

## **Appendix 5.7. Model Archive Summary for Atrazine Concentration at U.S. Geological Survey station 07144100; Little Arkansas River near Sedgwick, Kansas, during December 2014 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: 07144100

Site Name: Little Arkansas River near Sedgwick, Kansas

Location: Latitude 37°52'59", longitude 97°25'27" referenced to North American Datum of 1927, in NE 1/4 NW 1/4 NW 1/4 sec.15, T.25 S., R.1 W., Sedgwick 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 handrail. Check-bar elevation is 33.614 feet. The orifice is enclosed in a well-screen and attached to a concrete pier on the left downstream side of the bridge. Gage height was measured during May 1998 through December 2021. A YSI 6600 water-quality monitor equipped with water temperature, specific conductance, pH, dissolved oxygen, and turbidity (a YSI Model 6026 [September 1998 through December 2006] and YSI Model 6136 [July 2004 through March 2015]) sensors collected data during April 1998 through March 2015. 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 September 2014 through December 2021. A Hach Nitratax monitor collected nitrate data during March 2012 through December 2021.

Date model was developed: June 1, 2022

Model calibration data period: December 9, 2014 through August 25, 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 44 concomitant values of discretely collected atrazine and continuously measured turbidity during December 2014 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. One sample (collected on September 6, 2018) was not representative of the dataset and exceeded DFFITS, Cook's D, and studentized residual outlier criteria and was removed from the dataset. Removing data points based only on outlier criteria may overestimate the certainty of the model.

### **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 4 to 19 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 USGS National Water Quality Laboratory.

## Continuous Data

Concomitant turbidity values were time interpolated. If no concomitant continuous data were available within two 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 AMPA.

## Model Summary

Summary of atrazine regression analysis at station 07144100:

Atrazine-based model:

$$\log_{10}(ATR) = 0.779 \times \log_{10}(TBY) + 0.385 \times \sin(2\pi D) - 0.278 \times \cos(2\pi D) - 1.62$$

where,

$\log_{10}$  = logarithm base 10;

$ATR$  = atrazine, in micrograms per liter ( $\mu\text{g/L}$ );

$TBY$  = 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 1.03.

