Appendix 5. Model Archive Summary for Particulate Methylmercury Concentrations at Station 254543080405401: Tamiami Canal at S-12D Near Miami, Florida

Previous studies have documented relations between suspended sediment and suspended sediment mercury concentrations (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 methylmercury (PMeHg) at site S-12D. This model archive summary describes the PMeHg model developed to compute 15-minute frequency PMeHg 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 (Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government). 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, Mound 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 PMeHg and turbidity collected from September 11, 2013, through October 26, 2017. 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.

Methylmercury Suspended Sediment Data

Teflon equipment precleaned by the USGS Mercury Research Laboratory was used for the collection and transport of all mercury samples for analyses. 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 a 0.7-micron quartz-fiber filter to separate dissolved and suspended sediment mercury species, and shipped to the laboratory within 24 hours. Filtered samples were shipped on dry ice. Samples were analyzed for PMeHg using techniques and methods documented in DeWild and others (2004).

Concentrations of PMeHg in two of 26 samples collected (8 percent of samples) were below the detection limits for PMeHg, which were 0.007 and 0.009 nanograms per liter. Concentration values for these samples were censored to one-half their detection limits during the model development to avoid problems with using zero in the calculations.

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 9:41

a.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. 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 data was set to record at 2-minute intervals during sampling events. If no sensor data were recorded 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 turbidity data is used in the model and that the maximum measured value during sampling events is 16.0 FNU, while 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 analysis was done 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 PMeHg concentration. A variety of models that predict PMeHg were evaluated, including natural log (ln) transformed and log10 transformed data. The model inputs were not normally distributed; therefore, the ln transformed data was 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 PMeHg were examined for homoscedasticity. The ln transformed turbidity was selected as the best predictor of PMeHg based on residual plots, relatively high adjusted coefficient of determination (adjusted R^2), and relatively low model standard percentage error. Values for all statistics and metrics were computed and are included below, along with all relevant sample data and more in-depth statistical information. 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 PMeHg concentration at site number 254543080405401.

Methylmercury suspended sediment concentration-based model:

$$lnPMeHg = 0.752 * lnTurbidity - 3.82$$

where

ln = natural log;

PMeHg = methylmercury suspended sediment concentration, in nanograms per liter; and

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

Model Statistics, Data, and Plots

Model

$lnPMeHg = 0.752 * lnTurbidity - 3.82$

Variable Summary Statistics

	lnPMeHg	PMeHg	lnTurbidity	Turbidity
Minimum	-5.81	0.0030	-1.6100	0.2
1st Quartile	-4.61	0.0100	-0.5110	0.6
Median	-4.05	0.0175	-0.1120	0.9
Mean	-3.79	0.0406	0.0429	2.1
3rd Quartile	-3.02	0.0490	0.8330	2.3
Maximum	-1.63	0.1950	2.7700	16.0

