Methods for Computing Water-Quality Concentrations and Loads at Sites Operated by the U.S. Geological Survey Kansas Water Science Center

Open-File Report 2024–1049

Methods for Computing Water-Quality Concentrations and Loads at Sites Operated by the U.S. Geological Survey Kansas Water Science Center

By Mandy L. Stone, Casey J. Lee, Teresa J. Rasmussen, Thomas J. Williams, Ariele R. Kramer, and Brian J. Klager

Open-File Report 2024–1049

U.S. Geological Survey, Reston, Virginia: 2024

Contents

Abbreviations

AMLE   absolute maximum likelihood estimation

NRTWQ   National Real-Time Water Quality Data for the Nation Data Service

OLS   ordinary least squares

R^2 coefficient of determination

USGS   U.S. Geological Survey

WRTDS–K Weighted Regressions on Time, Discharge, and Season with Kalman filtering

Methods for Computing Water-Quality Concentrations and Loads at Sites Operated by the U.S. Geological Survey Kansas Water Science Center

By Mandy L. Stone, Casey J. Lee, Teresa J. Rasmussen, Thomas J. Williams, Ariele R. Kramer, and Brian J. Klager

Abstract

The U.S. Geological Survey (USGS) Kansas Water Science Center (KSWSC) has published time-series computations of water-quality concentrations and loads based on in situ sensor data since 1995. Water-quality constituent concentrations or densities are computed using regression models that relate in situ sensor values to laboratory analyses of periodically collected samples. These regression models currently (2024) follow no uniform published guidance and are individually documented through USGS reports. This report describes updated (2024) procedures designed to improve the consistency, quality, and timeliness of computed continuous water-quality data produced by the USGS KSWSC. Beginning in 2024, models developed by the USGS KSWSC that follow specific procedures and requirements related to sample collection, model fit, and model documentation outlined in this report are planned to be published and stored in the USGS National Real-Time Water Quality Data for the Nation Data Service. This report also describes USGS KSWSC procedures for evaluating and publishing time-series water-quality computations after initial model development and documentation. This guidance can be used to improve USGS KSWSC model development and data computation consistency and streamline the time-series water-quality computation process from model development to publication.

Introduction

Knowledge of water-quality concentrations and loads is needed to assess the health of streams, reservoirs, and receiving waters and to quantify the relative importance of upstream drainage basin contributions. Since 1995, the U.S. Geological Survey (USGS) Kansas Water Science Center (KSWSC) has been publishing time-series water-quality concentration and load computations using in situ sensor measurements, periodically collected discrete samples, and continuously computed discharge. Although sensors used and data produced by individual studies vary according to the requirements of cooperating agencies and USGS scientific priorities, a need exists to establish common practices within the USGS KSWSC to facilitate consistent project design, ensure data comparability, and develop and maintain mechanisms for data delivery.

The USGS KSWSC first published time-series water-quality concentrations and loads at sites in the Quivira National Wildlife Refuge from 1998 to 2001 (Christensen, 2001) and has published hundreds of models for this purpose through 2024 (Christensen and others, 2003, 2006; Mau and others, 2004; Rasmussen and others, 2005, 2008, 2016; Lee and others, 2008; Graham and others, 2010; Juracek, 2011, 2013; Lee and Foster, 2013; Stone and others, 2013a, b; Foster, 2014; Stone and Graham, 2014; Foster and Graham, 2016; Kramer and others, 2021a, b; Williams, 2021, 2023; Leiker, 2022; Stone and Klager, 2022, 2023; Kramer and Puls, 2023). Models have been developed to provide real-time, continuous computations of alkalinity, major ions, dissolved and solid-phase nutrient species, suspended-sediment and total suspended solids, pesticides, chlorophyll, fecal indicator bacteria, and compounds commonly related to harmful algal blooms, among others. Continuous computations of water-quality concentrations and loads are currently (2024) published on the USGS National Real-Time Water Quality Data for the Nation Data Service (NRTWQ) (https://nrtwq.usgs.gov).

Purpose and Scope

The purpose of this report is to document and describe the USGS KSWSC standard procedures for developing, publishing, maintaining, and updating continuous water-quality constituent concentration and load computations using models that relate in situ water-quality sensor values and concomitant discrete water-quality sample laboratory analysis values. These standard procedures are designed to improve the consistency and timeliness of computed continuous water-quality data published by the USGS KSWSC while also producing statistically valid models. Procedures described in this report are intended to improve product development and review

processes, enable consistent computation methods among studies, facilitate uniform communication among the USGS KSWSC and cooperating agencies, document procedures for data publication after models have been established, and streamline the time-series water-quality computation process from model development to publication. Procedures described herein were developed based on historical conditions observed at Kansas sites and thus apply only to models developed by the USGS KSWSC.

