Appendix 15. Model Archival Summary for Nitrate plus Nitrite Concentration at U.S. Geological Survey Site 06887500, Kansas River at Wamego, Kansas, during June 2014 through September 2019

This model archival summary summarizes the nitrate plus nitrite (NO_x; U.S. Geological Survey [USGS] parameter code 00631) concentration model developed to compute 15-minute NO_x concentrations from June 2014 onward. This model supersedes all previous models.

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: 06887500
Site name: Kansas River at Wamego, Kansas
Location: Lat 39°11'54", long 96°18'19" referenced to North American Datum of 1927, in SW 1/4 NW 1/4 SE 1/4 sec.9, T.10 S., R.10 E., Pottawatomie County, Kans., hydrologic unit 10270102.

Equipment: A YSI 6600 water-quality monitor equipped with sensors for water temperature, specific conductance (SC), dissolved oxygen, pH, and turbidity was installed from August 2012 through June 2014. A Xylem YSI EXO2 water-quality monitor equipped with sensors for water temperature, SC, dissolved oxygen, pH, turbidity, and chlorophyll (fCHL) and phycocyanin fluorescence was installed during June 2014 through September 2019. The monitor was housed in a 4-inch-diameter galvanized steel pipe. Readings from the water-quality monitor were recorded every 15 minutes and transmitted by way of satellite, hourly.

Date model was created: March 31, 2020

Model calibration data period: June 11, 2014, through September 23, 2019

Model application date: June 11, 2014, onward

Model-Calibration Dataset

All data were collected using USGS protocols (Wagner and others, 2006; U.S. Geological Survey, variously dated) and are stored in the National Water Information System (U.S. Geological Survey, 2020) database and available to the public. Ordinary least squares analysis was used to develop regression models using R programming language (R Core Team, 2020). Potential explanatory variables that were evaluated individually and in combination included streamflow, water temperature, SC, dissolved oxygen, pH, turbidity, fCHL, and phycocyanin fluorescence. The maximum time span between two continuous data points used for interpolation was 2 hours (in order 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 1 hour in the continuous data record resulted in missing interpolated data). Seasonal components (sine and cosine variables) were also evaluated as potential explanatory variables.

The final selected regression model was based on 73 concurrent measurements of NO_x concentration, sensor-measured SC, fCHL, and seasonal components (sine and cosine variables) during June 11, 2014, through September 23, 2019. Samples were collected throughout the range of continuously observed hydrologic conditions. No samples had concentrations below laboratory detection limits. Summary statistics and the complete model-calibration dataset are provided below. Potential outliers were identified using the methods described in Rasmussen and others (2009). Additionally, studentized residuals from the final model were inspected for values greater than three or less than negative three. Values outside of that range were considered potential outliers and were investigated. One of the sensor-measured SC samples, from July 28, 2014, was deemed an outlier and removed from the model calibration dataset. The removed sensor-measured SC value was significantly lower than the field monitor and laboratory result during this sample. All other potential outliers were not found to have errors associated with collection, processing, or analysis and were therefore considered valid.

This model is specific to the Kansas River at Wamego, Kans., during this study period and cannot be applied to data collected from other sites on the Kansas River or data collected from other waterbodies.

Nitrate plus Nitrite Sampling Details

Cross-section samples typically were collected either from the downstream side of the bridge or instream within 100 feet of the bridge. The equal-width-increment collection method was used (although multiple vertical, single vertical, and grab samples were occasionally collected), and samples typically were composited for analysis (U.S. Geological Survey, variously dated). During July 2012 through June 2017, cross-section samples were collected every 2 weeks during March through October, once a month during

November through February, and during selected reservoir release and runoff events. During July 2017 through September 2019, cross-section samples were collected on a monthly to bimonthly basis, depending on flow conditions. A FISP US DH–81, DH–95, D–95, D–96a, or D–96 depth integrating sampler was used. Additional detail on sample collection is available in Foster and Graham (2016) and Graham and others (2018). Samples were analyzed for NO_x concentration at the USGS National Water Quality Laboratory in Lakewood, Colorado.

Model Development

Ordinary least squares regression analysis was done using R programming language (R Core Team, 2020) to relate discretely collected NO_x concentration to sensor-measured SC, fCHL, and seasonal components (sine and cosine variables). The distribution of residuals was examined for normality, and the plots of residuals (the difference between the measured and computed values) were examined for homoscedasticity (departures from zero did not change substantially over the range of computed values). Previously published explanatory variables were also strongly considered for continuity.

SC, fCHL, and seasonal components (sine and cosine variables) were selected as good surrogates for NO_x based on residual plots, coefficient of determination (R^2), and model standard percentage error. Values for all the aforementioned statistics were computed and are included below along with all relevant sample data and additional statistical information.

Model Summary

The following is a summary of final regression analysis for NO_x concentration at USGS site 06887500:

NO_x concentration-based model:

$$NO_x = -0.00065 \times SC - 0.0995 \times fCHL + 0.163 \times \sin(2\pi D) - 0.0823 \times \cos(2\pi D) + 1.77$$

where

NO_x = nitrate plus nitrite concentration, in milligrams per liter;

SC = specific conductance, in microsiemens per centimeter at 25 degrees Celsius;

fCHL = chlorophyll fluorescence, in relative fluorescence units;

sin = sine;

D = date, in decimal years; and

cos = cosine.