## Model Statistics, Data, and Plots

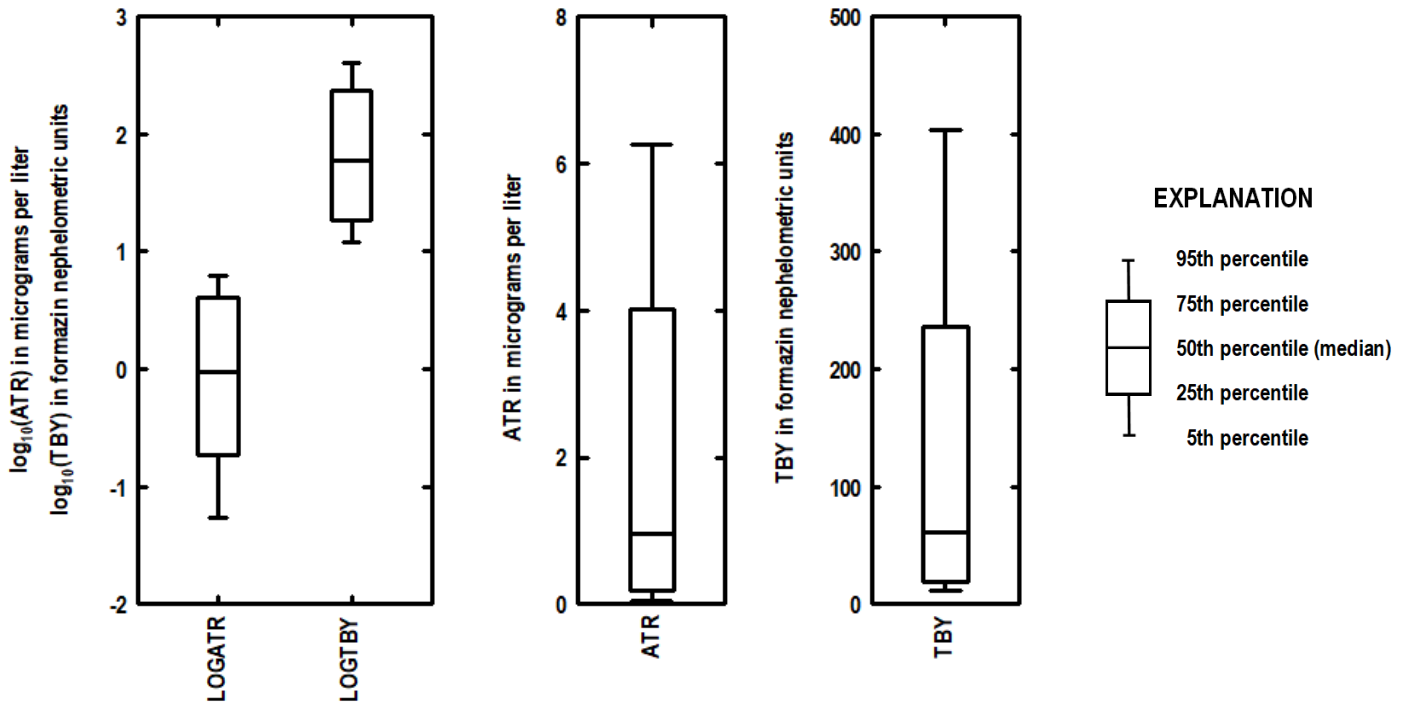
### Model

$$\text{LOGATR} = + 0.779 * \text{LOGTBY} + 0.385 * \text{SIN2PID} - 0.278 * \text{COS2PID} - 1.62$$

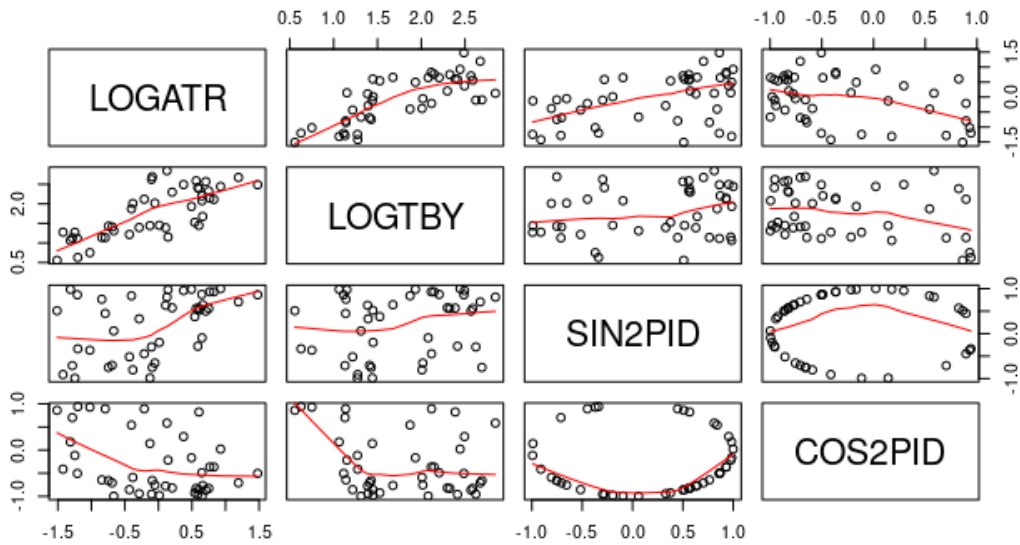
### Variable Summary Statistics

	LOGATR	ATR	LOGTBY	SIN2PID	COS2PID	TBY
Minimum	-1.5100	0.0308	0.556	-0.993	-0.998	3.6
1st Quartile	-0.7180	0.1920	1.270	-0.409	-0.829	18.8
Median	-0.0201	0.9580	1.770	0.505	-0.548	61.0
Mean	-0.0699	2.9300	1.780	0.203	-0.305	139.0
3rd Quartile	0.6040	4.0200	2.360	0.826	0.159	228.0
Maximum	1.4800	30.1000	2.850	1.000	0.941	702.0

## Box Plots



## Exploratory Plots



## Basic Model Statistics

Number of Observations	44
Standard error (RMSE)	0.437
Average Model standard percentage error (MSPE)	119
Coefficient of determination ( $R^2$ )	0.713
Adjusted Coefficient of Determination (Adj. $R^2$ )	0.692
Bias Correction Factor (BCF)	1.5

