

Prepared in cooperation with the North Dakota State Water Commission

Continuous Water-Quality Monitoring and Regression Analysis to Estimate Constituent Concentrations and Loads in the Sheyenne River, North Dakota, 1980–2006

Scientific Investigations Report 2007–5153

Continuous Water-Quality Monitoring and Regression Analysis to Estimate Constituent Concentrations and Loads in the Sheyenne River, North Dakota, 1980–2006

By Karen R. Ryberg

Prepared in cooperation with the North Dakota State Water Commission

Scientific Investigations Report 2007–5153

**U.S. Department of the Interior
U.S. Geological Survey**

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, Virginia: 2007

For product and ordering information:
World Wide Web: <http://www.usgs.gov/pubprod>
Telephone: 1-888-ASK-USGS

For more information on the USGS--the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:
World Wide Web: <http://www.usgs.gov>
Telephone: 1-888-ASK-USGS

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

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

Suggested citation:

Ryberg, K.R., 2007, Continuous water-quality monitoring and regression analysis to estimate constituent concentrations and loads in the Sheyenne River, North Dakota, 1980–2006: U.S. Geological Survey Scientific Investigations Report 2007–5153, 22 p.

Contents

Abstract.....	1
Introduction.....	1
Purpose and Scope	2
Description of Study Area	2
Previous Studies	4
Methods.....	4
Streamflow Measurements	4
Manual Water-Quality Measurements, Sample Collection, and Analysis.....	4
Development of Regression Equations to Estimate Constituent Concentrations	5
Calculation of Measured and Estimated Constituent Loads	7
Continuous Water-Quality Monitoring.....	9
Summary Statistics.....	10
Results of Regression Analysis	10
Hardness	10
Dissolved Solids	10
Calcium	14
Magnesium	14
Sodium	14
Sulfate	15
Measured and Estimated Constituent Loads	19
Summary.....	19
References Cited.....	20

Figures

1. Map showing location of the seven sampling sites on the Sheyenne River, N. Dak	3
2–5. Graphs showing:	
2. Comparison of measured and estimated sodium concentrations in the Sheyenne River, N. Dak., at site 1	8
3. Predicted sulfate concentration, 90-percent prediction interval, and individual samples used in regression analyses for sites 2 and 3 on the Sheyenne River, N. Dak., 2005–06	17
4. Predicted sulfate concentration, 90-percent prediction interval, and individual samples used in regression analyses for sites 5, 6, and 7 on the Sheyenne River, N. Dak., 2000–2006	18
5. Comparison of measured and estimated sulfate loads for site 7 on the Sheyenne River, N. Dak., 1997–2006	19

Tables

1. Sites used for regression analysis, Sheyenne River, N. Dak., 1980–2006	2
2. Constituents estimated using regression equations for the Sheyenne River, N. Dak., 1980–2006	6
3. Possible surrogate physical properties, or explanatory variables, used to develop regression relations to estimate constituent concentrations in Sheyenne River, N. Dak., 1980–2006	7
4. Conversion factors used in calculation of measured and estimated loads	9
5. Summary of statistics for water-quality sampling sites on the Sheyenne River, N. Dak., 1980–2006	11
6. Regression equations for estimates of hardness in the Sheyenne River, N. Dak., 1980–2006	13
7. Regression equations for estimates of dissolved solids in the Sheyenne River, N. Dak., 1980–2006	13
8. Regression equations for estimates of calcium in the Sheyenne River, N. Dak., 1980–2006	14
9. Regression equations for estimates of magnesium in the Sheyenne River, N. Dak., 1980–2006	15
10. Regression equations for estimates of sodium in the Sheyenne River, N. Dak., 1980–2006	16
11. Regression equations for estimates of sulfate in the Sheyenne River, N. Dak., 1980–2006	16

Conversion Factors, Abbreviations, and Datum

Multiply	By	To obtain
Length		
foot (ft)	0.3048	meter (m)
yard (yd)	.9144	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
square mile (mi ²)	259.0	hectare (ha)
square mile (mi ²)	2.590	square kilometer (km ²)
Flow rate		
cubic foot per second (ft ³ /s)	0.02832	cubic meter per second (m ³ /s)
Mass		
pound, avoirdupois (lb)	0.4536	kilogram (kg)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$^{\circ}\text{F}=(1.8\times^{\circ}\text{C})+32.$$

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

$$^{\circ}\text{C}=(^{\circ}\text{F}-32)/1.8.$$

Specific conductance is given in microsiemens per centimeter at 25 degrees Celsius ($\mu\text{S}/\text{cm}$ at 25°C).

Concentrations of chemical constituents in water are given in milligrams per liter (mg/L).

Horizontal coordinate information is referenced to North American Datum of 1927 (NAD 27).

Continuous Water-Quality Monitoring and Regression Analysis to Estimate Constituent Concentrations and Loads in the Sheyenne River, North Dakota, 1980–2006

By Karen R. Ryberg

Abstract

This report presents the results of a study by the U.S. Geological Survey, done in cooperation with the North Dakota State Water Commission, to estimate water-quality constituent concentrations at seven sites on the Sheyenne River, N. Dak. Regression analysis of water-quality data collected in 1980–2006 was used to estimate concentrations for hardness, dissolved solids, calcium, magnesium, sodium, and sulfate. The explanatory variables examined for the regression relations were continuously monitored streamflow, specific conductance, and water temperature. For the conditions observed in 1980–2006, streamflow was a significant explanatory variable for some constituents. Specific conductance was a significant explanatory variable for all of the constituents, and water temperature was not a statistically significant explanatory variable for any of the constituents in this study.

The regression relations were evaluated using common measures of variability, including R^2 , the proportion of variability in the estimated constituent concentration explained by the explanatory variables and regression equation. R^2 values ranged from 0.784 for calcium to 0.997 for dissolved solids. The regression relations also were evaluated by calculating the median relative percentage difference (*RPD*) between measured constituent concentration and the constituent concentration estimated by the regression equations. Median *RPDs* ranged from 1.7 for dissolved solids to 11.5 for sulfate. The regression relations also may be used to estimate daily constituent loads.

The relations should be monitored for change over time, especially at sites 2 and 3 which have a short period of record. In addition, caution should be used when the Sheyenne River is affected by ice or when upstream sites are affected by isolated storm runoff. Almost all of the outliers and highly influential samples removed from the analysis were made during periods when the Sheyenne River might be affected by ice.

Introduction

The U.S. Geological Survey (USGS) continuously records gage height, specific conductance, and water temperature at seven gaging sites on the Sheyenne River (fig. 1; table 1) in cooperation with several State and Federal agencies. Data from the continuous real-time water-quality monitors on the Sheyenne River are available on the Internet from the USGS National Water Information System (NWISWeb, <http://waterdata.usgs.gov/nd/nwis>). Continuous monitoring provides valuable information and may identify important variability that may not be evident in discrete water-quality sampling schemes. This information could be used to estimate concentrations of other water-quality constituents of interest, such as sulfate.

Devils Lake, in the adjoining Devils Lake Basin (fig. 1), began flooding in the 1990s and continues to the present (2007). According to the North Dakota State Water Commission's Devils Lake Quick Facts [n.d.], Devils Lake flooding "has destroyed hundreds of homes and businesses and inundated thousands of acres of productive farmland. The State of North Dakota and the U.S. Government have spent over \$450 million dollars in flood mitigation efforts including moving roads, rail and power lines, and building dikes." The State of North Dakota has constructed an outlet to allow water from Devils Lake to flow into the Sheyenne River. The maximum outlet discharge is based in part on the sulfate concentration at the Sheyenne River downstream from the outlet insertion point (North Dakota State Water Commission [n.d.], Devils Lake Flooding—Outlet). Many water-quality stakeholders are concerned about the effect that water from Devils Lake might have on the Sheyenne River. The Sheyenne River is a tributary of the Red River of the North (Red River), which flows into Canada, therefore, the water quality of the Sheyenne River is of interest to users of the Red River in North Dakota, Minnesota, and Canada.

2 Continuous Water-Quality Monitoring and Regression Analysis, Sheyenne River, 1980–2006

Table 1. Sites used for regression analysis, Sheyenne River, N. Dak., 1980–2006.

[--, not applicable]

Site number (figure 1)	U.S. Geological Survey station number	North Dakota State Water Commission location number	U.S. Geological Survey station name	Latitude	Longitude
1	05054500	--	Sheyenne River above Harvey, N. Dak.	47°42'10"	99°56'55"
2	05055300	15106807	Sheyenne River above Devils Lake State outlet near Flora, N. Dak.	47°54'28"	99°24'57"
3	05055400	15006708	Sheyenne River below Devils Lake State outlet near Bremen, N. Dak.	47°49'17"	99°16'34"
4	05056000	--	Sheyenne River near Warwick, N. Dak.	47°48'20"	98°42'57"
5	05057000	--	Sheyenne River near Cooperstown, N. Dak.	47°25'58"	98°01'38"
6	05058000	--	Sheyenne River below Baldhill Dam, N. Dak.	47°02'02"	98°05'00"
7	05059300	--	Sheyenne River above Sheyenne River Diversion near Horace, N. Dak.	46°45'01"	96°55'35"

For the study described in this report, the USGS, in cooperation with the North Dakota State Water Commission (NDSWC), analyzed the data collected at the seven sites on the Sheyenne River to determine whether streamflow and the continuously recorded physical properties could act as surrogates to estimate water-quality constituents that are important indicators of surface-water quality. The methods used in this study replicate those previously used in North Dakota for the Red River at Fargo (Ryberg, 2006) and in Kansas (Christensen and others, 2000; Christensen, 2001; Rasmussen and others, 2005; and Christensen and others, 2006) and may be replicated for other sites in North Dakota and the Nation to monitor water quality.

Estimation of water-quality constituents on the basis of surrogates provides several benefits. Although periodic Sheyenne River water-quality samples are collected manually, the delay associated with laboratory analysis does not permit immediate identification of undesirable levels of constituents. A relation between manually collected water-quality samples and continuously recorded water-quality physical properties allows immediate identification of potential water-quality problems. Examination of streamflow and physical properties of water that act as surrogates for constituents of interest also helps optimize visits for the collection of water-quality samples. For example, if specific conductance is used as a surrogate for estimating sulfate, it is desirable to collect water-quality samples representative of the range of specific-conductance values possible for the Sheyenne River at each site. In addition, the continuous estimation of water-quality information is useful in tracking changes in water quality as they occur.

Purpose and Scope

This report presents the results of regression analysis of water-quality constituents for seven sites on the Sheyenne River, N. Dak., using manually collected water-quality data and continuously recorded streamflow and physical property data from seven sites on the Sheyenne River. The data were collected during 1980–2006. The report provides regression equations that can be used to estimate concentrations and loads for hardness, dissolved solids, calcium, magnesium, sodium, and sulfate.

Description of Study Area

The Sheyenne River (fig. 1) originates in Sheridan County in central North Dakota and flows through southeastern North Dakota, ultimately emptying into the Red River, north of Fargo. The Sheyenne River has a drainage area of about 6,910 square miles, not including the closed Devils Lake Basin (Emerson, 2005).

At times there is no flow in the upper reaches of the Sheyenne River. Flows in lower reaches are regulated partly by release from Baldhill Dam, approximately 12 miles north of Valley City. Baldhill Dam, completed in 1951, provides water supply, flood control, recreation, and wildlife habitat (U.S. Army Corps of Engineers, 2005).