SC makes physical and statistical sense as an explanatory variable for NO_x because of its correlation with charged ionic species, such as NO_x (Hem, 1992). fCHL also was used as an explanatory variable in the model previously published by Foster and Graham (2016). Seasonal components (sine and cosine variables) are a logical addition to the model because NO_x increases are often associated with precipitation runoff events in the Kansas River.

Previous Models

Start Year	End Year	Model Equation	Reference
2012	2019	NO_x = -0.00102SC - 0.0176fCHL + 1.85	Foster and Graham (2016)
1999	2003	NO_x = 1.09logQ - 0.397logWT - 2.53 where Q = streamflow, in cubic feet per second; and, WT = water temperature, in degrees Celsius	Rasmussen and others (2005)

Model Statistics, Data, and Plots

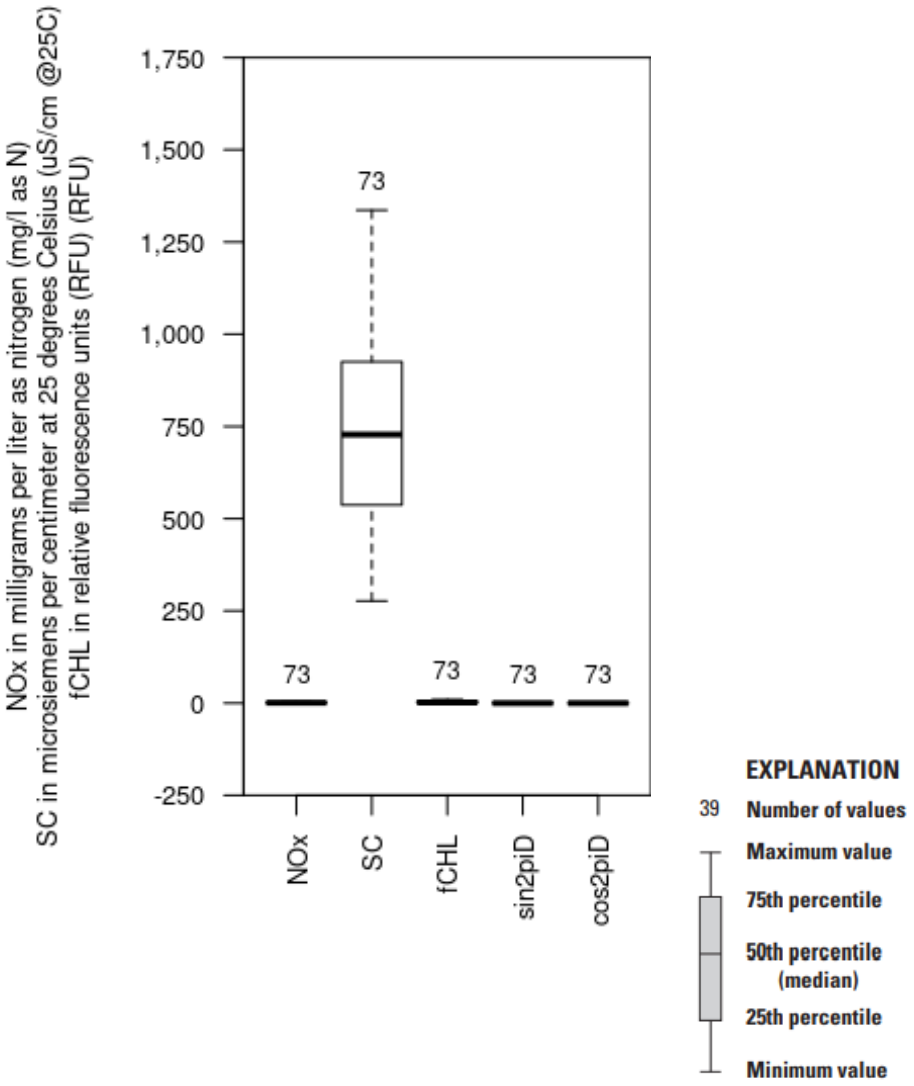
Model

NOx = - 0.00065 * SC - 0.0995 * fCHL + 0.163 * sin2piD - 0.0823 * cos2piD + 1.77

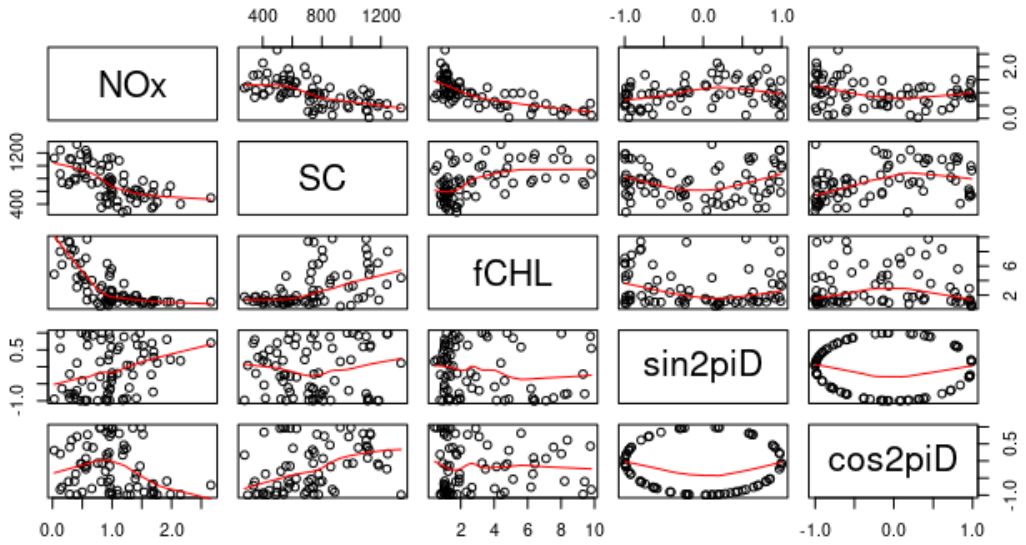
Variable Summary Statistics

	NOx	SC	fCHL	sin2piD	cos2piD
Minimum	0.030	277	0.44	-0.9980	-1.000
1st Quartile	0.580	537	1.15	-0.7910	-0.809
Median	0.960	728	1.77	-0.1970	-0.234
Mean	0.988	737	3.06	-0.0876	-0.175
3rd Quartile	1.310	925	4.62	0.5950	0.390
