Appendix 12. Model Archive Summary for Sodium Concentration at U.S. Geological Survey Station 07144790, Cheney Reservoir near Cheney, Kansas, during October 1, 2014, through September 30, 2021

This model archive summary summarizes the sodium (Na) concentration model developed to compute 15-minute, hourly, or daily sodium concentrations during October 1, 2014, onward. This model supersedes all prior models used during this period. The methods follow U.S. Geological Survey (USGS) guidance as referenced in relevant Office of Surface Water/Office of Water Quality Technical Memoranda and USGS Techniques and Methods, book 3, chapter C4 (Rasmussen and others, 2009; U.S. Geological Survey, 2016).

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

Site and Model Information

Site number: 07144790

Site name: Cheney Reservoir near Cheney, Kansas

Location: Lat 37°43'34", long 97°47'38" referenced to North American Datum of 1927, in SE 1/4 NE 1/4 NW 1/4 sec.06, T.27 S., R.4 W., Sedgwick County, Kans., hydrologic unit 11030014, in control house structure at outlet works of Cheney Dam on North Fork Ninnescah River, 6.0 mi north of Cheney, and at mile 15.9.

Equipment: A YSI, Inc., EXO water-quality monitor (YSI, Inc., 2017) equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, turbidity, chlorophyll fluorescence, and phycocyanin fluorescence was installed November 14, 2015. The EXO monitor is suspended from the dam intake tower walkway. The monitor is at a depth that fluctuates between three to six feet depending on reservoir elevation. Measurements from the EXO were recorded every 15 minutes to hourly and transmitted hourly via satellite. Reservoir elevation was measured using a Design Analysis H–350 nonsubmersible pressure transducer and H–355 gas system.

Date model was created: August 9, 2022

Model calibration data period: February 17, 2016, through August 31, 2021 (dataset consisted of 44 discrete water-quality samples).

Model application date: November 14, 2015, onward (date of EXO continuous water-quality monitor installation).

Model developed by: Ariele Kramer, USGS, Lawrence, Kans. (akramer@usgs.gov)

Model Calibration Dataset

All data were collected using USGS protocols (U.S. Geological Survey, 2006; Wagner and others, 2006; Bennett and others, 2014) and are stored in the USGS National Water Information System database (https://doi.org/10.5066/F7P55KJN; U.S. Geological Survey, 2022). Potential explanatory variables evaluated individually and in combination were water temperature, specific conductance, pH, dissolved oxygen, turbidity, chlorophyll fluorescence, phycocyanin fluorescence, seasonality (sine and cosine variables), and reservoir elevation.

The regression model is based on 44 concomitant values of discretely collected sodium concentration and continuously measured specific conductance during February 17, 2016, through August 31, 2021. Discrete samples were collected throughout the range of continuously observed hydrologic conditions. No samples had sodium concentrations that were less than laboratory minimum reporting level. All potential explanatory variables were time interpolated within the 15-minute to hourly continuous record based on the discrete sample time. The maximum time span between two continuous data points used for interpolation was 4 hours (to preserve the sample dataset, field monitor averages obtained during sample collection were used for model development data if no continuous data were available or if gaps larger than 4 hours in the continuous data record resulted in missing interpolated data). Summary statistics and the complete model-calibration dataset are provided below. Potential outliers were identified using the methods described in Rasmussen and others (2009) and Helsel and others (2020). All potential outliers were investigated by reviewing sample collection information sheets and laboratory reports; if there were no clear issues, explanations, or conditions that would cause a result to be invalid for model calibration, the sample was retained in the dataset. Three samples in the model calibration dataset were flagged as outliers but all were retained in the dataset after further review.

Sodium Sampling Details

Discrete water-quality samples were collected primarily by depth-integrating through the photic-zone (depth at which light is approximately 1 percent of that at the surface) using a double check-valve bailer (Lane and others, 2003). Vertical water-quality profiles collected during sampling indicated that thermal stratification rarely occurs, and water-quality conditions are typically uniform throughout the water column. Samples were collected from the walkway on the dam intake tower. Discrete samples were collected on a semifixed to event-based schedule six to eight times per year. All samples were collected between 9:15 a.m. and 12:20 p.m. Samples were analyzed for sodium concentration by the Wichita Municipal Water and Wastewater Laboratory in Wichita, Kans., according to standard methods (Eaton and others, 1995).

Continuous Water-Quality Data

Specific conductance was continuously measured (15 minutes to hourly) using a YSI, Inc., EXO multiparameter sonde (YSI, Inc., 2017). The water-quality monitor was operated and maintained according to standard USGS methods (Wagner and others, 2006; Bennett and others, 2014). All continuous water-quality data at Cheney Reservoir near Cheney, Kans. are available in near-real time (updated hourly) from the USGS National Water Information System database

(https://doi.org/10.5066/F7P55KJN; U.S. Geological Survey, 2022) using the site number 07144790.