Site 1 is USGS gaging station 05054500 Sheyenne River above Harvey, N. Dak. and is located 2 miles upstream from an unnamed tributary, and 4.5 miles south of Harvey. The recording gage was established in September 1955. Periodic water-quality sampling began at this site in 1971. Specific

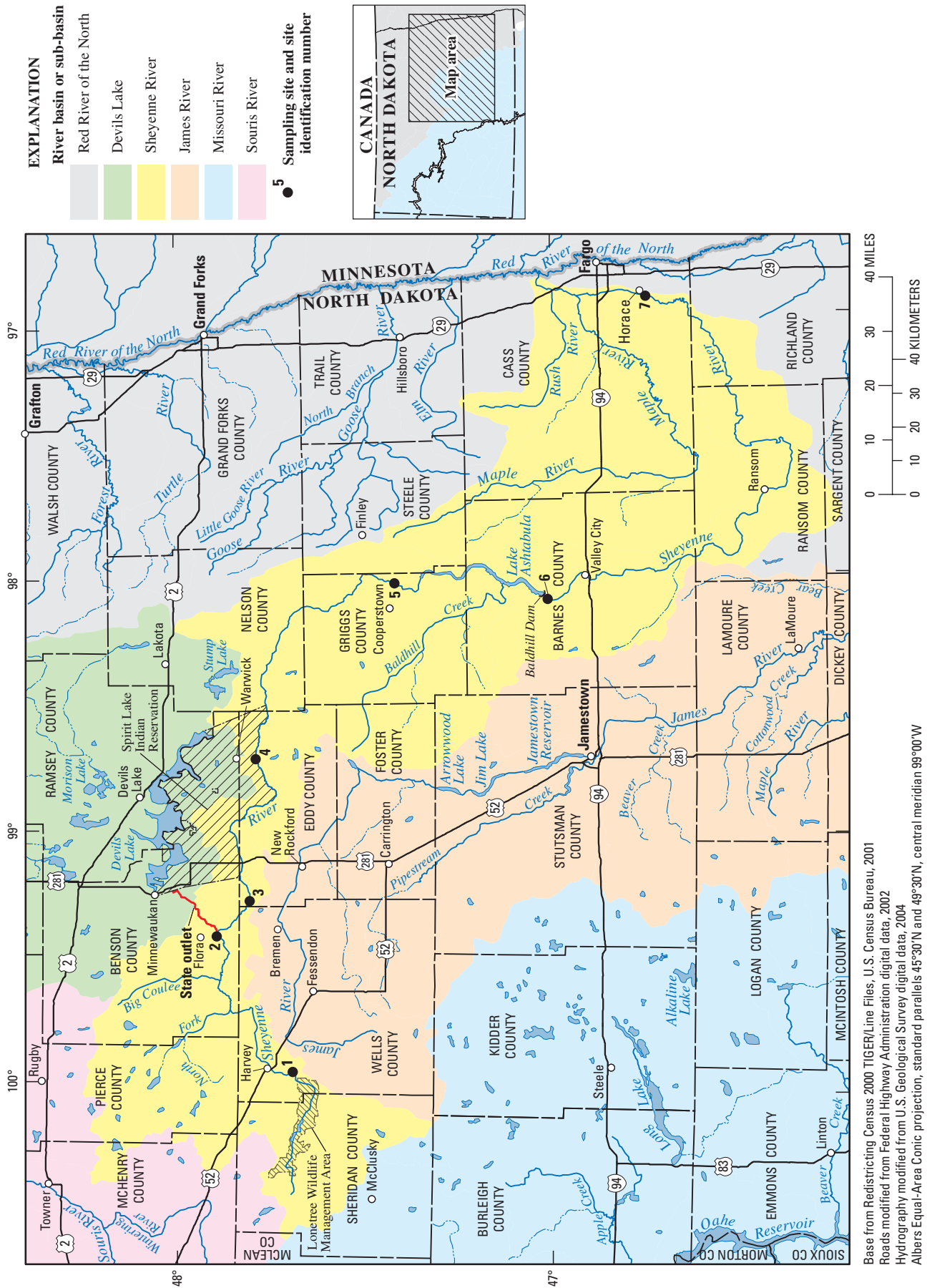


Figure 1. Location of the seven sampling sites on the Sheyenne River, N. Dak.

4 Continuous Water-Quality Monitoring and Regression Analysis, Sheyenne River, 1980–2006

conductance and water temperature have been continuously recorded since May 2006. This gage records little or no flow most winters and is heavily affected by ice (L.A. Cottengaim, USGS, station description, unpub. data, 2007).

Site 2 is USGS gaging station 05055300 Sheyenne River above Devils Lake State outlet near Flora, N. Dak. The gage began recording stage, specific conductance, and water temperature in October 2004, 3.5 mi southeast of Flora (R.A. Nustad, USGS, station description, unpub. data, 2006). Periodic water-quality sampling by the USGS and the NDSWC began at this site in 2005. Gage height is affected by ice and by growth of vegetation in the channel during periods of low flow (P.M. Scarpari, USGS, station analysis, unpub. data, 2006).

Site 3 is USGS gaging station 05055400 Sheyenne River below Devils Lake State outlet near Bremen, N. Dak. The seasonal (operated April through November) recording gage was established in March 2005, 10 mi northeast of Bremen. Specific conductance and water temperature have been continuously recorded, seasonally, since April 2005. Periodic water-quality sampling by the USGS and the NDSWC began at this site in 2005 (R.A. Nustad, USGS, station description, unpub. data, 2007).

Site 4 is USGS gaging station 05056000 Sheyenne River near Warwick, N. Dak. The recording gage was installed November 1949, 0.25 mi west and 3.3 mi south of Warwick. Periodic water-quality sampling at this site began in 1951. Specific conductance and water temperature have been continuously recorded since May 4, 2006 (P.M. Scarpari, USGS, station description, unpub. data, 2007).

Site 5 is USGS gaging station 05057000 Sheyenne River near Cooperstown, N. Dak. A recording gage was established in 1950 and moved to the current location in October 1985, at Ueland Dam, 5 mi east of Cooperstown. Periodic water-quality sampling at this site began in 1959. Specific conductance and water temperature have been continuously recorded since June 1997. Records may be poor during winter periods (R.A. Nustad, USGS, station description, unpub. data, 2006).

Site 6 is USGS gaging station 05058000 Sheyenne River below Baldhill Dam, N. Dak. The gage was established in 1949 and has been moved several times. Since September 2000, the gage has been located 0.1 mi downstream from Baldhill Dam, approximately 0.25 mi upstream from a bridge. Periodic water-quality sampling at this site began in 1959. Specific conductance and water temperature have been continuously recorded since April 1997. The station is generally not affected by ice because of releases from Baldhill Dam (R.A. Nustad, USGS, station description, unpub. data, 2007).

Site 7 is USGS gaging station 05059300 Sheyenne River above Sheyenne River Diversion near Horace, N. Dak. The recording gage was established in October 1992, 1 mi southwest of Horace, upstream from the diversion structure. Periodic water-quality sampling began at this site in 1992. Specific conductance and water temperature have been continuously recorded since June 1997. Gage height is affected by ice through most of the winter months and flow is regulated

to a large degree by Baldhill Dam, approximately 230 mi upstream (P.M. Scarpari, USGS, station description, unpub. data, 2007).

Additional information about these sites is available online at http://waterdata.usgs.gov/nd/nwis/nwisman/?site_no=XXXXXXXX, where XXXXXXXX is the 8-digit USGS station number listed in table 1.

Previous Studies

Regression relations have been developed previously for sites on the Sheyenne River to estimate monthly mean constituent concentrations (Guenther, 1991; Williams-Sether, 2004). Guenther's study (1991) preceded continuous recording of specific conductance and water temperature; therefore, the only continuously recorded variable available for regression analysis was streamflow. The only continuously recorded variable used by Williams-Sether (2004) was streamflow. Regression analysis to estimate constituent concentrations and loads was used by Ryberg (2006) for the Red River at Fargo, N. Dak., and has been used for numerous sites in Kansas (Christensen and others, 2000; Christensen, 2001; Rasmussen and others, 2005; and Christensen and others, 2006).

Methods

Data collection and analysis methods have varied over time and from site to site. This section describes how streamflow was measured, water-quality samples were collected and analyzed, and data were interpreted for this study.

Streamflow Measurements

Stage, or gage height, was measured to the nearest 0.01 ft at all stations. The data for all stations are stored and publicly available in the National Water Information System: Web Interface (NWISWeb), <http://waterdata.usgs.gov/nd/nwis>.

Methods used to determine streamflow are described in Buchanan and Somers (1969). Streamflow measurements were made periodically at each station. A stage-streamflow relation was developed on the basis of streamflow measurements and the stage of the stream at the time of measurement (Kennedy, 1984), and this relation was used to compute a continuous record of streamflow (Kennedy, 1983).

Manual Water-Quality Measurements, Sample Collection, and Analysis

During the collection of water-quality samples, physical properties of water, including specific conductance and water temperature, were measured. USGS water-quality samples were collected manually according to methods described in USGS techniques manuals (U.S. Geological Survey, variously

dated). After the manually measured physical property and water-quality data were checked and approved by the USGS, they were made publicly available in NWISWeb at <http://nwis.waterdata.usgs.gov/nd/nwis/qwdata>.

Water-quality samples collected by the USGS were analyzed by the NDSWC Laboratory or the North Dakota Department of Health Laboratory. The NDSWC Laboratory and the North Dakota Department of Health Laboratory followed U.S. Environmental Protection Agency approved methods and procedures. The NDSWC Laboratory closed July 2003. Water-quality samples were analyzed by the North Dakota Department of Health Laboratory using standardized methods (North Dakota Department of Health, 2003b) and quality assurance procedures (North Dakota Department of Health, 2003a). External agencies and customer organizations audited the North Dakota Department of Health Laboratory to assure the laboratory analytical methods and quality-assurance/quality-control procedures. The U.S. Environmental Protection Agency reviewed the laboratory procedures about every 3 years, and the Branch of Quality Systems of the USGS periodically reviewed the laboratory procedures.

Water-quality samples collected by the NDSWC were analyzed by the North Dakota Department of Health Laboratory using standardized methods (North Dakota Department of Health, 2003b) and quality-assurance procedures (North Dakota Department of Health, 2003a). Data are stored in the NDSWC's Ground and Surface Water database available online at <http://www.swc.nd.gov/4dlink2/4dcgi/wellsearchform/Map%20and%20Data%20Resources>. In the NDSWC database, the NDSWC identifies site 2, Sheyenne River above Devils Lake State outlet near Flora, N. Dak., with the location number 15106807, and site 3, Sheyenne River below Devils Lake State outlet near Bremen, N. Dak., with the location number 15006708. Only data from the NDSWC database that had field specific-conductance measurements were used for this report. The NDSWC's Ground and Surface Water database does not provide streamflow with the water-quality data. Therefore, the USGS automated data processing system (ADAPS, Revision NWIS-4.6.0-40) was used to obtain interpolated unit values for streamflow. ADAPS stores the streamflow values for the USGS Sheyenne River gaging sites in one-hour increments and interpolates a streamflow value based on the Water Commission's specified sampling time and the streamflow for the closest hour before and the closest hour after the sampling time.