Procedures for Publishing Continuous Water-Quality Data in the U.S. Geological Survey Kansas Water Science Center

This report documents (2024) USGS KSWSC procedures for computing time-series water-quality concentrations and loads. The USGS has developed models for several water-quality constituents (including alkalinity; ions; metals; nitrogen, phosphorus, and carbon species; pesticides; indicator bacteria; and sediment) and waterbodies (including streams, groundwater, and reservoirs). Updated procedures are designed to streamline the review process, facilitate clear communication among the USGS KSWSC and cooperating agencies, and document procedures for data publication after models have been established. Models that follow procedures described herein can be published as USGS data releases on NRTWQ. It is important to note that because of the wide range of complexity in relations among sensor data and water-quality constituents of interest, criteria defined in this report such as minimum sample counts, adequate representation of hydrologic conditions, measures of model fit, and so on represent minimum thresholds only. Additional sampling, analysis, or other requirements may be needed for specific projects.

The overarching objective of this report is to describe procedures for developing models among discretely sampled water-quality constituent concentrations and in situ water-quality sensor values that can be documented as USGS data releases. These "primary" models can be used to estimate water-quality concentrations and (or) loads from subdaily to annual time steps on NRTWQ. However, because in situ sensor measurements are not always available, due to instrument fouling, debris, or ice, this report also describes procedures for developing streamflow-based models designed to fill in sensor gaps for the purpose of estimating concentrations and (or) loads over monthly, seasonal, or annual time steps.

Primary Model Development

The USGS KSWSC uses regression-based methods to relate discretely sampled water-quality constituent concentrations to concomitant time-series data including, but not limited to, streamflow, water temperature, specific conductance, pH, dissolved oxygen, turbidity, time (including seasonality using periodic functions), and various iterations of fluorescence sensor measurements. The type of model used to relate constituent concentrations and in situ measurements is dictated by the presence or absence of censored data in the model dataset. If the discrete or time-series sample dataset does not have censored values, ordinary least squares (OLS) regression is used to model relations between time-series and discrete sample data. If one or more censored values are present, Tobit regression methods are used for fitting linear models using absolute maximum likelihood estimation (AMLE; Hald, 1949; Cohen, 1950; Tobin, 1958; Helsel and others, 2020). Although discrete sample results are typically related to in situ sensor values, they also may be related to field sensor readings during periods of in situ sensor malfunction, ice conditions, and so on.

Discrete Sample Collection Requirements

The number of samples needed to compute water-quality concentrations or loads varies by study objective, the constituent(s) of interest, site and hydrologic conditions, the waterbody being measured, sensor characteristics, and the time step being computed, among other factors (Lee and others, 2016, 2019). Although models cannot be evaluated until after samples have been collected, establishing minimum sampling criteria increases the potential for developing a useful model and facilitates consistency among studies. Within the USGS KSWSC, the number of samples used to establish models for new sites and constituents has varied (Christensen and others, 2003, 2006; Mau and others, 2004; Rasmussen and others, 2005, 2008, 2016; Juracek, 2011; Lee and Foster, 2013; Stone and others, 2013a, b; Stone and Graham, 2014; Foster and Graham, 2016; Kramer and others, 2021b; Williams, 2021, 2023; Leiker, 2022; Kramer and Puls, 2023; Stone and Klager, 2023). Among these studies, datasets with fewer than 10 and as many as 261 samples have been used to establish models for publishing continuous water-quality computations; among all sites and constituents, the median was 21 samples. The median number of samples used to establish models was similar among commonly computed constituents including suspended sediment (22), chloride (26), total dissolved solids (29), total phosphorus (30), and total nitrogen (28). Although the amount of sampling is an important component to developing an accurate model, it is important to note that the total number of samples used to compute a constituent of interest is often not as critical as the streamflow or water-quality conditions on the day of the year when those samples are collected (Rasmussen and others, 2009; Lee and others, 2019). Regression models with as few as 15 samples distributed across the range of hydrologic conditions (or other potential explanatory variables) have been described to provide more representative models than those with 50 samples collected during relatively similar hydrologic conditions (Rasmussen and others, 2009).

