Appendix 4.5. Model Archive Summary for Atrazine Concentration at U.S. Geological Survey station 07143672; Little Arkansas River at Highway 50 near Halstead, Kansas, during March 2017 through August 2021

This model archive summary summarizes the atrazine model developed to compute hourly or daily atrazine. Model development methods follow U.S. Geological Survey (USGS) guidance from Office of Surface Water/Office of Water Quality Technical Memoranda and USGS Techniques and Methods, book 3, chap. C4 (Rasmussen and others, 2009).

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: 07143672

Site Name: Little Arkansas River at Highway 50 near Halstead, Kansas

Location: Latitude 38°01'42.71", longitude 97°32'25.95" referenced to North American Datum of 1983, in NE 1/4 NE 1/4 sec.28, T.23 S., R.2 W., Harvey County, Kansas, hydrologic unit 11030012.

Equipment: A Sutron Satlink II High Data Rate Collection Platform and a Design Analysis Water Log H350/355 nonsubmersible pressure transducer transfers real-time stage and water-quality data via satellite. The primary reference gage is a Type-A wire-weight gage located on the downstream bridge guardrail. Check-bar elevation is 33.396 feet. The orifice tube is enclosed in 1.25-inch steel conduit trenched into the ground down to the edge of water, where the orifice emerges from the bank and culminates in a 2-inch open-end orifice tethered to a steel fencepost near the left edge of water. Gage height was measured during May 1998 through December 2019. A YSI 6600 water-quality monitor equipped with water temperature, specific conductance, pH, dissolved oxygen, and turbidity (a YSI Model 6026 [December 1998 through December 2006] and YSI Model 6136 [July 2004 through December 2017]) sensors collected data during May 1998 through December 2017. A YSI EXO2 water-quality monitor equipped with water temperature, specific conductance, pH, dissolved oxygen, turbidity, and fluorescent dissolved organic matter sensors collected data during January 2017 through December 2021. A Hach Nitratax monitor collected nitrate data during February 2017 through December 2021.

Date model was developed: June 1, 2022

Model calibration data period: March 30, 2017 through August 23, 2021

Model Data

All data were collected using USGS protocols (U.S. Geological Survey, variously dated; Wagner and others, 2006; Sauer and Turnipseed, 2010; Turnipseed and Sauer, 2010) and are stored in the National Water Information System (NWIS) database (U.S. Geological Survey, 2022). Explanatory variables were evaluated individually and in combination. Potential explanatory variables included streamflow, water temperature, specific conductance, pH, dissolved oxygen, YSI EXO2 turbidity, nitrate, and fluorescent dissolved organic matter. Seasonal components (sine and cosine variables) also were evaluated as explanatory variables.

The regression model is based on 37 concomitant values of discretely collected atrazine and continuously measured turbidity during March 2017 through August 2021. Discrete samples were collected over a range of streamflow and turbidity conditions. No samples had concentrations that were below laboratory detection limits. Summary statistics and the complete model-calibration dataset are provided below. Outliers and influential points were identified using studentized residuals, DFFITS, Cook's D (Cook, 1977), and leverage. All samples were retained in the dataset.

Atrazine

Discrete samples were collected from the downstream side of the bridge or instream within 50 feet of the bridge using equal-width-increment, multi-vertical, single vertical or grab-dip methods following U.S. Geological Survey (variously dated) and Rasmussen and others (2014). Discrete samples were collected on a semifixed to event-based schedule ranging from 7 to 8 samples per year with a depth-integrating FISP US DH–95, D–95, or DH–81 with a Teflon bottle, cap and nozzle or a grab sample with a Teflon bottle depending on sample location. Samples were analyzed for atrazine by the National Water Quality Laboratory according to standard methods (American Public Health Association and others, 1995).

Continuous Data

Concomitant turbidity values were time interpolated. If no concomitant continuous data were available within 2 hours of sample collection, the sample was not included in the dataset.

Model Development

Ordinary least squares regression analysis was done using R (version 4.0.0) programming language (R Core Team, 2020) to relate discretely collected atrazine to turbidity and other continuously measured data. The distribution of residuals was examined for normality and plots of residuals (the difference between the measured and model-calculated values) compared to model-computed atrazine were examined for homoscedasticity (departures from zero did not change substantially over the range of model-calculated values).

