



Prepared in cooperation with the city of Wichita, Kansas

Relations between Continuous Real-Time Physical Properties and Discrete Water-Quality Constituents in the Little Arkansas River, South-Central Kansas, 1998–2014

By Patrick P. Rasmussen, Patrick J. Eslick, and Andrew C. Ziegler

Open-File Report 2016–1057

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

U.S. Department of the Interior

SALLY JEWELL, Secretary

U.S. Geological Survey

Suzette M. Kimball, Director

U.S. Geological Survey, Reston, Virginia: 2016

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <http://www.usgs.gov/> or call 1–888–ASK–USGS (1–888–275–8747).

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod/>.

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

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Rasmussen, P.P., Eslick, P.J., and Ziegler, A.C., 2016, Relations between continuous real-time physical properties and discrete water-quality constituents in the Little Arkansas River, south-central Kansas, 1998-2014: U.S. Geological Survey Open-File Report 2016–1057, 20 p., <http://dx.doi.org/10.3133/ofr20161057>.

ISSN 2331-1258 (online)

Acknowledgments

The authors appreciate the assistance of Debra Ary, Terryl Pajor, and Vernon Strasser of the city of Wichita for providing much of the chemical analysis used for model development. The authors thank U.S. Geological Survey technical reviewers Charles Crawford (Columbus, Ohio), Teresa Rasmussen (Lawrence, Kansas), David Mueller (Denver, Colorado), Robert Hirsch (Reston, Virginia), and Amie Brady (Columbus, Ohio) for their reviews of the report. Special thanks to Aaron King for the initial data analysis. Lastly, the report and data in the report would not have been possible without the diligent data collection by past and current (2016) U.S. Geological Survey staff in the Wichita, Kansas Field Office including Trudy Bennett, Serena Castillo, Carlen Collins, and Barb Dague.

Contents

Abstract	1
Introduction	1
Purpose and Scope	2
Description of Study Area	2
Methods.....	4
Results of Regression Analysis for Selected Constituents	7
Hardness.....	8
Dissolved Solids.....	8
Total Suspended Solids	8
Calcium.....	9
Sodium.....	9
Alkalinity.....	9
Bicarbonate.....	9
Chloride.....	10
Sulfate.....	10
Total Organic Nitrogen.....	10
Total Phosphorus.....	11
<i>Escherichia Coli</i> Indicator Bacteria	11
Fecal Coliform Indicator Bacteria	11
Arsenic.....	11
Atrazine.....	12
Total Organic Carbon.....	12
Suspended-Sediment Concentration	12
Summary	13
References Cited.....	13
Appendix 1. Model Outputs for Little Arkansas River at Highway 50 near Halstead, KS (site 07143672).16	
Appendix 2. Model Outputs for Little Arkansas River near Sedgwick, KS (site 07144100).	16

Figures

Figure 1. *Map showing land use in Little Arkansas River Basin and location of pertinent boundaries and sites.* 3

Conversion Factors

[U.S. Customary units to International System of Units]

Multiply	By	To obtain
Length		
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
square mile (mi ²)	2.590	square kilometer (km ²)
Volume		
milliliter (mL)	0.0338	ounce, fluid (oz)
cubic foot (ft ³)	28.32	cubic decimeter (dm ³)
Flow rate		
cubic foot per second (ft ³ /s)	0.02832	cubic meter per second (m ³ /s)
milligram per liter (mg/L)	1	parts per million (ppm)
Mass		
milligram (mg)	0.001	gram (g)

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:

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

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

Specific conductance is given in microsiemens per centimeter at 25 degrees Celsius (μS/cm at 25 °C).

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

Relations between Continuous Real-Time Physical Properties and Discrete Water-Quality Constituent Concentrations in the Little Arkansas River, South-Central Kansas, 1998–2014

By Patrick P. Rasmussen, Patrick J. Eslick, and Andrew C. Ziegler

Abstract

Water from the Little Arkansas River is used as source water for artificial recharge of the *Equus* Beds aquifer, one of the primary water-supply sources for the city of Wichita, Kansas. The U.S. Geological Survey has operated two continuous real-time water-quality monitoring stations since 1995 on the Little Arkansas River in Kansas. Regression models were developed to establish relations between discretely sampled constituent concentrations and continuously measured physical properties to compute concentrations of those constituents of interest. Site-specific regression models were originally published in 2000 for the near Halstead and near Sedgwick U.S. Geological Survey streamgaging stations and the site-specific regression models were then updated in 2003. This report updates those regression models using discrete and continuous data collected during May 1998 through August 2014. In addition to the constituents listed in the 2003 update, new regression models were developed for total organic carbon. The real-time computations of water-quality concentrations and loads are available at <http://nrtwq.usgs.gov>. The water-quality information in this report is important to the city of Wichita because water-quality information allows for real-time quantification and characterization of chemicals of concern (including chloride), in addition to nutrients, sediment, bacteria, and atrazine transported in the Little Arkansas River. The water-quality information in this report aids in the decision making for water treatment before artificial recharge.

Introduction

The *Equus* Beds Aquifer Storage and Recovery (ASR) Program began in 1995 (Ziegler and others, 1999) with the primary goals of restoring capacity to the *Equus* Beds aquifer to ensure adequate water quantity for the city of Wichita and the surrounding area in south-central Kansas and of slowing or stopping the migration of saltwater contamination to the aquifer. Monitoring for the ASR Program was a cooperative effort between the city of Wichita and the U.S. Geological Survey (USGS). Additional participants include the Bureau of Reclamation (BOR, U.S. Department of the Interior); the *Equus* Beds Groundwater Management District No. 2 (GMD2); Kansas Water Office; Kansas Department of Agriculture, Division of Water Resources; Kansas Department of Health and Environment; and the U.S. Environmental Protection Agency (EPA). Water from the Little Arkansas River is used as source water for artificial recharge of the *Equus* Beds aquifer. The *Equus* Beds ASR Program became operational in two phases; Phase I began in 2007, and Phase II began in 2013. The

facilities have an overall design capacity to treat and artificially recharge about 40 million gallons per day (Stone and others, 2012).