Box Plots

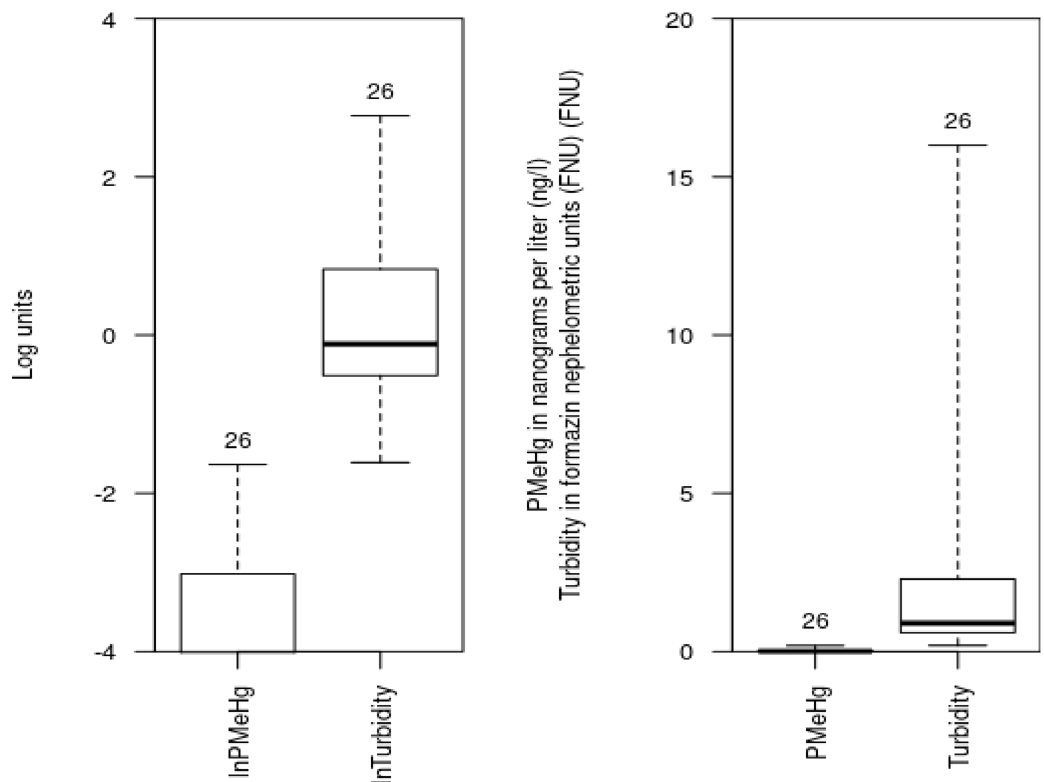


Figure 5.1. Boxplots of natural log (ln) transformed and nontransformed turbidity and particulate total methylmercury (PMeHg) collected during discrete sampling events.

Exploratory Plots

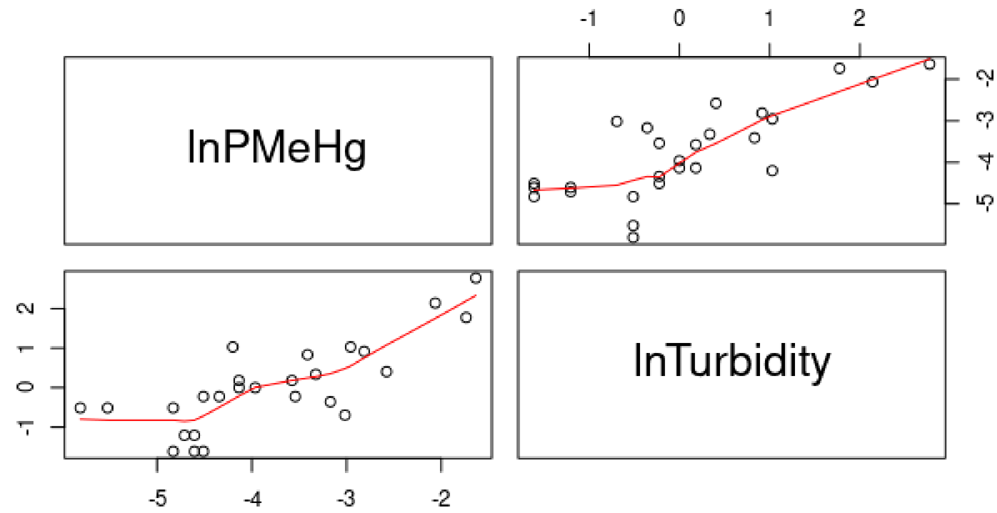


Figure 5.2. Exploratory plots of natural log (ln) transformed turbidity, in formazin nephelometric units, and particulate mercury (PMeHg), in nanograms per liter.

Basic Model Statistics

```
Number of Observations                                 26
Standard error (RMSE)                                 0.7
Average Model standard percentage error (MSPE)       75.9
Coefficient of determination (R²)                   0.603
Adjusted Coefficient of Determination (Adj. R²)     0.586
Bias Correction Factor (BCF)                         1.23
```

Explanatory Variables

```
            Coefficients Standard Error t value Pr(>|t|)
(Intercept)       -3.820          0.137  -27.80 8.98e-20
lnTurbidity        0.752          0.125    6.03 3.14e-06
```

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

Two observed samples were flagged as potential outliers. The sample collected on October 28, 2014, was flagged because the difference in fits (DFFITS) value was greater than 0.555, and the sample collected on September 23, 2015, was flagged because the leverage was greater than 0.231. All flagged observations were retained in the model.

