Project ID: 20-0085

Modeling of Mercury Bioaccumulation in Mosquitofish (*Gambusia* sp.) for the Everglades Regional Environmental Monitoring and Assessment Program

Miami, FL

Program Dates: May 1995 – September 2014

Project Leader: Peter I. Kalla, Ph.D.
Environmental Sampling Section
Applied Science Branch
Laboratory Services & Applied Science Division
USEPA – Region 4
980 College Station Road
Athens, Georgia 30605-2720

Final Report

Requestor:

Donatto Surratt, Ph.D.

Everglades National Park

Analytical Support:

Florida International University

Battelle Laboratories

Florida Department of Environmental Protection

United States Environmental Protection Agency, Region 4, Laboratory Services Branch

Approvals:

LSASD Project Leader:

PETER KALLA Digitally signed by PETER KALLA Date: 2020.06.04 10:01:41 -04'00'

Peter I. Kalla, Ph.D.
Environmental Sampling Section
Applied Science Branch

Date

Approving Official:

STACEY BOX Digitally signed by STACEY BOX Date: 2020.06.04 10:46:09 -04'00'

Stacey Box, Chief
Environmental Sampling Section
Applied Science Branch

Date

Associate Project Leader:

DANIEL SCHEIDT Digitally signed by DANIEL SCHEIDT Date: 2020.06.04 10:24:04 -04'00'

Daniel J. Scheidt, Water Division

Date

This report is a deliverable to Everglades National Park under Interagency Agreement # RW-14-92380601. It should be cited as:

Kalla, P., M. Cyterski, J. Minucci., and D. Scheidt. 2019. Modeling of mercury bioaccumulation in mosquitofish (*Gambusia* sp.) for the Everglades Regional Environmental Monitoring and Assessment Program. U.S. EPA, Region 4, Laboratory Services and Applied Science Division, Athens, GA. LSASD 20-0085. 35 pp.

Introduction

Data on mercury contamination from the Everglades Regional Environmental Monitoring and Assessment Program (REMAP) have been modelled by previous investigators, notably Kent Thornton (Stober et al. 2001) and Curtis Pollman (2014). These efforts were undertaken approximately midway through the history of Everglades REMAP, a program that has gone on for three decades. They featured structural equation modeling, and did not include any variables that implicitly reflect habitat quality or food web complexity. This report is based on a different modelling technique applied to all REMAP data to date, and it includes consideration of habitat and resultant efficiency of biomagnification.

The objective of this work was to characterize relationships between measured covariates and mosquitofish mercury (Hg) levels, using a general statistical modeling framework that could enable prediction of variations in mosquitofish (fish) Hg concentrations across the public Everglades. This model would help identify factors that are most influential on temporal and spatial fluctuations in fish Hg. The dataset was comprised of Hg measurements in fish and other ecosystem compartments, ambient water and sediment physico-chemistry, and categorical environmental factors. The data come from hundreds of random survey points sampled by the United States Environmental Protection Agency throughout the greater freshwater Everglades over the past three decades (Figure 1).

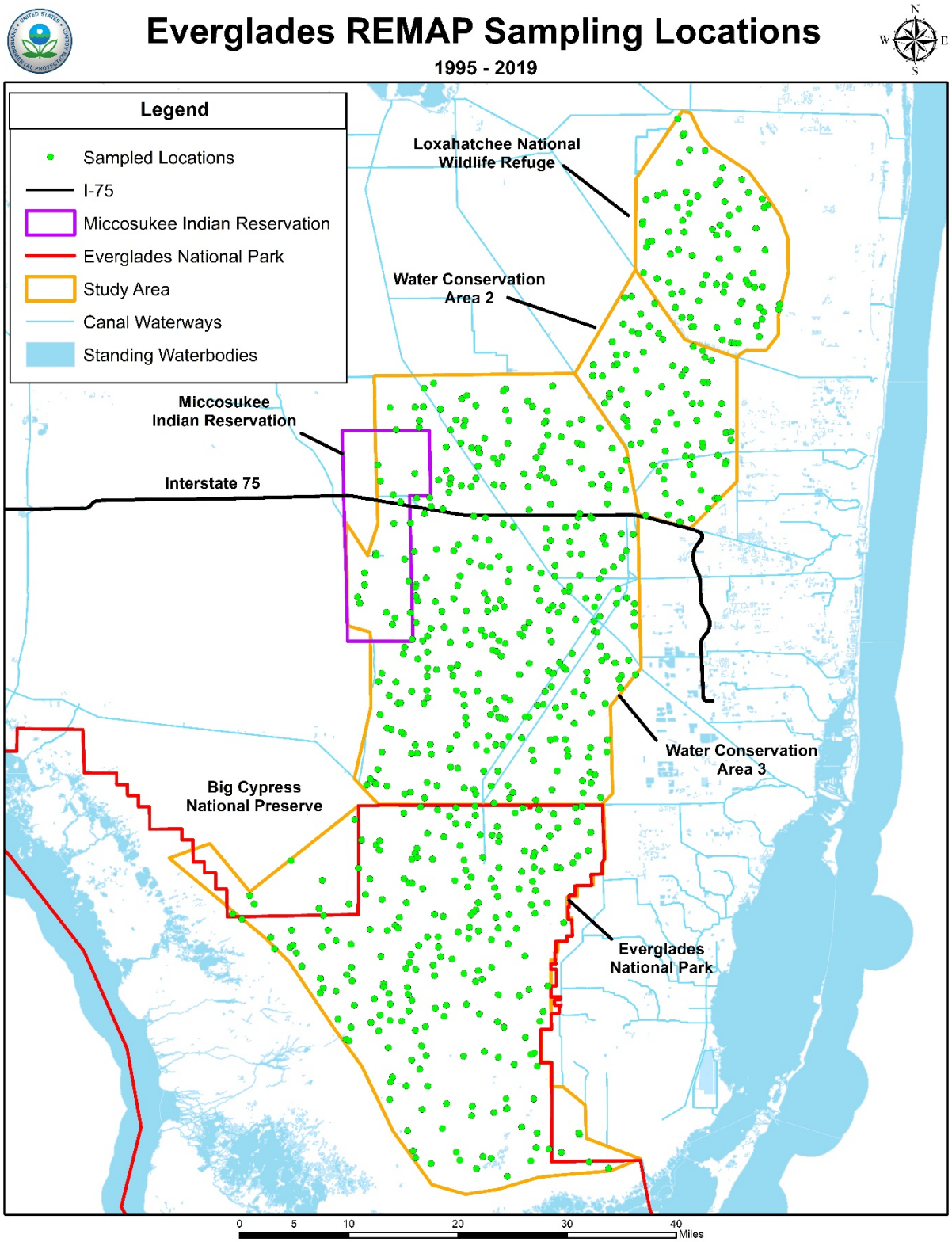

Figure 1. Survey points for Everglades REMAP where mosquitofish were sampled.