During the ASR Program, surface water at the USGS streamgaging stations on the Little Arkansas River at Highway 50 near Halstead (07143672, fig. 1) and the Little Arkansas River near Sedgwick (07144100, fig. 1), hereinafter near Halstead and near Sedgwick streamgaging station, respectively, was sampled to evaluate the quality of the source water. In addition, real-time water-quality monitors at the two stations provided continuous measures of specific conductance, pH, water temperature, dissolved oxygen, and turbidity. Regression models based on surrogate physical properties measured in real time are useful in the computation of water-quality constituent concentrations to support water treatment and recharge decisions, compare to water-quality criteria, and compute loads and yields to assess watershed transport.

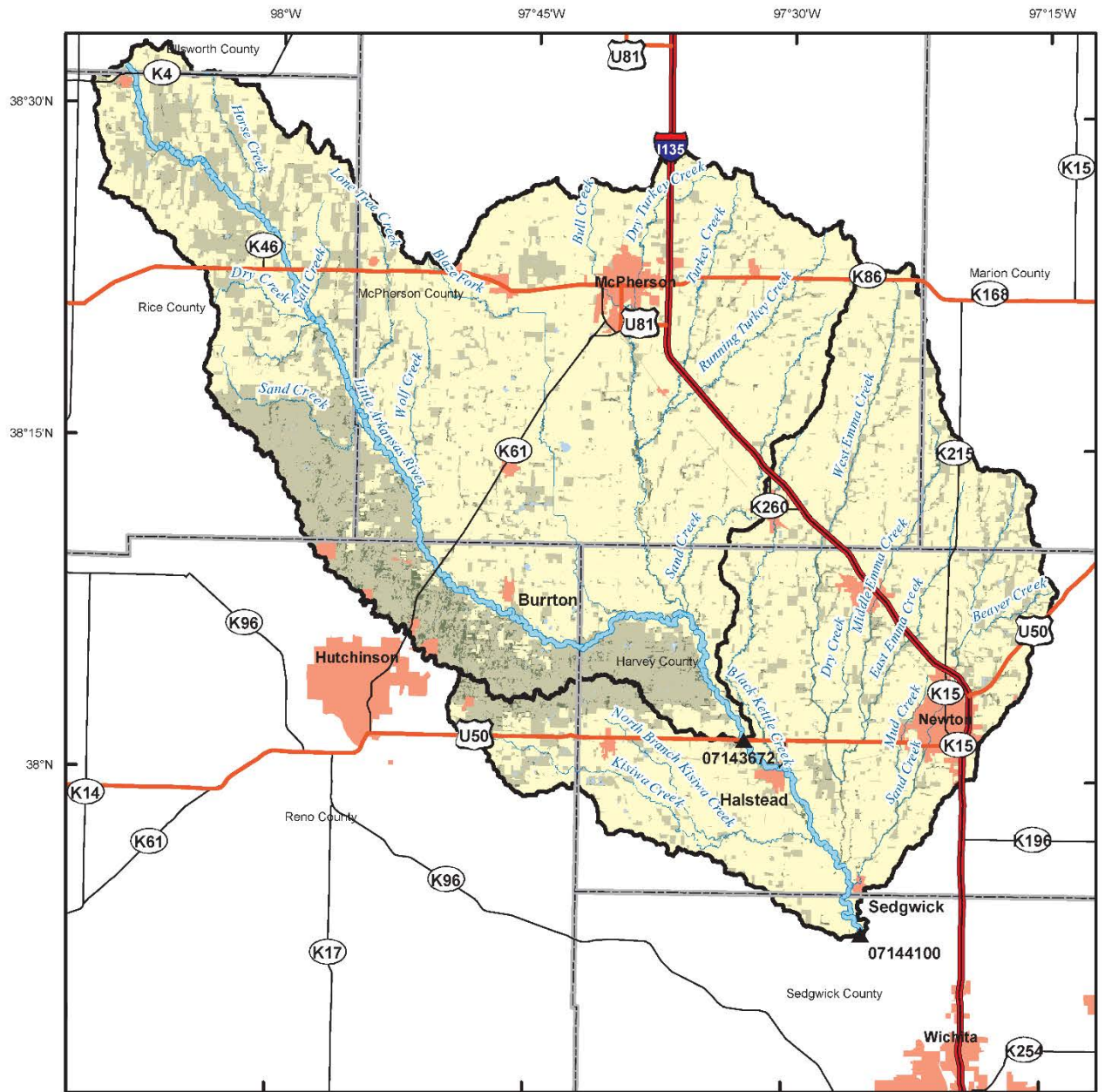
Purpose and Scope

The purpose of this report is to update and document site-specific regression models to compute concentrations of selected constituents using statistical relations between continuous and discrete water-quality data collected from the Little Arkansas River (USGS stations 07143672 and 07144100; fig. 1). Site-specific regression models were developed for the near Halstead and near Sedgwick streamgaging stations to assist with decision making for the city of Wichita to adjust water treatment and for reporting the quality of water used for aquifer recharge. The original regression models for these sites were published by Christensen and others (2000). The Christensen and others (2003) updates the models using continuous in-stream sensor measurements to compute selected constituent concentrations and loads in the Little Arkansas River. This report documents site-specific regression models for the computation of continuous concentrations of hardness, dissolved solids, total suspended solids, calcium, sodium, alkalinity, bicarbonate, chloride, sulfate, total organic nitrogen, total phosphorus, *Escherichia coli* bacteria (*E. coli*), fecal coliform bacteria (fecal coliform), arsenic, atrazine, and suspended-sediment from January 1, 2007 onward. Additionally, regression models for total organic carbon were developed. The real-time computations of water-quality concentrations and loads are available at <http://nrtwq.usgs.gov>. The water-quality information in this report is important to the city of Wichita because it allows for real-time quantification and characterization of chemicals of concern, including chloride, nutrients, sediment, bacteria, and atrazine transported in the Little Arkansas River.