Water-quality data were collected as long ago as 1951. However, using data from such a long period of time introduces many sources of variability. Field and laboratory methods change over time and three different laboratories were used for some sites with long periods of record (sites 1, 4, 5, and 6; fig. 1). In examining scatterplots and regression diagnostic plots, an inordinate number of outliers appeared in the data from the 1970s (data not shown). Many of the potential outliers may have been transposed during data entry, but the original data are no longer available to check. Water temperature was identified as a statistically significant

explanatory variable for many constituents, but this was an unexpected finding and did not follow from the results of previous studies (Christensen and others, 2000; Christensen, 2001; Ryberg, 2006). By limiting water-quality data to samples collected beginning in 1980 through 2006, some of the variability in laboratories, two rather than three analyzing laboratories, was eliminated, the number of potential outliers was reduced, and water temperature was no longer statistically significant as an explanatory variable for constituent concentration.

Data were removed from the regression analysis based on streamflow to eliminate samples collected from pooled and stagnant water. For sites 4 and 5, only water-quality samples with an associated streamflow of at least 5.00 ft³/s were used, based on structures at the gage and river characteristics (R.F. Lundgren and W.C. Damschen, USGS, oral commun., 2007). Site 1, generally having streamflow or being dry, is not as prone to pooling as sites 4 and 5 (R.F. Lundgren and W.C. Damschen, USGS, oral commun., 2007); however, the point of zero flow at site 1 is a gage height of 3.50 ± 0.05 ft (L.A. Cottengaim, USGS, station description, unpub. data, 2007), and 3.55 ft is equivalent to a streamflow of 0.34 ft³/s (U.S. Geological Survey, stage-discharge rating 10.0, November 21, 2006). Therefore, only water-quality samples with an associated streamflow greater than 0.34 ft³/s were used for site 1. For sites 2 and 3, only samples with an associated streamflow of at least 4.00 ft³/s were used. No data were used that were collected after July 13, 2006, for site 2 and July 7, 2006, for site 3 because streamflow was less than 4.0 ft³/s or was affected by ice for the remainder of 2006. Sites 6 and 7 have higher minimum streamflow caused by regulation at Baldhill Dam and tributary inflow; therefore, data for these sites were not removed based on streamflow.

The frequency of water-quality sampling varied from site to site and with time, resulting in some instances of serial correlation, especially with data collected in the 1980s. To reduce serial correlation, which may result in residuals that are not independent, the data were thinned (Helsel and Hirsch, 1995). After removal of data based on streamflow, the data were thinned for sites 1, 4, 5, 6, and 7 (fig. 1; table 1) so that no samples were closer than 4 weeks apart. Because of the shorter period of record, data for sites 2 and 3 (fig. 1; table 1) were thinned so that no samples were closer than 3 weeks apart.

Development of Regression Equations to Estimate Constituent Concentrations

All available water-quality constituents were considered for estimation using regression equations. However, some of them were removed from consideration due to various data issues, such as no data since the early 1980s (bicarbonate), not enough data (total phosphorus), and the data being too coarse, or discrete, for regression analysis (fluoride). Other constituents had enough data, but relations useful for prediction were

not found (potassium and chloride at sites 1, 4, 5, 6, and 7, and dissolved nitrite plus nitrate at site 1). The water-quality constituents selected for estimation are listed in table 2.

The rationale and methodology for expressing water-quality constituent concentrations in terms of other surrogate constituents or physical properties in a regression equation are explained in Helsel and Hirsch (1995). Helsel and Hirsch (1995) also detail the computations for regression estimation and identify measures commonly used to evaluate regression relations, including R^2 , the coefficient of determination. They explain the benefit of log transformation of variables, which was a technique used for some of the regression relations in this study.

R^2 is calculated as follows:

$$R^2 = 1 - \frac{SSE}{SS_y} \quad (1)$$

where

SSE is the error sum of squares, and
 SS_y is the sums of squares y, or total sums of squares, defined in Helsel and Hirsch (1995).

R^2 is a number, 0 through 1, that when multiplied by 100 is interpreted as the percentage of the variability in the response variable explained by the explanatory variables and the regression equation. Generally, the higher the R^2 , the better the regression relation. However, this does not guarantee the regression equation is useful (Neter and others, 1996). For example, if estimates require extrapolation outside the ranges of observed variables, the regression equation may not provide accurate estimates.

All-subset regression by leaps and bounds, a procedure that attempts to find the best regression relations using subsets of the given explanatory variables (Insightful Corporation, 2005), was used to select the best regression model for estimating a particular constituent. The possible explanatory variables were streamflow, specific conductance, and water temperature (table 3) as well as transformations of streamflow and specific conductance, $\log_{10}(\text{streamflow})$, and $\log_{10}(\text{specific conductance})$.

The criteria for determining the best regression relation were the adjusted coefficient of multiple determination (R_a^2) and Mallows's C_p (C_p). R_a^2 increases with the number of explanatory variables in the regression model, but R_a^2 allows for the comparison of models that have differing numbers of explanatory variables by penalizing models that have additional coefficients (Helsel and Hirsch, 1995). The C_p criterion is a measure of the total mean squared error and an indicator of model bias (Neter and others, 1996). In comparing models, the models with the lowest C_p values are considered those with the least bias.

Potential models with relatively high R_a^2 and low C_p were further examined using standard diagnostics for regression (Neter and others, 1996). The most common problems in the diagnostic residual plots were non-normality and heteroscedasticity, or non-constant variance, both violations of the assumptions underlying parametric regression. Transformations of the response variable (constituent concentration in this study) are effective fixes for both of these problems, which often occur together (Neter and others, 1996). Logarithmic transformations of the explanatory and response variables resulted in residuals that were approximately normally distributed. An explanatory variable was considered statistically significant and selected for regression relations if the p-value (attained significance level) for the variable was less than 0.05.

A resampling technique, jackknife after bootstrap, was used to examine the influence of individual observations in the development of the regression relations. "In bootstrap resampling, B [where B is a large number, such as 1,000] new samples, each of the same size as the observed data, are drawn with replacement from the observed data. The statistic is first calculated using the observed data and then recalculated using each of the new samples, yielding a bootstrap distribution. The resulting replicates are used to calculate the bootstrap estimates of bias, mean, and standard error for the statistic" (Insightful Corporation, 2001a). In this study the bootstrap statistics were the regression parameter estimates, the coefficients of the explanatory variables, and the intercept term. The jackknife after bootstrap was helpful in identifying very influential observations that were removed as outliers and indicated

Table 2. Constituents estimated using regression equations for the Sheyenne River, N. Dak., 1980–2006.

Constituent	Unit of measurement	U.S. Geological Survey parameter code
Hardness	Milligrams per liter as calcium carbonate	P00900
Dissolved solids, calculated	Milligrams per liter	P70301
Calcium	Milligrams per liter	P00915
Magnesium	Milligrams per liter	P00925
Sodium	Milligrams per liter	P00930
Sulfate	Milligrams per liter	P00945

Table 3. Possible surrogate physical properties, or explanatory variables, used to develop regression relations to estimate constituent concentrations in Shyenne River, N. Dak., 1980–2006.

Physical properties	Unit of measurement	U.S. Geological Survey parameter code
Streamflow	Cubic feet per second	P00060
Specific conductance	Microsiemens per centimeter at 25° Celsius	P00095
Temperature, water	Degrees Celsius	P00010

that the regression parameters estimated in this study were relatively unbiased in their mean and standard error.

An additional measure of influence, Cook's distance, was used along with residual plots to determine outliers or highly influential samples. Cook's distance quantifies the influence of individual samples on the predicted values (Neter and others, 1996).

In building a regression model, the final step of the process is validation of the selected regression model. Validation generally is done in one of three ways: (1) collection of new data to check the model and its predictions; (2) comparison of results with earlier observations, simulations, theoretical experiments, or physical theory; or (3) use of data that were held out of the model-building process to check the model and its prediction accuracy (Neter and others, 1996). Collection of new data is time-consuming and expensive and several years of data may be necessary to obtain samples representative of high and low streamflows, specific conductance, and constituent concentrations. Results of continued water-quality sampling should be compared with the regression relations presented in this report as hydrologic, climatologic, or human factors may change the relations. The regression relations were compared to previous studies and examined for scientific validity. Some data were withheld from the model building process and used to check the predictions made by regression relations.

As an indicator of the ability of the regression relations to estimate constituent concentrations, the measured concentrations were compared to the concentrations estimated by the regression relations by calculating relative percentage differences (*RPDs*) using the following equation:

$$RPD = \left| \frac{E - M}{M} \times 100 \right|, \quad (2)$$

where

- E is the constituent concentration estimated from the regression equation, and
- M is the measured constituent concentration.

M is assumed to be correct and the *RPD* is the relative difference of E from M , expressed as a percentage.

Logarithmic transformations of the explanatory and response variables resulted in improved *RPDs* over untransformed variables and in a better linear relation between the

measured and estimated concentrations. The improvement in the *RPD* when the explanatory and response variables were logarithmically transformed is shown in figure 2 for sodium concentrations at site 1. The top graph shows estimated and measured sodium concentrations with an *RPD* of 14.3. When the variables were logarithmically transformed, the *RPD* decreased to 7.5. In figure 2A, the relation between estimated and measured concentrations appears slightly curvilinear, with measured concentrations generally above the estimated concentration line for low concentrations, measured concentrations scattered on both sides of the line for concentrations in the middle, and high measured concentrations below the estimation line. In figure 2B the data generated with logarithmically transformed explanatory and response variables shows a much more linear relation between the measured and estimated concentrations.

Measured and estimated concentrations were compared graphically to evaluate the usefulness of the regression relations for prediction using continuously recorded streamflow and specific conductance data at each gage. The daily mean of the continuously recorded explanatory variables at each site and the regression relations for sulfate were used to predict sulfate concentrations along with 90-percent prediction intervals (Neter and others, 1996). The prediction intervals have a 90-percent probability of containing the true sulfate population (Helsel and Hirsch, 1995).

Calculation of Measured and Estimated Constituent Loads

Daily load is the total mass of a constituent that is transported past a gaging station in 1 day. Measured constituent loads were calculated by multiplying measured constituent concentrations by streamflow at the time the constituent concentrations were measured and multiplying by the conversion factor listed in table 4. Estimated mean daily constituent loads were calculated by multiplying the daily constituent concentrations estimated using the regression equations by daily streamflow and by a conversion factor (table 4).