## Variance Inflation Factors (VIF)

LOGTBY	SIN2PID	COS2PID
1.15	1.07	1.09

## Explanatory Variables

	Coefficients	Standard Error	t value	Pr(> t )
(Intercept)	-1.620	0.204	-7.94	9.45e-10
LOGTBY	0.779	0.115	6.80	3.54e-08
SIN2PID	0.385	0.102	3.78	5.20e-04
COS2PID	-0.278	0.106	-2.61	1.26e-02

## Correlation Matrix

	Intercept	LOGTBY	SIN2PID	COS2PID
Intercept	1.000	-0.928	0.135	-0.117
LOGTBY	-0.928	1.000	-0.253	0.287
SIN2PID	0.135	-0.253	1.000	-0.104
COS2PID	-0.117	0.287	-0.104	1.000

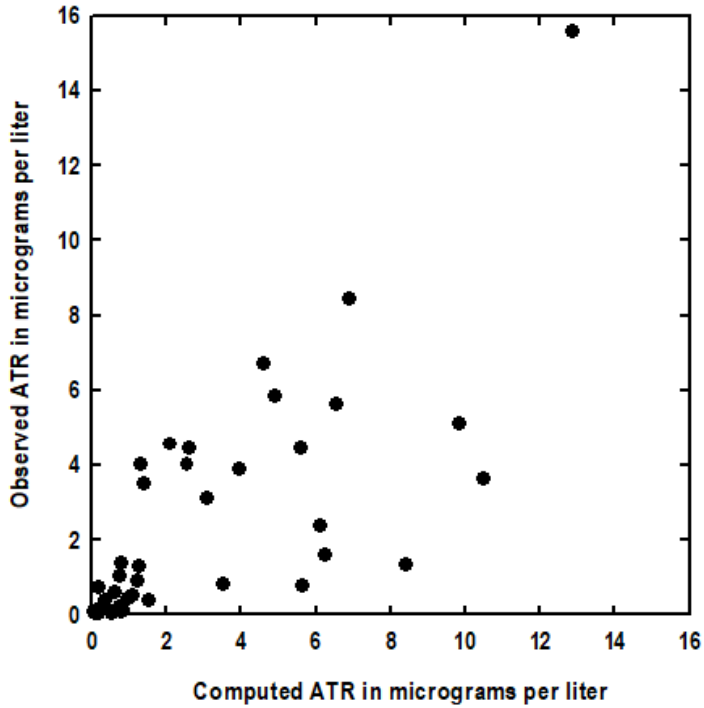
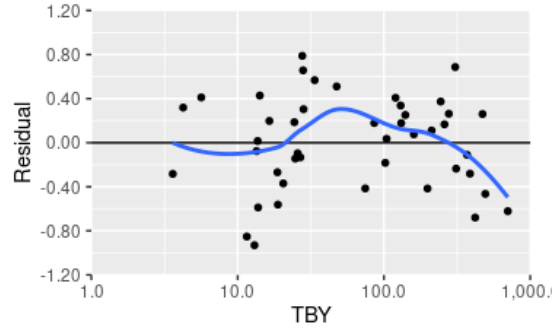
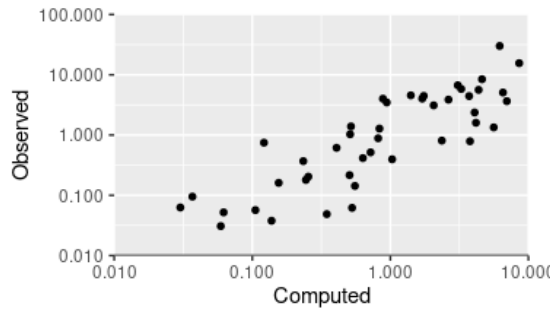
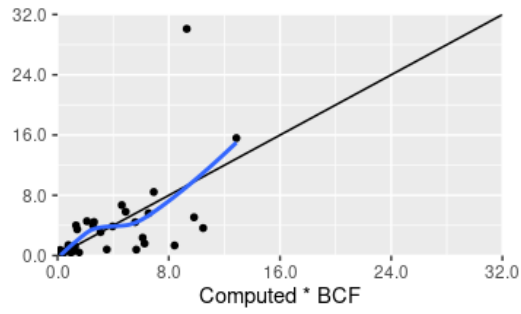
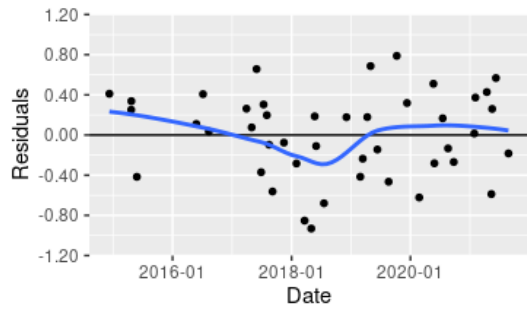
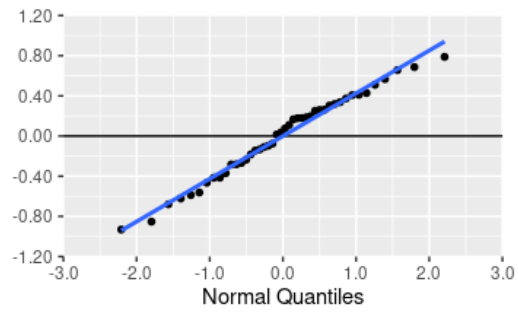
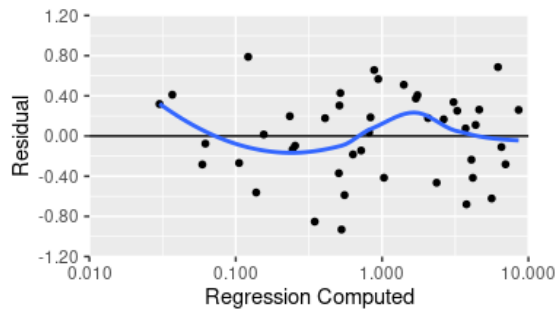
## Outlier Test Criteria