Methods

The response variable of interest was Log10 concentrations of total Hg(µg/kg) in mosquitofish from various locations in the Florida Everglades. These data (n=774) were collected in the dry and wet seasons in three different time periods: the mid- to late 1990's (1995, 1996 and 1999), 2005, and 2014 (wet season only). Our approach was to use Generalized Boosted Models (GBM), a non-linear multiple regression decision-tree-based machine learning technique (Friedman 2001) to determine which independent covariates in the collected data were most influential in determining fish mercury levels. We examined the following covariates:

- Alkaline Phosphatase: Surface Water (umole/L*hr)
- Ash Free Dry Weight: Soil (%) and Floc (%)
- Bulk Density: Soil (g/cc) and Floc (g/cc)
- Chlorophyll A: Surface Water (ug/l)
- Chloride: Surface Water (mg/l)
- Conductivity: Surface Water (uSiemens)
- Depth: Floc (m) and Surface Water (m)
- Dissolved Oxygen: Surface Water (mg/l)
- Filtered Ammonia: Surface Water (mg/l)
- Filtered Nitrate: Surface Water (mg/l)
- Filtered Nitrite: Surface Water (mg/l)
- HABCODE (nominal categorical factor): five different habitat characterizations about where the sample was collected:

  Sawgrass Marsh (n=430)
  Wet Prairie (n=279)
  Cattail (n=36)
  Slough (n=23)
  Other (n=6); 1 Pond, 4 Willow, 1 Brush

- Methyl Mercury: Soil (ug/kg), Surface Water (ng/l), Floc (ug/kg), and Periphyton (ug/kg)
- pH: Soil and Surface Water
- Redox Potential: Surface Water (mV) and Pore Water (mV)
- Season (binary categorical factor): Wet or Dry

- SUBAREA (nominal categorical factor): five different collection zones within the Everglades study area:

    Arthur R. Marshall Loxahatchee National Wildlife Refuge
    Water Conservation Area 2
    Water Conservation Area 3 north of Interstate 75
    Water Conservation Area 3 south of Interstate 75
    Everglades National Park

- Soluble Reactive Phosphorus: Surface Water(mg/l) and Pore Water (mg/l)
- Sulfate: Soil (ug/kg) and Surface Water (mg/l)
- Sulfide: Pore Water (mg/l)
- Temperature: Surface Water (C)
- Thickness: Soil (m)
- Total Carbon: Soil (%)
- Total Mercury: Soil (ug/kg), Surface Water (ng/l), Floc (ug/kg) and Periphyton (ug/kg)
- Total Nitrogen: Soil (%) and Surface Water (mg/l)
- Total Organic Carbon: Surface Water (mg/l)
- Total Phosphorus: Soil (ug/g), Surface Water (ug/l) and Floc (ug/g)
- Turbidity: Surface Water (NTU)
- Sampling Date

Using the GBM package (version 2.1.5, Greenwell et al. 2019) in R (R Core Team 2018) we developed seven different models:

- All data
- All time periods – wet season only
- All time periods excluding 2014 – dry season only
- 2005 (wet and dry season) and 2014 (wet season) only
- 1990's only
- 1990's, wet season only
- 1990's, dry season only

The data were parsed in the above manner to explore potential differences in the influence of covariates over time and season. Rainfall, discharge, and water level have varied widely during REMAP; and more stormwater runoff enters the public Everglades from the agricultural area to the north during the summer wet season (Scheidt and Kalla 2007).

Out of a total of 31734 cells in the covariate data matrix (774 samples x 41 covariates), there were 9699 (31%) blank cells, where a parameter had not been measured during a specific sampling event. The algorithm implemented in this GBM package directly incorporates these missing data into the fitting and prediction process (Greenwell et al. 2019). At each split in the regression tree, a third node is created for samples which are missing that covariate, which may lead to further splits or a leaf. Thus, the model accounts for potential correlations between absent data and the response variable (García-Laencina et al. 2010).

Due to the inherent stochasticity in the results of fitting a GBM model to a given dataset (i.e., slightly different models will be produced when fitting the same dataset), we developed bootstrapped estimates of model metrics by fitting 30 GBM models for each of these datasets. Bootstrapping was accomplished by running multiple iterations of the GBM model. In each iteration, 80% of the dataset was randomly placed into a training set to fit the model, and the remaining 20% was put into a testing set to examine model predictive capabilities.

The GBM package in R has an array of model parameters that can affect the fitting process and efficacy of the eventual solution:

- Error Term: Gaussian
- Maximum Number of Trees: 5000
- Shrinkage: 0.005
- Interaction Depth: 3
- Bag Fraction: 0.5
- Train Fraction: 1
- Number of Minimum Observations per Node: 10
- Cross-Validation Folds: 5

For an in-depth discussion of these parameters, see the GBM package documentation (Greenwell et al., 2019). Recommendations within Greenwell et

al. 2019 were followed for setting parameter values, with some modifications based on best professional judgement. Train Fraction was set to 1 because a true testing dataset was created at the start of each iteration for assessment of model predictive capabilities. In addition, out-of-sample error would be handled by examining cross-validation folds, as explained later. A Bag Fraction of 0.5 results in each successive tree in the iterative algorithm being fit to a random 50% of the observations in the training dataset, which mitigates overfitting of the training data. An Interaction Depth of 3 means that up to third-order interactions of model covariates can be captured by the model. We did not include interactions of an order greater than 3 to preserve model interpretability. The Minimum Number of Observations per Node value prevents the model from being unduly influenced by outliers or clusters of odd samples. For datasets with n>500, 10 was used for this parameter. For datasets with n between 100 and 500, a value of 5 was used. Smaller values of the Shrinkage parameter can increase model accuracy, but at the cost of increased computational time and more trees in the optimal solution. Values between 0.01 and 0.001 are recommended; we used 0.005.