Turbidity and seasonal components were selected as the best predictors of atrazine based on residual plots, high coefficient of determination (R^2), and low model standard percentage error (MSPE). Turbidity was positively correlated with atrazine.

Model Summary

Summary of final atrazine regression analysis at USGS station 07143672:

Atrazine-based model:

$$\log_{10}(ATR) = 0.721 \times \log_{10}(TBY) + \ 0.405 \times \sin(2\pi \mathrm{D}) - 0.38 \times \cos(2\pi \mathrm{D}) - 1.7$$

where,

$\log_{10}$ = logarithm base 10;
ATR = atrazine, in micrograms per liter (µg/L);
TBY = turbidity, in formazin nephelometric units (FNU); and
D = date in decimal years

The log-transformed model may be retransformed to original units so that ATR can be calculated directly. The retransformation introduces a bias in the calculated constituent. This bias may be corrected using Duan's bias correction factor (BCF; Duan, 1983). For this model, the calculated BCF is 1.9.

Model Statistics, Data, and Plots

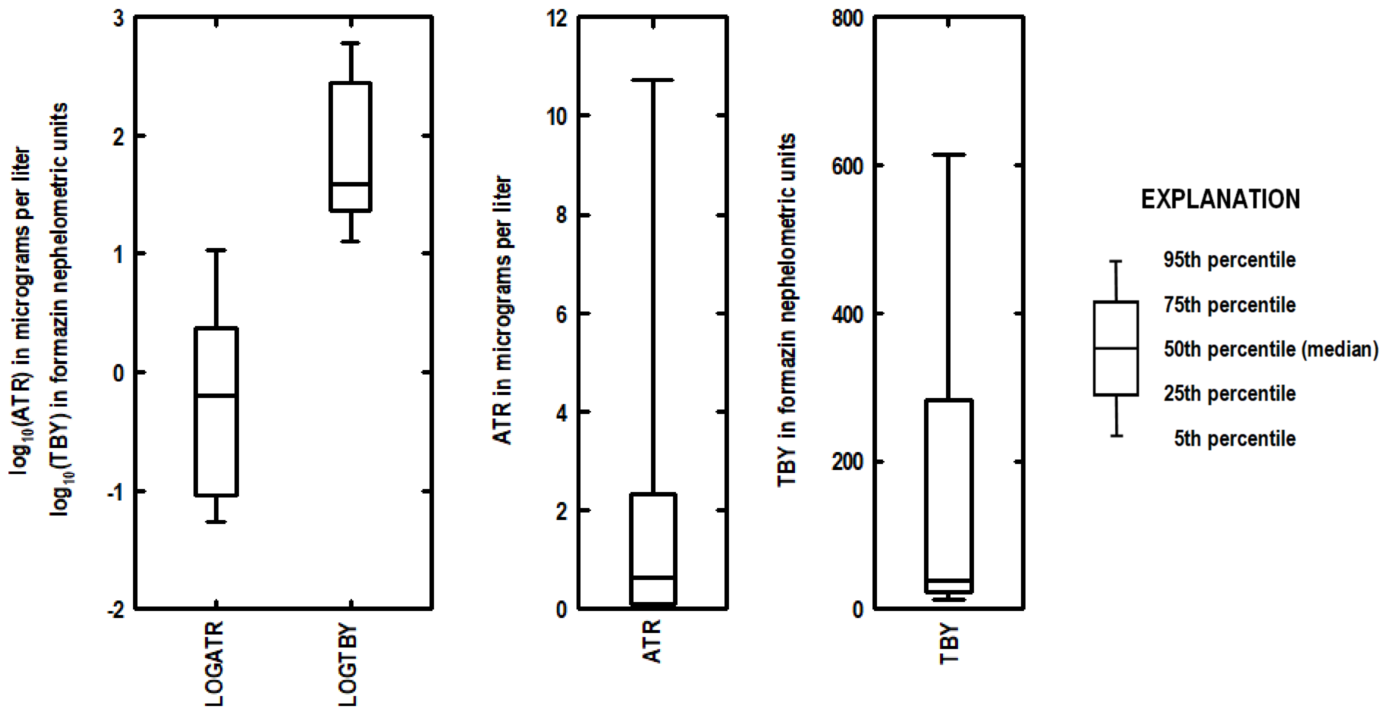
Model

LOGATR = + 0.721 * LOGTBY + 0.405 * SIN2PID - 0.38 * COS2PID – 1.7

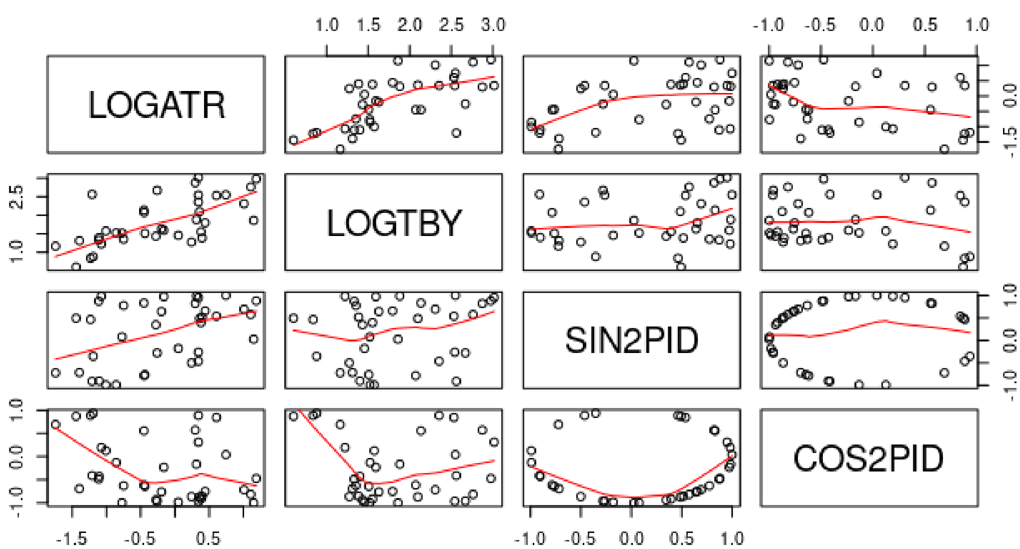
Variable Summary Statistics

```
              LOGATR     ATR LOGTBY SIN2PID COS2PID    TBY
Minimum       -1.740  0.0181  0.602  -0.992  -1.000    4.0