	lnPMeHg	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
10/28/2014 10:58	-5.81	-4.21	-1.600	-2.350	-2.620	0.0482	0.14000	-0.589
09/23/2015 09:41	-1.63	-1.74	0.103	0.173	0.169	0.2740	0.00565	0.104

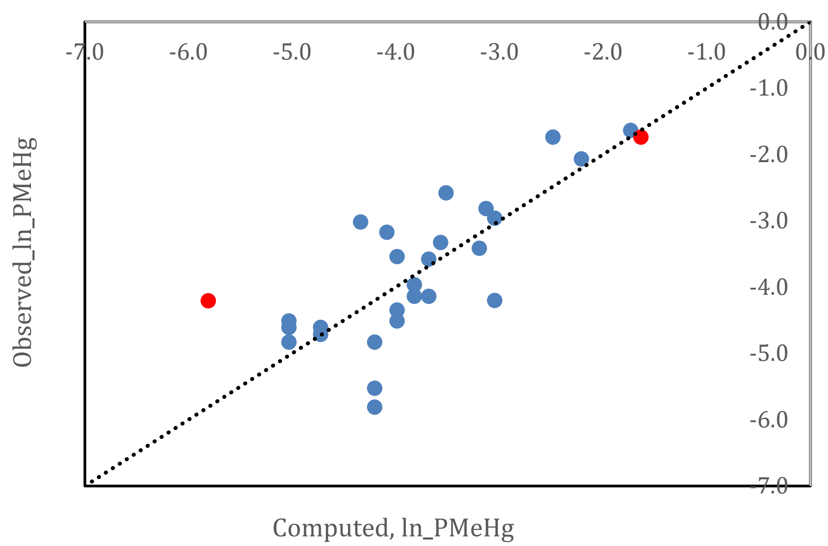


Figure 5.3. Relation between observed particulate methylmercury (PMeHg) and computed PMeHg; flagged observations are in red.

(ln, natural log)