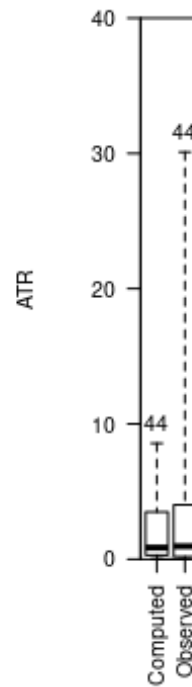
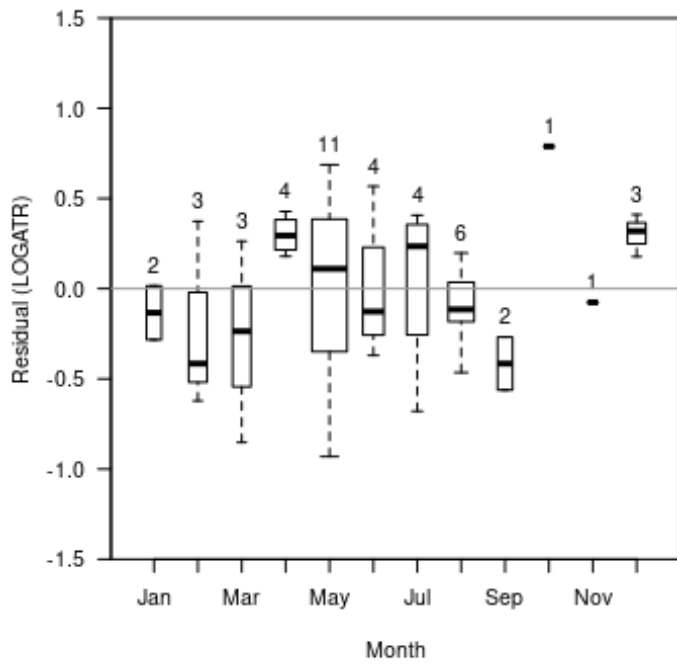
Leverage	Cook's D	DFFITS
0.273	0.317	0.603