Maximum	2.660	1340	9.80	0.9980	0.988

Box Plots



Exploratory Plots



Red line shows the locally weighted scatterplot smoothing (LOWESS).

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

Basic Model Statistics

```
Number of Observations                              73
Standard error (RMSE)                            0.326
Average Model standard percentage error (MSPE)      33
Coefficient of determination (R²)                0.636
Adjusted Coefficient of Determination (Adj. R²) 0.615

Variance Inflation Factors (VIF)
     SC    fCHL sin2piD cos2piD
   1.71    1.50    1.05    1.25
```

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	1.77000	0.128000	13.80	2.02e-21
SC	-0.00065	0.000188	-3.45	9.78e-04
fCHL	-0.09950	0.017900	-5.56	4.82e-07
sin2piD	0.16300	0.055500	2.94	4.45e-03
cos2piD	-0.08230	0.062500	-1.32	1.93e-01

Correlation Matrix

	Intercept	SC	fCHL	sin2piD	cos2piD
Intercept	1.000	-0.886	0.217	0.136	0.456
SC	-0.886	1.000	-0.567	-0.162	-0.442
fCHL	0.217	-0.567	1.000	0.204	0.266
sin2piD	0.136	-0.162	0.204	1.000	0.116
cos2piD	0.456	-0.442	0.266	0.116	1.000

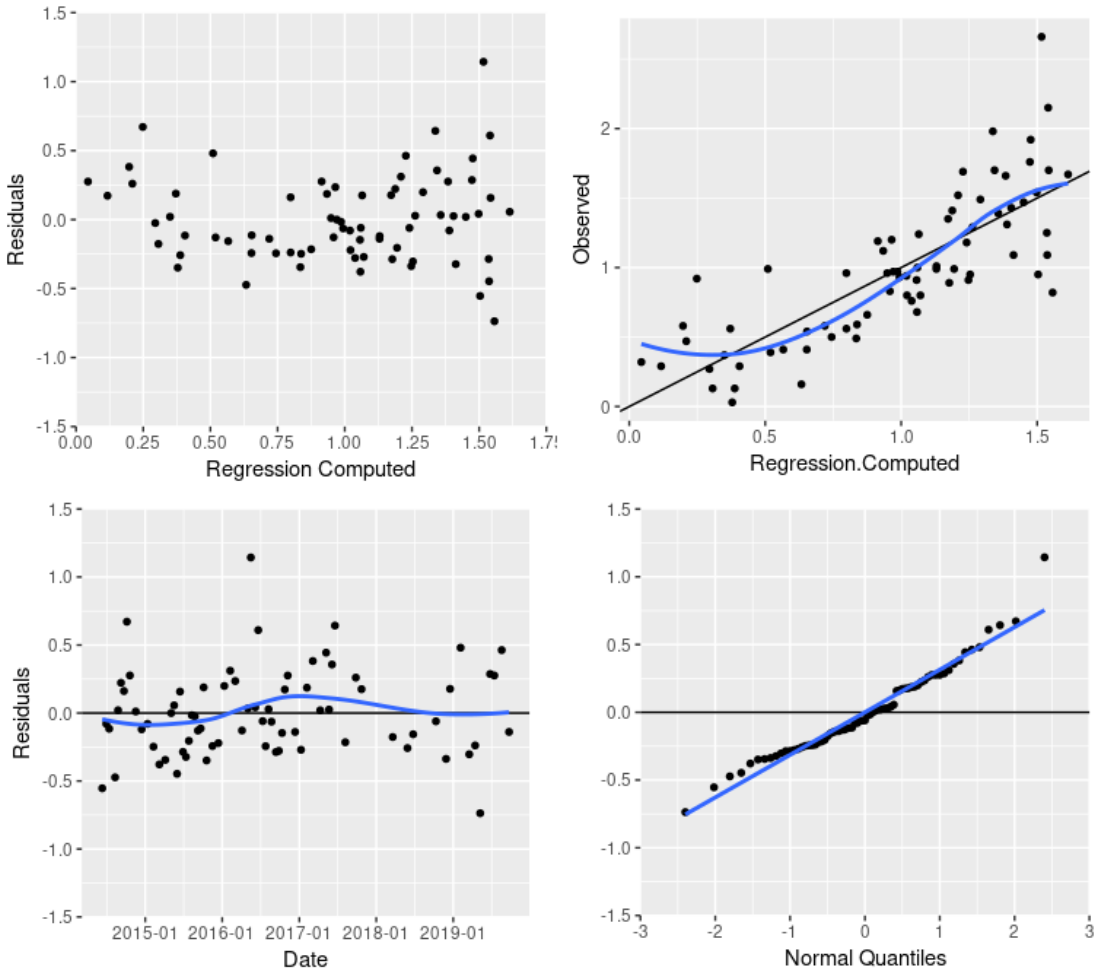
Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.205	0.363	0.523