Targeted Models

For purposes of this guidance, "targeted" models are among the most frequently developed models in the USGS KSWSC and are cases in which, in the absence of other inter-ferences, explanatory variables are known to directly respond to changes in the water-quality constituent of interest. Targeted relations include turbidity, suspended sediment, and (or) total suspended solids where, although factors such as sediment grain size and color affect the magnitude of light detected from the original source (Rasmussen and others, 2009), changes in the number of particles of a given size or color cause a corresponding change in turbidity. Other targeted relations include using specific conductance to compute total dissolved solids and major ions such as chloride, bromide, sodium, and calcium (Hem, 1992). In Kansas, these targeted relations typi-cally have the best metrics indicative of model fit among stud-ies that considered a broad array of water-quality constituents (Christensen, 2001; Christensen and others, 2006; Rasmussen and others, 2008; Stone and others, 2013b; Stone and Graham, 2014; Foster and Graham, 2016; Kramer and others, 2021a; Williams, 2021; Leiker, 2022; Kramer and Puls, 2023; Stone and Klager, 2023).

To build upon previous practices and ensure represen-tative model development, the USGS KSWSC requires at least 24 samples for simple linear regression and 36 samples at each site for multiple linear regression over 2–5 years for the development of targeted relations. Sample values should cover at least 80 percent of the range of explanatory variable measurements used to compute a given constituent. These sampling requirements represent minimum thresholds only for publication as a data release; additional samples may be required for a given site or constituent if a model does not meet project objectives, samples are not collected across the range of explanatory variables, and (or) the model does not conform to regression model assumptions. Samples should not be collected within 1 week of each other to minimize autocor-relation among residuals.

Exploratory and Secondary Models

In contrast to targeted relations, models are often devel-oped using in situ sensors, streamflow, time, or other factors that are not known to affect a constituent of interest across potential sampling sites. The use of streamflow and time to compute seasonally applied pesticides is an example of an exploratory model. In this example, stream or river pesticide concentrations do not vary directly because of changing sea-sons or flow conditions but by factors that are not (or cannot yet be) directly measured and are specific to the study loca-tion, such as upstream application timing, rates, and rainfall/runoff conditions, as well as specific compound sorption and transformation characteristics. Other examples of exploratory relations include using turbidity to compute indicator bacteria or of the use of most sensors, streamflow, and season to com-pute dissolved nutrients, chlorophyll, and properties related to harmful algal blooms.

In addition to exploratory models, sometimes "second-ary" models need to be developed when in situ sensor mea-surements are missing because of instrument malfunction and (or) environmental conditions that result in fouling, debris, or ice. Secondary models use streamflow, time, or other avail-able time-series data to compute constituent concentrations or loads during periods when in situ sensor measurements are not available; these data are required to compute water-quality concentrations and (or) loads over monthly, seasonal, or annual time steps.

Exploratory and secondary models are less likely to produce accurate estimates than those from targeted models and estimates from these models are likely to decrease in accu-racy from longer to shorter timescales (Lee and others, 2016, 2019). Therefore, if these models are used to compute con-centrations or loads at subdaily to daily time steps, the USGS KSWSC requires 36 samples for simple linear regression and 48 samples for multiple linear regression for publication as a data release. Samples must be collected over a 3–5-year time span and cover at least 80 percent of the range of explana-tory variables. More stringent sampling requirements for exploratory and secondary models are designed to increase the likelihood that these models continue to represent the constitu-ent of interest after the initial model publication. Exploratory or secondary models used to compute concentrations or loads at monthly or longer timespans are subject to the same criteria described previously for targeted models.

Model Development

For all models, relations between a constituent of interest and potential explanatory variables are evaluated to determine if variables seem to be linearly related, if sample data seem to represent the population being estimated, and that residuals are independent of explanatory variables, are normally distrib-uted, and have constant variance (Helsel and others, 2020). Because no universally accepted metric exists for determining the best model (Helsel and others, 2020), candidate models are chosen among those that maximize adjusted coefficient of determination (R^2; or pseudo R^2 for Tobit models; McK-elvey and Zavoina, 1975) and Mallow's C_p (which represents model bias and fit to the data; Helsel and others, 2020) and minimize root mean square error and prediction error sum of squares for OLS-estimated models or residual standard error for AMLE-estimated models. If either sine or cosine seasonal-ity variables are initially included in the model, the final model will include the corresponding counterpart, so both sine and cosine variables are included in the model. A bias correction factor (Duan, 1983; Cohn and others, 1989) is calculated for models with logarithmically transformed response variables to reduce the inherent negative bias (Helsel and others, 2020) during the retransformation of model computations back into their original units.