## Flagged Observations

	LOGATR	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
3/22/2018 10:50	-1.31	-0.46	-0.852	-2.06	-2.15	0.105	0.124	-0.737
5/2/2018 10:00	-1.21	-0.277	-0.931	-2.24	-2.37	0.0989	0.138	-0.785
10/9/2019 11:50	-0.127	-0.915	0.789	1.91	1.98	0.108	0.11	0.689
2/24/2020 10:30	0.127	0.749	-0.622	-1.57	-1.6	0.176	0.131	-0.739

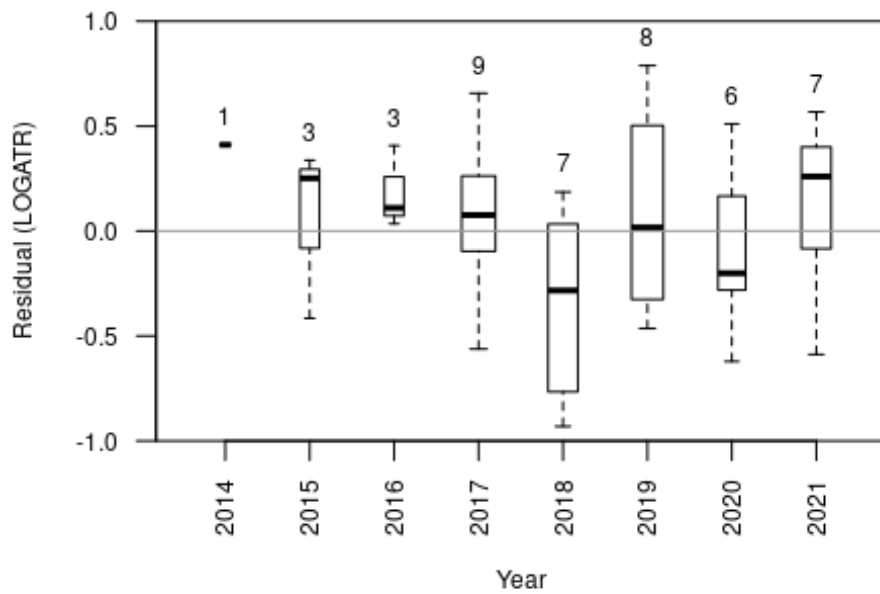
# Statistical Plots





### EXPLANATION

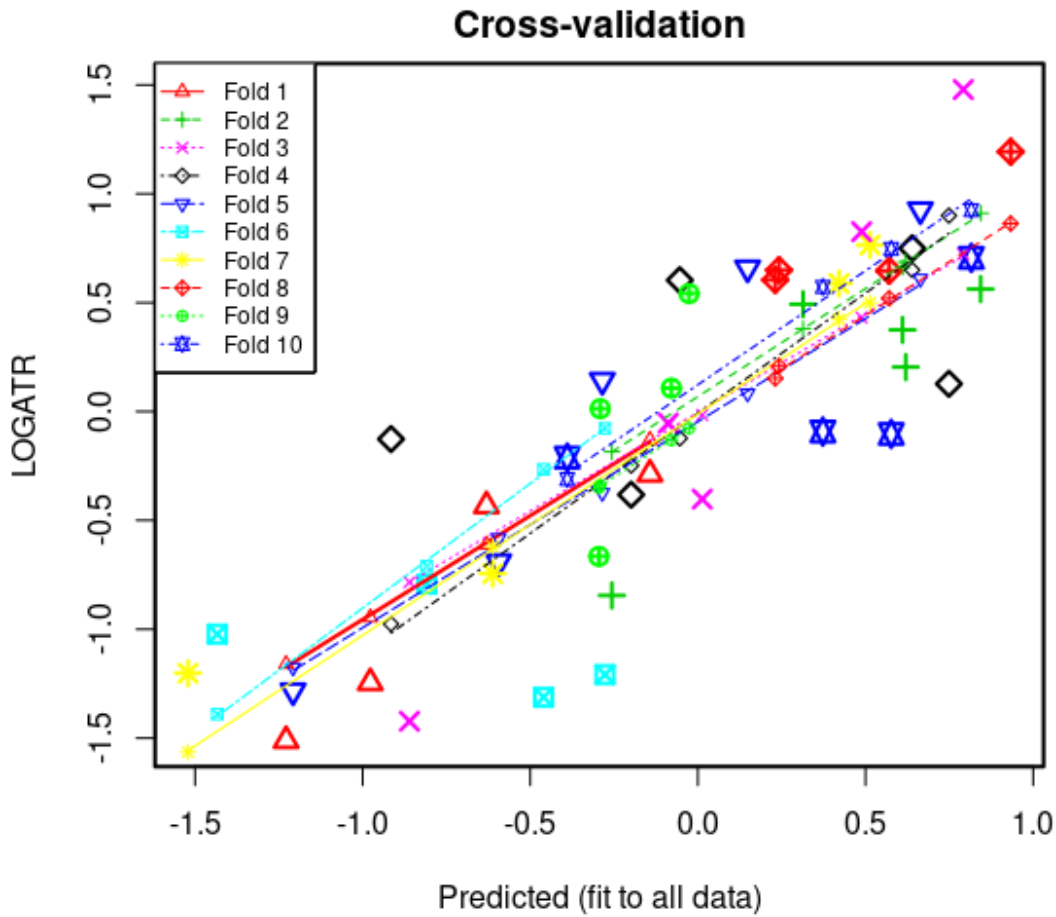
- 44 Number of values
- T Maximum value
- 75th percentile
- 50th percentile (median)
- 25th percentile
- Minimum value