For regression relations developed in terms of a logarithmically transformed constituent concentration, retransformation to the original units can cause an underestimation of chemical loads when adding individual load estimates over

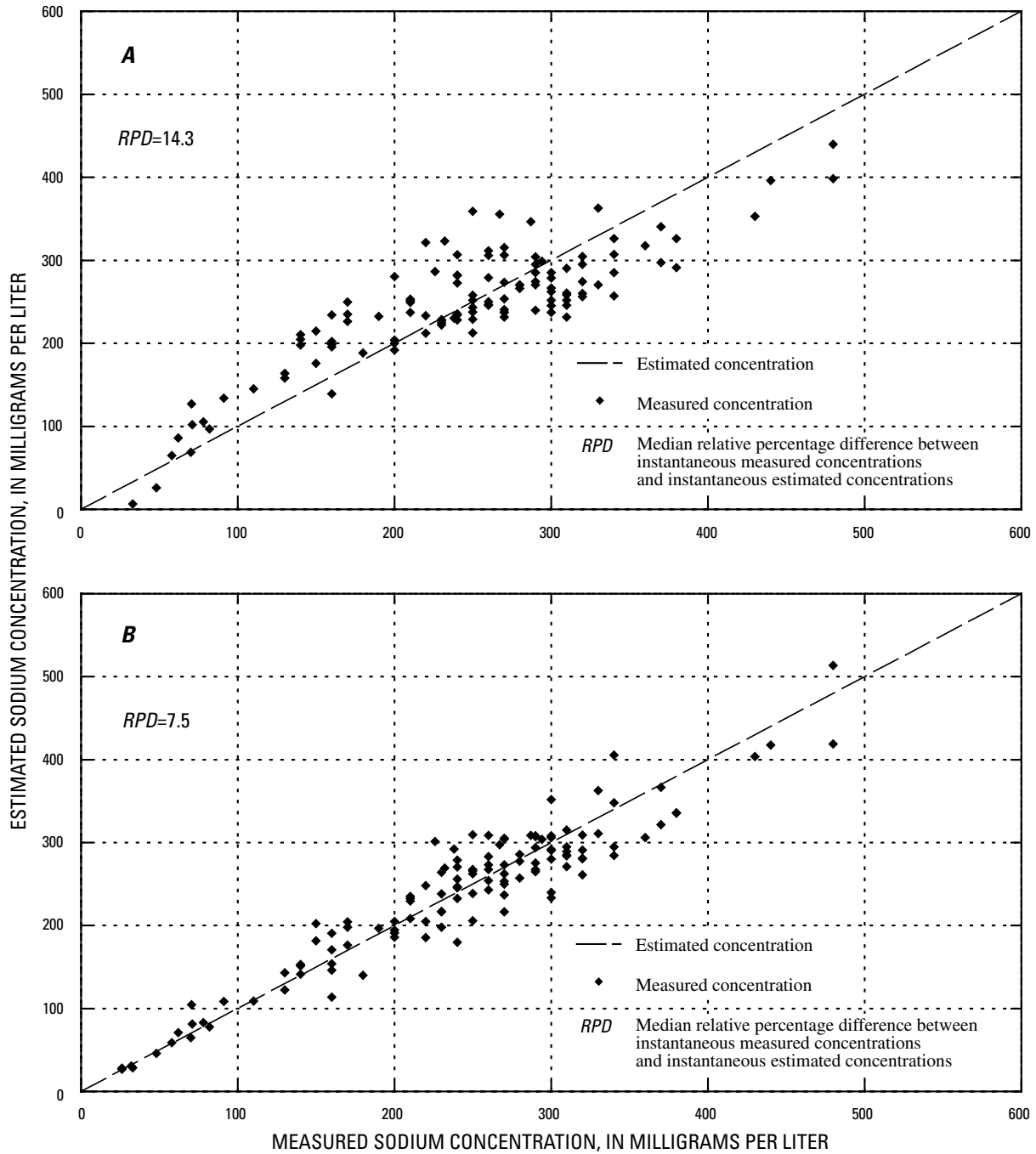


Figure 2. Comparison of measured and estimated sodium concentrations in the Sheyenne River, N. Dak., at site 1, *A* is a comparison using untransformed explanatory and response variables and *B* is a comparison using logarithmically transformed explanatory and response variables.

Table 4. Conversion factors used in calculation of measured and estimated loads.

[ft³/s, cubic feet per second]

Multiply	By	By	To obtain
milligrams per liter	streamflow, in ft ³ /s	5.39	pounds per day

a long period of time (Christensen, 2001). Multiplying the calculated load by a bias correction factor (*BCF*; Duan, 1983) corrects for this underestimation. Calculation of the *BCF* is shown in the equation below:

$$BCF = \frac{\sum_{i=1}^n 10^{e_i}}{n} \tag{3}$$

where

- e_i is the regression residual, in log units, and
- n is the number of samples used to develop the regression relation.

Continuous Water-Quality Monitoring

Continuous monitoring of specific conductance and water temperature began at site 1 (fig. 1; table 1) in May 2006 using a Campbell Scientific (CS) 547A conductivity/temperature probe. At site 2 (fig. 1; table 1), continuous monitoring of specific conductance and water temperature began in October 2004 with a CS 547 conductivity/temperature probe. On May 9, 2006, the probe was replaced with a YSI Incorporated (YSI) 600R conductivity/temperature probe. At site 3 (fig. 1; table 1), continuous monitoring of specific conductance and water temperature began in April 2005 with a CS 547 conductivity/temperature probe. On May 8, 2006, the probe was replaced with a YSI 600R conductivity/temperature probe.

At site 4 (fig. 1; table 1), continuous monitoring of specific conductance and water temperature began in May 2006, with a CS 547 conductivity/temperature probe. The probe was installed on the left edge of the water on the same side as a natural spring that discharges to the Sheyenne River 60–70 ft upstream from the probe. Based on field readings on October 24, 2006, the specific conductance in the spring may be lower than the specific conductance in the river and the spring may affect the specific conductance recorded by the probe. A hydrologic comparison of specific conductance at sites 2, 3, 4, and 5 showed that specific conductance at site 4 follows the same trend as that at site 5 but may be influenced by the spring during periods of low streamflow (1–10 ft³/s) (P.M. Scarpari, USGS, station analysis, unpub. data, 2007).

At site 5 (fig. 1; table 1), continuous monitoring of specific conductance and water temperature began in June 1997, with a CS 247 conductivity/temperature probe. At site 6 (fig. 1; table 1), continuous monitoring of specific conductance and water temperature began in April 1997, with a CS 547 conductivity/temperature probe. The probe was initially installed 0.3 mi downstream from the current position. The initial location was downstream from Baldhill Dam and downstream from return flow from the Baldhill Dam National Fish Hatchery. In the new location, since September 18, 2000, only water released from the dam is in contact with the probe (R.A. Nustad, USGS, station description, unpub. data, 2007). At site 7 (fig. 1; table 1), continuous monitoring of specific conductance and water temperature began in June 1997, with a CS 547 conductivity/temperature probe.

The CS probes were installed in buried, flexible poly-pipe intakes which extended from the gage houses to the river. The pipes were perforated at the end to allow river water to come in contact with the probes. In the CS probes, the electrical conductance sensor “consists of three stainless steel rings mounted in an epoxy tube...Resistance of water passing through the tube is measured by excitation of the center electrode with positive and negative voltage...Temperature is measured with a thermistor in a three wire half bridge configuration” (Campbell Scientific, Inc., 1996, 2000, and 2001). The accuracy of the conductivity sensor in standard ionized solutions is ±5 percent of reading in the range of 440 to 7,000 μS/cm at 25°C and ±10 percent of reading in the range of 5 to 440 μS/cm at 25°C (Campbell Scientific, Inc., 1996, 2000, and 2001). The accuracy of the temperature sensor is ±0.4°C in the range of –24°C to 48°C (Campbell Scientific, Inc., 1996, 2000, and 2001).

The YSI probes were installed in buried galvanized steel pipe. The pipes were perforated at the end to allow river water to come in contact with the probes. The YSI probes utilize “a cell with four pure nickel electrodes for the measurement of solution conductance. Two of the electrodes are current driven, and two are used to measure the voltage drop” (YSI Incorporated, 2006). For temperature measurement, the probes utilize “a thermistor of sintered metallic oxide that changes predictably in resistance with temperature variation” (YSI Incorporated, 2006). The accuracy of the conductivity sensor is ±0.5 percent of reading plus 1 μS/cm in the range of 0 to

100,000 $\mu\text{S}/\text{cm}$ and the accuracy of the temperature sensor is $\pm 0.15^\circ\text{C}$ in the range of -5 to 50°C (YSI Incorporated, 2006).

All probes were periodically cleaned and calibrated. For the CS 247, 547, and 547A models, calibration drift corrections were made when the specific conductance readings differed by more than 5 percent of the specific conductance of the standard solution used for calibration. Fouling corrections are made when readings before and after cleaning differed by more than 5 percent (L.A. Cottengaim, USGS, station analysis, unpub. data, 2007). The YSI 600R probes were calibrated and data were corrected according to criteria described by Wagner and others (2000).

The data for all sites were transmitted through satellite telemetry and downloaded into the USGS Automated Data Processing System (ADAPS), and then the data were made available on the Internet from NWISWeb.

Summary Statistics

Summary statistics of the manually measured physical properties and water-quality constituents are listed in table 5. The data collected by the USGS are stored in the USGS NWIS database and are available online at http://nwis.waterdata.usgs.gov/nd/nwis/qwdata?site_no=XXXXXXXX, where XXXXXXXX is the eight digit USGS station number identified in table 1. The data collected by the NDSWC are stored in the NDSWC's Ground and Surface Water database and are available online at <http://www.swc.nd.gov/4dlink2/4dcgi/wellsearchform/Map%20and%20Data%20Resources>.

Results of Regression Analysis

Regression relations between constituents and surrogate physical properties were examined, and a regression relation was developed for each constituent using one or more surrogate variables. A discussion of each constituent and the associated regression relations follows. Also included in the discussion are the samples, if any, that were removed from the regression analysis. Almost all of the samples removed were from February, March, and November with the majority being from March. These are periods when the sites may be affected by ice, especially sites 1–5 and 7; therefore, caution should be used in using the regression relations for concentration estimation during periods of ice.

Hardness

Hardness is a property of water resulting from the presence of calcium and magnesium that is associated with the reaction of water with soap (Hem, 1985). The higher the hardness, the more soap is required for cleansing, while water with hardness less than 30–50 mg/L may be corrosive to pipes,

depending on pH, alkalinity, and dissolved oxygen (North Dakota Department of Health, 2006). The hardness data used in this study were expressed in terms of calcium carbonate concentration.

Specific conductance “provides general indication of the content of dissolved matter for water that is not too saline or too dilute” (Hem, 1985); therefore, specific conductance may indicate the presence of calcium and magnesium. Specific conductance was a statistically significant predictor variable for hardness in the Sheyenne River at sites 4, 5, 6, and 7 (table 6). Site 1 showed nonconstant variance in the residual plots and had a high median *RPD* (greater than 20). Hardness was not analyzed for a regression relation at sites 2 and 3 because of insufficient data.

Regression diagnostics plots, Cook's distance (Neter and others, 1996), and jackknife after bootstrap (Insightful Corporation, 2001b), were all used to examine regression results for outliers and highly influential samples. For site 5, jackknife after bootstrap and Cook's distance indicated that the March 31, 2004, sample was highly influential in the estimation of regression parameters and therefore was removed from the regression analysis. For site 7, jackknife after bootstrap indicated that the February 6, 2006, sample was highly influential in the estimation of regression parameters. This sample also stood out in regression diagnostic plots of the residuals; therefore, the February 6, 2006, sample was removed from the regression analysis.

Dissolved Solids

Dissolved-solids concentration is used to evaluate water quality and to compare water from one location to the next. Dissolved-solids concentration may be determined in two ways: (1) by the weight of the dry residue remaining after evaporation of the water samples; or (2) by summing the concentrations of other dissolved constituents, if the concentrations of major ions are known (Hem, 1985). According to Hem, the computed value for dissolved solids, which was used in this study, “may give a more useful indication of total dissolved-ion concentration than the residue left by evaporation.” High concentrations of dissolved solids may have an adverse effect on taste and make the water appear cloudy (North Dakota Department of Health, 2006).