Model Development

Ordinary least squares linear regression was used to develop surrogate regression models that relate continuous water-quality conditions to discretely sampled constituent concentrations. All regressions were computed using the R software environment (R Core Team, 2020). The data and subsequent regression equation must meet the five assumptions necessary to apply ordinary least squares regression: the dependent variable is linearly related to the explanatory variables, data used to fit the model are representative of the data of interest, the variance of the residuals is constant (homoscedastic), the residuals are independent of the explanatory variables, and the residuals are normally distributed (Helsel and others, 2020). Previously published explanatory variables also were considered for continuity.

Specific conductance was selected as a good surrogate for sodium concentration based on residual plots, coefficient of determination (R^2), and model standard percentage error (MSPE). Values for the aforementioned statistics were computed and are included below along with all relevant sample data and additional statistical information.

Model Summary

Summary of final sodium concentration regression analysis at USGS site 07144790:

Sodium concentration-based model:

$$Na = 0.178 \times SPC - 48.9,$$

where,

Na = sodium, dissolved, in milligrams per liter (mg/L); and

SPC = specific conductance, in microsiemens per centimeter at 25 degrees Celsius.

SPC makes physical and statistical sense as an explanatory variable for sodium concentration because of its positive correlation with charged ionic species (Hem, 1985).

Extrapolation, defined as computation beyond the range of the model calibration dataset, may be used to extrapolate no more than 10 percent outside the range of the calibration data used to fit the model and is therefore limited. The extrapolation limit for sodium concentration using this model is 139.7 mg/L. Computed estimates outside that limit are not supported by the current model calibration dataset.

Model statistics, data, and plots

Definitions

Variable	Explanation
Cook’s D	Cook’s distance, a measure of influence (Helsel and others, 2020)

Variable	Explanation
DFFITS	Difference in fits, a measure of influence (Helsel and others, 2020)
E.vars	Explanatory variables
Leverage	An outlier's measure in the x direction (Helsel and others, 2020)
LOESS	Local polynomial regression fitting (Helsel and others, 2020)
MSE	Model standard error (Helsel and others, 2020)
MSPE	Model standard percentage error (Helsel and others, 2020)
Sodium	Sodium, dissolved, in milligrams per liter (mg/L) (USGS parameter code 00930)
Pr(>|*t*|)	The probability that the independent variable has no effect on the dependent variable (Helsel and others, 2020)
RMSE	Root mean square error (Helsel and others, 2020)
SPC	Specific conductance, in microsiemens per centimeter at 25 degrees Celsius (µS/cm at 25°C) (USGS parameter code 00095)
t value	Student's *t* value; the coefficient divided by its associated standard error (Helsel and others, 2020)