Outliers are identified following Rasmussen and others (2009) and Helsel and others (2020). Studentized residuals (which indicate outliers with high leverage), leverage, Cook's

distance (large values indicate influential observations; Cook, 1977), and difference in fit (large values indicate influential observations) values are used to identify influential data points for OLS-estimated models, and leverage and Cook's distance values are used to identify influential data points for AMLE-estimated models. Data points are only removed if a rationale supports that they are not representative of the dataset; supporting rationale may include hold-time violations, documented sampling issues, or other metadata indicative of potential bias to avoid erroneous inflation of model-computed values at the upper range of model relations. Outliers removed from models are documented in an associated data release. In addition, models are not published that:

1. Have an R^2 value smaller than 0.5,

2. Have explanatory variables with a variance inflation factor greater than 10 (to reduce the potential for multicollinearity),

3. Do not have at least four samples per year over the period of model development,

4. Have datasets representing less than 80 percent of the measured hydrologic condition range (for example, streamflow measurements) during the model-calibration time period,

5. Do not meet project objectives, or

6. Have substantial heteroscedasticity or nonnormality in residual plots.

In addition to single and multiple linear regression approaches, the Weighted Regressions on Time, Discharge, and Season with Kalman filtering (WRTDS–K; Hirsch, 2024) method may be used to develop secondary models to be paired with sensor-based models to compute monthly, seasonal, or annual estimates of constituent concentration and load. WRTDS–K was determined to produce the most accurate annual load estimates across various sampling regimes and constituents (Lee and others, 2019). Estimates of the uncertainty of monthly, seasonal, or annual time series are computed using USGS LOADEST for simple and multiple linear regression methods and EGRET software for WRTDS–K (Runkel and others, 2004; Hirsch and De Cicco, 2015; Hirsch and others, 2015). When longer term estimates among primary and secondary models are combined, the confidence or prediction intervals of primary and secondary models are summed in quadrature where the combined uncertainty is the square root of the sum of the squares of the individual uncertainties, as documented for data produced by the USGS National Water Quality Network (Lee and others, 2017).

Model Archive Summary

In addition to procedures described previously, the USGS KSWSC requires model documentation through a standard archive summary to be published and stored in the USGS National Real-Time Water Quality Data for the Nation Data Service. Model archive summaries must include site and model information; descriptions of the model-calibration dataset and discrete sampling details; model development information; a model summary; model statistics, data, and plots; and the model dataset. Model plots for OLS regressions include box and bivariate plots of independent and dependent variables, residual plots, seasonal and annual box plots, and cross-validation plots following appendix 1 in Williams (2023; appendix 1). Model plots for Tobit regressions include measured versus computed bivariate, residual versus computed, residual versus time, and independent versus dependent variable plots following appendix 12 in Williams (2023; appendix 2). The USGS KSWSC also requires a minimum of two USGS peer reviews, including one from outside the author's center, before the model archive can be published and stored in NRTWQ.

Changes In Sensor Technology

As technology improves, new sensors can report values differently, even from sensors produced by the same manufacturer that report in the same units. This difference primarily applies to sensors that report surrogate properties, such as turbidity, in which the number, wavelength, and angle of light sources and detectors can cause different readings in the same media (Rasmussen and others, 2009). To ensure that historical and more recently developed models based on different turbidity sensors report equivalent results, Rasmussen and others (2009) suggest applying correction factors based on side-by-side sensor deployments or the use of manufacturer-supplied correction factors.

The USGS KSWSC plans to develop correction factors to apply current models to historical sensor data; currently (2024), these cases are limited to historical changes in turbidity sensors. As new sensors are deployed, correction factors are documented as needed through the model archive published within the data release. Correction factors are based on side-by-side in situ comparisons where factors affecting sensor readings are expected to be similar to site locations where models are in use. For example, the number, size, shape, and color of particles in water are the primary factors that affect turbidity readings; thus, the development of a correction factor for a potential change in turbidity sensors could be developed and applied across sites with similar characteristics (for example, drainage basin soil types, stream size, and substrate composition) to allow comparative use of historical data. If sufficient side-by-side data are not available in the same waterbody across 80 percent of the range of streamflows and sensor data, models are required to be redeveloped for new sensors that

report apparent properties, such as turbidity. Sensor values will not be changed within the USGS Water Data for the Nation (U.S. Geological Survey, 2024) database; correction factors are documented on NRTWQ and incorporated into models that use historical data.