Flagged Observations

	NOx	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
201410060900	0.92	0.249	0.671	2.13	2.19	0.0678	0.0661	0.591
201605160850	2.66	1.520	1.140	3.59	3.96	0.0466	0.1260	0.875
201703060950	0.58	0.198	0.382	1.28	1.29	0.1660	0.0658	0.576
201905091520	0.82	1.560	-0.737	-2.34	-2.43	0.0710	0.0839	-0.671

Statistical Plots

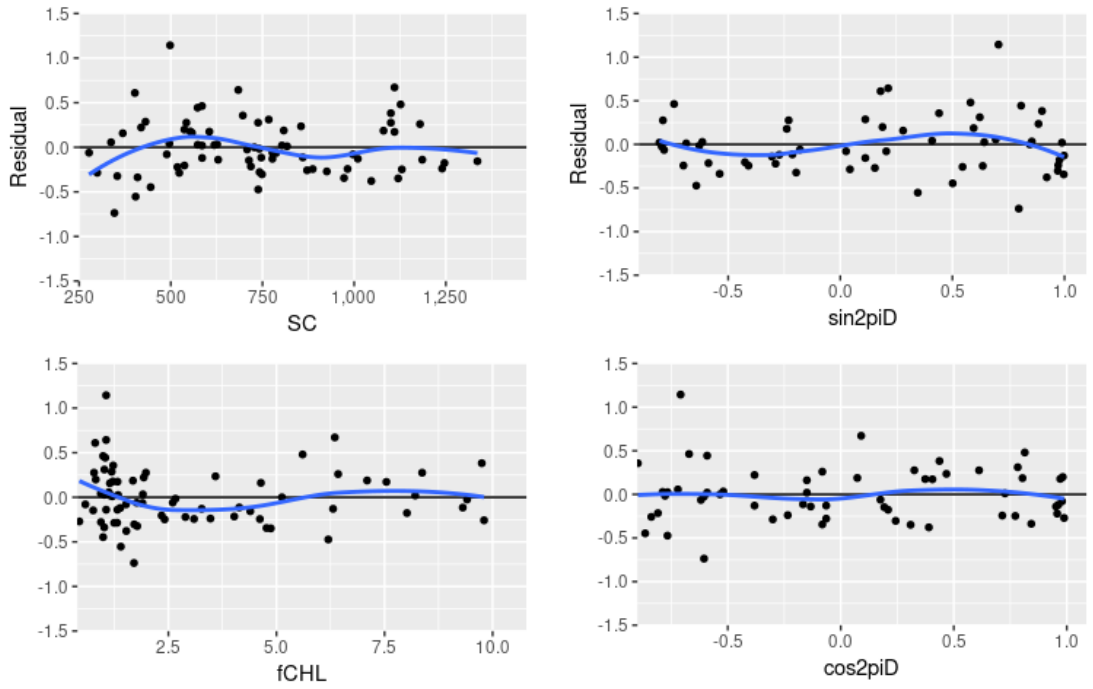


First row (left): Residual NOx related to regression computed NOx.

First row (right): Observed NOx related regression computed NOx with local polynomial regression fitting, or locally estimated scatterplot smoothing (LOESS), indicated by the blue line.