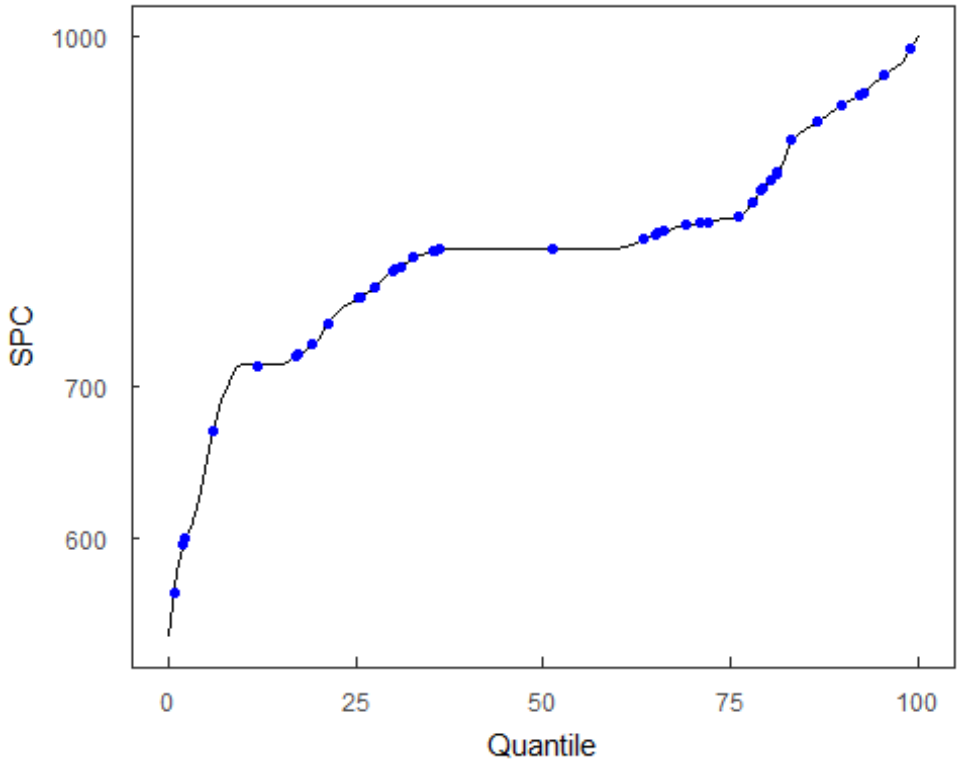
Model

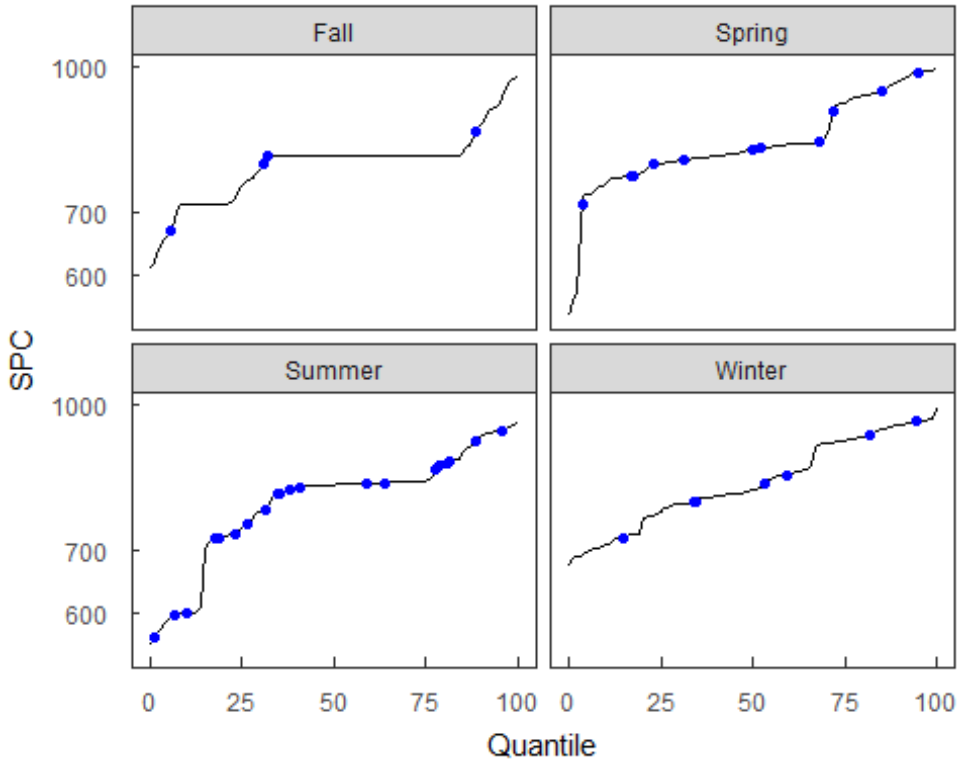
$$Na = 0.178 \times SPC - 48.9$$

Variable summary statistics

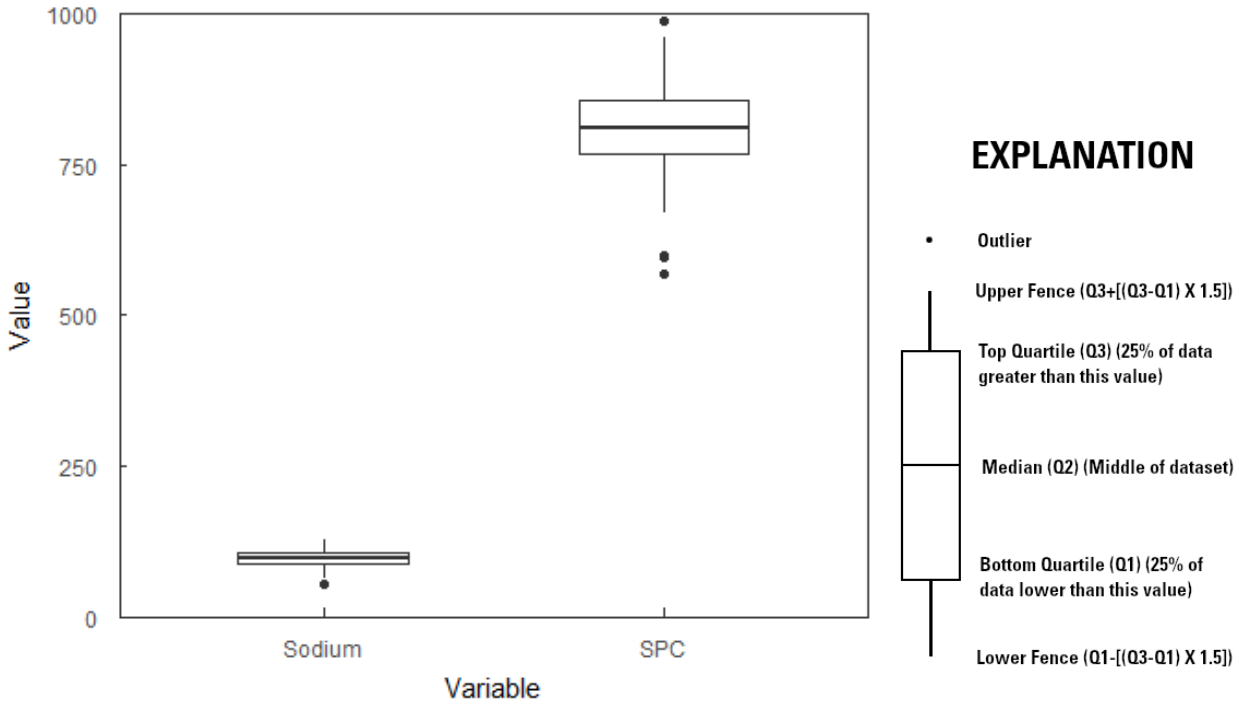
Variable	Minimum	Q1	Median	Mean	Q3	Maximum
Sodium	54.8	87	96.9	94.2	107	127
SPC	568	767	810	805	855	988

Duration plots

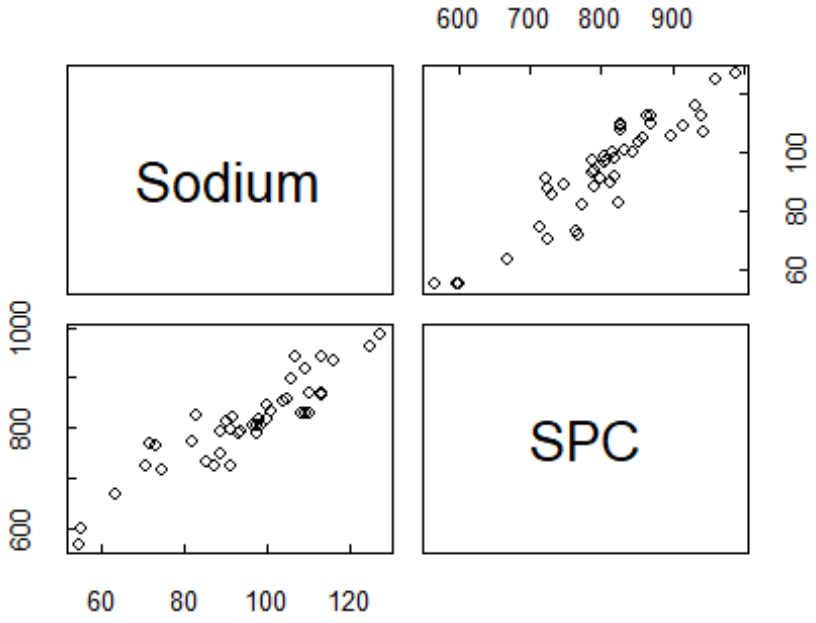




Box plots



Scatter plots



The x- and y-axis labels for a given bivariate plot are defined by the intersecting row and column labels.

Basic model statistics

Statistic	Value
Observations	44
R^2	0.849
Adjusted R^2	0.845
RMSE	6.92
Upper MSPE (90%)	7.34
Lower MSPE (90%)	-7.34

Model coefficients

	Estimate	Standard error	t value	Pr(>|t|)
(Intercept)	-48.8825242	9.3888220	-5.20646	5.4e-06
SPC	0.1777432	0.0115877	15.33896	0.0e+00