Statistical Plots

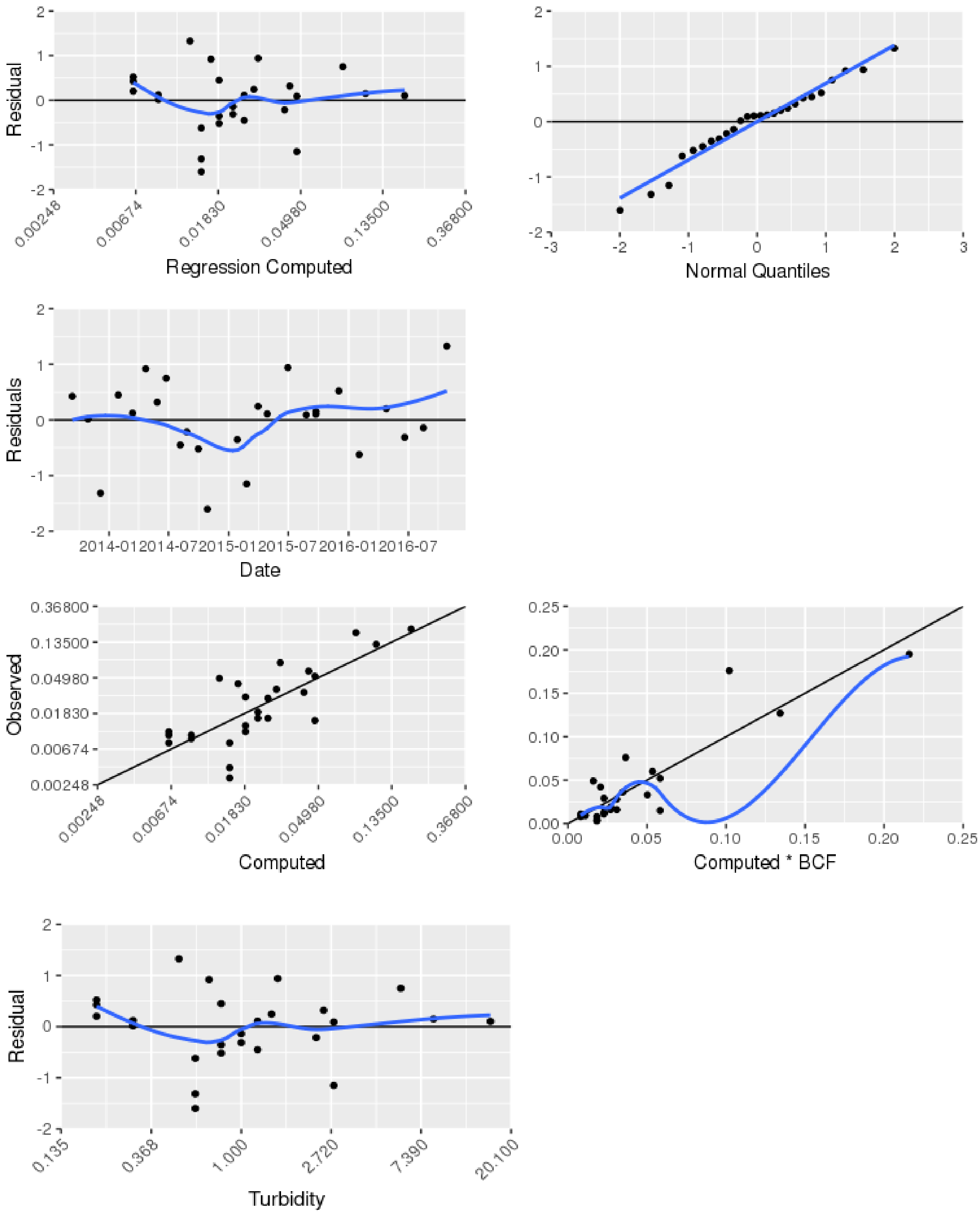


Figure 5.4. Residual and observed versus computed plots.

(BCF, bias correction factor)

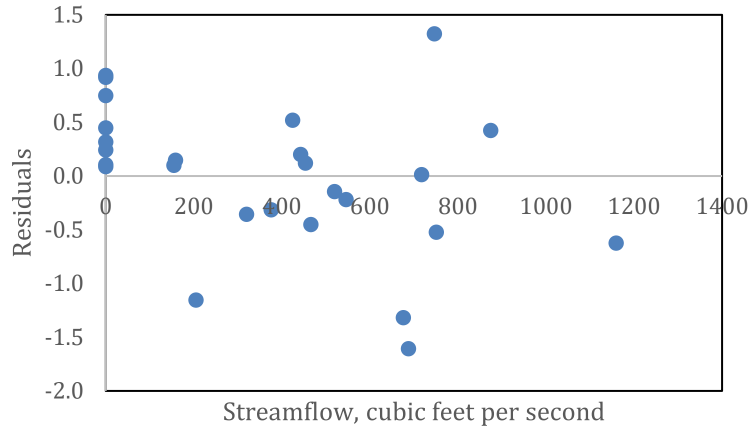


Figure 5.5. Relation between residuals and streamflow showing that the residuals had no systematic bias with respect to streamflow.