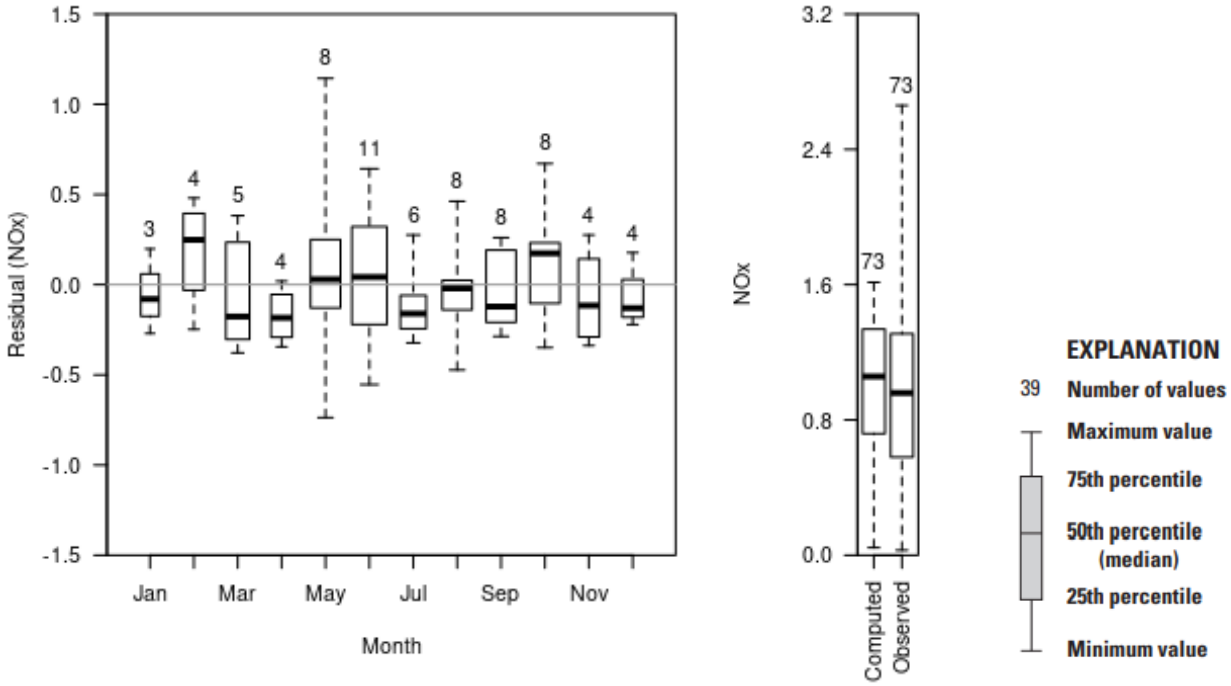
Second row (left): Residual NOx related to date with LOESS, indicated by the blue line.

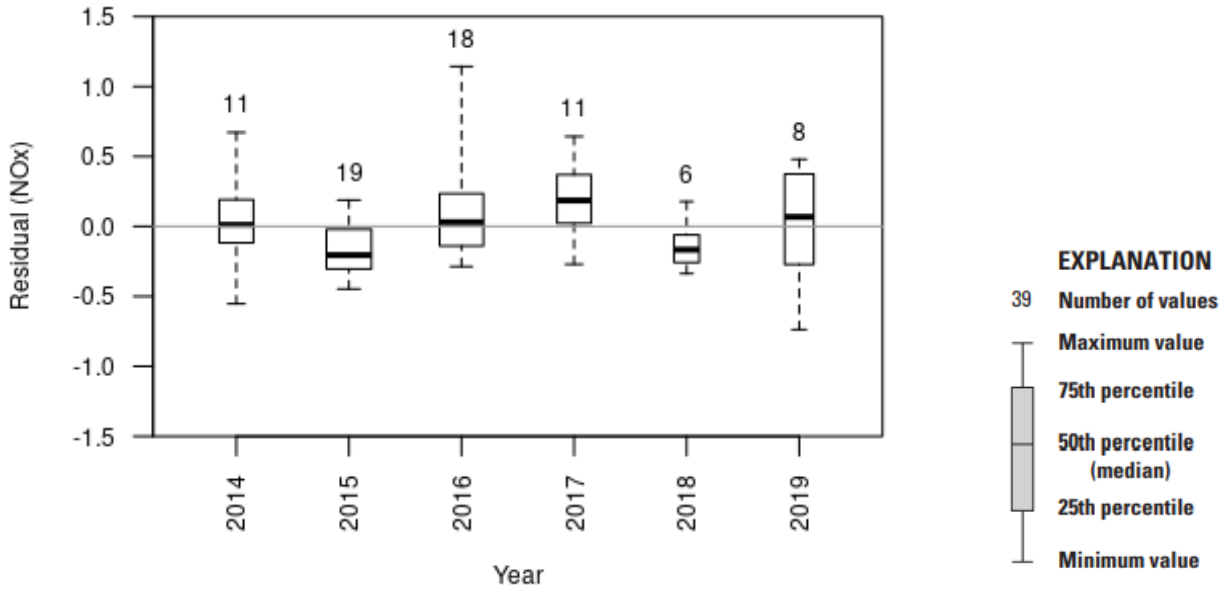
Second row (right): Residual NOx related to the corresponding normal quantile of the residual with simple linear regression, indicated by the blue line.