Specific conductance has been an effective predictor of dissolved solids in previous studies (Ryberg, 2006; and Christensen and others, 2000). Streamflow also may affect dissolved solids, but the relation is not a simple one (Allan, 1995). For the Sheyenne River, specific conductance was a significant explanatory variable for estimating dissolved solids at sites 1, 4, 5, 6, and 7; streamflow was not (table 7). Dissolved solids was not analyzed for a regression relation at sites 2 and 3 because of insufficient data.

Regression diagnostics plots, Cook's distance (Neter and others, 1996), and jackknife after bootstrap (Insightful Corporation, 2001b), were used to examine regression results

Table 5. Summary of statistics for water-quality sampling sites on the Sheyenne River, N. Dak., 1980–2006.

[Samples for sites 2 and 3 collected by North Dakota State Water Commission, samples for all other sites collected by U.S. Geological Survey; streamflow in cubic feet per second; specific conductance in microsiemens per centimeter at 25 degrees Celsius; temperature in degrees Celsius; hardness in milligrams per liter as calcium carbonate; other constituents in milligrams per liter]

Water-quality physical property or constituent	Descriptive statistics			Percent of samples in which values were less than or equal to those shown					
	Sample size	Maximum	Minimum	Mean	0.95	0.75	(Median) 0.50	0.25	0.05
Site 1, Sheyenne River above Harvey, N. Dak.									
Streamflow	224	498	0.19	36	199	28	6.6	2.1	0.79
Specific conductance	215	2,300	280	1,270	1,860	1,540	1,320	1,110	435
Temperature, water	221	28.2	0	9.5	24.5	17.3	8.0	.5	0
Hardness	147	550	61	220	440	290	190	140	87
Dissolved solids, calculated	146	1,610	192	869	1,240	1,010	892	793	327
Calcium	146	74	14	36	62	45	34	27	20
Magnesium	147	92.6	5.6	30.1	65.8	42	23	16	8.6
Sodium	147	480	20	231	370	300	240	170	57.2
Sulfate	147	610	44	240	460	280	220	180	85
Chloride	146	32	3.6	17	28	20	16	13	6.0
Site 2, Sheyenne River above Devils Lake State outlet near Flora, N. Dak.									
Streamflow	125	600	4.0	59	264	67	20	8.3	4.4
Specific conductance	125	1,880	570	1,550	1,830	1,700	1,600	1,530	806
Sulfate	125	520	110	370	490	420	380	330	200
Site 3, Sheyenne River below Devils Lake State outlet near Bremen, N. Dak.									
Streamflow	125	579	4.0	61	171	84	30	13	5.5
Specific conductance	125	1,880	625	1,560	1,810	1,710	1,620	1,530	940
Sulfate	125	540	160	380	500	440	380	350	220
Site 4, Sheyenne River near Warwick, N. Dak.									
Streamflow	247	3,160	0.29	206	997	156	29	11	2.2
Specific conductance	238	1,900	218	903	1,420	1,130	909	655	440
Temperature, water	243	30.0	0	9.9	25.9	18.0	8.0	1.0	0
Hardness	60	430	87	260	380	320	270	190	100
Dissolved solids, calculated	60	911	184	530	851	713	527	363	223
Calcium	60	79	20	48	68	56	49	38	23
Magnesium	60	59	9.0	33	52	43	33	22	12
Sodium	60	190	14	92	160	130	81	51	25
Sulfate	60	310	37	140	280	210	130	80.5	46.8
Chloride	60	37	3.5	14	22	17	14	10	5.5

Table 5. Summary of statistics for water-quality sampling sites on the Sheyenne River, N. Dak., 1980–2006.—Continued

[Samples for sites 2 and 3 collected by North Dakota State Water Commission, samples for all other sites collected by U.S. Geological Survey; streamflow in cubic feet per second; specific conductance in microsiemens per centimeter at 25 degrees Celsius; temperature in degrees Celsius; hardness in milligrams per liter as calcium carbonate; other constituents in milligrams per liter]

Water-quality physical property or constituent	Descriptive statistics			Percent of samples in which values were less than or equal to those shown					
	Sample size	Maximum	Minimum	Mean	0.95	0.75	(Median) 0.50	0.25	0.05
Site 5, Sheyenne River near Cooperstown, N. Dak.									
Streamflow	327	5,290	0.4	395	1,810	321	88.8	34.7	7.64
Specific conductance	314	2,000	280	921	1,360	1,110	948	686	410
Temperature, water	321	28.0	0	9.6	24.5	17.2	8.0	1.0	0
Hardness	102	510	99	290	430	350	310	220	130
Dissolved solids, calculated	103	1,130	182	558	801	686	587	401	261
Calcium	102	110	24	61	97	72	64	46	29
Magnesium	102	58	9.5	33	50	40	33	23	13
Sodium	102	190	19	86	140	110	89	57	31
Sulfate	103	360	51	160	270	180	150	120	70
Chloride	112	39	5.2	15	24	17	15	11	6.6
Site 6, Sheyenne River below Baldhill Dam, N. Dak.									
Streamflow	285	5,510	0.05	450	2,360	346	107	33	10
Specific conductance	280	1,630	366	866	1,270	1,020	861	684	529
Temperature, water	283	26.2	0	10.5	24.4	19.3	7.5	3.0	1.0
Hardness	77	430	150	290	410	360	280	230	200
Dissolved solids, calculated	79	842	262	566	781	685	542	455	382
Calcium	77	78	33	58	78	69	58	47	39
Magnesium	77	60	17	36	55	46	35	27	23
Sodium	77	130	27	83	120	100	78	65	49
Sulfate	79	310	78	180	300	230	160	130	100
Chloride	89	31	8.2	16	24	18	15	13	9.3
Site 7, Sheyenne River above Sheyenne River Diversion near Horace, N. Dak.									
Streamflow	175	4,080	40	857	3,110	1,030	402	209	78
Specific conductance	168	1,510	418	908	1,270	1,050	945	749	490
Temperature, water	169	28.5	0	11.8	25.0	21.0	12.0	1.0	0
Hardness	32	500	170	350	440	400	370	300	190
Dissolved solids, calculated	34	810	291	608	795	692	642	524	320
Calcium	32	110	40	76	99	88	76	65	43
Magnesium	32	54	18	39	50	45	40	32	20
Sodium	32	110	31	75	100	91	80	59	33
Sulfate	34	300	98	220	290	250	220	190	110
Chloride	34	40	2.9	23	35	28	22	18	11

for outliers or highly influential points. For site 5, the sample from March 28, 1991, appeared to be an outlier in regression diagnostic plots and the sample from March 31, 2004, was highly influential in the estimation of regression parameters based on Cook's distance and jackknife after bootstrap; both samples were removed. For site 6, the sample from November 21, 2005, appeared to be an outlier in residual plots and had a large Cook's distance; therefore, the sample was removed from the regression analysis.

Table 6. Regression equations for estimates of hardness in the Sheyenne River, N. Dak., 1980–2006.

[n, numbers of samples used to develop regression equation; $CaCO_3$, hardness, in milligrams per liter as calcium carbonate; SC , specific conductance, in microsiemens per centimeter at 25 degrees Celsius; R^2 , coefficient of multiple determination; median RPD , median relative percentage difference; BCF , bias correction factor; --, not applicable]

Site number	n	Equation	Range of explanatory variable	R^2	Median RPD	BCF
1	--	--	--	--	--	--
2	--	--	--	--	--	--
3	--	--	--	--	--	--
4	45	$\log_{10}(CaCO_3) = 0.991 \log_{10}(SC) - 0.501$	$SC: 311-1,372$	0.915	7.7	1.0081
5	64	$\log_{10}(CaCO_3) = 1.040 \log_{10}(SC) - 0.601$	$SC: 295-1,600$.947	6.1	1.0041
6	65	$\log_{10}(CaCO_3) = 1.003 \log_{10}(SC) - 0.489$	$SC: 438-1,320$.936	3.9	1.0020
7	30	$\log_{10}(CaCO_3) = 1.016 \log_{10}(SC) - 0.482$	$SC: 449-1,293$.942	5.8	1.0020

Table 7. Regression equations for estimates of dissolved solids in the Sheyenne River, N. Dak., 1980–2006.

[n, numbers of samples used to develop regression equation; DS , dissolved solids concentration, in milligrams per liter; SC , specific conductance, in microsiemens per centimeter at 25 degrees Celsius; R^2 , coefficient of multiple determination; median RPD , median relative percentage difference; BCF , bias correction factor; --, not applicable]

Site number	n	Equation	Range of explanatory variable	R^2	Median RPD	BCF
1	127	$\log_{10}(DS) = 1.054 \log_{10}(SC) - 0.352$	$SC: 320-2,300$	0.986	2.2	1.0011
2	--	--	--	--	--	--
3	--	--	--	--	--	--
4	45	$\log_{10}(DS) = 1.098 \log_{10}(SC) - 0.488$	$SC: 311-1,372$.993	1.7	1.0008
5	64	$\log_{10}(DS) = 1.065 \log_{10}(SC) - 0.386$	$SC: 320-2,300$.991	2.2	1.0007
6	65	$\log_{10}(DS) = 1.046 \log_{10}(SC) - 0.328$	$SC: 295-1,600$.997	2.6	1.0007
7	30	$\log_{10}(DS) = 1.032 \log_{10}(SC) - 0.290$	$SC: 449-1,380$.966	2.7	1.0012

Calcium

Calcium contributes to hardness, is “a major component of the solutes in most natural waters,” and it is generally “the predominant cation in river water” (Hem, 1985). Previous studies have shown a positive linear relation between calcium and specific conductance (Christensen and others, 2003; Rasmussen and others, 2005; Christensen and others, 2006). Specific conductance was a statistically significant predictor variable for calcium in the Sheyenne River at sites 4, 5, 6, and 7 (table 8). Streamflow was also a statistically significant predictor variable for calcium in the Sheyenne River at sites 5, 6, and 7. Despite logarithmic transformation of the explanatory and response variables, site 1 had a nonlinear regression relation; therefore, a relation is not reported for calcium at site 1. This supports the lack of relation found for hardness as well. Calcium was not analyzed for a regression relation at sites 2 and 3 because of insufficient data.

Magnesium

Magnesium contributes to hardness and is “a common element essential in plant and animal nutrition” (Hem, 1985). Previous studies have shown a positive linear relation between magnesium and specific conductance (Rasmussen and others, 2004; Christensen and others, 2006). Specific conductance was a statistically significant predictor variable for magnesium in the Sheyenne River at sites 4, 5, 6, and 7 (table 9). Streamflow was also a statistically significant

predictor variable for magnesium in the Sheyenne River at site 5. Despite logarithmic transformation of the explanatory and response variables, site 1 showed nonconstant variance in the residuals; therefore a relation is not reported for magnesium at site 1. This supports the lack of relation found for hardness as well. Magnesium was not analyzed for a regression relation at sites 2 and 3 because of insufficient data.

Sodium

Sodium in drinking water may be a concern for individuals on sodium restricted diets and high concentrations of sodium may make water unsuitable for irrigation (North Dakota Department of Health, 2006). Human activities, such as deicing of highways in the winter, may influence the concentrations of sodium in ground and surface water (Hem, 1985).