Description of Study Area

The study area is located in south-central Kansas (fig. 1). The two data-collection sites involved in this study are USGS streamgaging stations on the Little Arkansas River at Highway 50 near Halstead and the Little Arkansas River near Sedgwick. The near Halstead streamgaging station has a contributing drainage area of about 685 square miles. The near Sedgwick streamgaging station has a contributing drainage area of 1,165 square miles (Christensen and others, 2000). Streamflow at both stations is affected by groundwater withdrawals, surface-water diversions, and return flow from irrigated areas (Putnam and others, 2000).

Land use in the Little Arkansas River Basin is primarily agricultural. In the study area, land use is about 70 percent cropland and 24 percent grassland (Peterson and others, 2005). The percentages of each crop type for the Little Arkansas River at Highway 50 near Halstead subbasin are about 66 percent wheat, 6 percent sorghum, 4 percent soybeans, 13 percent corn, and 8 percent alfalfa. Common agricultural chemicals applied to these crops include fertilizers (such as nitrate, ammonia, and phosphorus) and herbicides (such as alachlor and atrazine). Livestock raised in the area include cattle and hogs (Kansas Department of Agriculture, 2006).



Base map from U.S. Geological Survey digital data, 1:2,000,000, 1994
 Albers Conic Equal-Area projection,
 Standard parallels 29°30' and 45°30', central meridian 96

Land-use data from Kansas Applied Remote Sensing Program (2009)

EXPLANATION

- ▲ U.S. Geological Survey stream-gaging station and identifier
07143672
- ▭ Watershed Boundary
- Land Use**
- Urban
- Cropland
- Grassland
- Woodland
- Water



Figure 1. Map showing land use in Little Arkansas River Basin and location of pertinent boundaries and sites.

The study area is underlain by the *Equus* Beds aquifer, the easternmost extent of the larger High Plains aquifer. The *Equus* Beds aquifer (as much as 300-feet thick) consists of alluvial deposits of sand and gravel interbedded with clay or silt and is an important source of groundwater because of the good water quality, shallow depth to the water table, and large saturated thickness (Williams and Lohman, 1949). The general direction of groundwater movement within the study area is to the east-northeast (Aucott and others, 1998) except where the hydraulic gradient is altered by pumping wells and near a low-head dam on the Little Arkansas River at Halstead. The Little Arkansas River generally is a gaining stream within the study area as indicated by higher water levels in wells adjacent to the stream compared to water levels in the stream (Myers and others, 1996; Aucott and others, 1998). The Little Arkansas River is not a gaining stream near Halstead where a low-head dam about 2 river miles downstream from the Near Halstead streamgaging station causes higher water levels in the stream than in the adjacent aquifer resulting in stream-water recharge to the aquifer in this area.

Methods

Regression models for real-time computation of concentrations of hardness, dissolved solids, total suspended solids, calcium, sodium, alkalinity, bicarbonate, chloride, sulfate, total organic nitrogen, total phosphorus, *Escherichia coli* bacteria (*E. coli*), fecal coliform bacteria (fecal coliform), arsenic, atrazine, and suspended-sediment were updated from Christensen and others (2000, 2003). Additionally, regression models for total organic carbon were developed in this study. Methods for development of models, quantifying uncertainty, and determining probability of exceeding criteria are given in Rasmussen and others (2009) and at <http://nrtwq.usgs.gov/ks/methods/>. Model calibration periods were from 1998 to 2014 for all constituents, except for models that contained turbidity (calibration period 2004–2014), and except for models that contained total suspended solids (model calibration period 2005–2014). The updated and new total organic carbon models are applied to data collected from January 1, 2007 onward (appendixes 1 and 2). Methods for the collection and quality assurance of continuous streamflow and continuous and discrete water-quality data are available in detail in the following references: Sauer and Turnipseed (2010), Turnipseed and Sauer (2010), Wagner and others (2006), Wilde and Radke (1998), Ziegler and Combs (1997), and Ziegler and others (2010).

Constituent concentrations in surface water commonly are strongly statistically and physically related to other constituent concentrations and factors like streamflow, turbidity, and specific conductance. Expressing one constituent concentration in terms of a physical property and other stream variables can be done using simple or multiple linear regression models (Helsel and Hirsch, 2002). In most cases in this report, the explanatory variables have to have a physical basis to be included in the regression model. General examples of explanatory variables having a physical basis include the specific conductance of water caused by dissolved cations and anions in water or suspended sediment causes water to be turbid. Others are less direct, like total phosphorus, total organic nitrogen, total organic carbon, and indicator bacteria are associated with and are part of suspended sediment and atrazine concentrations are greatest during high streamflow or periods of low specific conductance and during the spring after application of herbicides. Regression models provide the continuous and real-time computation of constituent concentrations that are not measured continuously using variables measured continuously, such as streamflow, water temperature, dissolved oxygen, specific conductance, pH, and turbidity. The method of model development using simple or multiple linear regression is described by Helsel and Hirsch (2002).

Additionally, Helsel and Hirsch (2002) describe the calculation and use of many of the regression statistics reported here or used in the selection of the preferred model, including root mean square error (*RMSE*) or standard error of the regression, coefficient of determination (R^2), adjusted

coefficient of determination ($adjR^2$), Mallow's C_p , prediction error sum of squares ($PRESS$), and variance inflation factor (VIF) (Marquardt, 1970). The calculation and use of model standard percentage error ($MSPE$) is described by Rasmussen and others (2009). The bias correction factor or smear factor (BCF) is described in Duan (1983). The regression statistics and metrics used in this study are summarized and discussed in the Model Archive Summaries (MAS) in appendixes 1 and 2.

Continuously measured in-stream properties were interpolated to match the same mean time of the discrete sample collection (Rasmussen and others, 2009). In the case that the continuous measurement of a physical property (water temperature, specific conductance (SC), or turbidity) corresponding to the same time and date of a discrete water-quality sample was unavailable or unable to be accurately computed from continuous data, the discrete water-quality sample was excluded from the regression. Individual samples with a studentized residual greater than 3.0 or less than -3.0 and a high Cook's D, as determined in Helsel and Hirsch (2002), were identified as outliers and removed.