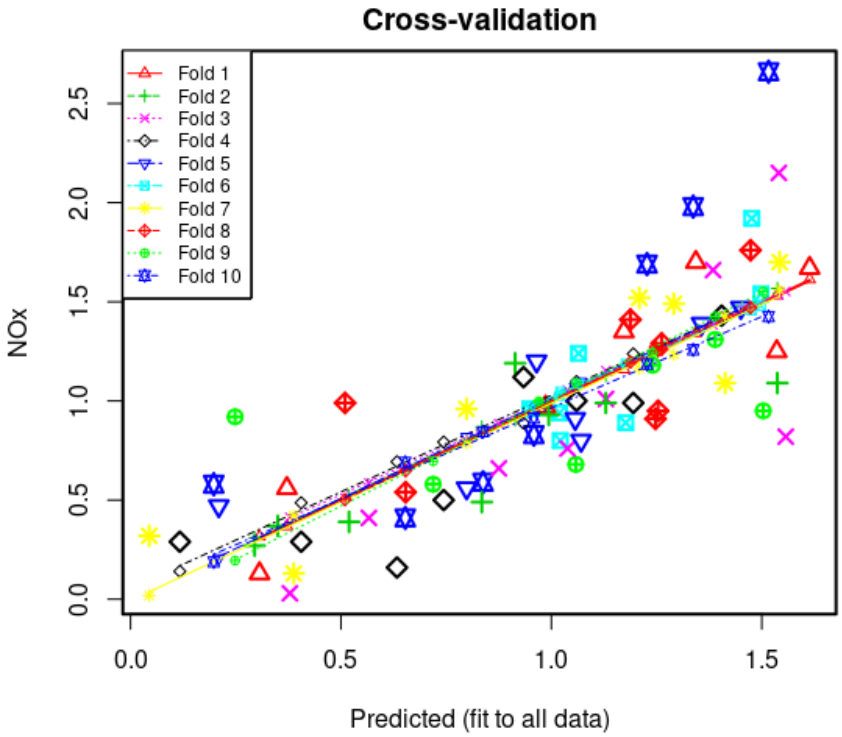
First row: Residual NOx related to SC (left) and sin2piD (right) with LOESS, indicated by the blue line.

Second row: Residual NOx related to fCHL (left) and cos2piD (right) with LOESS, indicated by the blue line.