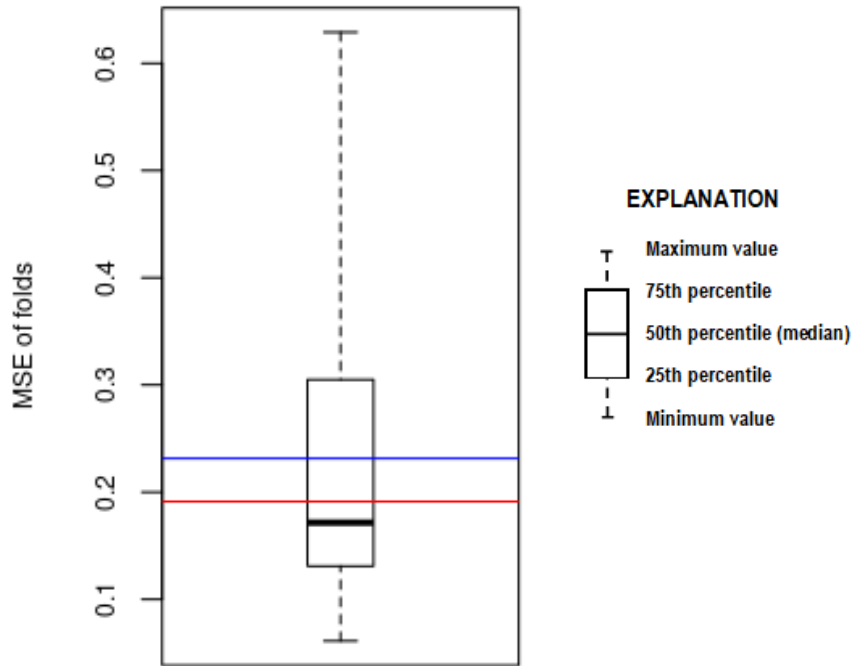
### EXPLANATION

- 1 Number of values
- T Maximum value
- 75th percentile
- 50th percentile (median)
- 25th percentile
- Minimum value

# Cross Validation



Minimum MSE of folds: 0.0611  
Mean MSE of folds: 0.2310  
Median MSE of folds: 0.1720  
Maximum MSE of folds: 0.6290  
(Mean MSE of folds) / (Model MSE): 1.2100



