Appendix 4. Model Archive Summary for Particulate Mercury Concentrations at Station 254543080405401: Tamiami Canal at S-12D Near Miami, Florida

Previous studies have documented relations between suspended sediment and particulate mercury (Schoellhamer and others, 2007; Horowitz, 2009; Etheridge, 2015). With documented techniques and methods on the use of turbidity for computing suspended sediment concentrations (Rasmussen and others, 2009), turbidity data can also be used as a proxy for suspended sediment mercury concentrations. This study focuses on the development of surrogate models for continuous monitoring of mercury (Hg) in the Florida Everglades.

High density and long-term data will aid with the description of short- and long-term variability of carbon and mercury concentrations, which will improve understanding of carbon input and transport. Prior to this study, no continuous and long-term time-series data on carbon concentrations were available for the freshwater wetlands of the Florida Everglades.

The objectives of this study were to develop and document a surrogate model to calculate concentration and loads of particulate total mercury (PTHg) at site S-12D. This model archive summary describes the PTHg model developed to compute 15-minute frequency PTHg concentrations from turbidity data collected from September 5, 2013, to April 3, 2017, at site S-12D. The methods used follow U.S. Geological Survey (USGS) guidance as referenced in Rasmussen and others (2009).

Site and Model Information

USGS site number: 254543080405401
https://waterdata.usgs.gov/fl/nwis/inventory/?site_no=254543080405401&agency_cd=USGS&

Site name: Tamiami Canal at S-12D near Miami, Florida

Location: lat 25°45'43" N., long 80°40'54" W., referenced to North American Datum of 1927, in T. 54 S., R. 36 E., Miami-Dade County, Florida, hydrologic unit 03090202, on south bank 100 feet southwest of structure 12-D, near east boundary of Indian reservation on U.S. Highway 41.

Equipment: A YSI EXO water-quality monitoring system equipped with sensors for water temperature, specific conductance, turbidity, and an fDOM sensor. The monitor is housed in an 8-inch-diameter polyvinyl chloride (PVC) pipe on a diagonal extending off the end of the structure into the stream. Readings from the YSI EXO were recorded every 15 minutes and transmitted hourly by way of satellite. The model applies only to this site (254543080405401) and specified time period (September 5, 2013, to April 3, 2017).

Model number: 1.0

Date model was created: January 4, 2018

Model calibration data period: September 11, 2013, to October 26, 2016

Model application date: September 5, 2013, to April 3, 2017

Computed by: Amanda Booth, USGS Caribbean-Florida Water Science Center

Reviewed by: Mark Brigham, USGS Minnesota Water Science Center, Mounds View, Minnesota, and Brian Downing, USGS California Water Science Center, Sacramento, California

Approved by: David Sumner, USGS Caribbean-Florida Water Science Center

Model Data

All data were collected using USGS protocols and are stored in the National Water Information System (NWIS) database (U.S. Geological Survey, 2019). The regression model is based on 26 concurrent measurements of PTHg and turbidity, collected from September 11, 2013, through October 26, 2016. Samples were collected throughout the range of observed hydrologic and water-quality conditions. Summary statistics and the complete model-calibration data are provided in the dataset.

Mercury Suspended Sediment Data

Teflon equipment precleaned by the USGS Mercury Research Laboratory was used for the collection and transport of all mercury samples. A hydrokinetic nozzle, 200-milliliter bottle, and nozzle-bottle holder were used to collect enough water to fill a 2-liter bottle for each sample. Initially, two water samples were collected for mercury analysis: (1) a point sample next to the water-quality sensors and (2) a single vertically integrated sample at the location of highest water velocity. Samples were placed on ice for transport, filtered through 47-millimeter 0.7-micron quartz-fiber filters to separate dissolved and mercury suspended sediment species, and shipped to the laboratory within 24 hours. Filtered samples were shipped on dry ice. Samples were analyzed for PTHg using techniques and methods documented in Olund and others (2004).

Elevated levels of turbidity are rare at this site other than immediately after a gate opening. To try to capture a wide range of conditions, a sampling event was coordinated on September 23, 2015, to collect a sample just after the gates were opened. Multiple samples were collected on this date at 8:02 a.m., 8:19 a.m., and 9:41a.m., Eastern Standard Time. Because of the similarity in the lab values and the field sensor data between the 8:02 a.m. and 8:19 a.m. samples, only the 8:02 a.m. sample was included in model development, to avoid statistically overrepresenting that condition. The 9:41 a.m. data was included in model development because there was a significant change in field sensor data; this sample also had the highest turbidity of any sample collected.