Model Application and Maintenance

The ultimate purpose of publishing model archives is to document methods used to publish time-series water-quality computations for USGS stakeholders. NRTWQ is used to provide time-series water-quality computations with associated uncertainty from models published by the USGS. Unlike the USGS Water Data for the Nation database (U.S. Geological Survey, 2024), NRTWQ is designed to serve model computations that can be overwritten as new models are developed. Model computations are accompanied by graphics and information on model form, error, and fit. After initial model publication, continued evaluation of model fit is required to ensure that time-series computations continue to be representative of environmental conditions. This report establishes minimum criteria for sampling and model review to continue to publish data from previously documented models on NRTWQ.

Determining procedures for model validation necessitates weighing the number of samples needed to represent changing environmental conditions with the resources required to collect those samples. To maintain consistency with minimum requirements used for national-scale USGS publications on water-quality conditions (Rasmussen and others, 2009; Lee and others, 2017; Oelsner and others, 2017), models published by the USGS are planned to require at least four samples per year that span a range of annual streamflow conditions to be collected. Samples are initially planned to represent typical seasonal conditions but will be adjusted as necessary to represent hydrologic or water-quality conditions outside the range of previously developed models whenever possible.

The USGS KSWSC plans to complete, at a minimum, model reviews annually and every 3 years to ensure that operational NRTWQ models continue to represent environmental conditions since model publication. Annual reviews are qualitative evaluations of whether samples from the most recent year seem to maintain the same relations with explanatory variables as defined in published models. These reviews are expected to evaluate if the recent samples are within existing model uncertainty, check for outliers, and ensure results are properly displayed on NRTWQ. After 3 years of ongoing model operation, model reviews are planned to be completed to determine if samples since publication differ significantly from published relations. As described in Rasmussen and others (2009) for turbidity and suspended-sediment relations, an analysis of covariance is planned to be used to evaluate the slopes and intercepts (if slopes are not significantly different) among (1) the regression model on the basis of the original model and additional data, (2) the original regression model,

and (3) the regression model solely on the basis of the additional data to determine if the existing model is still suitable for use, provided that the assumptions for analysis of covariance are met. If the slope or intercept significantly differs (probability value less than 0.05) for any of the three cases, the model should be discontinued from NRTWQ and reevaluated before publishing new model computations. New models are also planned to be developed if newly collected discrete sample concentrations or measured explanatory variables exceed those in the initial model by more than 20 percent. New models are planned to be developed for all operational NRTWQ models 6 years after the initial publication. Unless changes in the relation among constituents of interest and explanatory variables have been observed through time, updated models are planned to be developed using all available concomitant in situ data. Updated models are planned to be documented as new data releases on NRTWQ. Within NRTWQ, change logs are planned to cite previous model forms and document the date the models were updated.

Summary

The U.S. Geological Survey (USGS) Kansas Water Science Center (KSWSC) has published time-series computations of water-quality concentrations and loads based on in situ sensor data since 1995. Water-quality constituent concentrations or densities are computed using regression models that relate in situ sensor values to laboratory analyses of periodically collected samples. These regression models currently (2024) follow no uniform published guidance and are individually documented through USGS reports. The purpose of this report is to document and describe the USGS standard procedures used by the KSWSC for developing, publishing, maintaining, and updating continuous water-quality constituent concentration and load computations using models that relate in situ water-quality sensor values and concomitant discrete water-quality sample laboratory analysis values. This report describes updated (2024) procedures designed to improve the consistency, quality, and timeliness of computed continuous water-quality data produced by the USGS KSWSC. Beginning in 2024, models developed by the USGS KSWSC following specific procedures and requirements related to sample collection, model fit, and model documentation outlined in this report are planned to be published and stored in the USGS National Real-Time Water Quality Data for the Nation Data Service. This report also describes USGS KSWSC procedures for evaluating and publishing time-series water-quality computations after initial model development and documentation. This guidance can be used to improve model development and data computation consistency and streamline the time-series water-quality computation process from model development to publication.