Previous studies have shown a positive linear relation between sodium and specific conductance, as well as multiple regression relations that include pH and streamflow in addition to specific conductance (Christensen and others 2003; Rasmussen and others, 2004; Christensen and others, 2006). Specific conductance was a statistically significant predictor variable for sodium in the Sheyenne River at sites 1, 4, 5, 6, and 7 (table 10). Streamflow was also a statistically significant predictor variable for sodium in the Sheyenne River at sites 1 and 6. Sodium was not analyzed for a regression relation at sites 2 and 3 because of insufficient data.

Table 8. Regression equations for estimates of calcium in the Sheyenne River, N. Dak., 1980–2006.

[n, numbers of samples used to develop regression equation; *Ca*, calcium concentration, in milligrams per liter; *Q*, streamflow, in cubic feet per second; *SC*, specific conductance, in microsiemens per centimeter at 25 degrees Celsius; *R*², coefficient of multiple determination; median *RPD*, median relative percentage difference; *BCF*, bias correction factor; --, not applicable]

Site number	n	Equation	Range of explanatory variable(s)	<i>R</i> ²	Median <i>RPD</i>	<i>BCF</i>
1	--	--	--	--	--	--
2	--	--	--	--	--	--
3	--	--	--	--	--	--
4	45	$\log_{10}(Ca) = 0.770 \log_{10}(SC) - 0.585$	<i>SC</i> : 311–1,372	0.784	9.5	1.0148
5	65	$\log_{10}(Ca) = -0.040 \log_{10}(Q) + 0.818 \log_{10}(SC) - 0.548$	<i>Q</i> : 7.6–4,260 <i>SC</i> : 280–1,600	.909	7.6	1.0065
6	65	$\log_{10}(Ca) = 0.014 \log_{10}(Q) + 0.856 \log_{10}(SC) - 0.790$	<i>Q</i> : 8.2–3,050 <i>SC</i> : 438–1,362	.839	5.7	1.0040
7	27	$\log_{10}(Ca) = -0.061 \log_{10}(Q) + 0.708 \log_{10}(SC) - 0.080$	<i>Q</i> : 60.2–3,810 <i>SC</i> : 449–1,380	.884	5.3	1.0037

Table 9. Regression equations for estimates of magnesium in the Sheyenne River, N. Dak., 1980–2006.

[n, numbers of samples used to develop regression equation; *Mg*, magnesium concentration, in milligrams per liter; *Q*, streamflow, in cubic feet per second; *SC*, specific conductance, in microsiemens per centimeter at 25 degrees Celsius; *R*², coefficient of multiple determination; median *RPD*, median relative percentage difference; *BCF*, bias correction factor; --, not applicable]

Site number	n	Equation	Range of explanatory variable(s)	<i>R</i> ²	Median <i>RPD</i>	<i>BCF</i>
1	--	--	--	--	--	--
2	--	--	--	--	--	--
3	--	--	--	--	--	--
4	45	$\log_{10}(Mg) = 1.205 \log_{10}(SC) - 2.016$	<i>SC</i> : 311–1,372	0.953	6.6	1.0062
5	65	$\log_{10}(Mg) = 0.038 \log_{10}(Q) + 1.232 \log_{10}(SC) - 2.181$	<i>Q</i> : 7.6–4,260 <i>SC</i> : 280–1,600	.949	6.8	1.0054
6	65	$\log_{10}(Mg) = 1.144 \log_{10}(SC) - 1.814$	<i>SC</i> : 438–1,320	.944	4.9	1.0023
7	28	$\log_{10}(Mg) = 1.079 \log_{10}(SC) - 1.627$	<i>SC</i> : 449–1,380	.955	4.9	1.0018

Sulfate

Sulfate sources include the weathering of rocks, agricultural runoff, fuel combustion, municipal and industrial effluent, and precipitation (Hem, 1985; Allan, 1995). The presence of too much sulfate has three main undesirable effects in water used by humans: (1) sulfate can have a laxative effect with excessive intake, (2) water with high sulfate concentrations can form hard scales in boilers, and (3) sulfate can negatively affect taste (North Dakota Department of Health, 2006).

Sulfate is negatively charged and increased sulfate concentrations increase specific conductance (Hem, 1985). Streamflow was statistically significant at sites 1, 4, and 5 only (table 11).

Jackknife after bootstrap, Cook's distance, and regression plots identified the March 30, 2005, sample at site 1, the November 21, 2005, sample at site 6, and the September 16, 1993, sample at site 7 as highly influential points. They were removed from the regression analyses.

For sites 2 and 3, the period of record is small compared to the other sites and it is a period of record with relatively low flow and high specific conductance. Additional high flow data might result in the inclusion of the streamflow term in the regression equation, which is present in the equations for the upstream and downstream sites (1 and 4). The equations presented in table 11 should not be extrapolated beyond the range of explanatory variables used to develop the regression equations.

Additional high flow data might include concentrations affected by storm runoff and sites 1, 2, and 3 may be more affected by storm runoff than sites farther downstream. The storm effect would depend on the location of the storm. The north fork of the Sheyenne River, which joins the Sheyenne

River between Harvey and Flora (fig. 1) is less sulfatic than the Sheyenne River (W.M. Schuh, North Dakota State Water Commission, written commun., 2007). Therefore, increased streamflow at sites 2 and 3 caused by storm runoff may have varying effects on sulfate because of the location of the storm.

In addition to the numeric measurements used to evaluate the regression relations, measured and estimated concentrations were compared graphically to evaluate the usefulness of the regression relations for prediction using continuously recorded streamflow and specific conductance data at each gage. The permit to release water from the Devils Lake outlet into the Sheyenne River is based in part on sulfate concentration; therefore, there is a great deal of interest in sulfate in the Sheyenne River Basin and sulfate was used to illustrate the graphical comparison of measured and estimated concentrations in figures 3 and 4. For sites 1 and 4, continuous recording of specific conductance began in May 2006, resulting in little data for continuous prediction in 2006; therefore, sites 1 and 4 are not compared graphically. For sites 2 and 3, the daily mean of the continuously recorded specific conductance and streamflow data at each site from January 1, 2005, through December 31, 2006, and the regression relations for sulfate were used to predict sulfate concentrations along with 90-percent prediction intervals, representing the error in the prediction of sulfate. The sample concentrations used to develop the regression relations were plotted as points as well as the sample concentrations that were removed when the data were thinned. The regression relations for sites 2 and 3 were developed using samples collected by the NDSWC; however, the USGS also collected samples at these sites as an independent verification of the NDSWC's sampling program and the USGS sample concentrations also are plotted as points.

Table 10. Regression equations for estimates of sodium in the Sheyenne River, N. Dak., 1980–2006.

[n, numbers of samples used to develop regression equation; Na , sodium concentration, in milligrams per liter; Q , streamflow, in cubic feet per second; SC , specific conductance, in microsiemens per centimeter at 25 degrees Celsius; R^2 , coefficient of multiple determination; median RPD , median relative percentage difference; BCF , bias correction factor; --, not applicable]

Site number	n	Equation	Range of explanatory variable(s)	R^2	Median RPD	BCF
1	125	$\log_{10}(Na) = -0.120 \log_{10}(Q) + 1.168 \log_{10}(SC) - 1.217$	Q : 0.35–362 SC : 320–2,300	0.952	7.5	1.0076
2	--	--	--	--	--	--
3	--	--	--	--	--	--
4	43	$\log_{10}(Na) = 1.379 \log_{10}(SC) - 2.069$	SC : 311–1,372	.958	9.2	1.0072
5	65	$\log_{10}(Na) = 1.282 \log_{10}(SC) - 1.857$	SC : 280–1,600	.938	9.0	1.0083
6	64	$\log_{10}(Na) = -0.029 \log_{10}(Q) + 1.235 \log_{10}(SC) - 1.671$	Q : 8.2–3,050 SC : 438–1,362	.886	7.5	1.0058
7	28	$\log_{10}(Na) = 1.242 \log_{10}(SC) - 1.827$	SC : 449–1,380	.876	8.8	1.0075

Table 11. Regression equations for estimates of sulfate in the Sheyenne River, N. Dak., 1980–2006.

[n, numbers of samples used to develop regression equation; SO_4 , sulfate concentration, in milligrams per liter; Q , streamflow, in cubic feet per second; SC , specific conductance, in microsiemens per centimeter at 25 degrees Celsius; R^2 , coefficient of multiple determination; median RPD , median relative percentage difference; BCF , bias correction factor; --, not applicable]

Site number	n	Equation	Range of explanatory variable(s)	R^2	Median RPD	BCF
1	125	$\log_{10}(SO_4) = 0.161 \log_{10}(Q) + 1.492 \log_{10}(SC) - 2.414$	Q : 0.35–362 SC : 320–2,300	0.894	8.0	1.0108
2	11	$\log_{10}(SO_4) = 1.251 \log_{10}(SC) - 1.434$	SC : 575–1,760	.922	8.9	1.0059
3	11	$\log_{10}(SO_4) = 1.021 \log_{10}(SC) - 0.692$	SC : 625–1,812	.850	8.2	1.0072
4	45	$\log_{10}(SO_4) = 0.126 \log_{10}(Q) + 1.475 \log_{10}(SC) - 2.420$	Q : 9–2,120 SC : 311–1,372	.898	7.9	1.0145
5	66	$\log_{10}(SO_4) = 0.096 \log_{10}(Q) + 1.308 \log_{10}(SC) - 1.848$	Q : 7.6–4,260 SC : 280–1,600	.893	9.8	1.0108
6	65	$\log_{10}(SO_4) = 1.318 \log_{10}(SC) - 1.635$	SC : 438–1,362	.820	11.5	1.0107
7	28	$\log_{10}(SO_4) = 1.026 \log_{10}(SC) - 0.722$	SC : 449–1,380	.928	5.2	1.0026