Surrogate Data

The turbidity data used in this analysis were measured using a YSI EXO V2 (Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government). Turbidity values at S-12D ranged from 0.0 formazin nephelometric units (FNU) on many days throughout the 2013 water year to 87.0 FNU on December 10, 2014. The sensor was set to record data at 2-minute intervals during discrete sampling events; otherwise, sensor data were collected at 15-minute intervals. If no sensor data were collected the minute the sample was collected, the sensor data from the minute before and minute after were averaged together.

It is necessary to highlight that the maximum value of turbidity measured during sampling events and used in development of the model is 16.0 FNU, whereas the maximum recorded value is 87.0 FNU. As evident in measured field data, these turbidity spikes are not common and were short in duration, usually lasting less than half an hour. Daily mean values for days when spikes occurred never exceeded 10.0 FNU. All turbidity values greater than 30.0 FNU are excluded from the computations of concentrations and load data.

Model Development

Regression analyses were made using Microsoft Excel and the USGS Surrogate Analysis and Index Developer (SAID) tool (Domanski and others, 2015) by examining turbidity, fDOM, and other continuously measured data as explanatory variables for estimating PTHg concentration. A variety of models that predict PTHg were evaluated, including natural log (ln) transformed and log10 transformed data. The model inputs were not normally distributed, therefore, the ln transformed data were used. The distribution of residuals was examined for normality, and plots of residuals (the difference between the observed and computed values) as compared to computed PTHg were examined for homoscedasticity. The ln transformed turbidity was selected as the best predictor of PTHg based on residual plots, relatively high adjusted coefficient of determination (adjusted R^2), and relatively low model standard percentage error. Values for all aforementioned statistics and metrics were computed and are included below, along with all relevant sample data and more in-depth statistical information. While the ln transformed model underestimates the highest values used in the model, these values are not typical. The data are not normal and the model using the ln transformed data does a better job at modeling the values that are typical. When discharge (Q) equaled zero, a Q value of 0.001 was entered for the program to create the graphics.

Model Summary

Summary of final regression analysis for PTHg concentration at site number 254543080405401.

Particulate mercury concentration-based model:

$$lnPTHg = 0.701 * lnTurbidity - 0.906$$

where

ln = natural log;

PTHg = mercury suspended sediment concentration, in nanograms per liter; and

turbidity = turbidity, YSI EXO model, in formazin nephelometric units.

Model Statistics, Data, and Plots

Model

lnPTHG = 0.701 * lnTurbidity −0.906

Variable Summary Statistics

	lnPTHg	PTHg	lnTurbidity	Turbidity
Minimum	-2.290	0.101	-1.6100	0.2
1st Quartile	-1.600	0.201	-0.5110	0.6
Median	-1.020	0.359	-0.1120	0.9
Mean	-0.876	0.759	0.0429	2.1
3rd Quartile	-0.524	0.592	0.8330	2.3
Maximum	1.870	6.470	2.7700	16.0