Some constituents had sample results that were less than the laboratory reporting level or minimum reporting level (MRL). Such results are reported as "less-than" the MRL when the analyte is either not detected or detected at a concentration less than the level (Childress and others, 1999). Constituent datasets with nondetects (left-censored data) near Halstead station include 4 percent of total suspended sediment (TSS) samples, 1.4 percent of sulfate samples, 3 percent of arsenic samples, and 3.2 percent of atrazine samples. Constituent datasets with nondetects at near Sedgwick station include 3.7 percent of TSS samples, 2.5 percent of chloride samples, 2.2 percent of sulfate samples, and less than 1 percent of atrazine samples. The data initially were analyzed using tobit regression (Hald, 1949; Cohen, 1959). After the tobit regressions were completed, the nondetects arbitrarily were assigned a value of one-half of the censoring level and were analyzed in the same manner as the constituent datasets without nondetects (Helsel, 2005). The resulting regression coefficients were compared with the tobit regression coefficients. The comparison of the regression coefficients using an assigned value at one-half the detection limit and tobit were nearly identical because there were few nondetections for the models developed. In datasets with a substantial percentage of nondetections, a tobit regression model would be preferred.

The use of regression models has limitations. A regression model between the response and explanatory variables generally is site specific and potentially changes throughout time if the constituent sources change or improved sensor technology becomes available. For example, turbidity measurements are affected by the physical properties of suspended-sediment particles such as size, color, and density (Anderson, 2004). These suspended-sediment properties are affected by complex watershed properties such as stream morphologic characteristics, land-use distribution, and so forth. All of these complexities are simplified and combined into the regression model coefficients. It is unlikely that any two streams or even one stream at two locations would have identical suspended-sediment properties. Suspended-sediment properties at one location may change over a range of time; therefore, regression analysis is site specific, and the regression model must be verified by continued sample collection and refined as needed (Rasmussen and others, 2009).

Multiple regression models were developed and evaluated to find the one that was best for continuously computing each constituent of interest. The regression statistics $RMSE$, $PRESS$, and Mallow's C_p were used to narrow the list of possible models. In general, only the models with the lowest $RMSE$, $PRESS$, and Mallow's C_p (only comparable for models with the same response variable, that is, response variable TSS is not comparable with $\log[TSS]$) were given further consideration. Models containing multiple transformations of the same physical property (for example, streamflow [Q] and $\log_{10}[Q]$) were not considered. The preferred model was selected using the following process and criteria:

- 1) Response variable transformation was based on plots of residuals as compared to computed values (provided in appendixes 1 and 2); if the pattern of the residuals from a simple-linear regression (SLR) model was heteroscedastic (non-uniform residuals), then a log-transformed model was preferred.
- 2) If the p -value for an additional explanatory variable was less than 0.05, then it was retained in a multiple-linear regression (MLR) model; otherwise, the additional explanatory variable was rejected.
- 3) Additional explanatory variables were considered if the average of the upper and lower MSPE of a model improved greater than 10 percent.
- 4) If a MLR model is selected, the VIF must be less than 10 indicating minimal multicollinearity.

Log₁₀ transformation of the response variable has a consequence that must be considered when computing concentrations. Computed values must be retransformed back to the original units, a step that introduces a bias (typically negative) in computed concentration values (Miller, 1951; Koch and Smillie, 1986) unless the data are perfectly and positively correlated. The bias arises because regression predicts the mean of a normal distribution in log units, and the retransformed value results in a value closer to the median value than the mean in linear space. To correct for retransformation bias, the retransformed values must be multiplied by a bias correction factor (BCF; equation 1).

Duan (1983) introduced a nonparametric bias-correction factor called the "smearing" estimator (Helsel and Hirsch, 2002, p. 256). The smearing factor is calculated from the mean of the residual values (equation 1; Duan, 1983)

$$BCF = \sum_{i=1}^n \frac{10^{\log_{10}(y_i) - \log_{10}(\hat{y}_i)}}{n} \quad (1)$$

where

BCF is the bias correction factor,

y_i is the i th measured suspended-sediment concentration,

\hat{y}_i is the i th regression-computed suspended-sediment concentration, and

n is the number of measured suspended-sediment concentrations in the model-calibration dataset.

All regression models have uncertainty inherently associated with each computation. Uncertainty can be defined in a number of different ways, including relative percent difference, absolute error, and prediction intervals. Prediction intervals, used on the National Real Time Water Quality (NRTWQ; <http://nrtwq.usgs.gov>) Web page, define a range of values for the response variable for a given level of certainty. The level of certainty presented with the modeled data is the 90-percent prediction interval. The larger the range, the more uncertainty associated with the regression computed value.

Calculating prediction intervals for regression models with two or more explanatory variables involves matrices (equation 2; Draper and Smith, 1998).

$$\hat{y} \pm \left[t_{\alpha/2, n-p} \sqrt{s^2(1 + x'(X'X)^{-1}x)} \right] \quad (2)$$

where

t is the value of Student's t-distribution having $n-p$ degrees of freedom with an exceedance probability of $\alpha/2$,

α is the level of certainty for the prediction interval (1-0.9, or 0.1 for 90-percent prediction interval),

p is the number of explanatory variables plus one,

s is the variance of the residuals calculated during model development,

$(X'X)^{-1}$ is the X prime X inverse matrix calculated during model development—an expression of the covariance among all explanatory variables,

x' is the mean of the independent variable measurements, and

x is the independent variable measurement.

The models presented in this report are applied to continuous data and computations starting on January 1, 2007. Data will continue to be collected and evaluated annually. At least six samples per year will be collected over the range of hydrologic and sensor conditions. A fundamental characteristic of hydrology is variability, with periods of floods and periods of droughts. Additionally, watershed conditions can change seasonally or from other factors such as through changes in land use, implementation of best management practices, or by wildfire. Therefore, regression models at a site should never be considered static, but rather considered to represent a set period in a continually dynamic system in which additional data will help verify any change in constituent concentration and source. Models that are updated will be used to compute concentrations after a thorough independent technical review and approval process (described in detail in Rasmussen and others, 2009) and updated annually, usually applying the new model at the beginning of a calendar year. All models and computations that are displayed also are archived on <http://nrtwq.usgs.gov>. Previously computed and archived values will not be changed as new models are developed.