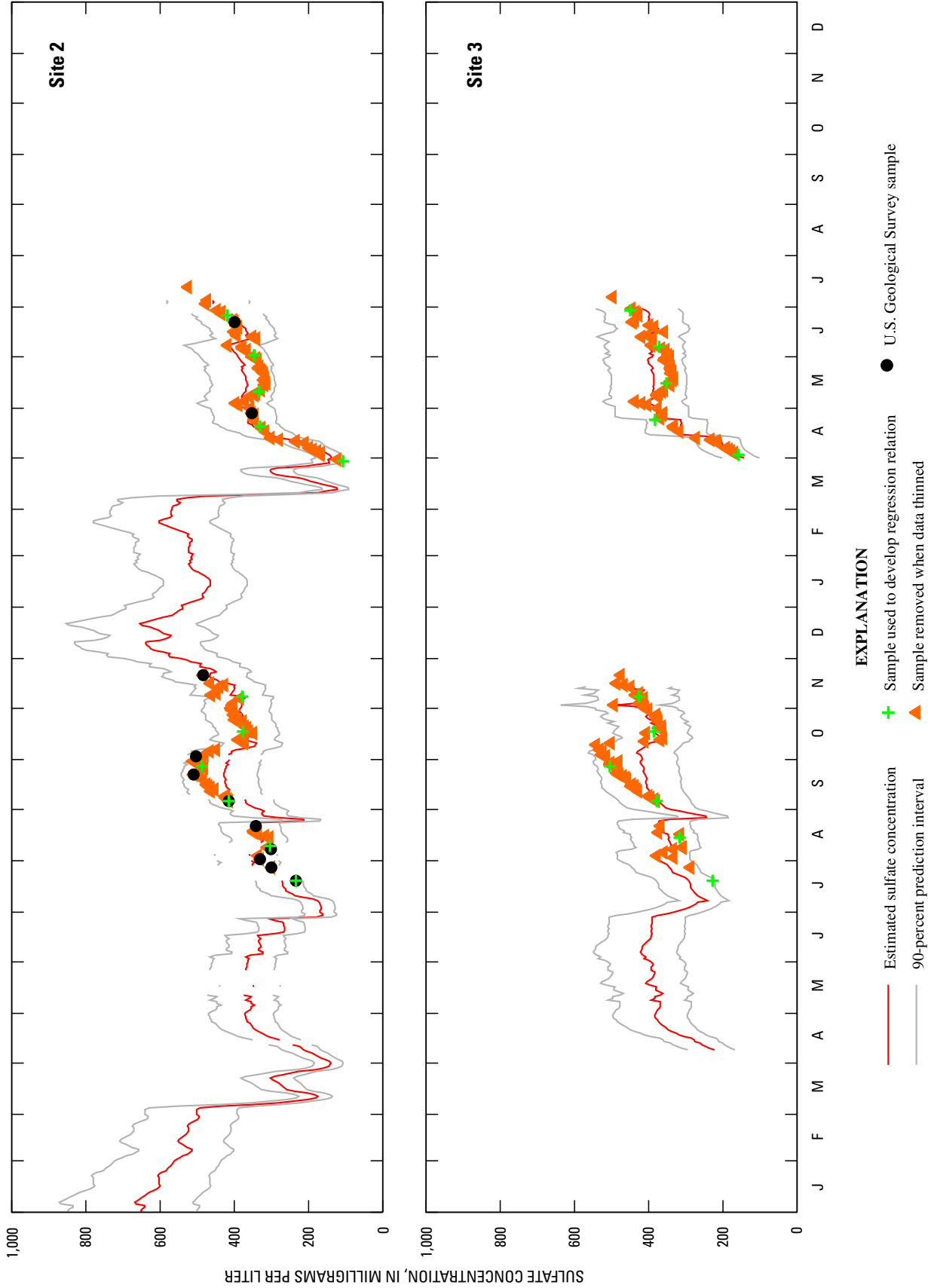


Figure 3. Predicted sulfate concentration, 90-percent prediction interval, and individual samples used in regression analyses for sites 2 and 3 on the Sheyenne River, N. Dak., 2005–06.

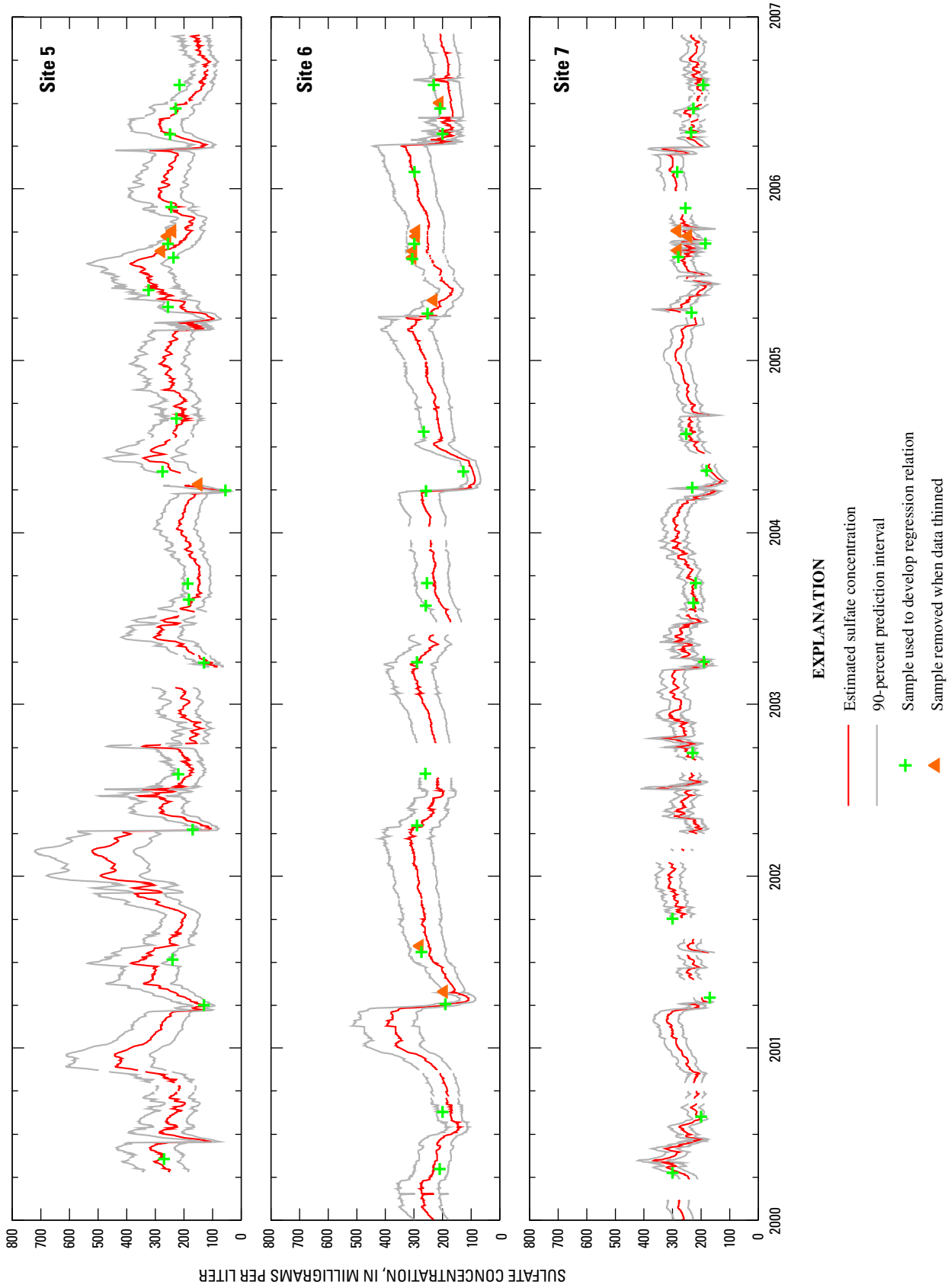


Figure 4. Predicted sulfate concentration, 90-percent prediction interval, and individual samples used in regression analyses for sites 5, 6, and 7 on the Sheyenne River, N. Dak., 2000–2006.

For sites 5, 6, and 7, the daily mean of the continuously recorded specific conductance and streamflow data from January 1, 2000, through December 31, 2006, and the regression relations for sulfate were used to predict sulfate concentrations at each site along with 90-percent prediction intervals, representing the error in the prediction of sulfate. The sample concentrations used to develop the regression relations were plotted as points, as well as the sample concentrations that were removed when the data were thinned. Breaks in the prediction lines (figs. 3 and 4) indicate periods in which one or more of the predictor variables were unavailable or periods in which streamflow fell outside the bounds defined in the “Manual Water-Quality Measurements, Sample Collection, and Analysis” section of this report.

Measured and Estimated Constituent Loads

Because the explanatory variables used in the regression equations (tables 6–11) are continuously monitored properties, the regression relations developed in this study may be

used to estimate daily constituent loads in the Sheyenne River. As an example of this application of the relations, the daily load for sulfate at site 7 is shown in figure 5. Measured daily load was computed and the daily load was estimated using the regression equation (table 11) and the daily mean streamflow and specific conductance. Breaks in the line representing estimated daily load indicate periods in which one or more of the explanatory variables was unavailable.

The estimated loads and measured loads compare well and the estimated load illustrates the high degree of variability in daily sulfate load. The peak load generally occurs in the second quarter of the year during spring runoff, April–June.

Summary

The U.S. Geological Survey, in cooperation with the North Dakota State Water Commission (NDSWC), analyzed data collected at seven sites on the Sheyenne River to determine whether streamflow and the continuously recorded physical properties could act as surrogates to estimate water-quality constituents that are important indicators of surface-water quality. Estimation of water-quality constituents on

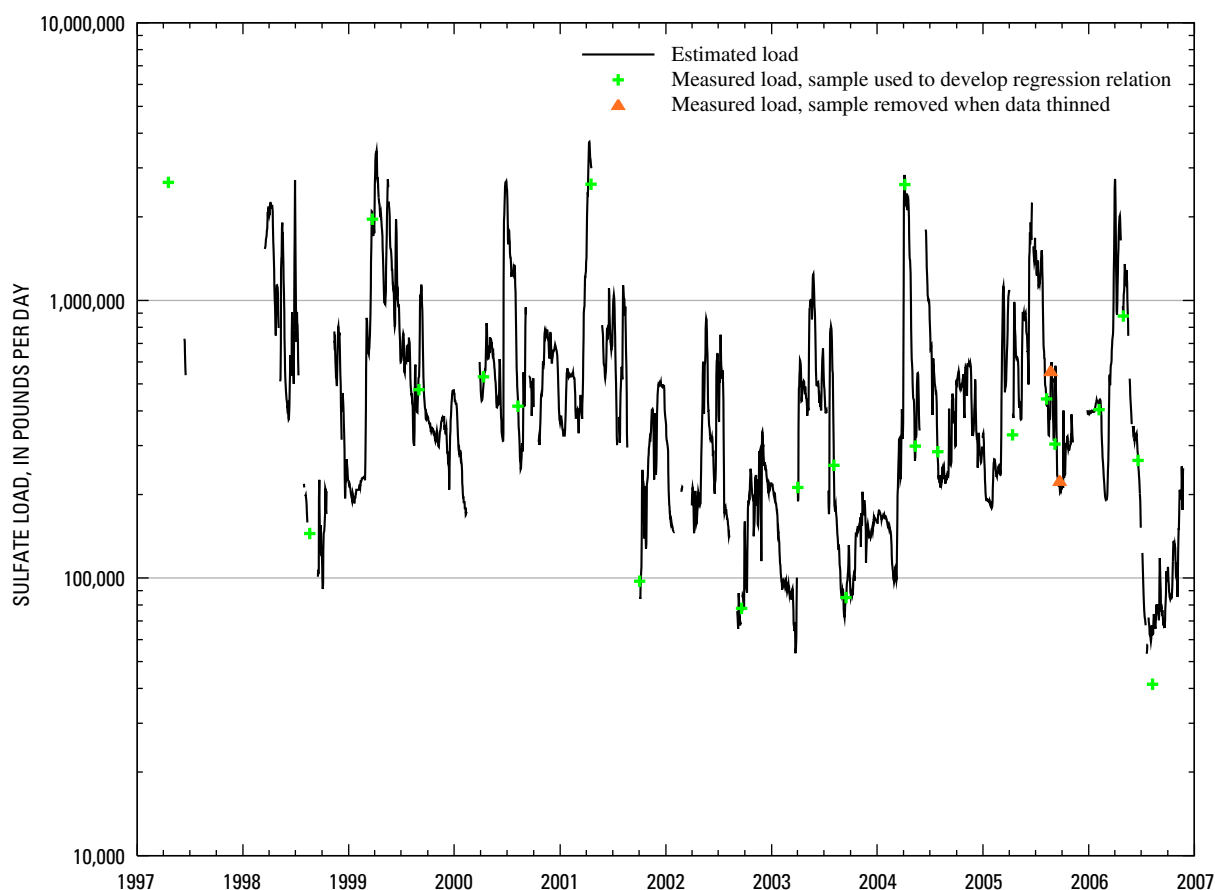


Figure 5. Comparison of measured and estimated sulfate loads for site 7 on the Sheyenne River, N. Dak., 1997–2006.

the basis of surrogates provides several benefits. Although periodic Sheyenne River water-quality samples are collected manually and analyzed, the delay associated with laboratory analysis does not permit immediate identification of undesirable concentrations of constituents. A relation between manually collected water-quality samples and real-time water-quality measurements allows immediate identification of potential water-quality problems. Examination of streamflow and physical properties of water that act as surrogates for constituents of interest also helps optimize visits for the collection of water-quality samples.