Box Plots

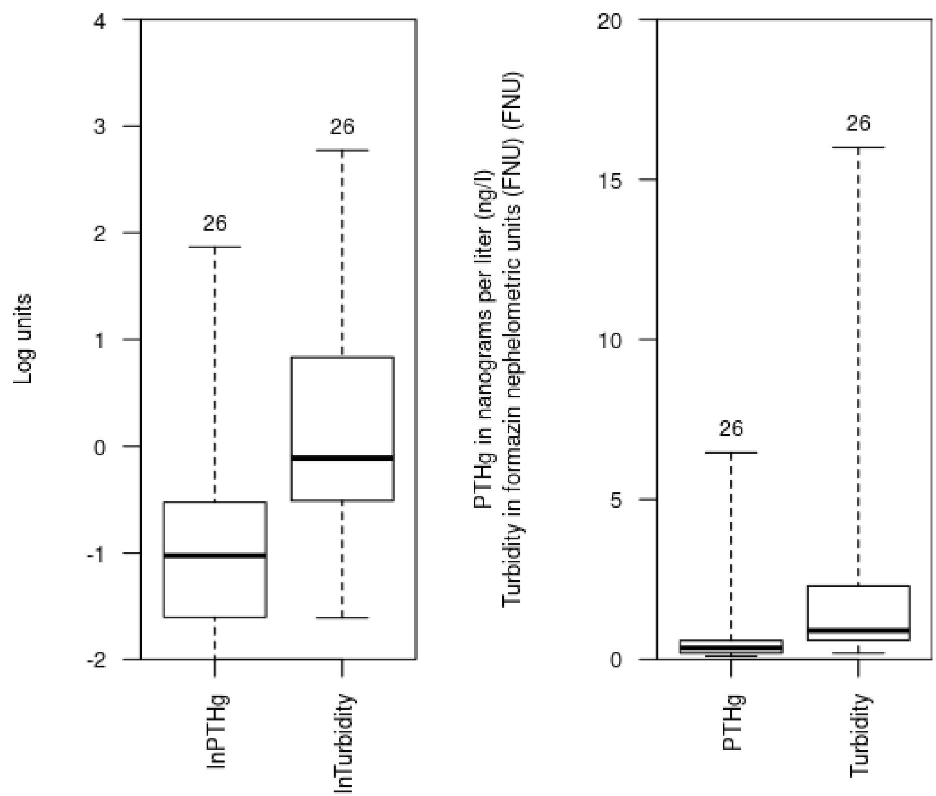


Figure 4.1. Boxplots of natural log (ln) transformed and nontransformed turbidity and particulate total mercury (PTHg) collected during discrete sampling events.

Exploratory Plots

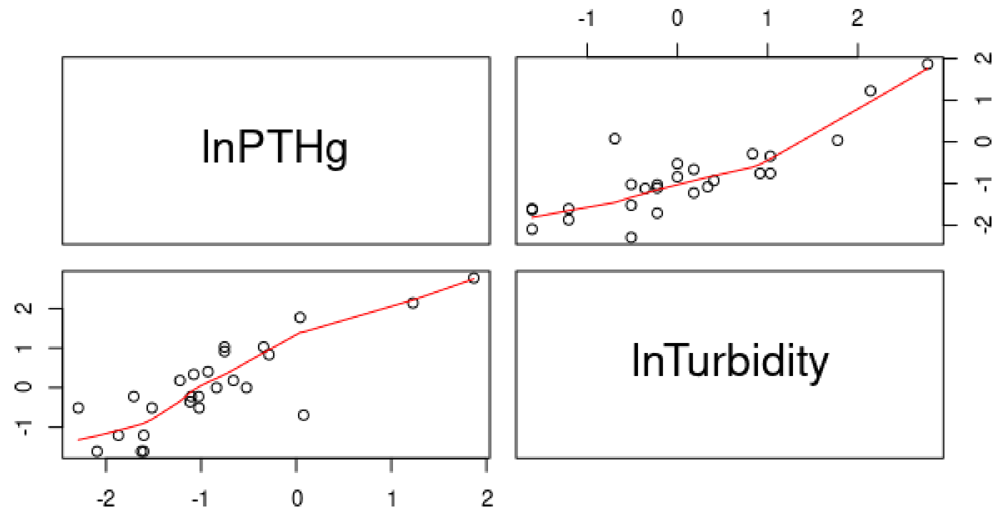


Figure 4.2. Exploratory plots of natural log (ln) transformed turbidity, in formazin nephelometric units, and particulate total mercury (PTHg), in nanograms per liter.

Basic Model Statistics

```
Number of Observations                                  26
Standard error (RMSE)                                0.522
Average Model standard percentage error (MSPE)        54.6
Coefficient of determination (R²)                    0.703
Adjusted Coefficient of Determination (Adj. R²)      0.691
Bias Correction Factor (BCF)                          1.15
```

Explanatory Variables

```
            Coefficients Standard Error t value Pr(>|t|)
(Intercept)       -0.906         0.1020   -8.84 5.12e-09
lnTurbidity        0.701         0.0929    7.54 8.79e-08
```

Correlation Matrix

```
          Intercept  E.vars
Intercept    1.0000 -0.0389
E.vars      -0.0389  1.0000
```

Outlier Test Criteria

```
Leverage Cook's D   DFFITS
   0.231    0.193    0.555
```

Flagged Observations

Three observed samples were flagged as potential outliers. All three samples were flagged because the difference in fits (DFFITS) values were greater than 0.555. The samples collected on September 23, 2015, and October 26, 2015, were also flagged because the Cook's D value was greater than 0.193. The only sample that was also flagged for leverage was from September 23, 2015. All flagged observations were retained in the model.

	lnPTHG	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
09/23/2015 08:02	1.230	0.594	0.633	1.34	1.36	0.1780	0.193	0.632
09/23/2015 09:41	1.870	1.040	0.830	1.87	1.98	0.2740	0.658	1.210
10/26/2016 09:00	0.077	-1.390	1.470	2.90	3.51	0.0556	0.247	0.853