As more trees are added to a GBM solution, the training data error (RMSE) will continue to decline; the RMSE of out-of-sample data also initially declines as more trees are added, but then rises if too many are used, i.e., the model becomes overfit. There are several ways to determine the optimum number of trees in the GBM solution; we used 5-fold cross-validation to measure the point at which out-of-sample RMSE began to rise.

Due to the stochasticity in creating training/testing datasets, a Bag Fraction < 1, and the random aliquoting of the training data into cross-validation folds, sometimes a GBM model can produce a poor solution. In each iteration of the bootstrap process, we used two metrics to ensure a GBM model was “valid”:

- The optimum solution had less than 5000 trees (maximum allowable, meaning convergence was achieved).
- The number of unique fitted values produced by the model was at least 25% of the total number of fitted values.

One hallmark of a poorly-fit GBM model is a solution with few trees, leading to a very low number of unique fitted values. However, it is also possible to

sometimes reach a good solution with a relatively small number of trees. Therefore, it is more robust to assess the quality of the model by examining the number of unique fitted values than the number of trees in the solution.

The bootstrap algorithm was run until 30 "valid" GBM models were produced, and then examined their characteristics as follows:

- R^2 of Actual Observations versus Model Fits for the Training Data
- R^2 of Actual Observations versus Model Fits for the Testing Data
- The Influence of the Covariates

Partial Dependence Plots

In order to visually inspect how each covariate influences the response variable in a GBM model, the GBM package in R provides the ability to create Partial Dependence Plots (PDPs), which show the univariate relationship between a covariate (values plotted on the X-axis) and the response variable (values plotted on the Y-axis), while factoring out the effect of all other covariates. In a default GBM model, the PDP for any given covariate can be very complex, rising and falling abruptly over the entire range of covariate values. However, the modeler may invoke a parameter named "var.monotone" in the GBM model call, which can constrain each covariate such that its relationship to the response variable must be monotonically increasing, monotonically decreasing, or unconstrained (the default case). This can produce a simplified, more interpretable model, and may or may not reduce the fit of the model (R^2) to a significant extent. Because an aggregate PDP over multiple GBM models cannot be created, we had a different approach for developing PDPs for a single model fit to the Overall dataset:

1) Fit a model using the covariates with highest average influence across the 30 GBM models created previously for the Overall dataset.
2) Examine the PDPs for these covariates to determine if their response relationship could be simplified as monotonically increasing or monotonically decreasing.

3) Develop a second GBM model using the "var.monotone" parameter to indicate what type of relationship to use for each covariate, as identified in Step 2.
4) Plot the PDPs from this model.

Results

Table 1 summarizes the results of 30 GBM models developed on each of the seven different datasets.

Table 1. R-squared values for training and testing data for each of the modeled datasets, with standard deviations in parentheses.

Model	n	R-Squared Training	R-Squared Testing
Overall	774	0.88 (0.02)	0.59 (0.05)
Overall, Wet Season	520	0.91 (0.03)	0.61 (0.06)
Overall, Dry Season	254	0.85 (0.04)	0.56 (0.07)
2005 and 2014	275	0.85 (0.03)	0.38 (0.09)
1990's, All Data	499	0.91 (0.03)	0.51 (0.09)
1990's, Wet Season	305	0.88 (0.04)	0.47 (0.08)
1990's Dry Season	194	0.87 (0.05)	0.46 (0.12)

For each of these datasets, the average R^2 values for the training data were, as expected, higher than the average R^2 values for the testing data, indicating some overfitting of the training data. In Figure 1 (the models fit to the Overall dataset), the GBM model for each iteration was used to produce a fitted value for all data points, regardless of whether they were in the training or testing data for that iteration. Then, the average prediction for all data points across the 30 models was computed.

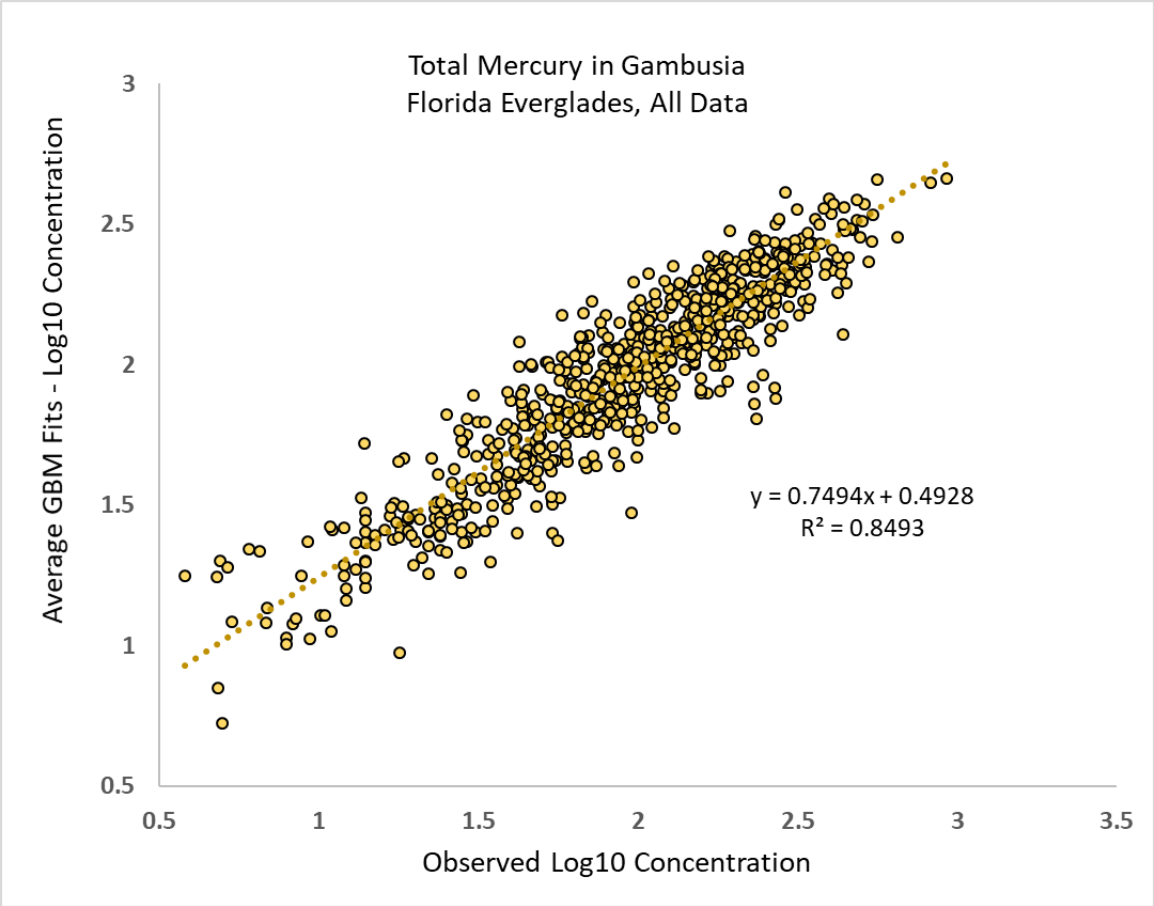

Figure 2. Plot of GBM average fitted values (across 30 models) versus observations for the overall dataset (n=774).