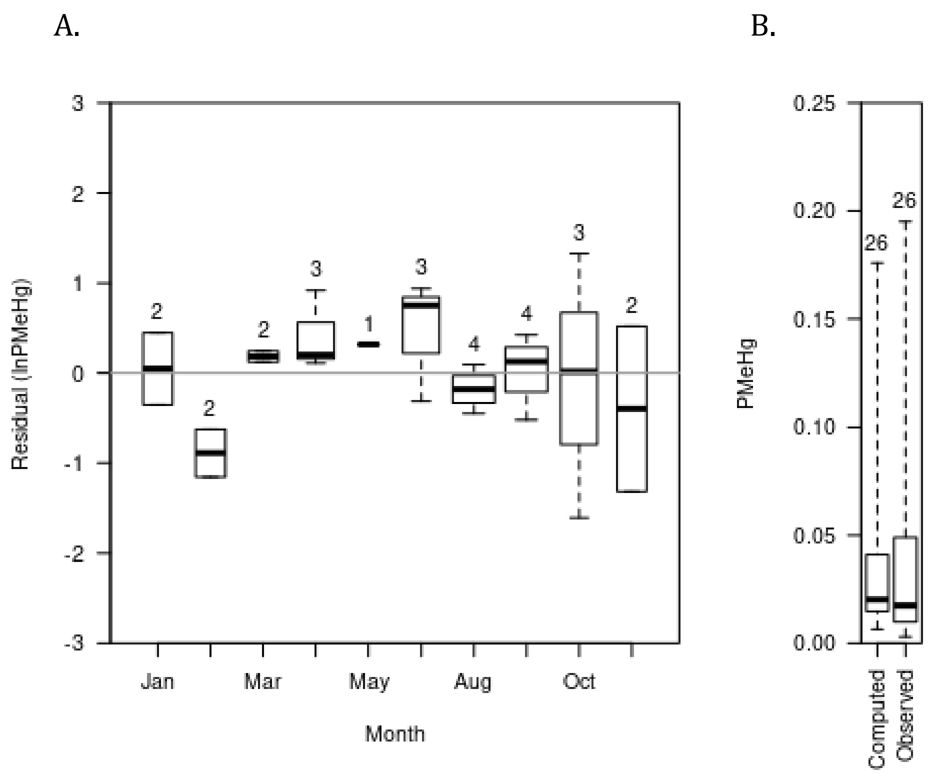


Figure 5.6. *A*, Seasonal variation in residuals of natural log of particulate methylmercury (lnPMeHg,), and *B*, computed