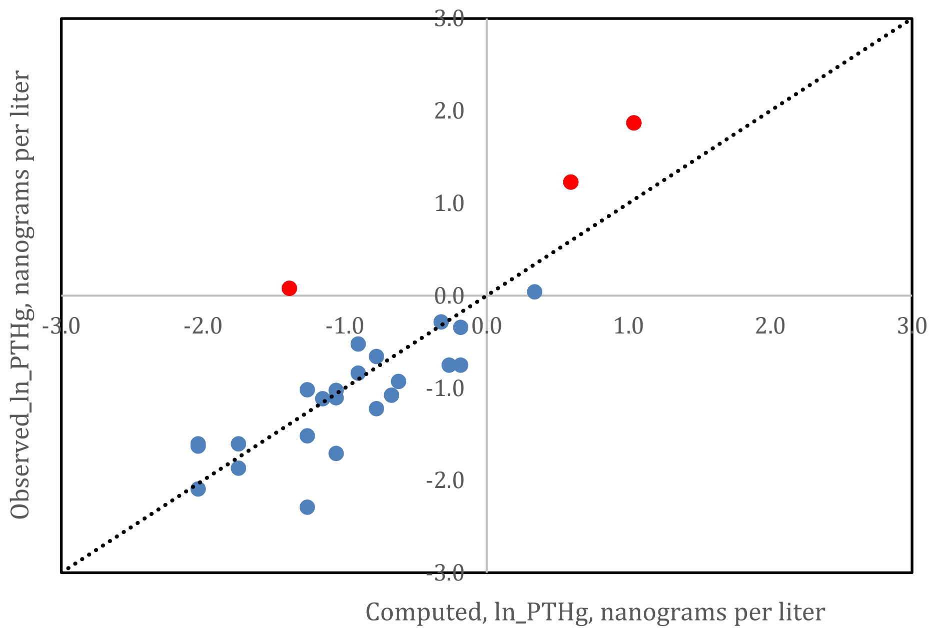


Figure 4.3. Relation between observed particulate total mercury (PTHg) and computed PTHg; flagged observations are in red.

(ln, natural log)