Correlation matrix

	Sodium	SPC
Sodium	1.0000000	0.9211572
SPC	0.9211572	1.0000000

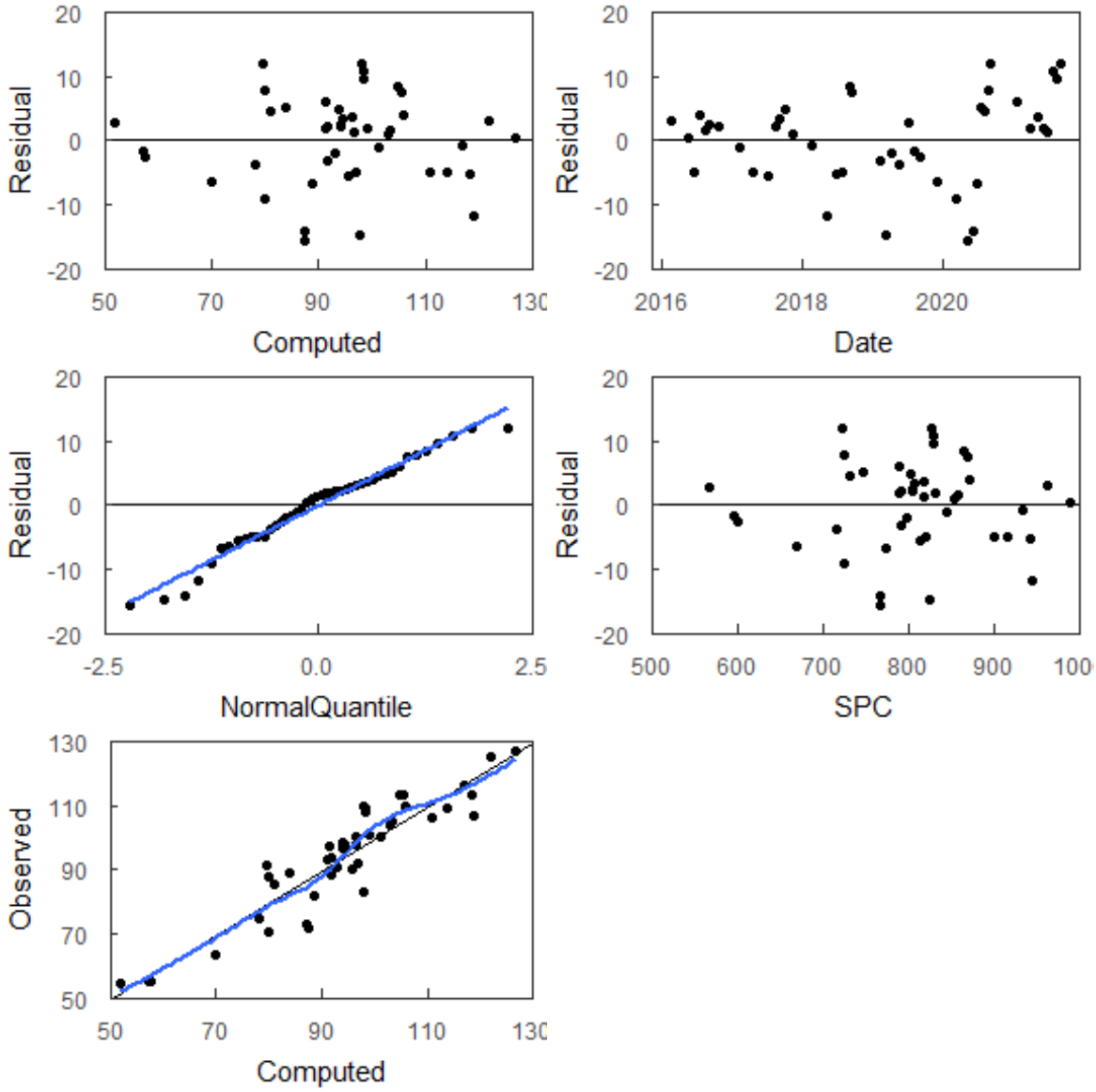
Outlier test criteria

Leverage	DFFITS	CooksD
0.1364	0.4264	0.1939

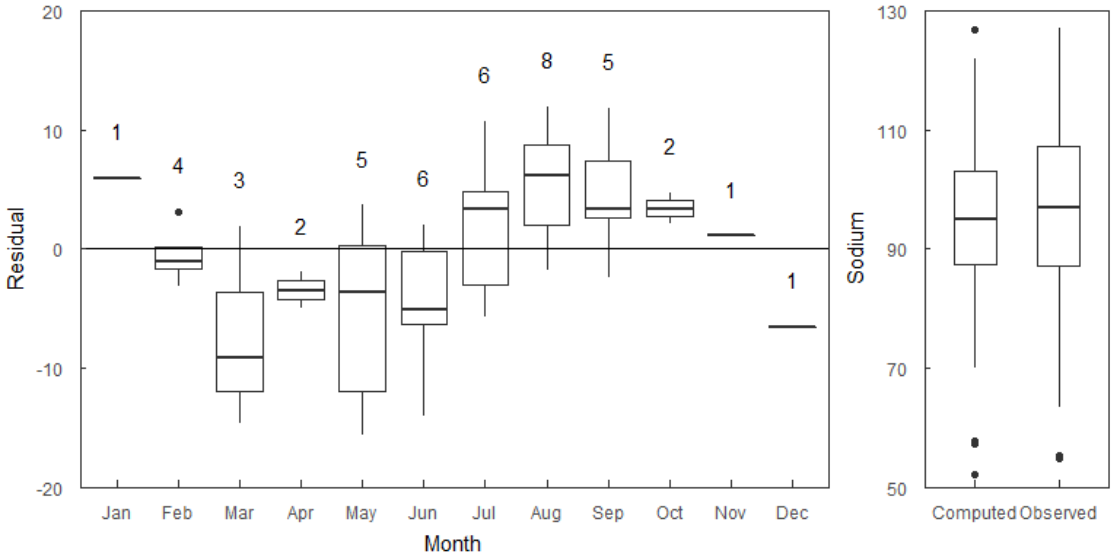
Flagged observations

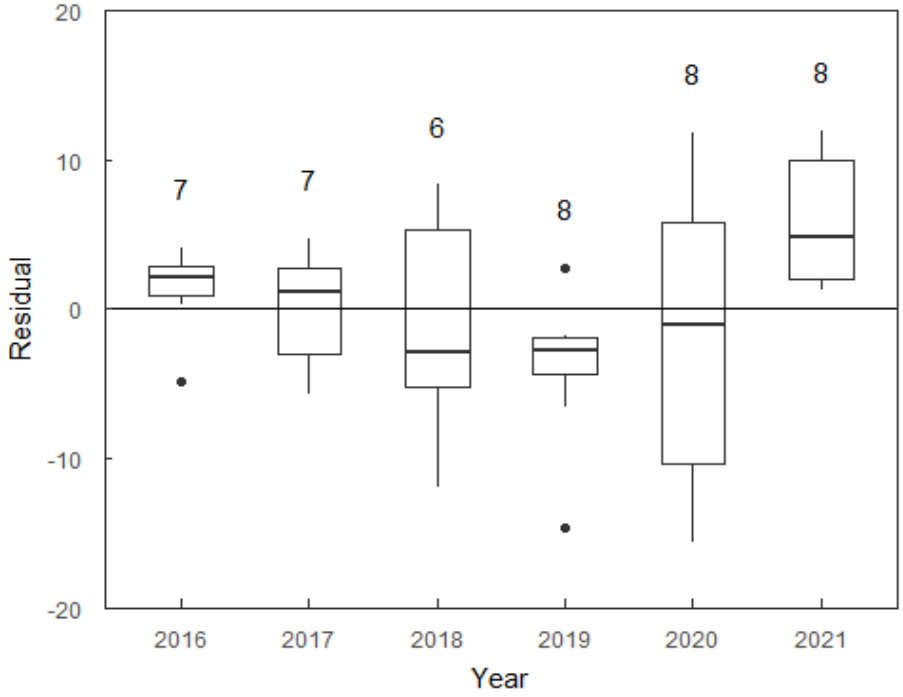
datetime	Sodium	CooksD	DFFITS	Leverage	Studentized Residual
2019-07-09 10:15:00	54.8	0.0209	0.202	0.181	0.431
2019-08-06 11:00:00	55.4	0.0069	-0.116	0.144	-0.283