References Cited

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 2001–4248, 28 p. [Also available at https://doi.org/10.3133/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 Ninnescah River and Cheney Reservoir, south-central Kansas, 1997–2003: U.S. Geological Survey Scientific Investigations Report 2006–5095, 43 p. [Also available at https://doi.org/10.3133/sir20065095.]

Christensen, V.G., Ziegler, A.C., Rasmussen, P.P., and Jian, X., 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, Va., American Water Resources Association, AWRA Technical Publication Series no. TPS–03–1, compact disc. [Also available at https://nrtwq.usgs.gov/ks/methods/christensen2003.]

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

Cohn, T.A., Delong, L.L., Gilroy, E.J., Hirsch, R.M., and Wells, D.K., 1989, Estimating constituent loads: Water Resources Research, v. 25, no. 5, p. 937–942. [Also available at https://doi.org/10.1029/WR025i005p00937.]

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

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

Foster, G.M., 2014, Relations between continuous real-time turbidity data and discrete suspended-sediment concentration samples in the Neosho and Cottonwood Rivers, east-central Kansas, 2009–2012: U.S. Geological Survey Open-File Report 2014–1171, 20 p. [Also available at https://doi.org/10.3133/ofr20141171.]

Foster, G.M., and Graham, J.L., 2016, Logistic and linear regression model documentation for statistical relations between continuous real-time and discrete water-quality constituents in the Kansas River, Kansas, July 2012 through June 2015: U.S. Geological Survey Open-File Report 2016–1040, 27 p., accessed December 12, 2023, at https://doi.org/10.3133/ofr20161040.

Graham, J.L., Stone, M.L., Rasmussen, T.J., and Poulton, B.C., 2010, Effects of wastewater effluent discharge and treatment facility upgrades on environmental and biological conditions of the Upper Blue River, Johnson County, Kansas and Jackson County, Missouri, January 2003 through March 2009: U.S. Geological Survey Scientific Investigations Report 2010–5248, 85 p. [Also available at https://doi.org/10.3133/sir20105248.]

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

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J., 2020, Statistical methods in water resources: U.S. Geological Survey Techniques and Methods, book 4, chap. A3, 458 p. [Also available at https://doi.org/10.3133/tm4A3. Supersedes USGS Techniques of Water-Resources Investigations, book 4, chap. A3, version 1.1.]

Hem, J.D., 1992, Study and interpretation of the chemical characteristics of natural water: U.S. Geological Survey Water-Supply Paper 2254 (3d ed.), 263 p., 4 pls. [Also available at https://doi.org/10.3133/wsp2254.]

Hirsch, R.M., 2024, WRTDS Kalman: U.S. Geological Survey web page, accessed March 29, 2024, at https://doi-usgs.github.io/EGRET/articles/WRTDSK.html.

Hirsch, R.M., Archfield, S.A., and De Cicco, L.A., 2015, A bootstrap method for estimating uncertainty of water quality trends: Environmental Modelling & Software, v. 73, p. 148–166. [Also available at https://doi.org/10.1016/j.envsoft.2015.07.017.]

Hirsch, R.M., and De Cicco, L.A., 2015, User guide to exploration and graphics for RivEr Trends (EGRET) and dataRetrieval—R packages for hydrologic data (ver. 2.0, February 2015): U.S. Geological Survey Techniques and Methods, book 4, chap. A10, 93 p., accessed December 12, 2023, at https://doi.org/10.3133/tm4A10.

Juracek, K.E., 2011, Suspended-sediment loads, reservoir sediment trap efficiency, and upstream and downstream channel stability for Kanopolis and Tuttle Creek Lakes, Kansas, 2008–10: U.S. Geological Survey Scientific Investigations Report 2011–5187, 35 p., accessed December 12, 2023, at https://doi.org/10.3133/sir20115187.

Juracek, K.E., 2013, Suspended-sediment loads and reservoir sediment trap efficiency for Clinton Lake, Kansas, 2010–12: U.S. Geological survey Scientific Investigations Report 2013–5153, 10 p., accessed December 12, 2023, at https://doi.org/10.3133/sir20135153.

Kramer, A.R., Klager, B.J., Stone, M.L., and Eslick-Huff, P.J., 2021a, Regression relations and long-term water-quality constituent concentrations, loads, yields, and trends in the North Fork Ninnescah River, south-central Kansas, 1999–2019: U.S. Geological Survey Scientific Investigations Report 2021–5006, 51 p., accessed December 12, 2023, at https://doi.org/10.3133/sir20215006.