Statistical Plots

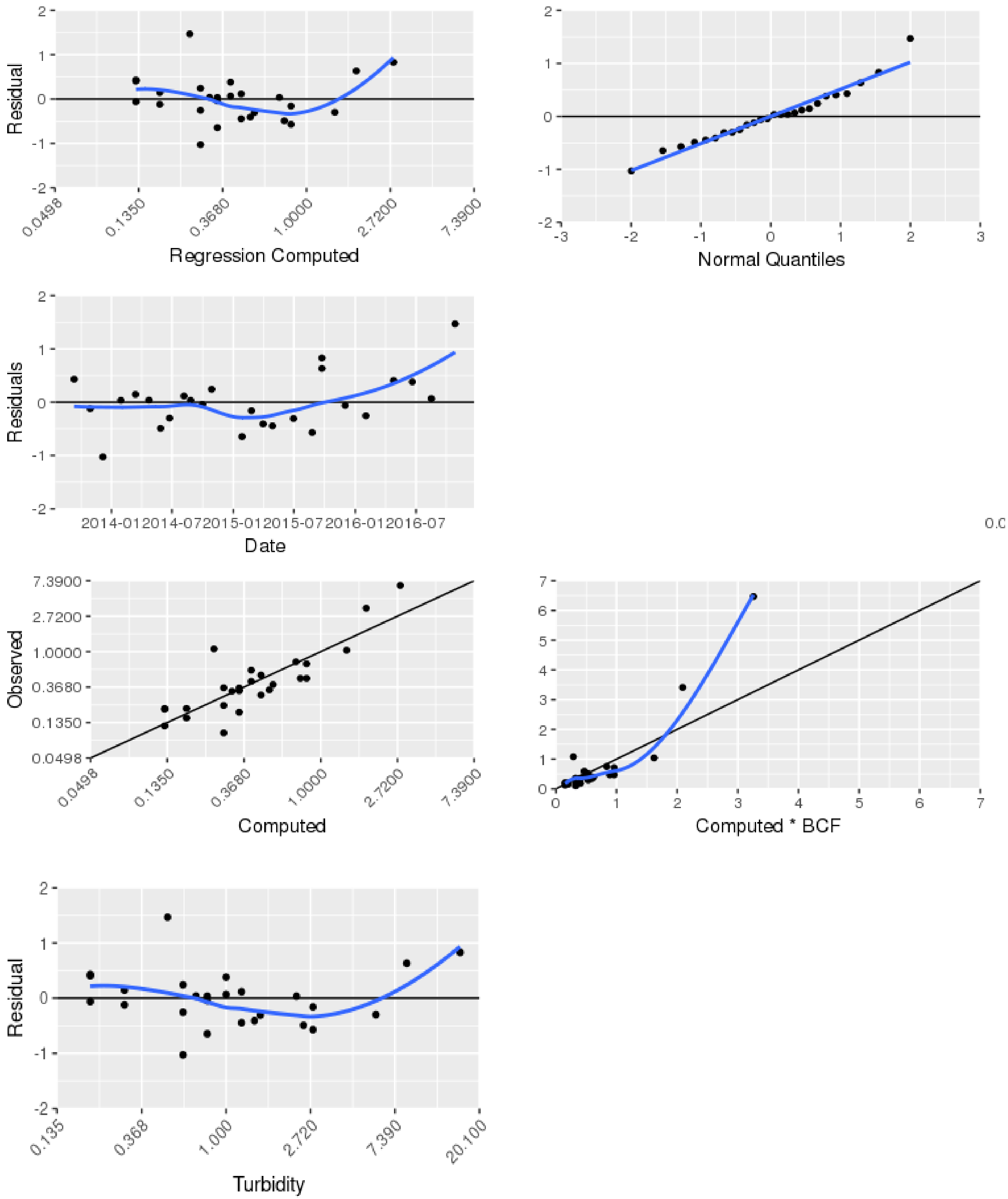


Figure 4.4. Residual and observed versus computed plots.

(BCF, bias correction factor)

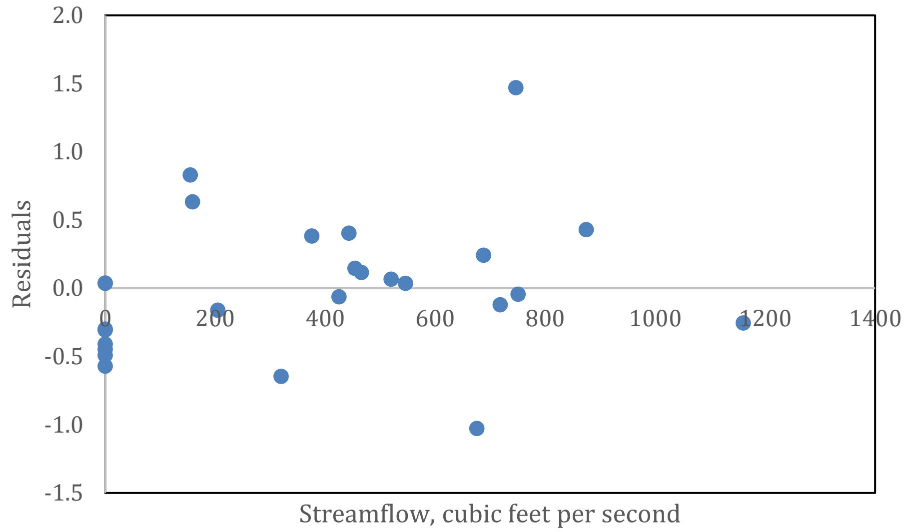


Figure 4.5. Relation between residuals and streamflow showing that the residuals had no systematic bias with respect to streamflow.