The R^2 of this relationship is close to the average R^2 of the training data, even though these points represent a mixture of models for which any data point had a 20% probability of being in the testing data.

One common occurrence when using GBM to model data is the muted scale of the GBM fitted values, as seen by the slope <1 in the regression line shown in Figure 1. GBM generally does a good job of fitting the pattern of the observations, but it has difficultly with their magnitudes. The largest GBM predictions typically are not as large as the largest observations and are not as small as the smallest observations. The slope and intercept of the best-fit linear regression line can be used to rectify this situation by rotating the data ellipse, expanding the scale of the GBM fitted values. If the regression line intercept in Figure 1 (0.4928) is

subtracted from every y-value, and then the result is divided by the slope (0.7494), an "adjusted" GBM fit is calculated. When these adjusted values are plotted versus observations (the blue dots in Figure 2), the R^2 value remains unchanged, but the slope and intercept become 1 and 0, as would be ideal for a plot of model fitted values versus observations. The data ellipse has simply been rotated about the mean fitted GBM value. Any original GBM fitted value above the mean becomes larger, while any original GBM fitted value below the mean becomes smaller.

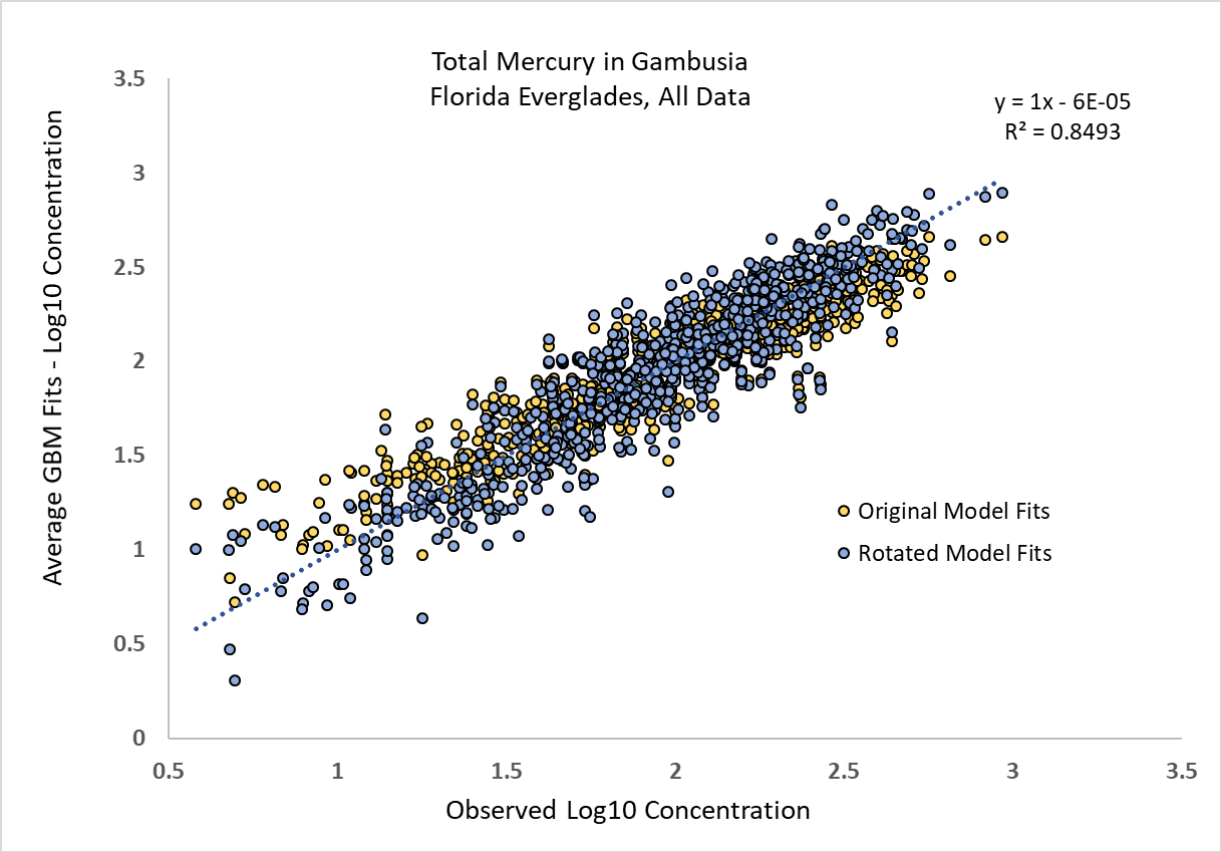

Figure 3. Plot of the average fitted GBM values of the Overall model, rotated using the slope and intercept of the regression line of Figure 1.

Finally, we examined the influence of the various covariates on these GBM models (Table 2). There was variability from iteration to iteration within a dataset, as well as between-dataset variability in the most influential covariates. For each

of the seven datasets, we first calculated the mean influence of each covariate across the 30 GBM models. Then, we filtered the covariates by their mean influence, retaining only those that accounted for at least 2% of the total influence for that dataset. Finally, we summed up covariate influence across the 7 datasets (700 would be the maximum possible value of this sum) as a measure of overall significance.

Table 2. Summary of the influence of covariates to the various fitted GBM models. The last two columns are the summed and average % influence of each covariate in the seven different datasets for which models were developed. This sum could be 700 at most if a certain covariate had complete influence in all seven models. The dotted line separates those variables that had a summed influence > 10 across the seven datasets.