2019-09-03 10:40:00	55.3	0.0121	-0.154	0.141	-0.38

Statistical plots



The blue line shows the locally estimated scatterplot smoothing (LOESS). The black dots correspond to observed values. The black line represents the 1:1 line.





EXPLANATION

7 Number of values

• Outlier

Upper Fence (Q3+[(Q3-Q1) X 1.5])

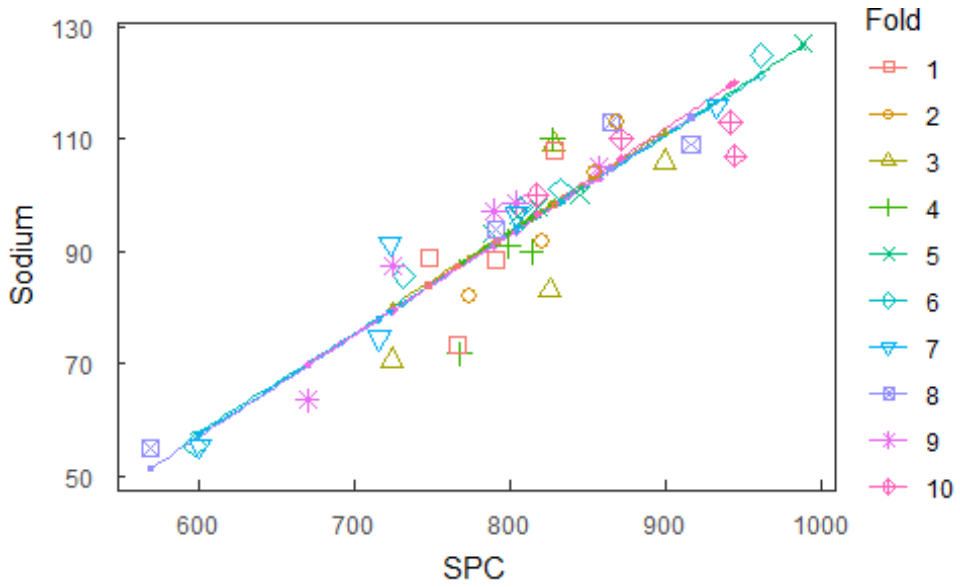
Top Quartile (Q3) (25% of data greater than this value)