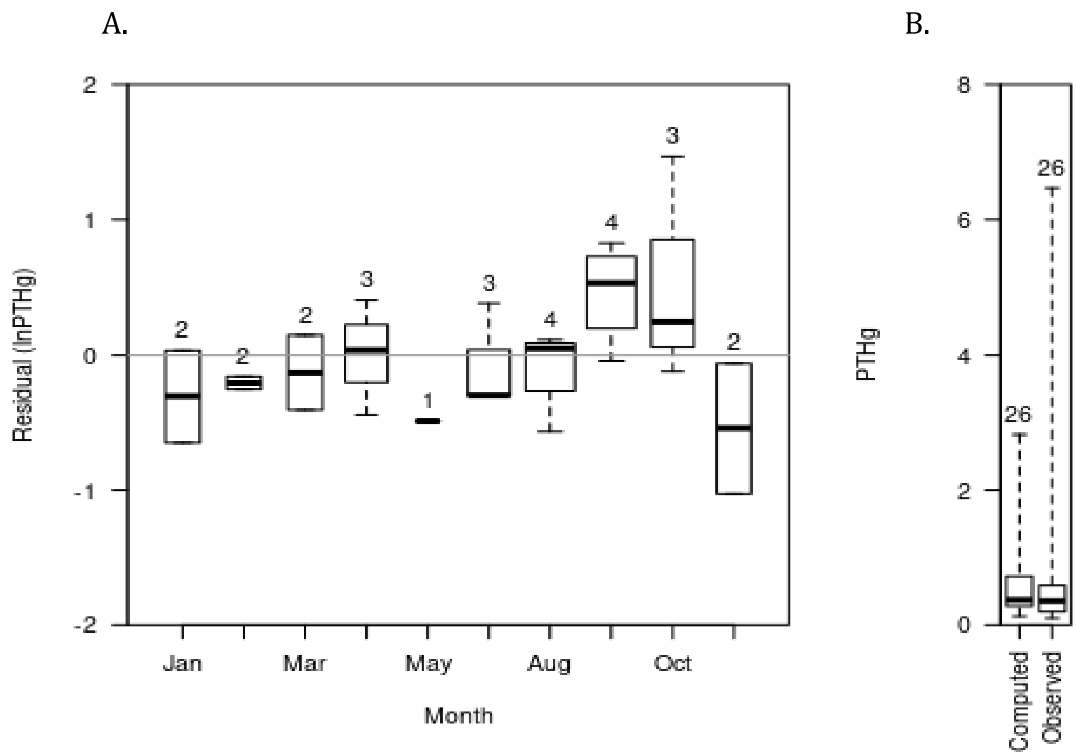


Figure 4.6. *A*, Seasonal variation in residuals of natural log of particulate total mercury (lnPTHg), and *B*, computed and observed particulate total mercury (PTHg), in nanograms per liter.

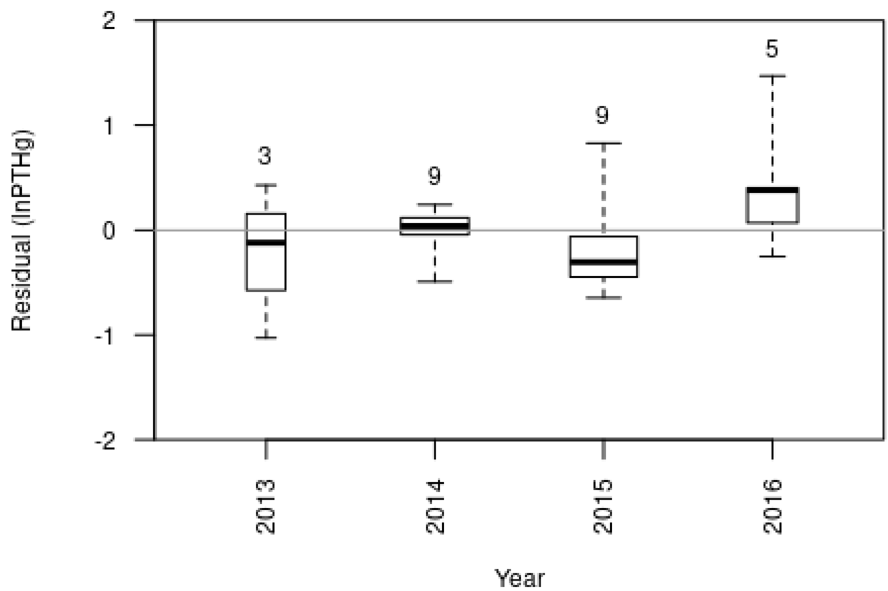
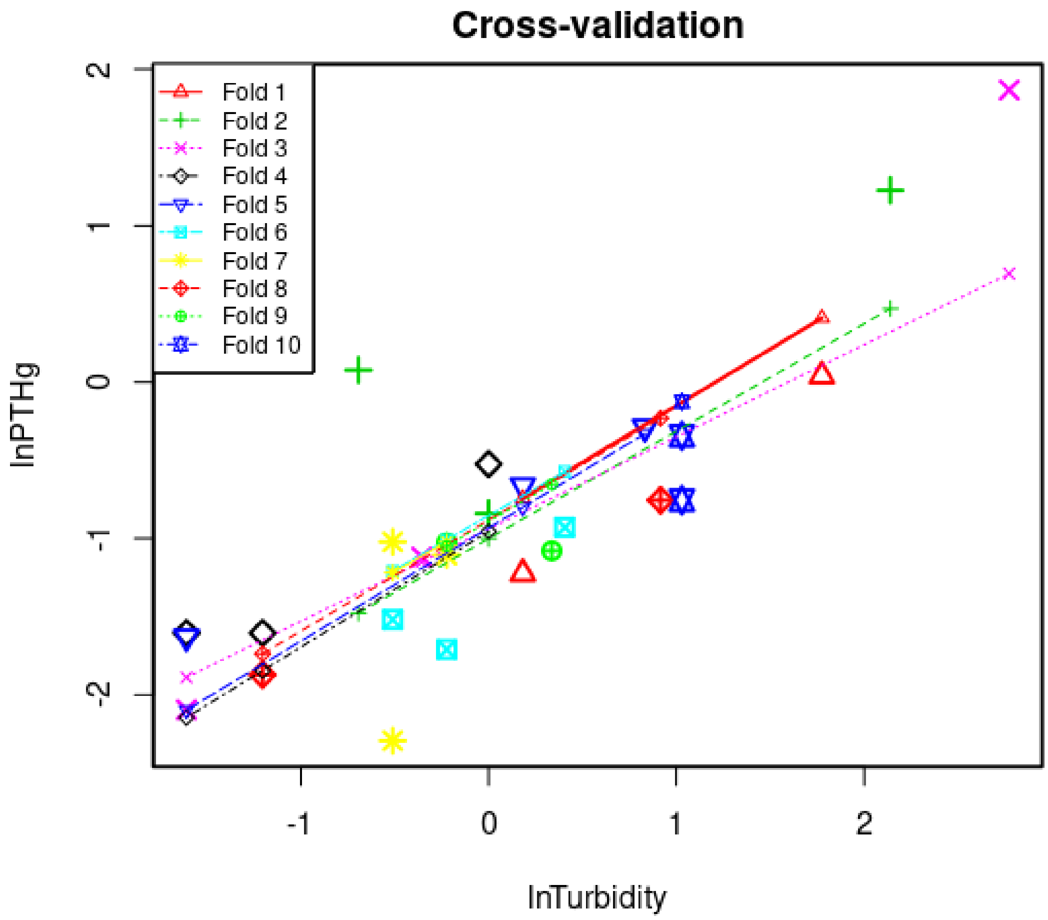