Variable	Overall	Overall, Wet	Overall, Dry	2005, 2014	1990s	1990s, Wet	1990s, Dry	# of Models	Summed Influence	Average Influence
Alkaline Phosphatase, Surface Water	9.5	25.1	5.9	7.1	8.6	8.9	2.3	7	67.5	9.6
Methyl Mercury, Surface Water	7.1	8.5	7.2	4.3	9.5	15.4	6.0	7	58.0	8.3
Methyl Mercury, Soil	3.9	3.5	15.0		7.0	3.7	19.4	6	52.5	7.5
Methyl Mercury, Periphyton	7.2	3.6	6.7	6.5	6.9	3.5	9.6	7	43.9	6.3
Conductivity, Surface Water	5.1	3.1	5.1	2.1	7.1	7.3	5.2	7	35.1	5.0
Sulfate, Surface Water	4.7	2.5	5.4	5.2	4.6	3.9	7.8	7	34.2	4.9
Total Mercury, Surface Water	3.2	2.2	5.5	4.0	3.4	2.6	5.0	7	25.9	3.7
Temperature, Surface Water	7.5	3.3		3.0	5.7	5.2		5	24.7	3.5
Total Phosphorus, Surface Water	2.5	2.9		3.9	3.8	6.6	3.1	6	22.8	3.3
Sulfate, Soil	6.0		7.8		2.2		6.2	4	22.2	3.2
Dissolved Oxygen, Surface Water		3.1		2.0	4.4	8.6		4	18.1	2.6
Total Organic Carbon, Surface Water	2.9				5.9	3.1	4.7	4	16.6	2.4
Redox Potential, Pore Water			3.7		4.2	3.6	3.9	4	15.4	2.2
Total Phosphorus, Soil		2.0		2.1	2.7	2.8	2.2	5	11.9	1.7
HABCODE	3.4	2.5			3.8	2.1		4	11.8	1.7
Chlorophyll A, Surface Water	3.8	2.4		4.7				3	10.8	1.5
Sampling Date			5.5				4.9	2	10.4	1.5
SUBAREA			4.2				4.6	2	8.8	1.3
Redox Potential, Surface Water		2.4			3.4	2.3		3	8.1	1.2
Total Mercury, Soil		2.7			2.0	3.0		3	7.8	1.1
Water Depth		3.2		2.0		2.5		3	7.7	1.1
Total Nitrogen, Soil			4.0	2.2				2	6.2	0.88
Soluble Reactive Phosphorus, Pore Water				5.1				1	5.1	0.72
Filtered Ammonia, Surface Water		2.0		2.5				2	4.6	0.65
Sulfide, Pore Water			2.3	2.1				2	4.4	0.63
pH, Soil	3.2							1	3.2	0.45
Filtered Nitrite, Surface Water				3.1				1	3.1	0.45
Filtered Nitrate, Surface Water				2.9				1	2.9	0.41
Methyl Mercury, Floc				2.7				1	2.7	0.39
Total Phosphorus, Floc				2.5				1	2.5	0.35
Total Mercury, Periphyton				2.3				1	2.3	0.33
Turbidity, Surface Water						2.0		1	2.0	0.29

Alkaline Phosphatase in Surface Water had the most influence summed across the 7 datasets (67.5 of possible 700, or 9.6% average total influence). The 25.1% influence for the Overall wet season was the highest for any covariate in the seven datasets. In addition to that covariate, there were 5 others that had > 2% influence in all seven models: Methyl Mercury in Surface Water (8.3% average influence), Methyl Mercury in Periphyton (6.3%), Surface Water Conductivity (5.0%), Sulfate in Surface Water (4.9%), and Total Mercury in Surface Water (3.7%). Those covariates above the dotted line in Table 2 had a summed influence of at least 10.

Partial Dependence Plots

For the PDP analysis, 14 covariates were used that had >2% average influence across the 30 GBM models fit to the Overall dataset (i.e., the "Overall" column of Table 2, n = 774). After preliminary examination of the PDPs in a model where all parameters were unconstrained, we retained an unconstrained definition for 6 of the 14 parameters in the final model, because their preliminary PDP was deemed to be ill-suited to fitting via a monotonically-increasing or decreasing function. The plots below (Figures 4a-n) show the PDPs for these 14 most influential covariates in the final model.

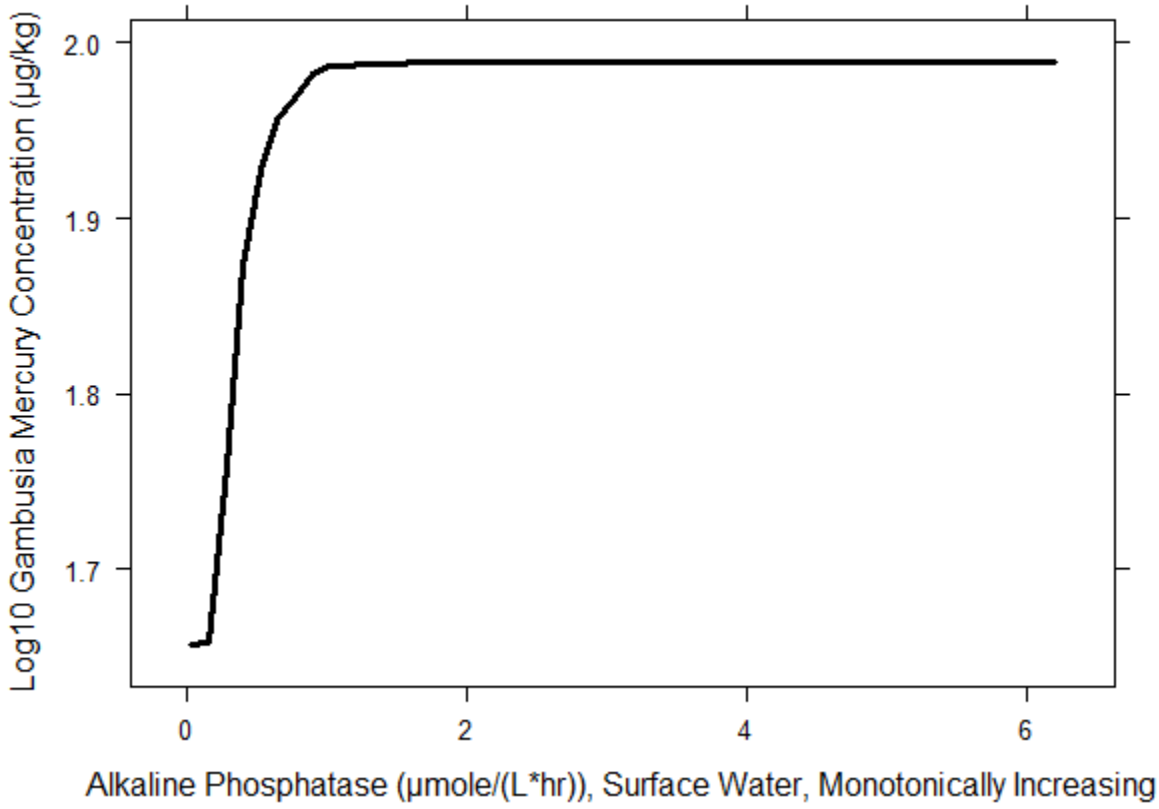

Figure 4a. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on alkaline phosphatase activity (µmole L^{-1} hr^{-1}) in surface water. The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