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	12/9/2014	-1.02	0.751	0.095	5.64	-1.43	0.0553	0.411	1.04
2	4/22/2015	0.765	2.15	5.82	140	0.513	4.9	0.251	0.439
3	4/23/2015	0.826	2.11	6.7	130	0.488	4.62	0.338	0.782
4	5/27/2015	0.204	2.3	1.6	198	0.62	6.26	-0.416	-0.947
5	5/26/2016	0.75	2.33	5.62	212	0.639	6.54	0.11	0.0852
6	7/6/2016	0.649	2.08	4.46	120	0.242	2.62	0.407	0.947
7	8/11/2016	-0.0531	2.02	0.885	104	-0.0884	1.22	0.0353	-0.0284
8	3/30/2017	0.926	2.44	8.44	277	0.664	6.92	0.263	0.568
9	5/1/2017	0.646	2.2	4.43	160	0.57	5.58	0.0759	0.0284
10	5/31/2017	0.603	1.45	4.01	28	-0.0539	1.33	0.657	1.56
11	6/28/2017	-0.666	1.31	0.216	20.5	-0.295	0.76	-0.37	-0.782
12	7/13/2017	0.0128	1.45	1.03	28.2	-0.291	0.767	0.304	0.636
13	8/2/2017	-0.433	1.22	0.369	16.5	-0.631	0.351	0.198	0.377
14	8/16/2017	-0.69	1.41	0.204	25.7	-0.594	0.382	-0.0962	-0.2
15	9/6/2017	-1.42	1.27	0.0378	18.8	-0.86	0.207	-0.562	-1.14
16	11/15/2017	-1.28	1.13	0.0521	13.5	-1.21	0.0931	-0.0756	-0.142
17	1/31/2018	-1.51	0.556	0.0308	3.6	-1.23	0.0887	-0.283	-0.707
18	3/22/2018	-1.31	1.06	0.0487	11.5	-0.46	0.52	-0.852	-1.8
19	5/2/2018	-1.21	1.11	0.0619	13	-0.277	0.793	-0.931	-2.21
20	5/23/2018	0.107	1.39	1.28	24.4	-0.0791	1.25	0.186	0.317
21	6/1/2018	0.706	2.57	5.08	370	0.816	9.82	-0.11	-0.258
22	7/19/2018	-0.103	2.62	0.789	420	0.577	5.66	-0.68	-1.56
23	12/4/2018	-0.212	2.12	0.614	131	-0.39	0.612	0.178	0.2



24	2/27/2019	-0.402	1.87	0.396	74.5	0.0129	1.55	-0.415	-0.861
25	3/14/2019	0.375	2.49	2.37	311	0.61	6.12	-0.236	-0.502
26	4/11/2019	0.493	1.93	3.11	85.8	0.314	3.09	0.179	0.258
27	5/1/2019	1.48	2.49	30.1	307	0.792	9.29	0.687	1.8
28	6/12/2019	-0.287	1.39	0.516	24.7	-0.143	1.08	-0.144	-0.377
29	8/20/2019	-0.0921	2.69	0.809	493	0.373	3.54	-0.465	-1.04
30	10/9/2019	-0.127	1.44	0.747	27.7	-0.915	0.182	0.789	2.21
31	12/11/2019	-1.2	0.627	0.0628	4.24	-1.52	0.0452	0.319	0.707
32	2/24/2020	0.127	2.85	1.34	702	0.749	8.42	-0.622	-1.4
33	5/21/2020	0.659	1.68	4.56	47.5	0.148	2.11	0.511	1.26
34	5/27/2020	0.562	2.59	3.65	387	0.844	10.5	-0.281	-0.636
35	7/17/2020	0.588	2.41	3.87	259	0.421	3.96	0.167	0.142
36	8/19/2020	-0.745	1.43	0.18	26.8	-0.612	0.367	-0.133	-0.317
37	9/23/2020	-1.25	1.27	0.0568	18.7	-0.977	0.158	-0.268	-0.568
38	1/27/2021	-0.793	1.14	0.161	13.7	-0.809	0.233	0.0155	-0.0852
39	2/3/2021	0.604	2.39	4.02	244	0.231	2.55	0.374	0.861
40	4/14/2021	0.143	1.15	1.39	14.2	-0.285	0.779	0.428	1.14
41	5/12/2021	-0.845	1.14	0.143	13.8	-0.257	0.831	-0.588	-1.26
42	5/17/2021	1.19	2.67	15.6	472	0.933	12.9	0.26	0.502
43	6/9/2021	0.542	1.53	3.48	33.6	-0.0262	1.41	0.568	1.4
44	8/25/2021	-0.382	2.01	0.415	102	-0.199	0.949	-0.183	-0.439

## 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](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>.]