Kramer, A.R., Peterman-Phipps, C.L., Mahoney, M.D., and Lukasz, B.S., 2021b, Sediment concentrations and loads upstream from and through John Redmond Reservoir, east-central Kansas, 2010–19: U.S. Geological Survey Scientific Investigations Report 2021–5037, 50 p., accessed December 12, 2023, at https://doi.org/10.3133/sir20215037.

Kramer, A.R., and Puls, K.A., 2023, Documentation of linear regression models for computing water-quality constituent concentrations using continuous real-time water-quality data for the North Fork Ninnescah River and Cheney Reservoir, Kansas, 2014–21: U.S. Geological Survey Scientific Investigations Report 2023–5037, 20 p., accessed December 12, 2023, at https://doi.org/10.3133/sir20235037.

Lee, C., and Foster, G., 2013, Assessing the potential of reservoir outflow management to reduce sedimentation using continuous turbidity monitoring and reservoir modelling: Hydrological Processes, v. 27, no. 10, p. 1426–1439. [Also available at https://doi.org/10.1002/hyp.9284.]

Lee, C.J., Hirsch, R.M., and Crawford, C.G., 2019, An evaluation of methods for computing annual water-quality loads: U.S. Geological Survey Scientific Investigations Report 2019–5084, 59 p., accessed December 12, 2023, at https://doi.org/10.3133/sir20195084.

Lee, C.J., Hirsch, R.M., Schwarz, G.E., Holtschlag, D.J., Preston, S.D., Crawford, C.G., and Vecchia, A.V., 2016, An evaluation of methods for estimating decadal stream loads: Journal of Hydrology, v. 542, p. 185–203. [Also available at https://doi.org/10.1016/j.jhydrol.2016.08.059.]

Lee, C.J., Murphy, J.C., Crawford, C.G., and Deacon, J.R., 2017, Methods for computing water-quality loads at sites in the U.S. Geological Survey National Water Quality Network (ver. 1.3, August 2021): U.S. Geological Survey Open-File Report 2017–1120, 20 p., accessed December 12, 2023, at https://doi.org/10.3133/ofr20171120.

Lee, C.J., Rasmussen, P.P., and Ziegler, A.C., 2008, Characterization of suspended-sediment loading to and from John Redmond Reservoir, east-central Kansas, 2007–2008: U.S. Geological Survey Scientific Investigations Report 2008–5123, 25 p. [Also available at https://doi.org/10.3133/sir20085123.]

Leiker, B.M., 2022, Linear regression model documentation for computing water-quality constituent concentrations using continuous real-time water-quality data for the Republican River, Clay Center, Kansas, July 2018 through March 2021: U.S. Geological Survey Scientific Investigations Report 2022–5016, 13 p., accessed December 12, 2023, at https://doi.org/10.3133/sir20225016.

Mau, D.P., Ziegler, A.C., Porter, S.D., and Pope, L.M., 2004, Surface-water-quality conditions and relation to taste-and-odor occurrences in the Lake Olathe watershed, northeast Kansas, 2000–02: U.S. Geological Survey Scientific Investigations Report 2004–5047, 95 p. [Also available at https://doi.org/10.3133/sir20045047.]

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

Oelsner, G.P., Sprague, L.A., Murphy, J.C., Zuellig, R.E., Johnson, H.M., Ryberg, K.R., Falcone, J.A., Stets, E.G., Vecchia, A.V., Riskin, M.L., De Cicco, L.A., Mills, T.J., and Farmer, W.H., 2017, Water-quality trends in the Nation's rivers and streams, 1972–2012—Data preparation, statistical methods, and trend results (ver. 2.0, October 2017): U.S. Geological Survey Scientific Investigations Report 2017–5006, 136 p., accessed December 12, 2023, at https://doi.org/10.3133/sir20175006.

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., accessed December 12, 2023, at https://doi.org/10.3133/ofr20161057.

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. [Also available at https://doi.org/10.3133/tm3C4.]

Rasmussen, T.J., Lee, C.J., and Ziegler, A.C., 2008, Estimation of constituent concentrations, loads, and yields in streams of Johnson County, northeast Kansas, using continuous water-quality monitoring and regression models, October 2002 through December 2006: U.S. Geological Survey Scientific Investigations Report 2008–5014, 103 p. [Also available at https://doi.org/10.3133/sir20085014.]