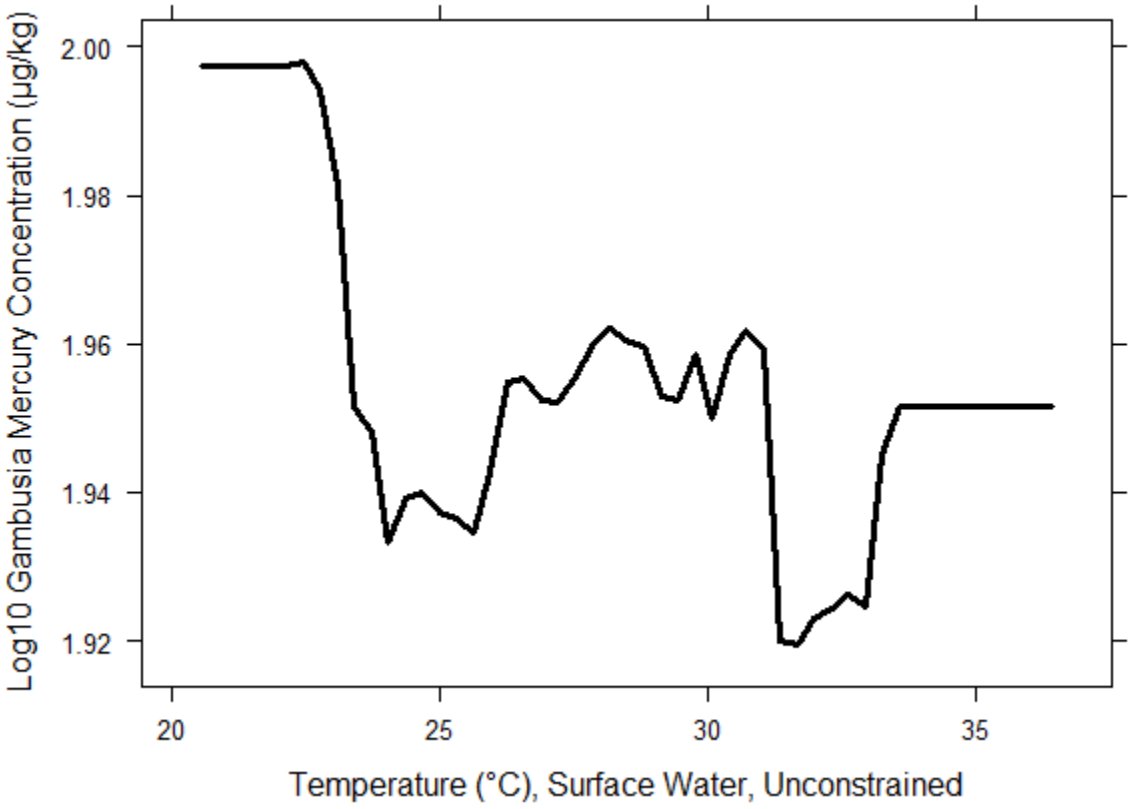

Figure 4b. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on surface water temperature (°C). The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

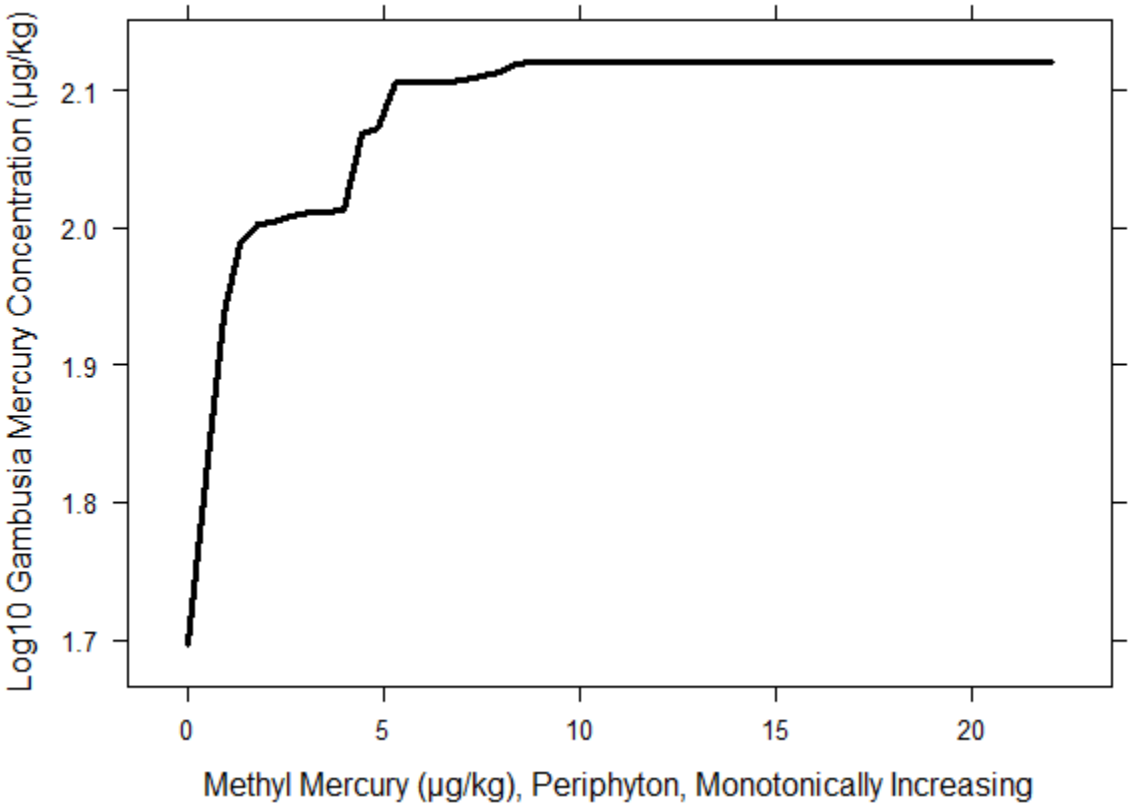

Figure 4c. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on methyl mercury in periphyton (ug/kg). The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