Results of Regression Analysis for Selected Constituents

The relation between in-stream continuous measurements and constituents of concern for artificial recharge was determined at the near Halstead and near Sedgwick streamgaging stations. For every constituent, several regression models with various explanatory variables were developed and evaluated using *RMSE*, adjusted R^2 , Mallows' C_p , *PRESS*, *VIF*, and *MSPE* (see the "Models Considered" sections of appendixes 1 and 2). A single model was selected for each constituent. Additionally, plots of residuals were compared to computed values and normal probability plots are provided for each selected model in appendix 1 and appendix 2.

In general, models shown in appendixes 1 and 2 were updated from the previous models (Christensen and others, 2000; Christensen and others, 2003) to include results from newly collected

samples. The primary reason for updating models is a better descriptive sampling range as a result of additional data. This manifests as a change in the model coefficients and in some cases, model form. A change in model form (hardness, calcium, sodium, bicarbonate) typically was a move to \log_{10} transformation or an additional explanatory variable that improved the relation between constituents and their continuous surrogates. No significance tests between the new and old models were done.

Other model modifications are a result of changing sensors used to measure explanatory variables. The most common model modification occurred when turbidity measured with the YSI 6026 sensor that was used in the previous models as an explanatory variable was changed to turbidity from the YSI 6136 sensor beginning in 2004; therefore, all updated and total organic carbon models with turbidity used only readings from the 6136 sensor for calibration. The following is a discussion of each constituent and the associated regression models for the near Halstead and near Sedgwick streamgaging stations. Models evaluated and selected are shown in appendixes 1 and 2.

For models with censored data, regression coefficients were compared with the tobit regression coefficients. None of the coefficients differences were greater than 10 percent, suggesting that the regression models developed using the assigned values of one-half of the censoring level were valid. All computations were completed using R, 3.0.1 for Windows (R Core Team, 2013).

Hardness

Because hardness is closely related to dissolved calcium and magnesium, which are both charged cations, specific conductance is a statistically significant surrogate for hardness. Hardness models were evaluated for the near Halstead and near Sedgwick streamgaging stations (appendixes 1 and 2). Untransformed models generally resulted in a lower MSPE than log-transformed models, but also produce heteroscedastic residual plots; therefore, only log-transformed models, which produced homoscedastic residual plots, were considered further. The upper and lower *MSPEs* of the log-transformed SLR are below 20 percent, so the SLR model is preferred. Both models use SC as an explanatory variable; the positive value of the $\log(\text{SC})$ coefficient indicates that specific conductance increases as hardness increases, which is intuitive.

Dissolved Solids

Dissolved solids in surface water are highly correlated with specific conductance (Hem, 1992). Dissolved solids models at near Halstead and near Sedgwick streamgaging stations have an untransformed response variable and specific conductance as an explanatory variable. The residual plots for the untransformed models show a homoscedastic pattern (appendixes 1 and 2), so no transformation of the response variable is necessary. The MSPE is below 20 percent, so no additional explanatory variables are necessary. The positive value of the SC coefficient indicates that specific conductance increases as dissolved solids concentrations increase, which is intuitive.

Total Suspended Solids

Turbidity is an indicator of sediment and other solid material transported in a stream and, thus, a physical property related to total suspended solids. Near Halstead and near Sedgwick streamgaging station total suspended solids models were considered with log-transformed YSI model 6136 turbidity as an explanatory variable. Residual plots for the log-transformed models are homoscedastic (appendixes 1 and 2). The preferred model for both streamgages has a log-transformed TSS and an explanatory variable of log-transformed YSI model 6136 turbidity.

Calcium

Calcium is a charged ionic species, and therefore, should be related to specific conductance. All calcium models considered, near Halstead and near Sedgwick streamgaging stations, included specific conductance as an explanatory variable. Untransformed models were considered, as previous calcium models for these sites were untransformed. The results produced heteroscedastic residual plots, indicating that a log-transform of the response variable was preferred. The positive value of the $\log(\text{SC})$ coefficient indicates that specific conductance increases as dissolved calcium concentrations increase, which is intuitive.

Sodium

Sodium is also a charged ionic species and, therefore, related to specific conductance. All regression models considered for computing continuous sodium concentrations near Halstead and near Sedgwick streamgaging stations contain a form with specific conductance as an explanatory variable. Previous sodium models used untransformed sodium and SC. Untransformed models with this dataset generally resulted in heteroscedastic residual plots (appendixes 1 and 2); therefore, only log-transformed models were considered further. Specific conductance produced the best model, as additional explanatory variables did not improve the regression statistics. The positive value of the $\log(\text{SC})$ coefficient indicates that specific conductance increases as dissolved sodium concentrations increase, which is intuitive.

Alkalinity

Hem (1992) suggests that a relation between alkalinity and specific conductance can be established. Models considered for the computation of continuous alkalinity include a form with specific conductance as an explanatory variable. Untransformed models resulted in heteroscedastic residual plots (appendixes 1 and 2), indicating that a log-transform of the response variable is preferred. The average of the upper and lower MSPE for the SLR model with the log-transform of specific conductance is more than 20 percent, so an additional explanatory variable (log-transform of streamflow) was considered for the model. The p -value of $\log(Q)$ is less than 0.05, so the variable was retained. The preferred alkalinity model, therefore, contains a log-transformed response variable and explanatory variables of log-transformed specific conductance and log-transformed streamflow. Log transformation is consistent with the previous models used at near Halstead and near Sedgwick streamgaging stations. The positive value of the $\log(\text{SC})$ coefficient makes intuitive sense, and the negative value of the $\log(Q)$ may indicate that higher flows tend to dilute constituents that affect alkalinity.