Cross-Validation



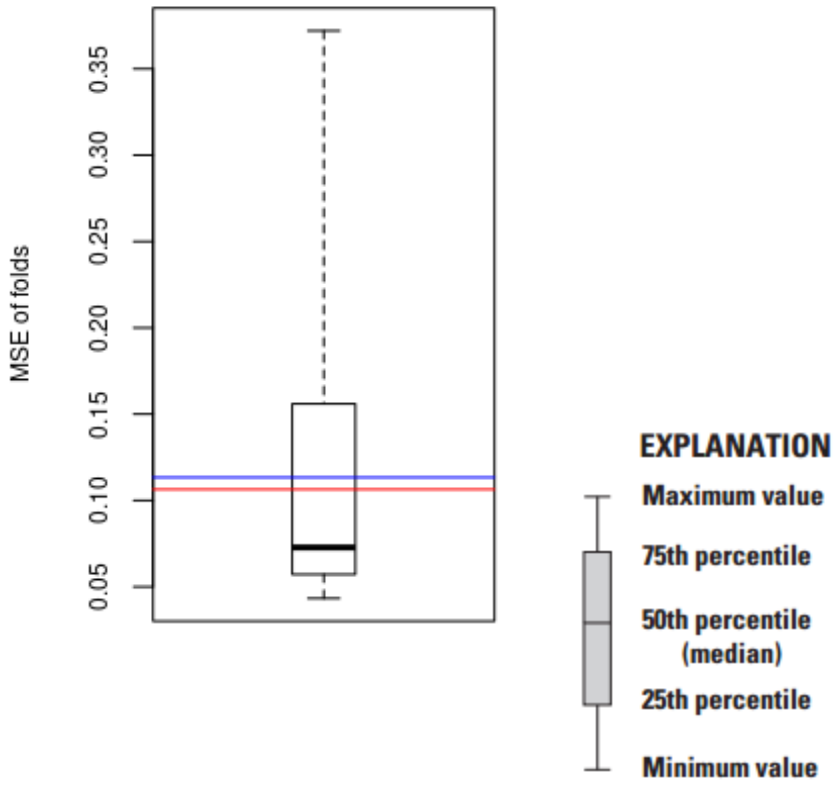

```
Fold - equal partition of the data (10 percent of the data).

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.


              Minimum MSE of folds:  0.0433
                 Mean MSE of folds:  0.1130
               Median MSE of folds:  0.0728
              Maximum MSE of folds:  0.3720
 (Mean MSE of folds) / (Model MSE):  1.0700
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Model-Calibration Dataset

	Date	NOx	SC	fCHL	sin2piD	cos2piD	Computed NOx	Residual	Normal Quantiles	Censored Values
1	2014-06-11	0.95	404	1.4	0.346	-0.938	1.5	-0.553	-2.02	--
2	2014-06-30	1.31	489	1.52	0.0258	-1	1.39	-0.0792	-0.103	--
3	2014-07-14	0.29	747	9.31	-0.214	-0.977	0.405	-0.115	-0.207	--
4	2014-08-11	0.16	738	6.2	-0.642	-0.767	0.633	-0.473	-1.81	--
5	2014-08-25	0.37	803	8.21	-0.806	-0.591	0.35	0.0202	0.207	--
6	2014-09-08	1.41	419	1.93	-0.924	-0.382	1.19	0.222	0.75	--
7	2014-09-22	0.96	558	4.64	-0.989	-0.15	0.798	0.162	0.461	--
8	2014-10-06	0.92	1110	6.35	-0.996	0.0903	0.249	0.671	2.02	--
9	2014-10-20	1.19	738	1.98	-0.946	0.325	0.914	0.276	1	--
10	2014-11-17	0.96	817	1.21	-0.687	0.727	0.949	0.0112	0.137	--
11	2014-12-15	1.01	585	1.39	-0.272	0.962	1.13	-0.12	-0.242	--
12	2015-01-12	0.94	997	0.58	0.205	0.979	1.02	-0.0795	-0.137	--
13	2015-02-09	0.59	1130	2.41	0.635	0.772	0.838	-0.248	-0.796	--
14	2015-03-09	0.68	1050	1.53	0.921	0.39	1.06	-0.378	-1.53	--
15	2015-04-06	0.49	972	4.77	0.997	-0.0817	0.835	-0.345	-1.34	--
16	2015-05-04	0.97	728	5.13	0.845	-0.534	0.97	-0.000498	0.103	--
17	2015-05-18	1.67	337	1.12	0.693	-0.721	1.61	0.0562	0.386	--
18	2015-06-01	1.09	445	0.99	0.501	-0.865	1.54	-0.447	-1.65	--
19	2015-06-15	1.7	369	1.15	0.28	-0.96	1.54	0.157	0.423	--
20	2015-06-29	1.25	299	1.31	0.043	-0.999	1.54	-0.286	-1	--

21	2015-07-13	1.09	354	1.78	-0.197	-0.98	1.41	-0.323	-1.19	--
22	2015-07-27	0.99	537	2.34	-0.425	-0.905	1.19	-0.205	-0.539	--
23	2015-08-10	0.97	741	2.66	-0.629	-0.778	0.986	-0.0163	0.0685	--
24	2015-08-24	0.27	708	9.41	-0.796	-0.605	0.295	-0.0249	0.0343	--
25	2015-09-08	0.39	777	6.31	-0.924	-0.382	0.519	-0.129	-0.313	--
26	2015-09-21	0.54	860	4.14	-0.986	-0.167	0.654	-0.114	-0.172	--
27	2015-10-05	0.56	808	7.1	-0.997	0.0731	0.372	0.188	0.662	--
28	2015-10-19	0.03	1120	4.87	-0.951	0.309	0.379	-0.349	-1.43	--
29	2015-11-16	0.41	982	3.09	-0.699	0.715	0.653	-0.243	-0.705	--
30	2015-12-14	0.8	518	2.89	-0.288	0.957	1.02	-0.221	-0.62	--
31	2016-01-11	1.49	537	0.82	0.188	0.982	1.29	0.199	0.705	--
32	2016-02-08	1.52	767	1.02	0.622	0.783	1.21	0.311	1.12	--
33	2016-03-03	1.2	855	3.59	0.884	0.467	0.965	0.235	0.796	--
34	2016-04-04	0.83	1010	3.27	0.998	-0.0645	0.958	-0.128	-0.277	--
35	2016-05-02	1.39	626	1.91	0.854	-0.52	1.36	0.033	0.313	--
36	2016-05-16	2.66	498	1.06	0.706	-0.709	1.52	1.14	2.4	--
37	2016-06-06	1.54	496	0.937	0.409	-0.912	1.5	0.0419	0.349	--
38	2016-06-20	2.15	402	0.807	0.18	-0.984	1.54	0.61	1.65	--
39	2016-07-11	1	777	2.6	-0.18	-0.984	1.06	-0.0597	0	--
40	2016-07-25	0.5	887	4.62	-0.409	-0.912	0.744	-0.244	-0.75	--
41	2016-08-08	1.29	573	1.02	-0.615	-0.788	1.26	0.028	0.277	--
42	2016-08-22	0.93	787	1.9	-0.786	-0.619	0.994	-0.0639	-0.0685	--
43	2016-09-12	0.89	523	1.25	-0.954	-0.301	1.18	-0.287	-1.06	--
44	2016-09-26	0.76	742	0.94	-0.998	-0.0645	1.04	-0.278	-0.946	--
45	2016-10-11	0.91	713	0.76	-0.981	0.192	1.06	-0.147	-0.423	--
46	2016-10-24	0.29	1110	7.54	-0.914	0.405	0.117	0.173	0.499	--
47	2016-11-07	0.32	1100	8.37	-0.791	0.612	0.0445	0.275	0.894	--
48	2016-12-12	0.99	629	1.06	-0.305	0.952	1.13	-0.139	-0.386	--
49	2017-01-09	0.8	925	0.44	0.154	0.988	1.07	-0.271	-0.894	--
50	2017-02-06	1.12	1080	1.68	0.595	0.804	0.934	0.186	0.62	--
51	2017-03-06	0.58	1100	9.75	0.9	0.437	0.198	0.382	1.26	--
52	2017-04-10	1.47	585	1.16	0.989	-0.15	1.45	0.0197	0.172	--
53	2017-05-08	1.92	573	1.04	0.806	-0.591	1.48	0.444	1.34	--