1st Quartile  -1.010  0.0984  1.390  -0.464  -0.857   24.3
Median        -0.196  0.6370  1.590   0.464  -0.490   39.2
Mean          -0.211  2.4400  1.830   0.166  -0.283  181.0
3rd Quartile   0.360  2.2900  2.350   0.775   0.192  226.0
Maximum        1.190 15.4000  3.020   0.999   0.935 1040.0
```

Box Plots



Exploratory Plots



Basic Model Statistics

```
Number of Observations                               37
Standard error (RMSE)                              0.53
Average Model standard percentage error (MSPE)      154
Coefficient of determination (R²)                 0.619
Adjusted Coefficient of Determination (Adj. R²) 0.584
Bias Correction Factor (BCF)                        1.9
```

```
Variance Inflation Factors (VIF)
 LOGTBY SIN2PID COS2PID
   1.06    1.06    1.02
```

Explanatory Variables

```
            Coefficients Standard Error t value Pr(>|t|)
(Intercept)       -1.700          0.269   -6.32 3.75e-07
LOGTBY             0.721          0.142    5.07 1.51e-05
SIN2PID            0.405          0.132    3.07 4.23e-03
COS2PID           -0.380          0.134   -2.84 7.66e-03
```

Correlation Matrix

```
          Intercept LOGTBY SIN2PID COS2PID
Intercept    1.0000 -0.932  0.1130  0.0416
LOGTBY      -0.9320  1.000 -0.2160  0.1110
SIN2PID      0.1130 -0.216  1.0000 -0.0994
COS2PID      0.0416  0.111 -0.0994  1.0000
```