and observed particulate methylmercury (PMeHg), in nanograms per liter.

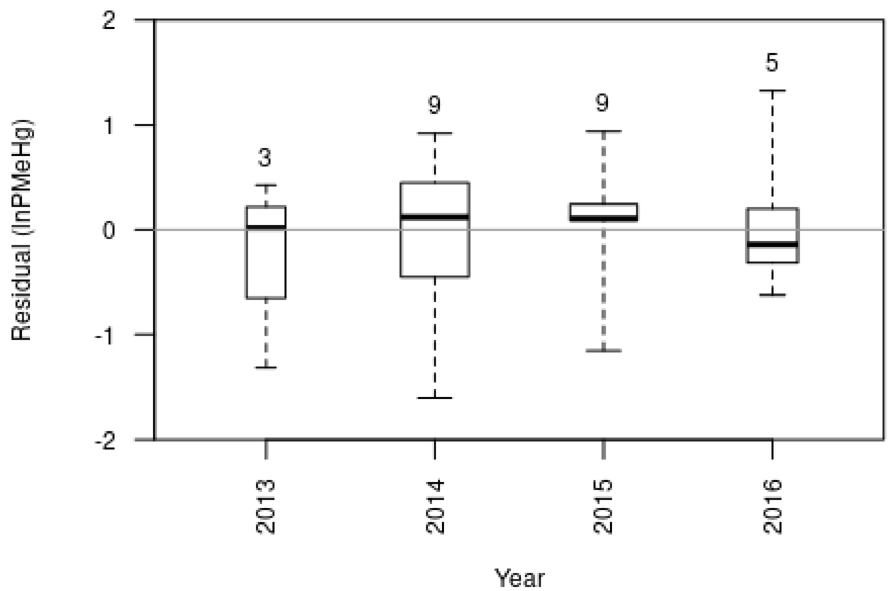
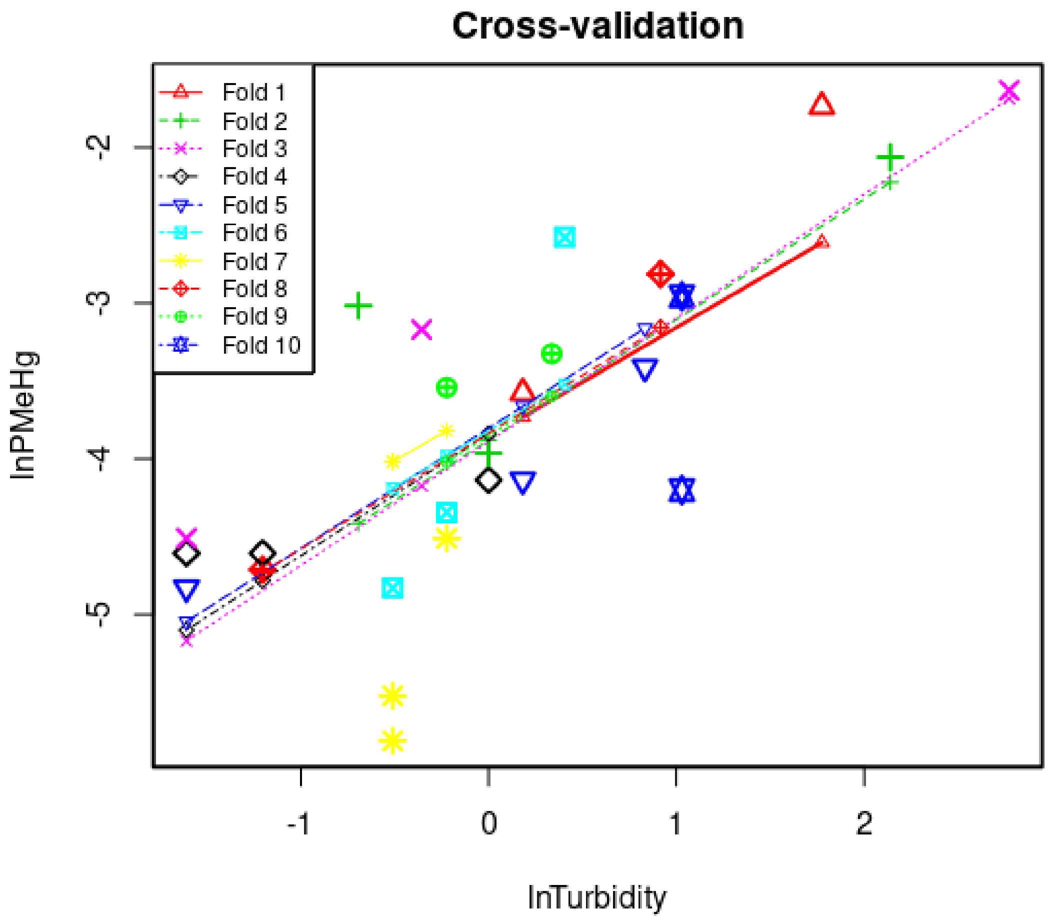