Bicarbonate

Like alkalinity, models considered for the computation of continuous bicarbonate include a form with specific conductance as an explanatory variable. Untransformed models resulted in heteroscedastic residual plots (appendixes 1 and 2), indicating that a log-transform of the response variable is preferred. The average of the upper and lower MSPE for the SLR model with the log-transform of specific conductance is more than 20 percent, so an additional explanatory variable (log-transform of streamflow) was considered for the model. The p -value of $\log(Q)$ is less than 0.05, so the variable was retained in the model. The preferred bicarbonate model, therefore, contains a log-transformed response variable and explanatory variables of log-transformed specific conductance and log-transformed streamflow. This is a departure from the previous models at these two sites. At the near Halstead streamgaging station, log-transformed bicarbonate previously was expressed as a function of linear SC

and log-transformed streamflow. At the near Sedgwick streamgaging station, log-transformed bicarbonate was expressed as a function of log-transformed SC alone. The positive value of the log(SC) coefficient makes intuitive sense, and the negative value of the log(Q) may indicate that higher flows tend to dilute bicarbonate concentrations.

Chloride

With increasing chloride concentration, the specific conductance value of water will also increase. However, larger specific conductance and chloride concentrations generally correspond with low-flow conditions, indicating that streamflow may be a significant explanatory variable. Chloride concentrations are likely smaller at the downstream Sedgwick streamgaging station because of dilution from groundwater inflow containing smaller chloride concentrations (Ziegler and others, 2010). Chloride models considered for near Halstead and near Sedgwick streamgaging stations included a form with specific conductance as an explanatory variable. The untransformed models generally have higher MSPE values and heteroscedastic residual plots (appendixes 1 and 2), so log transformation of the response variable is necessary. The average of the upper and lower MSPE for the SLR model with the log-transform of specific conductance is above 20 percent, so an additional explanatory variable (log-transform of streamflow) was considered for the model. The *p*-value of log(Q) is less than 0.05, so the variable was retained in the Sedgwick model. Adding log(Q) as an explanatory variable for the near Halstead model did not reduce MSPE, so it was not retained in that model. The preferred chloride model contains a log-transformed response variable and, as explanatory variables, log-transformed specific conductance and, at the Sedgwick streamgaging station, log-transformed streamflow. This is a departure from the previous models at these two sites. Previously, the response variable, untransformed chloride was expressed with the explanatory variable, untransformed SC at the near Halstead streamgaging station and log-transformed chloride was expressed with log-transformed streamflow and untransformed SC at the Sedgwick streamgaging station. The positive value of the log(SC) coefficient makes intuitive sense.

Sulfate

Sulfate is a charged ionic species and directly related to specific conductance; an increase in sulfate concentration results in a corresponding increase in specific conductance in surface water. Sulfate models were considered for the Sedgwick and near Halstead streamgaging stations. Based on plots of residuals, log-transformed sulfate was chosen instead of untransformed sulfate as the response variable, because it produced a homoscedastic pattern. Because average MSPE for log-transformed sulfate models at both sites was more than 20 percent, log-transformed streamflow was considered as an additional explanatory variable. In the Sedgwick model, log-transformed streamflow was retained as an explanatory variable, but in the near Halstead model, the addition of log(Q) did not reduce MSPE and it was not retained. The positive value of the log(SC) coefficient makes intuitive sense.

Total Organic Nitrogen

Total organic nitrogen is likely associated with suspended sediment and runoff containing nitrogenous organic matter (Juracek and Rasmussen, 2008). Organic material is one of the major components of total suspended solids; thus, turbidity is a suitable surrogate for total organic nitrogen. All total organic nitrogen models considered include a form with YSI model 6136 turbidity as an explanatory variable. The untransformed models generally have higher MSPE values and heteroscedastic residual plots (appendixes 1 and 2), so log transformation of the response variable is

necessary. The largest concentrations of total organic nitrogen are in the spring and summer, presumably when it is applied to crops and when rainfall is most abundant (Kansas Center for Agriculture Resources and the Environment and others, 2011). Total organic nitrogen concentrations, therefore, are highly seasonal. To account for seasonality in the regression models, day of the year (a number between 1 and 365) is included as a variable using the multiple regression with periodic functions method described by Helsel and Hirsch (1992). For the near Halstead and near Sedgwick streamgaging stations, turbidity and day of the year are the explanatory variables used to compute total organic nitrogen concentrations. Time of year shows a strong physical correlation with total organic nitrogen concentrations. The older total organic nitrogen models did not include day of the year as an explanatory variable, likely because there were not enough samples to define it as a significant variable. The positive value of the $\log(\text{Turb}_{6136})$ coefficient makes intuitive sense.

Total Phosphorus

Because phosphorus is likely to sorb to suspended sediment, turbidity is a good surrogate for the computation of continuous total phosphorus concentrations. All total phosphorus models considered include a form with YSI model 6136 turbidity as an explanatory variable. The untransformed models generally have somewhat heteroscedastic residual plots (appendixes 1 and 2), so log transformation of the response variable is necessary. The MSPE of the SLR model is less than 20 percent, so no additional explanatory variables are necessary. This is the first model to use log-transformed variables to compute total phosphorus at these two sites. The positive value of the $\log(\text{TURB}_{6136})$ coefficient makes intuitive sense.

***Escherichia Coli* Indicator Bacteria**

Turbidity can be an important surrogate for the computation of *E. coli* density, because bacteria may sorb to suspended particles. Based on the comparison of Mallows' *C_p* and *PRESS* values for all models output (appendix 1), the preferred model contains a log-transformed response variable and, as an explanatory variable, log-transformed turbidity. The residual plot shows a homoscedastic pattern. The positive value of the $\log(\text{TURB}_{6136})$ coefficient makes intuitive sense.

Fecal Coliform Indicator Bacteria

Because *E. coli* is a specific species of fecal coliform, it is plausible that the regressor for computing continuous fecal coliform densities is similar to the model for computing continuous *E. coli* densities. Residual plots for the log-transformed models are homoscedastic (appendixes 1 and 2). The calculated MSPE is much greater than 20 percent, so an additional explanatory variable (log-transformed specific conductance) was considered for the model. The *p*-value of the $\log(\text{SC})$ term in the MLR model is greater than 0.05, so the term was rejected. The preferred model, therefore, contains a log-transformed response variable and, as an explanatory variable, log-transformed turbidity. The positive value of the $\log(\text{TURB}_{6136})$ coefficient makes intuitive sense.