54	2017-05-22	1.43	619	1.34	0.642	-0.767	1.4	0.0253	0.242	--
55	2017-06-05	1.7	696	1.22	0.441	-0.898	1.34	0.357	1.19	--
56	2017-06-19	1.98	684	1.06	0.214	-0.977	1.34	0.643	1.81	--
57	2017-08-07	0.66	718	4.02	-0.588	-0.809	0.875	-0.215	-0.579	--
58	2017-09-26	0.47	1180	6.43	-0.997	-0.0817	0.21	0.26	0.844	--
59	2017-10-23	1.24	605	1.32	-0.928	0.374	1.06	0.175	0.539	--
60	2018-03-20	0.13	1250	8.02	0.978	0.209	0.307	-0.177	-0.499	--
61	2018-05-29	0.13	872	9.8	0.545	-0.838	0.388	-0.258	-0.844	--
62	2018-06-25	0.41	1340	4.39	0.112	-0.994	0.566	-0.156	-0.461	--
63	2018-10-11	1.18	277	1.77	-0.984	0.176	1.24	-0.0607	-0.0343	--
64	2018-11-28	0.91	409	1.02	-0.538	0.843	1.25	-0.337	-1.26	--
65	2018-12-17	1.35	553	1.22	-0.239	0.971	1.17	0.178	0.579	--
66	2019-02-05	0.99	1130	5.61	0.581	0.814	0.51	0.48	1.53	--
67	2019-03-18	0.95	749	1.71	0.97	0.243	1.25	-0.304	-1.12	--
68	2019-04-15	0.56	1240	3.47	0.972	-0.234	0.799	-0.239	-0.662	--
69	2019-05-09	0.82	346	1.7	0.796	-0.605	1.56	-0.737	-2.4	--
70	2019-06-25	1.76	431	1.19	0.112	-0.994	1.47	0.287	1.06	--
71	2019-07-15	1.66	542	0.78	-0.23	-0.973	1.38	0.276	0.946	--
72	2019-08-19	1.69	585	0.99	-0.741	-0.671	1.23	0.463	1.43	--
73	2019-09-23	0.58	1190	1.32	-0.991	-0.133	0.719	-0.139	-0.349	--

Definitions

D: Date, in decimal years.

Cook's D: Cook's distance (Helsel and others, 2020).

DFFITS: Difference in fits statistic (Helsel and others, 2020).

E.vars: Explanatory variables.

fCHL: Chlorophyll fluorescence, in relative fluorescence units (32320).

Leverage: An outlier's measure in the x direction (Helsel and others, 2020).

LOESS: Local polynomial regression fitting, or locally estimated scatterplot smoothing (Helsel and others, 2020).

LOWESS: Locally weighted scatterplot smoothing (Cleveland, 1979; Helsel and others, 2020).

MSE: Model standard error (Helsel and others, 2020).

MSPE: Model standard percentage error (Helsel and others, 2020).

NOx: Inorganic nitrogen (nitrate and nitrite), in milligrams per liter as nitrogen (00631).

Probability(>|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).

SC: Specific conductance, in microsiemens per centimeter at 25 degrees Celsius (00095).

t value: Student's t value; the coefficient divided by its associated standard error (Helsel and others, 2020).

References Cited

Cleveland, W.S., 1979, Robust locally weighted regression and smoothing scatterplots: Journal of the American Statistical Association, v. 74, no. 368, p. 829-836.

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 July 2020 at https://doi.org/10.3133/ofr20161040.

Graham, J.L., Foster, G.M., Williams, T.J., Mahoney, M.D., May, M.R., and Loftin, K.A., 2018, Water-quality conditions with an emphasis on cyanobacteria and associated toxins and taste-and-odor compounds in the Kansas River, Kansas, July 2012 through September 2016: U.S. Geological Survey Scientific Investigations Report 2018–5089, 55 p. [Also available at https://doi.org/10.3133/sir20185089.]

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, ver. 1.1.]

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. [Also available at https://doi.org/10.3133/wsp2254.]

R Core Team, 2020, R—A language and environment for statistical computing, version 4.0.3: Vienna, Austria, R Foundation for Statistical Computing, accessed December 2020 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, 53 p. [Also available at https://doi.org/10.3133/tm3C4.]

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.]

U.S. Geological Survey, 2020, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 2020 at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey, variously dated, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1–A9 [variously paged], accessed July 2020 at https://water.usgs.gov/owq/FieldManual/.

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.] [Supersedes USGS Water-Resources Investigations Report 2000–4252.]