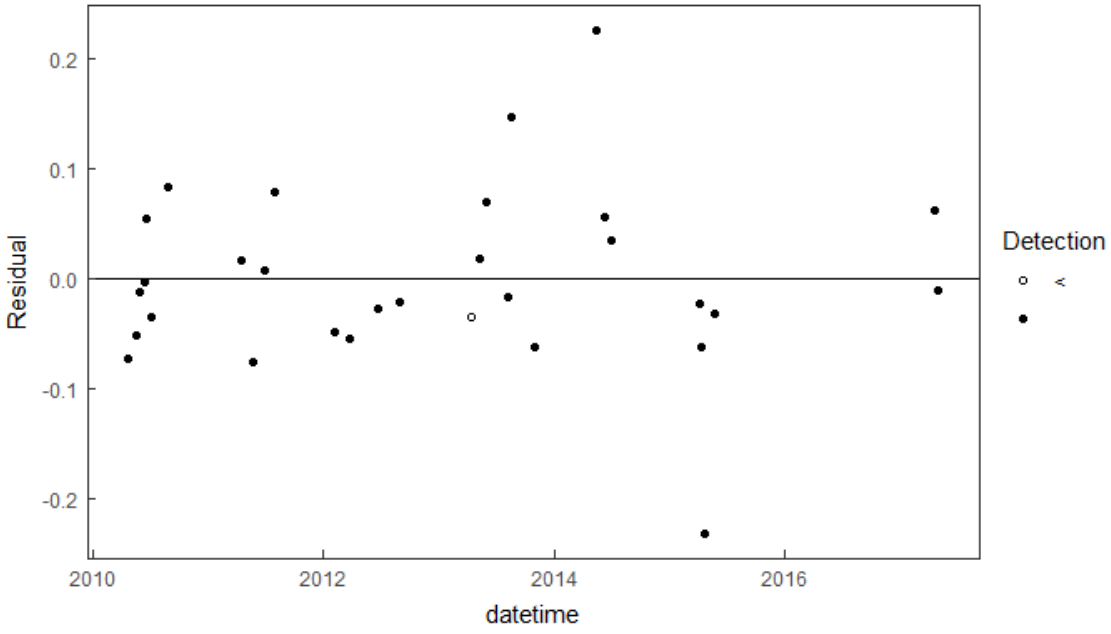
95% Confidence Intervals

```
                 2.5 %     97.5 %
(Intercept) -2.2007363 -1.6751259
logTBY6136   0.5669229  0.8664864
```

Plots







Variable Summary Statistics

Independent Variable (xvar) - Turbidity (TBY6136)

```
   Min. 1st Qu.  Median    Mean  3d Qu.    Max.
   6.50   22.38   66.00   72.22   96.12  230.00
```

Standard Deviation

```
[1] 54.64463
```

Dependent Variable (yvar) - Total Phosphorus

```
   Min. 1st Qu.  Median    Mean  3d Qu.    Max.
  <0.03    0.10   0.235  0.2409    0.33    0.57
```

Standard Deviation

```
[1] 0.143
```

Model-Calibration Data Set

	datetime	logTP	logTBY6136	Phosphorus	TBY6136	Computed_logTP	Computed_TP	residuals
1	2010-04-23 10:00:00	-0.658	1.881	0.22	76.0	-0.590	0.2751	-0.07221
2	2010-05-17 16:40:00	-0.523	2.079	0.3	120.0	-0.448	0.3816	-0.05136
3	2010-05-27 10:00:00	-0.456	2.114	0.35	130.0	-0.423	0.4042	-0.01173
4	2010-06-14 11:30:00	-0.495	1.985	0.32	96.5	-0.516	0.3264	-0.00314
5	2010-06-16 10:15:00	-0.495	1.792	0.32	62.0	-0.653	0.2377	0.05415
6	2010-07-06 10:30:00	-0.42	2.290	0.38	195.0	-0.297	0.5404	-0.03423
7	2010-08-25 11:00:00	-0.284	2.362	0.52	230.0	-0.245	0.6083	0.08439
8	2011-04-13 10:00:00	-1.15	1.079	0.07	12.0	-1.164	0.0733	0.01681
9	2011-05-23 10:20:00	-0.854	1.622	0.14	41.8	-0.776	0.1793	-0.07490
10	2011-06-28 10:00:00	-1.3	1.041	0.05	11.0	-1.192	0.0688	0.00808
11	2011-07-27 11:20:00	-0.745	1.241	0.18	17.4	-1.049	0.0956	0.07869
12	2012-02-06 09:45:00	-0.796	1.599	0.16	39.8	-0.792	0.1729	-0.04829
13	2012-03-23 10:15:00	-0.62	1.888	0.24	77.3	-0.585	0.2784	-0.05441
14	2012-06-20 09:15:00	-1.22	1.190	0.06	15.5	-1.085	0.0880	-0.02633
15	2012-08-27 09:30:00	-1	1.305	0.1	20.2	-1.002	0.1064	-0.02063
16	2013-04-11 10:10:00	<-1.52	0.813	<0.03	6.5	-1.355	0.0472	-0.03463
17	2013-05-10 10:00:00	-1.22	1.041	0.06	11.0	-1.192	0.0688	0.01808
18	2013-05-31 10:00:00	-0.481	1.771	0.33	59.0	-0.669	0.2294	0.07057
19	2013-08-05 10:05:00	-0.523	1.961	0.3	91.4	-0.532	0.3140	-0.01613
20	2013-08-16 08:30:00	-0.509	1.447	0.31	28.0	-0.901	0.1345	0.14709
21	2013-10-31 10:00:00	-0.658	1.845	0.22	70.0	-0.616	0.2593	-0.06157
22	2014-05-13 10:00:00	-0.244	2.054	0.57	113.2	-0.466	0.3660	0.22619
23	2014-06-10 10:30:00	-0.409	2.021	0.39	105.0	-0.489	0.3468	0.05593
24	2014-07-02 09:10:00	-0.456	1.958	0.35	90.8	-0.535	0.3125	0.03474
25	2015-04-08 09:45:00	-1	1.312	0.1	20.5	-0.998	0.1076	-0.02254
26	2015-04-14 09:55:00	-0.602	1.946	0.25	88.3	-0.543	0.3064	-0.06169
27	2015-04-21 10:15:00	-1.05	1.978	0.09	95.0	-0.520	0.3228	-0.23111
28	2015-05-26 10:45:00	-0.658	1.746	0.22	55.8	-0.686	0.2203	-0.03209
29	2017-04-20 12:00:00	-0.367	2.135	0.43	136.6	-0.407	0.4188	0.06186
30	2017-05-02 09:50:00	-0.638	1.708	0.23	51.0	-0.714	0.2067	-0.01056

Definitions

TP: phosphorus, unfiltered, in milligrams per liter as phosphorus (00665)
TBY6136: turbidity, YSI model 6136, in formazin nephelometric units (63680)
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.]

Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: Journal of the American Statistical Association, v. 78, no. 383, p. 605–610. [Also available at https://doi.org/10.1080/01621459.1983.10478017.]

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.