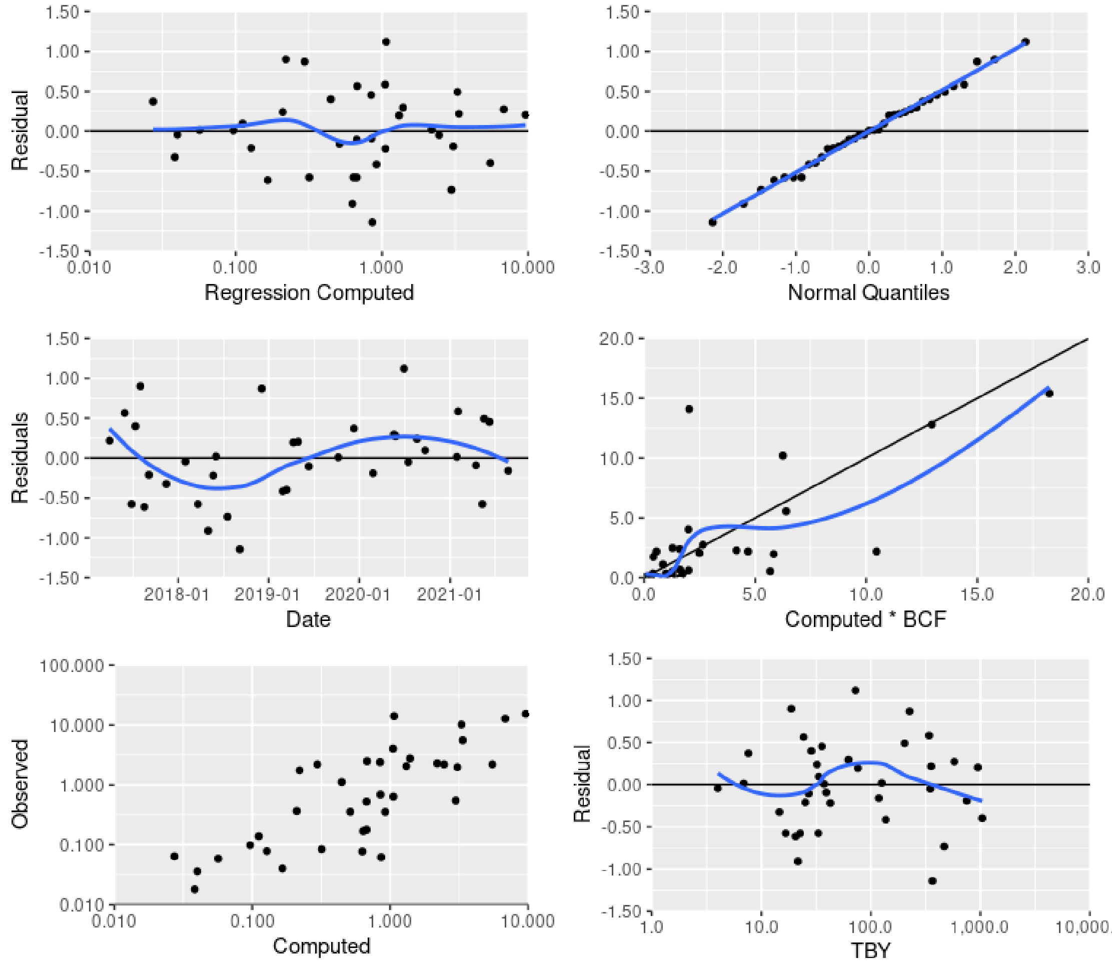
Outlier Test Criteria

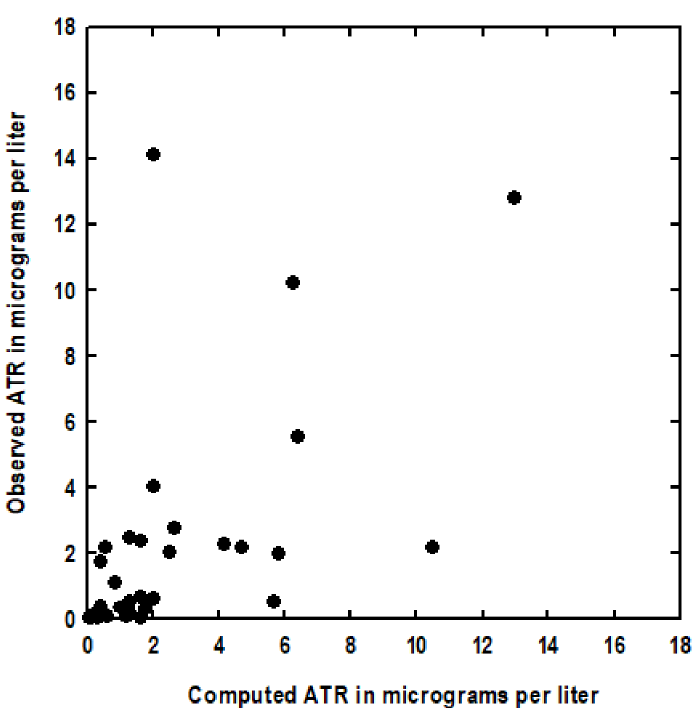
```
Leverage Cook's D   DFFITS
   0.324    0.316    0.658
```

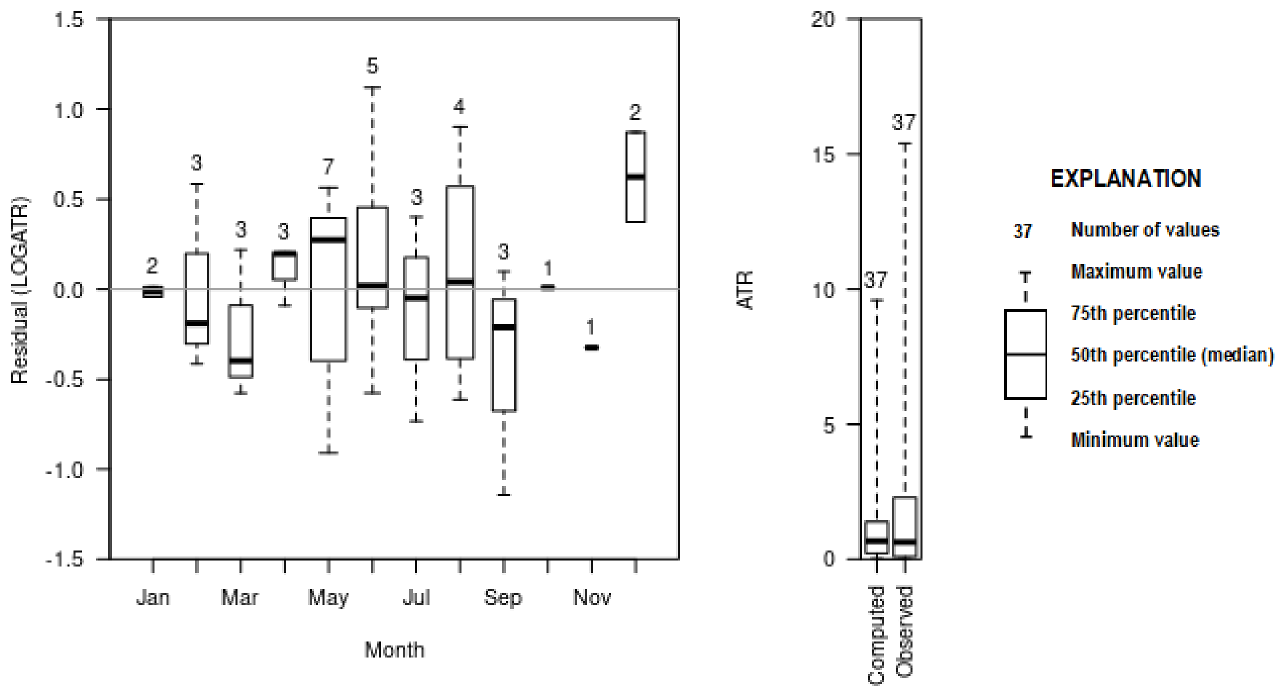
Flagged Observations

		LOGATR	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
9/6/2018	10:00	-1.21	-0.0661	-1.14	-2.35	-2.54	0.159	0.261	-1.1
12/3/2018	11:05	0.342	-0.529	0.871	1.82	1.89	0.187	0.191	0.909