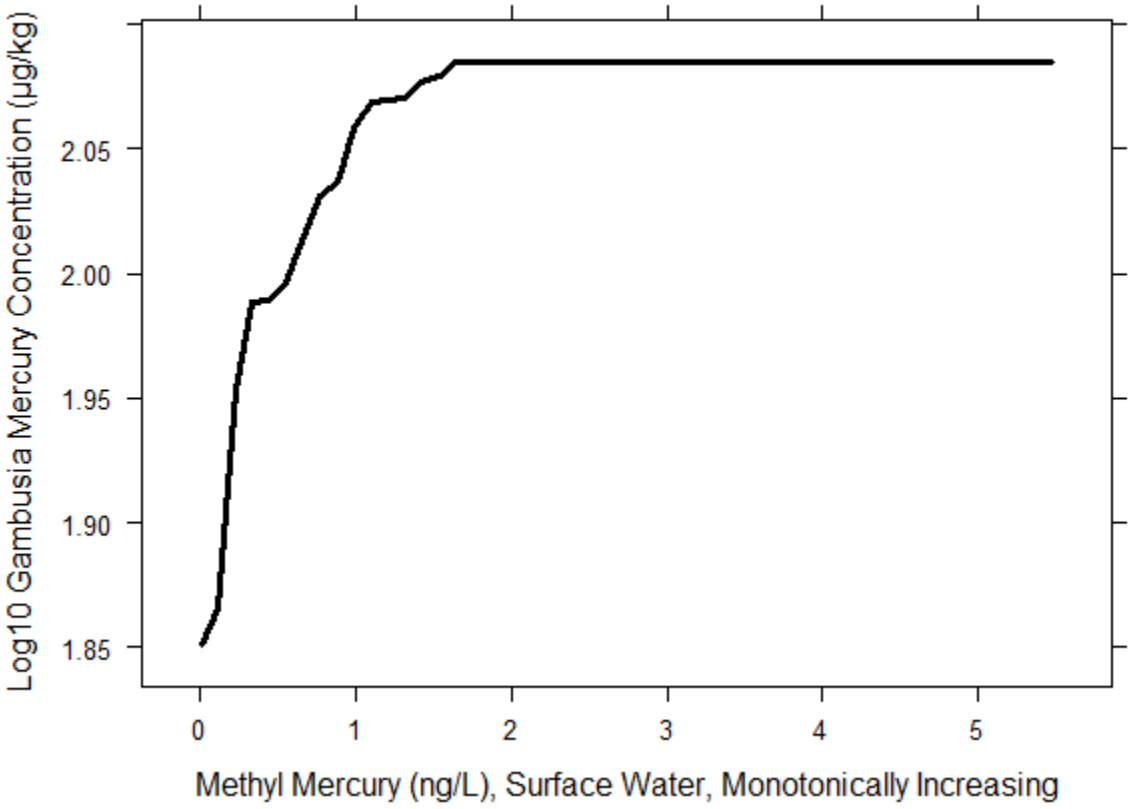

Figure 4d. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on methyl mercury in surface water (ng L^{-1}). The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

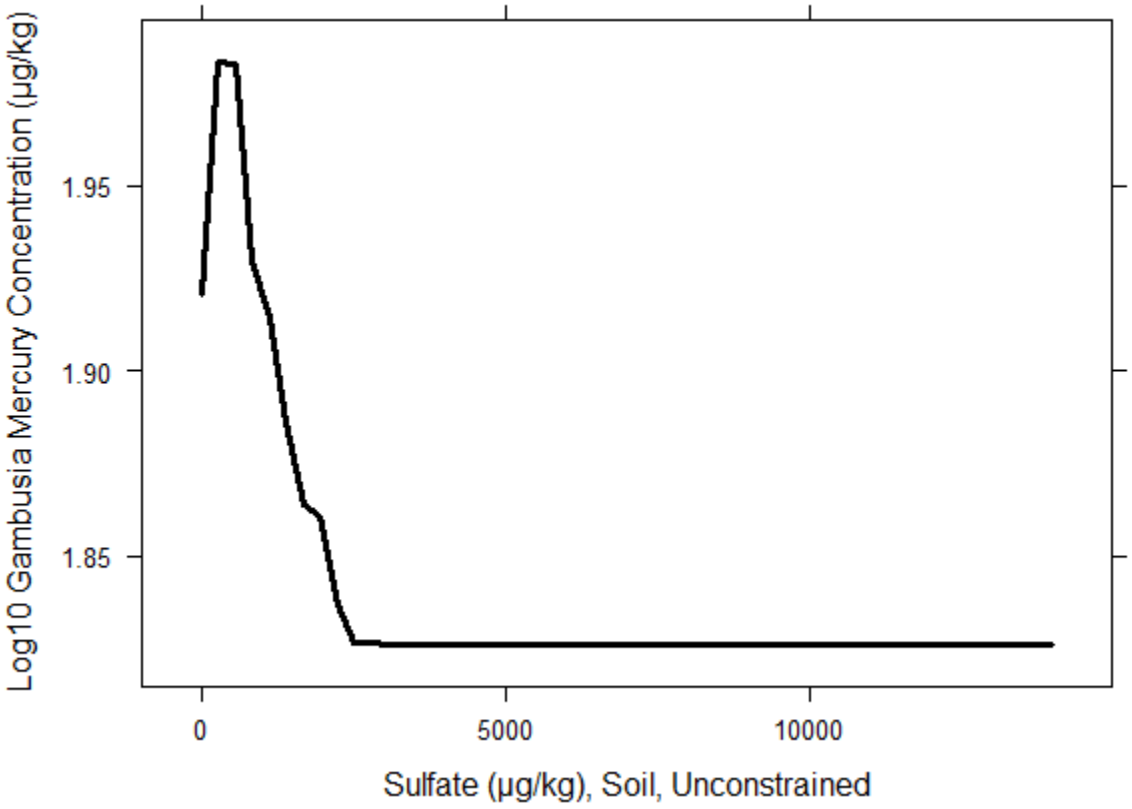

Figure 4e. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on sulfate (µg kg^{-1}) in soil. The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