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, 117 p. [Also available at https://doi.org/10.3133/sir20055165.]

Runkel, R.L., Crawford, C.G., and Cohn, T.A., 2004, Load estimator (LOADEST)—A FORTRAN program for estimating constituent loads in streams and rivers: U.S. Geological Survey Techniques and Methods, book 4, chap. A5, 69 p., accessed December 12, 2023, at https://doi.org/10.3133/tm4A5.

Stone, M.L., and Graham, J.L., 2014, Model documentation for relations between continuous real-time and discrete water-quality constituents in Indian Creek, Johnson County, Kansas, June 2004 through May 2013: U.S. Geological Survey Open-File Report 2014–1170, 70 p., accessed December 12, 2023, at https://doi.org/10.3133/ofr20141170.

Stone, M.L., Graham, J.L., and Gatotho, J.W., 2013a, Model documentation for relations between continuous real-time and discrete water-quality constituents in Cheney Reservoir near Cheney, Kansas, 2001–2009: U.S. Geological Survey Open-File Report 2013–1123, 100 p., accessed December 12, 2023, at https://doi.org/10.3133/ofr20131123.

Stone, M.L., Graham, J.L., and Gatotho, J.W., 2013b, Model documentation for relations between continuous real-time and discrete water-quality constituents in the North Fork Ninnescah River upstream from Cheney Reservoir, south-central Kansas, 1999–2009: U.S. Geological Survey Open-File Report 2013–1014, 101 p., accessed December 12, 2023, at https://doi.org/10.3133/ofr20131014.

Stone, M.L., and Klager, B.J., 2022, Documentation of models describing relations between continuous real-time and discrete water-quality constituents in the Little Arkansas River, south-central Kansas, 1998–2019: U.S. Geological Survey Open-File Report 2022–1010, 34 p., accessed December 12, 2023, at https://doi.org/10.3133/ofr20221010.

Stone, M.L., and Klager, B.J., 2023, Long-term water-quality constituent trends in the Little Arkansas River, south-central Kansas, 1995–2021: U.S. Geological Survey Scientific Investigations Report 2023–5102, 103 p., accessed December 12, 2023, at https://doi.org/10.3133/sir20235102.

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

U.S. Geological Survey, 2024, USGS water data for the Nation: U.S. Geological Survey National Water Information database, accessed March 25, 2024, at https://doi.org/10.5066/F7P55KJN.

Williams, T.J., 2021, Linear regression model documentation and updates for computing water-quality constituent concentrations or densities using continuous real-time water-quality data for the Kansas River, Kansas, July 2012 through September 2019: U.S. Geological Survey Open-File Report 2021–1018, 18 p., accessed December 12, 2023, at https://doi.org/10.3133/ofr20211018.

Williams, T.J., 2023, Linear regression model documentation for computing water-quality constituent concentrations or densities using continuous real-time water-quality data for the Kansas River above Topeka Weir at Topeka, Kansas, November 2018 through June 2021: U.S. Geological Survey Scientific Investigations Report 2022–5130, 14 p., accessed December 12, 2023, at https://doi.org/10.3133/sir20225130.

Appendix 1.   Model Archive Summary Example—Ordinary Least Squares

An example summary of a model archive that uses ordinary least squares analysis (appendix 1 in Williams [2023]) is available for download at https://doi.org/10.3133/ofr20241049.

Reference Cited

Williams, T.J., 2023, Linear regression model documentation for computing water-quality constituent concentrations or densities using continuous real-time water-quality data for the Kansas River above Topeka Weir at Topeka, Kansas, November 2018 through June 2021: U.S. Geological Survey Scientific Investigations Report 2022–5130, 14 p., accessed January 18, 2024, at https://doi.org/10.3133/sir20225130.

Appendix 2.   Model Archive Summary Example—Tobit

An example summary of a model archive that uses Tobit (appendix 12 in Williams [2023]) is available for download at https://doi.org/10.3133/ofr20241049.

Reference Cited

Williams, T.J., 2023, Linear regression model documentation for computing water-quality constituent concentrations or densities using continuous real-time water-quality data for the Kansas River above Topeka Weir at Topeka, Kansas, November 2018 through June 2021: U.S. Geological Survey Scientific Investigations Report 2022–5130, 14 p., accessed January 18, 2024, at https://doi.org/10.3133/sir20225130.