Data collection and analysis methods, as well as the periods of record for manual and continuous water-quality monitoring, vary among the sites. This variability was documented as part of the study and should be considered when applying the regression relations.

Constituents for which regression relations were developed were hardness, dissolved solids, calcium, magnesium, sodium, and sulfate. Potential explanatory variables were the continuously recorded water properties streamflow, specific conductance, and water temperature. Each regression relation was evaluated for usefulness based on the R^2 and median relative percentage differences (*RPDs*).

The base-10 logarithm of hardness was a function of the base-10 logarithm of specific conductance for sites 4–7. The hardness relations were characterized by R^2 of 0.915–0.947 and median *RPDs* of 3.9–7.7.

The base-10 logarithm of dissolved solids was a function of the base-10 logarithm of specific conductance for sites 1 and 4–7. The dissolved solids relations were characterized by R^2 of 0.966–0.997 and median *RPDs* of 1.7–2.7.

The base-10 logarithm of calcium was a function of the base-10 logarithm of specific conductance for site 4 and a function of the base-10 logarithm of streamflow and the base-10 logarithm of specific conductance for sites 5–7. The calcium relations were characterized by R^2 of 0.784–0.909 and median *RPDs* of 5.3–9.5.

The base-10 logarithm of magnesium was a function of the base-10 logarithm of specific conductance for sites 4, 6, and 7 and a function of the base-10 logarithm of streamflow and the base-10 logarithm specific conductance for site 5. The magnesium relations were characterized by R^2 of 0.944–0.955 and median *RPDs* of 4.9–6.8.

The base-10 logarithm of sodium was a function of the base-10 logarithm of streamflow and the base-10 logarithm of specific conductance for sites 1 and 6 and a function of the base-10 logarithm of specific conductance for sites 4, 5, and 7. The sodium relations were characterized by R^2 of 0.876–0.958 and median *RPDs* of 7.5–9.2.

The base-10 logarithm of sulfate was a function of the base-10 logarithm of streamflow and the base-10 logarithm of specific conductance for sites 1, 4, and 5 and a function of the base-10 logarithm of specific conductance for sites 2, 3, 6, and 7. The sulfate relations were characterized by R^2 of 0.820–0.928 and median *RPDs* of 5.2–11.5.

The regression relations may be used to continuously estimate constituent concentrations in the Sheyenne River and these estimates may be used to continuously estimate concentration loads. The relations should be monitored for change over time, especially at sites 2 and 3 which have a short period of record. In addition, caution should be used when the Sheyenne River is affected by ice and when the upstream sites (sites 1, 2, and 3) are affected by isolated storm runoff. Almost all of the outliers and highly influential samples removed from the analysis were from periods when the Sheyenne River might be affected by ice.

References Cited

- Allan, J.D., 1995, Stream ecology—Structure and function of running waters: London, Chapman & Hall, 388 p.
- Buchanan, T.J., and Somers, W.P., 1969, Discharge measurements at gaging stations: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. A8, 65 p., accessed June 2, 2006, at <http://pubs.usgs.gov/twri/twri3a8/>
- Campbell Scientific, Inc., 1996, 247 conductivity & temperature probes; Logan, Utah, Campbell Scientific, Inc., 11 p., accessed April 9, 2007, at <http://www.campbellsci.com/documents/manuals/retired/247.pdf>
- Campbell Scientific, Inc., 2000, CS547 Conductivity and temperature probe and A547 interface; Logan, Utah, Campbell Scientific, Inc., 20 p., accessed April 9, 2007, at <http://www.campbellsci.com/documents/manuals/cs547.pdf>
- Campbell Scientific, Inc., 2001, CS547A Conductivity and temperature probe and A547 interface; Logan, Utah, Campbell Scientific, Inc., 20 p., accessed April 9, 2007, at <http://www.campbellsci.com/documents/manuals/cs547a.pdf>
- Christensen, V.G., 2001, Characterization of surface-water quality based on real-time monitoring and regression analysis, Quivira National Wildlife Refuge, south-central Kansas, December 1998 through June 2001: U.S. Geological Survey Water-Resources Investigations Report 01–4248, 28 p., accessed March 2, 2006, at <http://pubs.water.usgs.gov/wri014248>
- 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 Ninescaw River and Cheney Reservoir, south-central Kansas, 1997–2003: U.S. Geological Survey Scientific Investigations Report 2006–5095, 43 p., accessed April 24, 2007, at <http://pubs.usgs.gov/sir/2006/5095/>

- Christensen, V.G., Jian, Xiaodong, and Ziegler, A.C., 2000, Regression analysis and real-time water-quality monitoring to estimate constituent concentrations, loads, and yields in the Little Arkansas River, south-central Kansas, 1995–99: U.S. Geological Survey Water-Resources Investigations Report 00–4126, 36 p., accessed March 2, 2006, at <http://pubs.water.usgs.gov/wri004126>
- Christensen, V.G., Ziegler, A.C., Rasmussen, P.P., and Jian, Xiaodong, 2003, Continuous real-time water-quality monitoring of Kansas streams, in Proceedings of 2003 Spring Specialty Conference on Agricultural Hydrology and Water Quality, May 12–14, 2003, Kansas City, MO.: Middleburg, VA., American Water Resources Association, AWRA Technical Publication Series No. TPS–03–1.
- Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: *Journal of the American Statistical Association*, v. 78, p. 605–610.
- Emerson, D.G., 2005, Historic and naturalized monthly streamflow for selected sites in the Red River of the North Basin in North Dakota, Minnesota, and South Dakota, 1931–2001: U.S. Geological Survey Scientific Investigations Report 2005–5092, 228 p., accessed April 4, 2007, at <http://pubs.water.usgs.gov/sir2005-5092/>
- Guenther, R.S., 1991, Methods for estimating monthly mean concentrations of selected water-quality constituents for stream sites in the Red River of the North Basin, North Dakota and Minnesota: U.S. Geological Survey Water-Resources Investigations Report 91–4086, 113 p., accessed April 4, 2007, at <http://pubs.er.usgs.gov/pubs/wri/wri914086>
- Helsel, D.R., and Hirsch, R.M., 1995, *Statistical methods in water resources*: New York, Elsevier Science B.V., 529 p.
- Hem, J.D., 1985, *Study and interpretation of the chemical characteristics of natural water*, (3d ed.): U.S. Geological Survey Water-Supply Paper 2254, 264 p., accessed March 7, 2006, at <http://pubs.er.usgs.gov/usgspubs/wsp/wsp2254>
- Insightful Corporation, 2001a, Resampling techniques—Bootstrap and jackknife, in *S-PLUS 6 for Windows guide to statistics*, v. 2: Seattle, Insightful Corporation, p. 537–562.
- Insightful Corporation, 2001b, *S-PLUS 6 for Windows user's guide*: Seattle, Insightful Corporation, 688 p.
- Insightful Corporation, 2005, *S-PLUS language reference*, in *S-PLUS 7.0 for Windows professional developer*; Seattle, Insightful Corporation [variously paged].
- Kennedy, E.J., 1983, Computation of continuous records of streamflow: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. A13, 53 p., accessed June 2, 2006, at <http://pubs.usgs.gov/twri/twri3-a13/>
- Kennedy, E.J., 1984, Discharge ratings at gaging stations: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. A10, 59 p., accessed June 2, 2006, at <http://pubs.usgs.gov/twri/twri3-a10/>
- Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W., 1996, *Applied linear statistical models* (4th ed.): Boston, WCB/McGraw-Hill, 1,408 p.
- North Dakota Department of Health, 2003a, North Dakota Department of Health Chemistry Division quality assurance plan: Bismarck, North Dakota Department of Health [variously paged].
- North Dakota Department of Health, 2003b, North Dakota Department of Health environmental methods manual: Bismarck, North Dakota Department of Health [variously paged].
- North Dakota Department of Health, 2006, How minerals affect water supplies: Miscellaneous publication, about 3 p., accessed June 9, 2006, at <http://www.health.state.nd.us/WQ/GW/pubs/mineral.htm>
- North Dakota State Water Commission [n.d.], Devils Lake quick facts: accessed July 28, 2006, at http://www.swc.state.nd.us/4DLink9/4dcgi/GetContentPDF/PB-206/DL_Quick_Facts.pdf
- North Dakota State Water Commission [n.d.], Devils Lake flooding—Outlet: accessed July 28, 2006, at <http://www.swc.state.nd.us/4dlink9/4dcgi/GetSubCategoryRecord/Devils%20Lake%20Flooding/Outlet>
- Rasmussen, T.J., Ziegler, A.C., and Rasmussen, P.P., 2005, Estimation of constituent concentrations, densities, loads, and yields in lower Kansas River, northeast Kansas, using regression models and continuous water-quality monitoring, January 2000 through December 2003: U.S. Geological Survey Scientific Investigations Report 2005–5165, 126 p., accessed April 24, 2007, at <http://pubs.usgs.gov/sir/2005/5165/>
- Ryberg, K.R., 2006, Continuous water-quality monitoring and regression analysis to estimate constituent concentrations and loads in the Red River of the North, Fargo, North Dakota, 2003–05: U.S. Geological Survey Scientific Investigations Report 2006–5241, 35 p., accessed April 24, 2007, at <http://pubs.water.usgs.gov/SIR20065241>

- U.S. Army Corps of Engineers, 2005, Lake Ashtabula–Baldhill Dam–Valley City, North Dakota: U.S. Army Corps of Engineers brochure [variously paged], accessed March 14, 2007, at <http://www.mvp.usace.army.mil/docs/rec/ashtabula.pdf>
- U.S. Geological Survey [variously dated], National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1–A9, [variously paged], accessed June 2, 2006, at <http://pubs.water.usgs.gov/twri9A>
- Wagner, R.J., Boulger, R.W., Jr., Oblinger, C.J., and Smith, B.A., 2000, Guidelines and standard procedures for continuous water-quality monitors—Station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1-D3 [variously paged], accessed April 11, 2007, at <http://pubs.usgs.gov/tm/2006/tm1D3/>
- Williams-Sether, T.J., 2004, Regression equations for estimating concentrations of selected water-quality constituents for selected gaging stations in the Red River of the North Basin, North Dakota, Minnesota, and South Dakota: U.S. Geological Survey Water-Resources Investigations Report 03–4291, 33 p., accessed April 26, 2006, at <http://pubs.water.usgs.gov/wri034291/>
- YSI Incorporated, 2006, 6-Series multiparameter water quality sondes user manual: Yellow Springs, Ohio, YSI Incorporated, 372 p., accessed April 9, 2007, at https://www.y.si.com/DocumentServer/DocumentServer?docID=EMS_FULL