Arsenic

Larger dissolved arsenic concentrations in stream water generally occurred at both sites during low streamflow when base flow to the stream was supplied from groundwater, indicating that the arsenic source is the groundwater inflow. In other words, the arsenic concentration is expected to decrease as streamflow increases. Various arsenic models were considered, all indicating that log-transformed streamflow is the strongest explanatory variable. The untransformed models have

heteroscedastic residual plots (appendixes 1 and 2), so log transformation of the response variable is required for the model. The calculated MSPE for the SLR model is much greater than 20 percent, so an additional explanatory variable (water temperature) was considered for the model. The p -value of the water temperature term was less than 0.05, so the term was retained. The preferred model, therefore, has a log-transformed response variable and explanatory variables of log-transformed streamflow and water temperature. Water temperature may just be a surrogate for the seasonal fluctuation in arsenic concentrations. This is a departure from the previous arsenic models when water temperature was not an explanatory variable.

Atrazine

In a study by Ziegler and others (1999), atrazine was determined to have an inverse relation with chloride, and therefore an inverse relation with specific conductance as well. Like total organic nitrogen, Christensen and Ziegler (1998) observed the largest concentrations of atrazine in the spring and summer, presumably when atrazine is applied to crops and when rainfall is most abundant. Atrazine concentrations, therefore, are also highly seasonal. To account for this in the regression models, day of the year is included as a variable using the multiple regression with periodic functions method described by Helsel and Hirsch (2002). For the near Halstead and near Sedgwick streamgaging stations, specific conductance and day of the year are the explanatory variables used to compute atrazine concentrations. Time of year shows a strong physical correlation with atrazine concentrations. The largest observed concentrations are in the spring and summer, when it is applied to crops and rainfall is more abundant. The differences from the older atrazine models likely are caused by changes in application practices of atrazine that may have caused a change in the timing and distribution within the watersheds, and relations of these variables (Kansas Center for Agriculture Resources and the Environment and others, 2011). Unfortunately, because of the large uncertainties (MSPEs greater than 200) in these models, changes in application practices cannot be easily quantified.

Total Organic Carbon

Organic matter is one of the major components of total suspended solids (Juracek and Rasmussen, 2008; Hem, 1992); turbidity may be a suitable surrogate for total organic carbon. All total organic carbon models considered include YSI model 6136 turbidity as an explanatory variable. The positive value of the $\log(\text{TURB6136})$ coefficient makes intuitive sense. The untransformed models generally have higher MSPE values and heteroscedastic residual plots (appendixes 1 and 2), so log transformation of the response variable is necessary. Both models for computing total organic carbon concentrations use turbidity only.

Suspended-Sediment Concentration

Like total suspended solids, turbidity is likely a suitable surrogate for the computation of continuous suspended-sediment concentration. Suspended-sediment models considered include YSI model 6136 turbidity as an explanatory variable. The two untransformed models have heteroscedastic residual plots (appendixes 1 and 2), so log transformation of the response variable is necessary. The average of the upper and lower MSPE for the SLR model with the log-transform of specific conductance is more than 20 percent, so an additional explanatory variable (log-transformed streamflow) was considered for the model. The p -value of $\log(Q)$ is greater than 0.05, so the variable was retained in the model. The preferred suspended-sediment model contains a log-transformed

response variable and, as explanatory variable, log-transformed turbidity and streamflow. The positive value of the log(TURB6136) coefficient makes intuitive sense.

Summary

Water from the Little Arkansas River is used as source water for artificial recharge of the *Equus* Beds aquifer. The U.S. Geological Survey has operated two continuous real-time water-quality monitoring stations since 1995 on the Little Arkansas River in Kansas. Continuously measured water-quality physical properties include streamflow, specific conductance, pH, water temperature, dissolved oxygen, and turbidity. Discrete water-quality samples were collected during 1995 through 2014 and analyzed for sediment, nutrients, bacteria, atrazine, and other water-quality constituents. Regression models were developed to establish relations between discretely sampled constituent concentrations and continuously measured physical properties to compute concentrations of selected constituents of interest.

Regression models at USGS streamgaging stations on the Little Arkansas River at Highway 50 near Halstead (07143672) and the Little Arkansas River near Sedgwick (07144100) were originally published in 2000 and were updated in 2003. This report updates those models using discrete and continuous data collected during May 1998 through August 2014. This report presents the regression models. The real-time computations of water-quality concentrations and loads are available at <http://nrtwq.usgs.gov>. The water-quality information in this report is important to the city of Wichita because water-quality information allows for real-time quantification and characterization of chemicals of concern (including chloride), in addition to nutrients, sediment, bacteria, and atrazine transported in the Little Arkansas River. The water-quality information aids in the decision making for water treatment before artificial recharge.

References Cited

- Anderson, C.W., 2004, Turbidity (version 2.0): U.S. Geological Survey Techniques of Water Resources Investigations, book 9, chap. A6, section 6.7, 64 p., accessed July 1, 2009, at http://water.usgs.gov/owq/FieldManual/Chapter6/6.7_contents.html.
- Aucott, W.R., Myers, N.C., and Dague, B.J., 1998, Status of groundwater levels and storage in the Wichita well field area, south-central Kansas, 1997: U.S. Geological Survey Water-Resources Investigations Report 98-4095, 15 p.
- Childress, C.J.O., Foreman, W.T., Connor, B.F., and Maloney, T.J., 1999, New report procedures based on long-term method detection levels and some considerations for interpretations of water-quality data provided by the U.S. Geological Survey National Water Quality Laboratory: U.S. Geological Survey Open-File Report 99-193, 19 p.
- Christensen, V.G., and Ziegler, A.C., 1998, Atrazine in source water intended for artificial ground-water recharge, south-central Kansas: U.S. Geological Survey Fact Sheet 074-98, 4 p.
- Christensen, V.G., Jian Xiaodong, and Ziegler, A.C., 2000, Regression analysis and real-time water-quality monitoring to compute 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.
- 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, Missouri: Middleburg,