Median (Q2) (Middle of dataset)

Bottom Quartile (Q1) (25% of data lower than this value)

Lower Fence (Q1-[(Q3-Q1) X 1.5])

Cross Validation



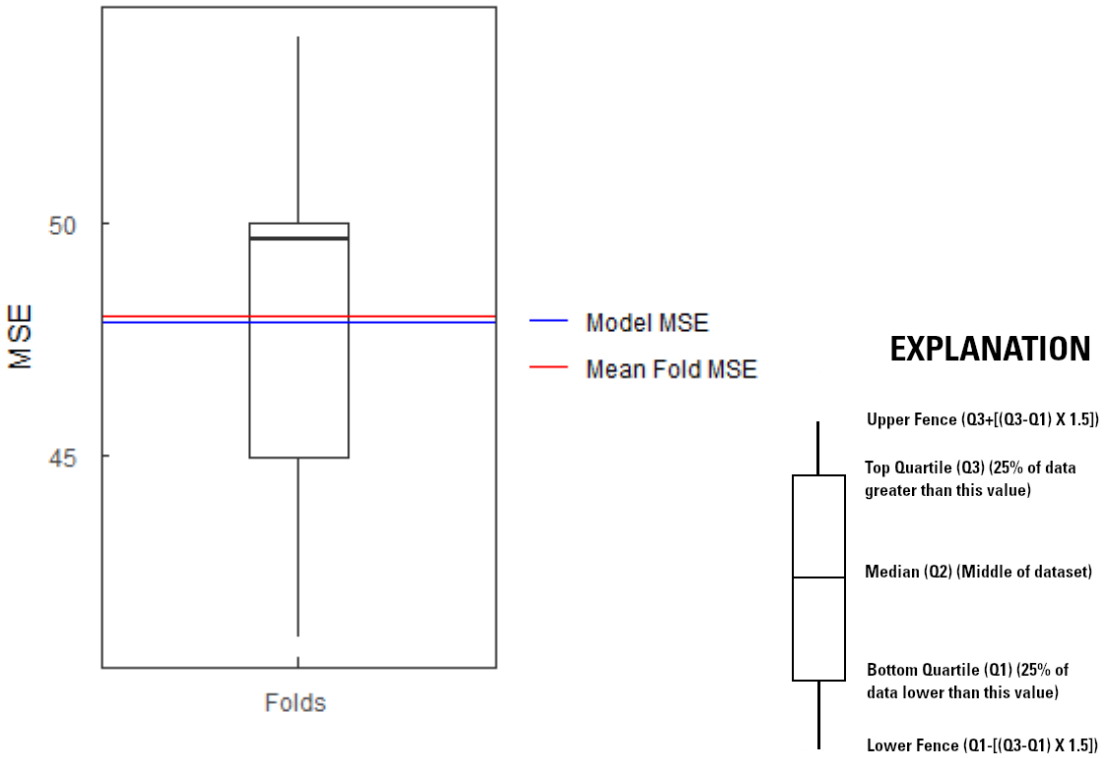
Fold - equal partition of the data (10 percent of the data).

Large symbols – observed value of a data point removed in a fold.

Small symbols – recomputed value of a data point removed in a fold.

Recomputed regression lines – adjusted regression line with one fold removed.