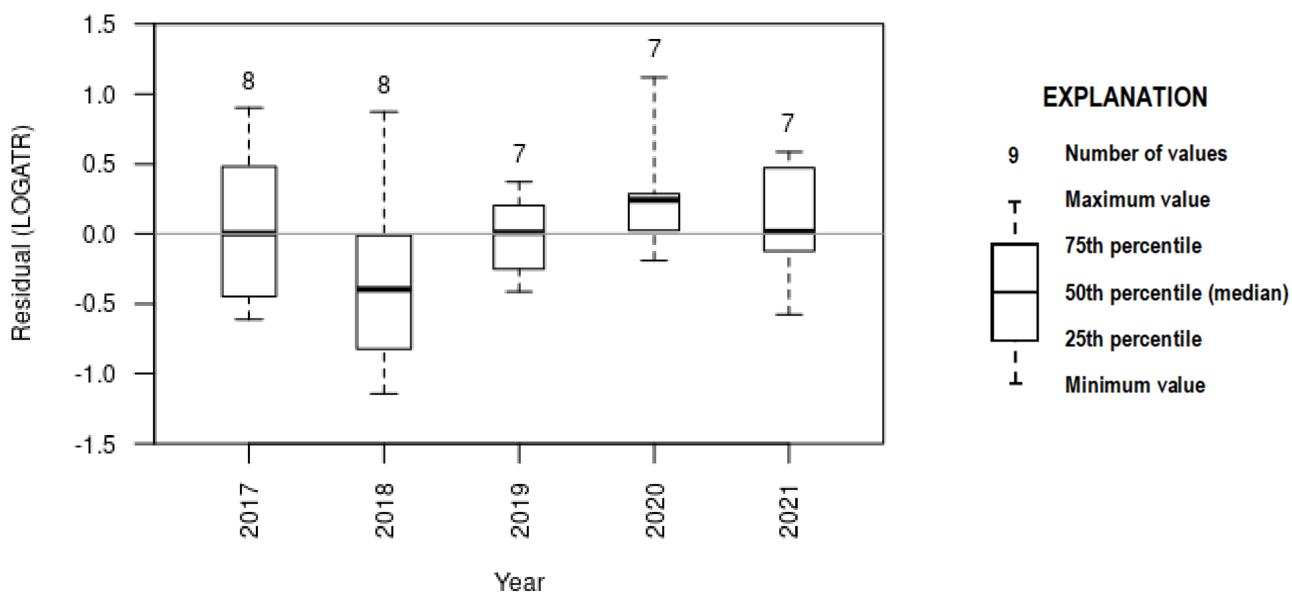
Statistical Plots





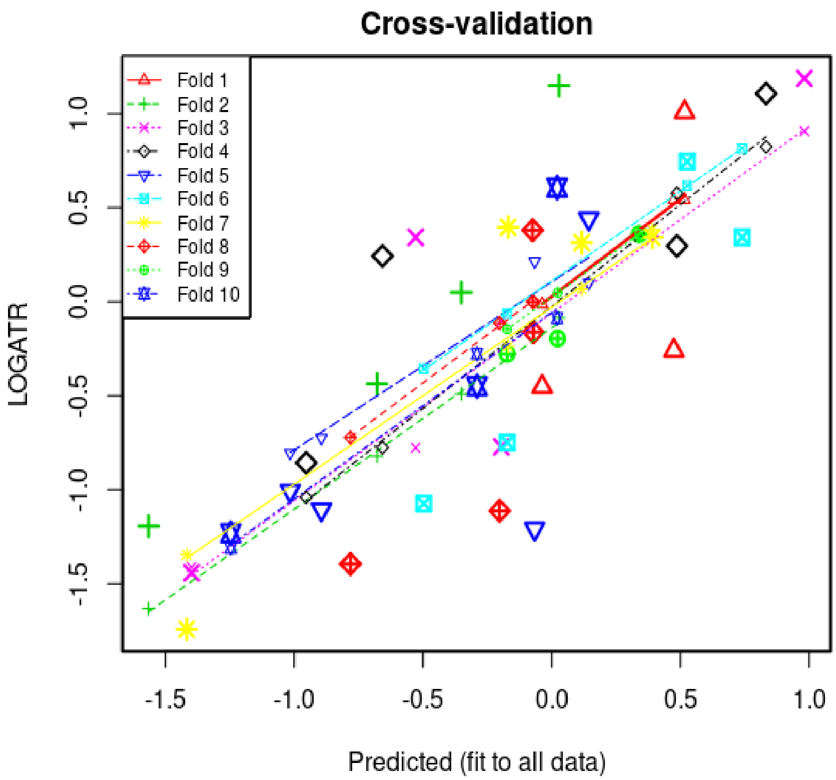


Residual (LOGATR)
Month
Jan
Mar
May
Jul
Sep
Nov
ATR
Computed
Observed
EXPLANATION
37
Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value

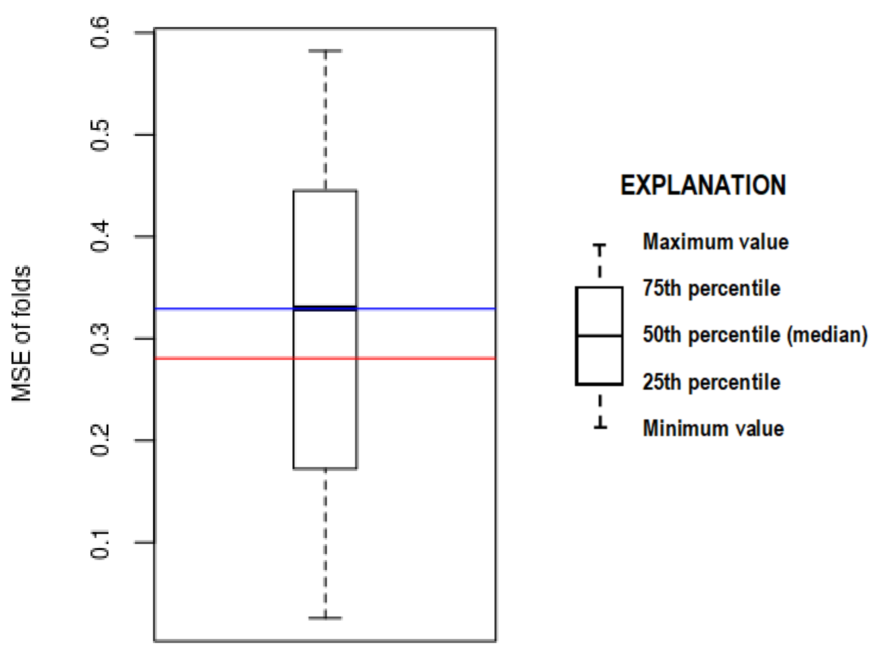


Residual (LOGATR)
Year
2017
2018
2019
2020
2021
EXPLANATION
9
Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value