Figure 4.7. Annual variation in residuals.

(lnPTHg, natural log of particulate total mercury)

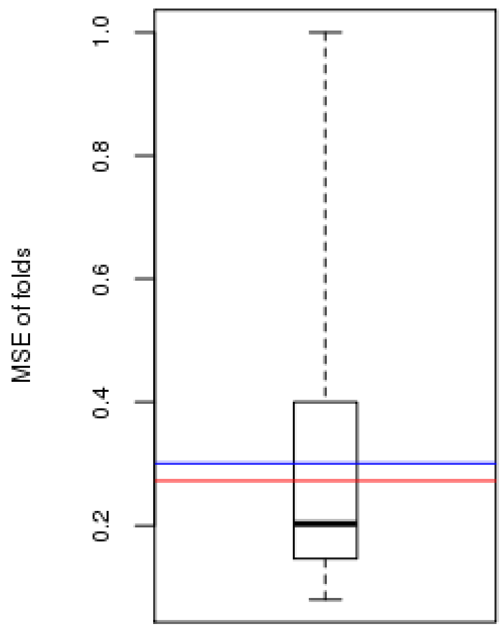
Cross Validation

K-fold cross validation was used to validate the model. The advantage of K-fold cross validation is that all the examples in the dataset are eventually used for both training and testing. The data were split randomly into 10 experiments or folds.




```
          Minimum MSE of folds:   0.0799
             Mean MSE of folds:   0.3010
           Median MSE of folds:   0.2030
          Maximum MSE of folds:   1.0000
(Mean MSE of folds) / (Model MSE):   1.1000
```

Figure 4.8. Cross validation plot. (lnPTHg, natural log of particulate total mercury; lnTurbidity, natural log of turbidity; MSE, mean standard of error)




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Figure 4.9. Mean standard of error (MSE) of folds boxplot.

Model Calibration Dataset

0	Date	lnPTHg	lnTurbidity	PTHg	Turbidity	Computed lnPTHg	Computed PTHg	Residual	Normal Quantiles	Censored Values
1	2013-09-11	-1.6	-1.61	0.201	0.2	-2.03	0.151	0.43	1.09	--
2	2013-10-29	-1.87	-1.2	0.154	0.3	-1.75	0.201	-0.12	-0.242	--
3	2013-12-06	-2.29	-0.511	0.101	0.6	-1.26	0.326	-1.03	-2	--
4	2014-01-29	-1.03	-0.223	0.358	0.8	-1.06	0.399	0.0356	0.0479	--
5	2014-03-14	-1.6	-1.2	0.201	0.3	-1.75	0.201	0.146	0.553	--
6	2014-04-23	-1.12	-0.357	0.327	0.7	-1.16	0.363	0.0386	0.242	--
7	2014-05-28	-0.755	0.916	0.47	2.5	-0.264	0.887	-0.491	-1.09	--
8	2014-06-24	0.0392	1.77	1.04	5.9	0.338	1.62	-0.299	-0.553	--
9	2014-08-06	-0.662	0.182	0.516	1.2	-0.779	0.53	0.117	0.445	--
10	2014-08-26	-0.286	0.833	0.751	2.3	-0.323	0.836	0.0362	0.144	--
11	2014-09-30	-1.11	-0.223	0.331	0.8	-1.06	0.399	-0.0428	-0.0479	--
12	2014-10-28	-1.02	-0.511	0.36	0.6	-1.26	0.326	0.243	0.668	--
13	2015-01-27	-1.71	-0.223	0.181	0.8	-1.06	0.399	-0.646	-1.55	--
14	2015-02-24	-0.344	1.03	0.709	2.8	-0.185	0.96	-0.159	-0.341	--
15	2015-03-31	-1.08	0.336	0.34	1.4	-0.671	0.59	-0.408	-0.794	--
16	2015-04-28	-1.22	0.182	0.294	1.2	-0.779	0.53	-0.446	-0.932	--
17	2015-06-30	-0.929	0.405	0.395	1.5	-0.622	0.62	-0.307	-0.668	--
18	2015-08-25	-0.755	1.03	0.47	2.8	-0.185	0.96	-0.57	-1.29	--
19	2015-09-23	1.23	2.14	3.41	8.5	0.594	2.09	0.633	1.29	--
20	2015-09-23	1.87	2.77	6.47	16	1.04	3.26	0.83	1.55	--
21	2015-12-02	-2.1	-1.61	0.123	0.2	-2.03	0.151	-0.0609	-0.144	--
22	2016-02-02	-1.52	-0.511	0.219	0.6	-1.26	0.326	-0.254	-0.445	--
23	2016-04-25	-1.63	-1.61	0.196	0.2	-2.03	0.151	0.405	0.932	--
24	2016-06-20	-0.524	0	0.592	1	-0.906	0.466	0.382	0.794	--
25	2016-08-16	-0.839	0	0.432	1	-0.906	0.466	0.0671	0.341	--
26	2016-10-26	0.077	-0.693	1.08	0.5	-1.39	0.287	1.47	2	--