Statistic	Value
Minimum MSE of folds	41.1
25th Percentile	45
Median MSE of folds	49.7
Mean MSE of folds	48
75th percentile	50
Maximum MSE of folds	54
Model MSE	47.9



Model calibration dataset

datetime	Sodium	SPC	Computed
2016-02-17 10:45:00	125	961	122
2016-05-17 10:20:00	127	988	127
2016-06-15 09:15:00	106	899	111
2016-07-18 10:40:00	110	871	106
2016-08-15 10:30:00	105	857	103
2016-09-06 10:40:00	96.6	804	94.0
2016-10-25 10:15:00	93.8	791	91.7
2017-02-09 10:40:00	100	844	101
2017-04-17 10:30:00	91.9	820	96.9
2017-07-10 11:40:00	90.0	813	95.7
2017-08-15 10:00:00	96.4	805	94.2
2017-09-07 10:00:00	97.7	806	94.4
2017-10-03 10:20:00	98.6	803	93.9
2017-11-13 12:00:00	104	854	103
2018-02-13 10:40:00	116	932	117
2018-05-08 10:30:00	107	944	119
2018-06-25 12:00:00	113	941	118
2018-07-26 11:40:00	109	916	114

datetime	Sodium	SPC	Computed
2018-08-29 11:00:00	113	864	105
2018-09-11 09:40:00	113	869	106
2019-02-05 11:20:00	88.6	791	91.7
2019-03-07 10:50:00	83.1	825	97.8
2019-04-09 10:30:00	91.0	798	93.0
2019-05-14 11:10:00	74.6	715	78.3
2019-07-09 10:15:00	54.8	568	52.1
2019-08-06 11:00:00	55.4	597	57.2
2019-09-03 10:40:00	55.3	600	57.8
2019-12-04 10:50:00	63.5	669	70.1
2020-03-04 11:00:00	70.6	724	79.8
2020-05-06 10:30:00	71.8	767	87.4
2020-06-03 10:20:00	73.2	766	87.3
2020-06-25 11:30:00	82.0	774	88.7
2020-07-15 10:00:00	89.0	747	83.9
2020-08-04 11:30:00	85.6	731	81.0
2020-08-18 11:40:00	87.5	724	79.8
2020-09-01 10:50:00	91.4	723	79.6
2021-01-13 10:30:00	97.3	789	91.3
2021-03-31 10:30:00	93.1	788	91.2
2021-05-04 10:20:00	100	817	96.3
2021-06-03 10:30:00	101	832	99.0
2021-06-21 11:00:00	97.8	818	96.5
2021-07-20 10:40:00	109	828	98.3
2021-08-10 10:00:00	108	828	98.3
2021-08-31 11:40:00	110	827	98.1

References Cited

Bennett, T.J., Graham, J.L., Foster, G.M., Stone, M.L., Juracek, K.E., Rasmussen, T.J., and Putnam, J.E., 2014, U.S. Geological Survey quality-assurance plan for continuous water-quality monitoring in Kansas, 2014: U.S. Geological Survey Open-File Report 2014–1151, 34 p. plus appendixes, accessed September 7, 2022, at https://doi.org/10.3133/ofr20141151.

Eaton, A.D., Clesceri, L.S., and Greenberg, A.E., eds., 1995, Standard methods for the examination of water and wastewater (19th ed.): New York, American Public Health Association, 905 p.

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., 1985, Study and interpretation of the chemical characteristics of natural water (3d. ed): U.S. Geological Survey Water-Supply Paper 2254, 264 p. [Also available at https://doi.org/10.3133/wsp2254.]

Lane, S.L., Flanagan, S., and Wilde, F.D., 2003, Selection of equipment for water sampling (ver. 2.0): U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A2, accessed September 2022 at http://pubs.water.usgs.gov/twri9A2/

R Core Team, 2020, R—A language and environment for statistical computing: R Foundation for Statistical Computing software release (version 4.0.2), accessed September 7, 2022, at https://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, 52 p. [Also available at https://doi.org/10.3133/tm3C4.]

U.S. Geological Survey, 2006, Collection of water samples (ver. 2.0, September 2006): U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A4 [variously paged]. [Also available at https://doi.org/10.3133/twri09A4.]

U.S. Geological Survey, 2016, Policy and guidance for approval of surrogate regression models for computation of time series suspended-sediment concentration and loads: U.S. Geological Survey Office of Surface Water Technical Memorandum 2016.07, Office of Water Quality Technical Memorandum 2016.10, 40 p., accessed September 7, 2022, at https://water.usgs.gov/water-resources/memos/memo.php?id=467.

U.S. Geological Survey, 2022, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed September 7, 2022, at https://doi.org/10.5066/F7P55KJN.

Wagner, R.J., Boulger, R.W., Jr., Oblinger, C.J., and Smith, B.A., 2006, Guidelines and standard procedures for continuous water-quality monitors—Station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods, book 1, chap. D3, 51 p. plus 8 attachments. [Also available at https://doi.org/10.3133/tm1D3.]

YSI, Inc., 2017, EXO user manual—Advanced water quality monitoring platform (rev. G): Yellow Springs, Ohio, YSI, Inc., 154 p., accessed September 7, 2022, at https://www.ysi.com/file%20library/documents/manuals/exo-user-manual-web.pdf.