Figure 5.7. Annual variation in residuals.

(lnPMeHg, natural log of particulate methylmercury)

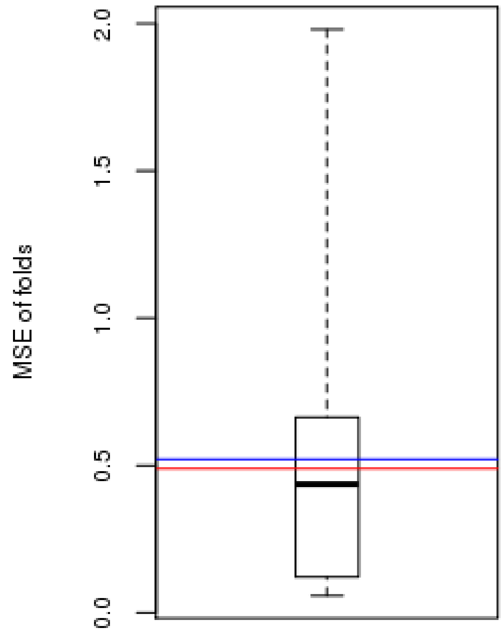
Cross Validation

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:  0.0583
                Mean MSE of folds:  0.5200
              Median MSE of folds:  0.4360
             Maximum MSE of folds:  1.9800
(Mean MSE of folds) / (Model MSE):  1.0600
```

Figure 5.8. Cross validation plot. (lnPMeHg, natural log of particulate methylmercury; lnTurbidity, natural log of turbidity; MSE, mean standard of error)




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Figure 5.9. Mean standard of error (MSE) of folds boxplot.

Model Calibration Dataset

0	Date	lnPMeHg	lnTurbidity	PMeHg	Turbidity	Computed lnPMeHg	Computed PMeHg	Residual	Normal Quantiles	Censored Values
1	2013-09-11	-4.61	-1.61	0.01	0.2	-5.03	0.00801	0.426	0.668	--
2	2013-10-29	-4.71	-1.2	0.009	0.3	-4.73	0.0109	0.016	-0.242	--
3	2013-12-06	-5.52	-0.511	0.004	0.6	-4.21	0.0183	-1.32	-1.55	< 0.009
4	2014-01-29	-3.54	-0.223	0.029	0.8	-3.99	0.0227	0.449	0.794	--
5	2014-03-14	-4.61	-1.2	0.01	0.3	-4.73	0.0109	0.121	0.144	--
6	2014-04-23	-3.17	-0.357	0.042	0.7	-4.09	0.0205	0.92	1.29	--
7	2014-05-28	-2.81	0.916	0.06	2.5	-3.13	0.0535	0.32	0.553	--
8	2014-06-24	-1.74	1.77	0.176	5.9	-2.49	0.102	0.75	1.09	--
9	2014-08-06	-4.14	0.182	0.016	1.2	-3.68	0.0308	-0.451	-0.794	--
10	2014-08-26	-3.41	0.833	0.033	2.3	-3.2	0.0502	-0.216	-0.445	--
11	2014-09-30	-4.51	-0.223	0.011	0.8	-3.99	0.0227	-0.521	-0.932	--
12	2014-10-28	-5.81	-0.511	0.003	0.6	-4.21	0.0183	-1.6	-2	< 0.007
13	2015-01-27	-4.34	-0.223	0.013	0.8	-3.99	0.0227	-0.353	-0.668	--
14	2015-02-24	-4.2	1.03	0.015	2.8	-3.05	0.0582	-1.15	-1.29	--
15	2015-03-31	-3.32	0.336	0.036	1.4	-3.57	0.0346	0.244	0.445	--
16	2015-04-28	-3.58	0.182	0.028	1.2	-3.68	0.0308	0.109	0.0479	--
17	2015-06-30	-2.58	0.405	0.076	1.5	-3.52	0.0364	0.94	1.55	--
18	2015-08-25	-2.96	1.03	0.052	2.8	-3.05	0.0582	0.0913	-0.144	--
19	2015-09-23	-2.06	2.14	0.127	8.5	-2.21	0.134	0.15	0.242	--
20	2015-09-23	-1.63	2.77	0.195	16	-1.74	0.216	0.103	-0.0479	--
21	2015-12-02	-4.51	-1.61	0.011	0.2	-5.03	0.00801	0.521	0.932	--
22	2016-02-02	-4.83	-0.511	0.008	0.6	-4.21	0.0183	-0.623	-1.09	--
23	2016-04-25	-4.83	-1.61	0.008	0.2	-5.03	0.00801	0.203	0.341	--
24	2016-06-20	-4.14	0	0.016	1	-3.82	0.0269	-0.314	-0.553	--
25	2016-08-16	-3.96	0	0.019	1	-3.82	0.0269	-0.142	-0.341	--
26	2016-10-26	-3.02	-0.693	0.049	0.5	-4.34	0.016	1.33	2	--

Model Limitations

Errors in the PMeHg surrogate model can be attributed to several factors. One source 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 PMeHg in the location of highest flow velocity. The difference between the point and average cross section turbidity ranged from 0.4 FNU 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 methylmercury, the percent recovery from the certified reference material and the check standards was provided by the USGS Mercury Research Laboratory. The average percent recovery from the certified reference material was 106 percent, ranging from 87 percent to 135 percent from October 2013 to June 2017. The average percent recovery from the check standards was 104 percent, ranging from 90 percent to 125 percent, from October 2013 to June 2017.

Definitions

```
PMeHg: Methylmercury(1+) in ng/l (62977)
Turbidity: Turbidity in FNU (63680)
```

App Version 1.0

References Cited

DeWild, J.F., Olund, S.D., Olson, M.L., and Tate, M.T., 2004, Methods for the preparation and analysis of solids and suspended solids for methylmercury: U.S. Geological Survey Techniques and Methods, book 5, chap. A7, 13 p., accessed October 1, 2017, at http://pubs.usgs.gov/tm/2005/tm5A7/.

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.

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.