Cross Validation




```
              Minimum MSE of folds:  0.0255
                 Mean MSE of folds:  0.3290
               Median MSE of folds:  0.3300
              Maximum MSE of folds:  0.5820
(Mean MSE of folds) / (Model MSE):  1.1700
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Model-Calibration Dataset

	Date	LOGATR	LOGTBY	ATR	TBY	Computed LOGATR	Computed ATR	Residual	Normal Quantiles
1	3/30/2017	0.745	2.55	5.56	355	0.526	6.39	0.219	0.416
2	5/30/2017	0.396	1.39	2.49	24.3	-0.169	1.29	0.565	1.16
3	6/27/2017	-0.772	1.52	0.169	33.2	-0.196	1.21	-0.576	-0.925
4	7/12/2017	0.0492	1.46	1.12	28.6	-0.351	0.847	0.4	0.826
5	8/1/2017	0.243	1.27	1.75	18.8	-0.657	0.418	0.901	1.72
6	8/17/2017	-1.39	1.31	0.0403	20.5	-0.782	0.314	-0.613	-1.3
7	9/5/2017	-1.11	1.4	0.0783	25.2	-0.895	0.242	-0.211	-0.491
8	11/14/2017	-1.74	1.16	0.0181	14.6	-1.42	0.0726	-0.324	-0.649
9	1/30/2018	-1.44	0.602	0.0361	4	-1.4	0.0758	-0.0432	-0.0674
10	3/21/2018	-1.07	1.22	0.0843	16.7	-0.497	0.605	-0.577	-1.03
11	5/1/2018	-1.11	1.33	0.0772	21.5	-0.202	1.19	-0.91	-1.72
12	5/22/2018	-0.196	1.63	0.637	42.6	0.0234	2.01	-0.219	-0.568
13	6/2/2018	0.36	2.1	2.29	126	0.34	4.16	0.0199	0.135
14	7/18/2018	-0.26	2.67	0.55	468	0.475	5.67	-0.734	-1.48
15	9/6/2018	-1.21	2.56	0.0619	365	-0.0661	1.63	-1.14	-2.14
16	12/3/2018	0.342	2.35	2.2	226	-0.529	0.563	0.871	1.48
17	2/26/2019	-0.452	2.14	0.353	137	-0.0369	1.75	-0.415	-0.826
18	3/14/2019	0.342	3.02	2.2	1040	0.741	10.5	-0.398	-0.735
19	4/10/2019	0.314	1.88	2.06	76.1	0.117	2.49	0.197	0.273
20	4/29/2019	1.19	2.98	15.4	950	0.982	18.2	0.206	0.344
21	6/11/2019	-0.277	1.43	0.528	27.1	-0.173	1.28	-0.105	-0.273
22	10/8/2019	-1.01	1.57	0.0984	37.5	-1.02	0.183	0.00929	0
23	12/10/2019	-1.19	0.881	0.064	7.6	-1.57	0.0516	0.372	0.735
24	2/25/2020	0.297	2.88	1.98	750	0.487	5.83	-0.19	-0.416
25	5/20/2020	0.441	1.79	2.76	62.4	0.144	2.65	0.297	0.649
26	5/26/2020	1.11	2.76	12.8	579	0.833	13	0.274	0.568
27	6/29/2020	1.15	1.86	14.1	72.3	0.0286	2.03	1.12	2.14
28	7/16/2020	0.342	2.54	2.2	349	0.391	4.68	-0.0488	-0.135
29	8/20/2020	-0.438	1.51	0.365	32.2	-0.679	0.398	0.241	0.491
30	9/22/2020	-0.857	1.52	0.139	33.4	-0.954	0.211	0.0969	0.204
31	1/28/2021	-1.23	0.839	0.0586	6.9	-1.25	0.108	0.0149	0.0674
32	2/2/2021	0.606	2.53	4.04	340	0.0211	2	0.585	1.3
33	4/15/2021	-0.162	1.59	0.688	39.2	-0.0711	1.61	-0.0913	-0.204
34	5/11/2021	-0.75	1.35	0.178	22.6	-0.173	1.28	-0.577	-1.16
35	5/18/2021	1.01	2.31	10.2	203	0.517	6.25	0.492	1.03
36	6/8/2021	0.38	1.55	2.4	35.7	-0.0739	1.6	0.454	0.925
37	8/23/2021	-0.45	2.07	0.355	118	-0.29	0.974	-0.159	-0.344

Definitions

ATR: Atrazine in ug/l (39632)
TBY: Turbidity in FNU (63680)

References Cited

American Public Health Association, American Water Works Association, and Water Environment Federation, 1995, Standard methods for the examination of water and wastewater (19th ed.): Washington D.C., American Public Health Association, 905 p.

Cook, D.R., 1977, Detection of influential observation in linear regression: Technometrics, v. 19, no. 1, p. 15–18. [Also available at https://www.jstor.org/stable/1268249?seq=4#metadata_info_tab_contents.]

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.]

R Core Team, 2020, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, version 4.0.0. [Also available 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., Bennett, T.J., Stone, M.L., Foster, G.M., Graham, J.L., and Putnam, J.E., 2014, Quality-assurance and data-management plan for water-quality activities in the Kansas Water Science Center, 2014: U.S. Geological Survey Open-File Report 2014–1233, 41 p. [Also available at https://doi.org/10.3133/ofr20141233.]

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. [Also available at https://doi.org/10.3133/tm3A7.]

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. [Also available at https://doi.org/10.3133/tm3A8.]

U.S. Geological Survey, 2022, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed December 8, 2021, 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]. [Also available 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, 96 p. [Also available at https://doi.org/10.3133/tm1D3.]