- Virginia, American Water Resources Association Technical Publication Series No. TPS-03-1, compact disc.
- Cohen, A. C., 1959, Simplified estimators for the normal distribution when samples are singly censored or truncated: *Technometrics* 1, 217–213.
- Draper, N.R., and Smith, H., 1998, *Applied regression analysis* (3d ed.): New York, John Wiley and Sons, Ltd., 706 p.
- Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: *Journal of the American Statistical Association*, v. 78, p. 605–610.
- Hald, A., 1949, Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point: *Skandinavisk Aktuarietidskrift*, v. 32, p. 119–34.
- Helsel, D.R., 2005, *Nondetects and Data Analysis: Statistics for Censored Environmental Data*: John Wiley and Sons, New York, 268 p.
- Helsel, D.R., and Hirsch, R.B., 1992, *Statistical methods in water resources*: Amsterdam, Elsevier, 522 p.
- Hem, J.D., 1992, *Study and interpretation of the chemical characteristics of natural water* (3d. ed): U.S. Geological Survey Water-Supply Paper 2254, 264 p.
- Juracek, K.E., and Rasmussen, P.P., 2008, Sediment quality and comparison to historical water quality, Little Arkansas River Basin, south-central Kansas, 2007: U.S. Geological Survey Scientific Investigations Report 2008–5187, 47 p.
- Kansas Center for Agriculture Resources and the Environment, Kansas State University Research and Extension, and Kansas department of Health and Environment, 2011, Little Arkansas River Watershed Restoration and Protection Strategy, information available on the Web, accessed October 20, 2013, at http://www.kswraps.org/files/attachments/littlearkansas_plansummary.pdf.
- Kansas Department of Agriculture, 2006, *Kansas farm facts 2006*: Topeka, Kansas, 103 p.
- Koch, R.W., and Smillie, G.M., 1986, Bias in hydrologic prediction using log-transformed regression models: *American Water Resources Association, Water Resources Bulletin*, v. 22, no. 5, p. 717–732.
- Marquardt, D.W., 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation: *Technometrics*, v. 12, p. 591–612.
- Miller, C.R., 1951, Analysis of flow-duration sediment rating curve method of computing sediment yield: U.S. Bureau of Reclamation, 55 p.
- Myers, N.C., Haragadine, G.D., and Gillespie, J.B., 1996, Hydrologic and chemical interactions of the Arkansas River and the *Equus* Beds aquifer between Hutchinson and Wichita, south-central Kansas: U.S. Geological Survey Water-Resources Investigations Report 95–4191, 100 p.
- Peterson, D.L., Whistler, J.L., Lomas, J.M., Dobbs, K.E., Jakubauskas, M.E., Egbert, S.E., and Martinko, E.A., 2005, 2005 Kansas Land Cover Patterns Phase I: Final Report: KBS Report 150, Kansas Biological Survey, University of Kansas, 29 p.
- Putnam, J.E., Lacock, D.L., Schneider, D.R., and Carlson, M.D., 2000, Water resources data, Kansas, water year 1999: U.S. Geological Survey Water-Data Report KS-99–1, p. 288 and 302.
- R Core Team, 2013, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, information available on the Web, accessed July 11, 2013, at <http://www.R-project.org>
- 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, 53 p.
- TIBCO Software, Inc., 2008, *TIBCO Spotfire S+® 8.1 for Windows® User’s Guide*, 582 p.

- 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.
- Stone, M.L., Rasmussen, T.J., Bennett, T.J., Poulton, B.C., and Ziegler, A.C., 2012, Protocols for collection of streamflow, water-quality, streambed-sediment, periphyton, macroinvertebrate, fish, and habitat data to describe stream quality for the Hydrobiological Monitoring Program, Equus Beds Aquifer Storage and Recovery Program, city of Wichita, Kansas: U.S. Geological Survey Open-File Report 2012-1055, 55 p.
- 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.
- 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.
- Wilde, F.D., and Radke, D.B., eds., 1998, Field measurements, in national field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A6, p. 3-20.
- Williams, C.C., and Lohman, S.W., 1949, Geology and groundwater resources of a part of south-central Kansas with a special reference to the Wichita municipal water supply: Kansas Geological Survey Bulletin 79, 455 p.
- Ziegler, A.C., and Combs, L.J., 1997, Baseline data-collection and quality control protocols and procedures for the *Equus* Beds Groundwater Recharge Demonstration Project near Wichita, Kansas, 1995-96: U.S. Geological Survey Open-File Report 97-235, 57 p.
- Ziegler, A.C., Christensen, V.G., and Ross, H.C., 1999, Baseline water-quality and preliminary effects of artificial recharge on groundwater, south-central Kansas, 1995-98: U.S. Geological Survey Water-Resources Investigations Report 99-4250, 74 p.
- Ziegler, A.C., Hansen, C.V., and Finn, D.A., 2010, Water quality in the *Equus* Beds aquifer and the Little Arkansas River before implementation of large-scale artificial recharge, south-central Kansas, 1995-2005: U.S. Geological Survey Scientific Investigations Report 2010-5023, 143 p.

Appendixes

Appendixes 1–2 contain the model archive summaries for the linear regression models. The previous models indicated in the summaries are documented in Christensen and others (2000) and Christensen and others (2003). See the National Real Time Water Quality (NRTWQ; <http://nrtwq.usgs.gov>) Web page for an explanation of graphs in the model archive summaries. Explanations are provided with each figure in the appendixes. For the “Normal Quantile” plot, the green line is a “theoretical” normal distribution of probabilities for the given residuals.

Appendix 1. Model Outputs for Little Arkansas River at Highway 50 near Halstead, KS (site 07143672).

Appendix 2. Model Outputs for Little Arkansas River near Sedgwick, KS (site 07144100).