Model Limitations

Errors in the PTHg surrogate model can be attributed to several factors. One of the sources of error is the turbidity data itself. Turbidity data were corrected as best as possible for electronic drift and fouling; however, there is unquantified error associated with the turbidity data itself.

Another limitation to this model is in the assumption that the sensor data collected at the station are representative of the channel. The point samples were more likely to be less than (or equal to) the concentrations observed in the vertically integrated samples at the point of highest velocity, with only a few samples that were higher. Because suspended particulate matter is expected to be positively correlated with flow velocity, it is expected that some values are higher at the location of highest velocity than at the location of the point sensor, which was not situated directly in the pathway of highest flow. The concentrations calculated from this model are expected to be representative of the location of the point sensor but might sometimes slightly underpredict the amount of PTHg in the location of highest flow velocity. The difference between the point and average cross section turbidity ranged from 0.4 to −0.8 FNU, with an average difference of 0.1 FNU.

An additional source of model error comes from processes used to collect and analyze discrete samples. For particulate total mercury, percent recovery for quality control samples at USGS Mercury Research Laboratory were provided and ranged from 89 percent to 110 percent, with an average of 101 percent from January 2014 to July 2017. The percent difference for analytical replicates was also provided for this time period and ranged from −7 percent to 17 percent. Matrix spikes were also provided and ranged from 78 percent to 107 percent.

Definitions

```
PTHg: Mercury in ng/l (62976)
Turbidity: Turbidity in FNU (63680)
```

App Version 1.0

References Cited

Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate analysis and index developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., accessed October 1, 2017, at https://doi.org/10.3133/ofr20151177.

Etheridge, A.B., 2015, Occurrence and transport of selected constituents in streams near the Stibnite mining area, Central Idaho, 2012–14: U.S. Geological Survey Scientific Investigations Report 2015–5166, 47 p., accessed April 1, 2019, at https://doi.org/10.3133/sir20155166.

Horowitz, A.J., 2009, Monitoring suspended sediments and associated chemical constituents in urban environments—Lessons from the city of Atlanta, Georgia, U.S.A, Water Quality Monitoring Program: Journal of Soils and Sediments, v. 9, p. 342–363, accessed April 9, 2020, at https://link.springer.com/article/10.1007/s11368-009-0092-y.

Olund, S. D., DeWild, J. F., Olson, M. L., Tate, M. T., 2004, Methods for the preparation and analysis of solids and suspended solids for total mercury: U.S. Geological Survey Techniques and Methods, book 5, chap. A8, 23 p., accessed October 1, 2017, at http://pubs.usgs.gov/tm/2005/tm5A8/.

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., accessed October 1, 2017, at https://doi.org/10.3133/tm3C4.

Schoellhamer, D.H., Mumley, T.E., Leatherbarrow, J.E., 2007, Suspended sediment and sediment-associated contaminants in San Francisco Bay: Environmental Research, v. 105, no. 1, p. 119–131, accessed January 15, 2019, at https://pubs.er.usgs.gov/publication/70031507.

U.S. Geological Survey, 2019, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed May 8, 2019, at https://doi.org/10.5066/F7P55KJN.