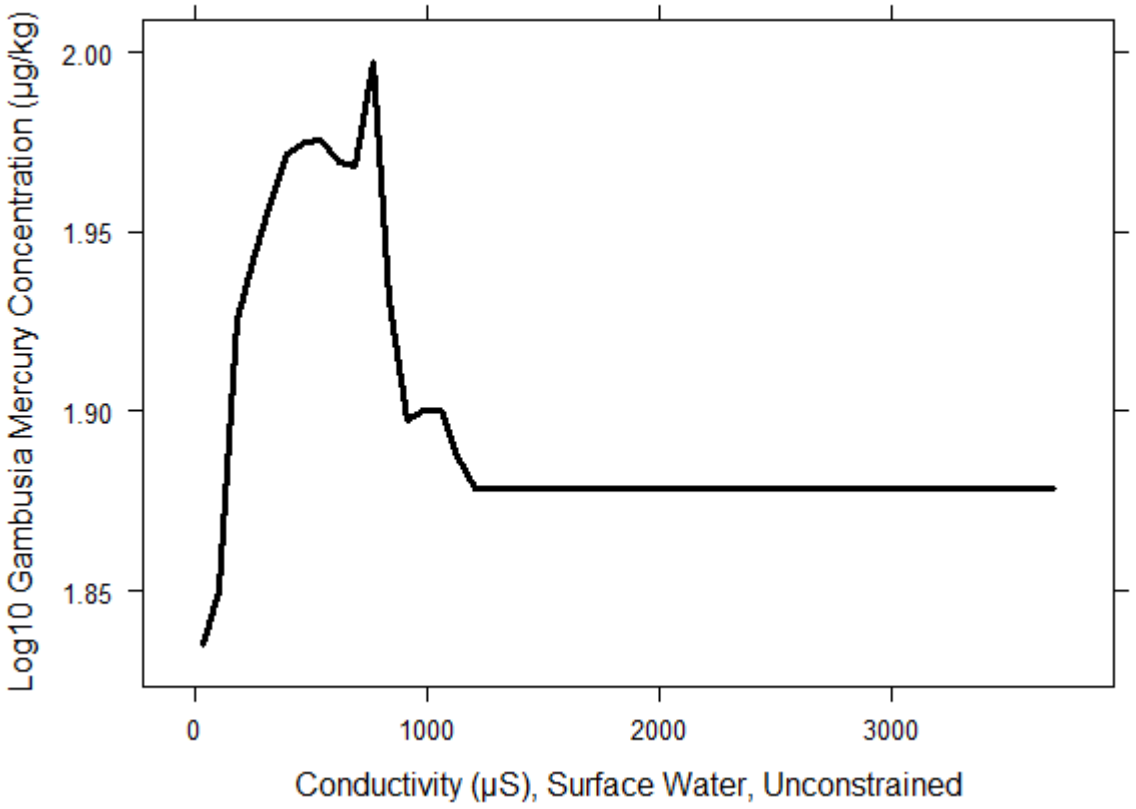

Figure 4f. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on surface water conductivity (uSiemans). The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

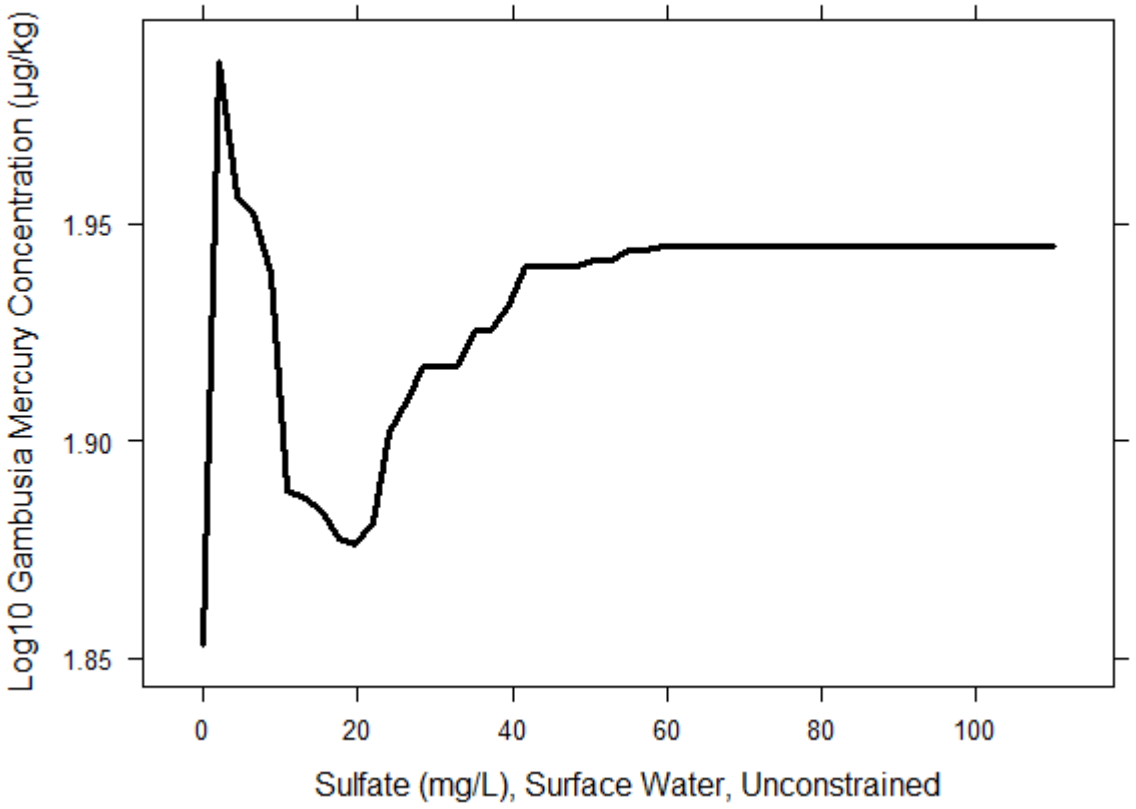

Figure 4g. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on sulfate(mg L^{-1}) in surface water. The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

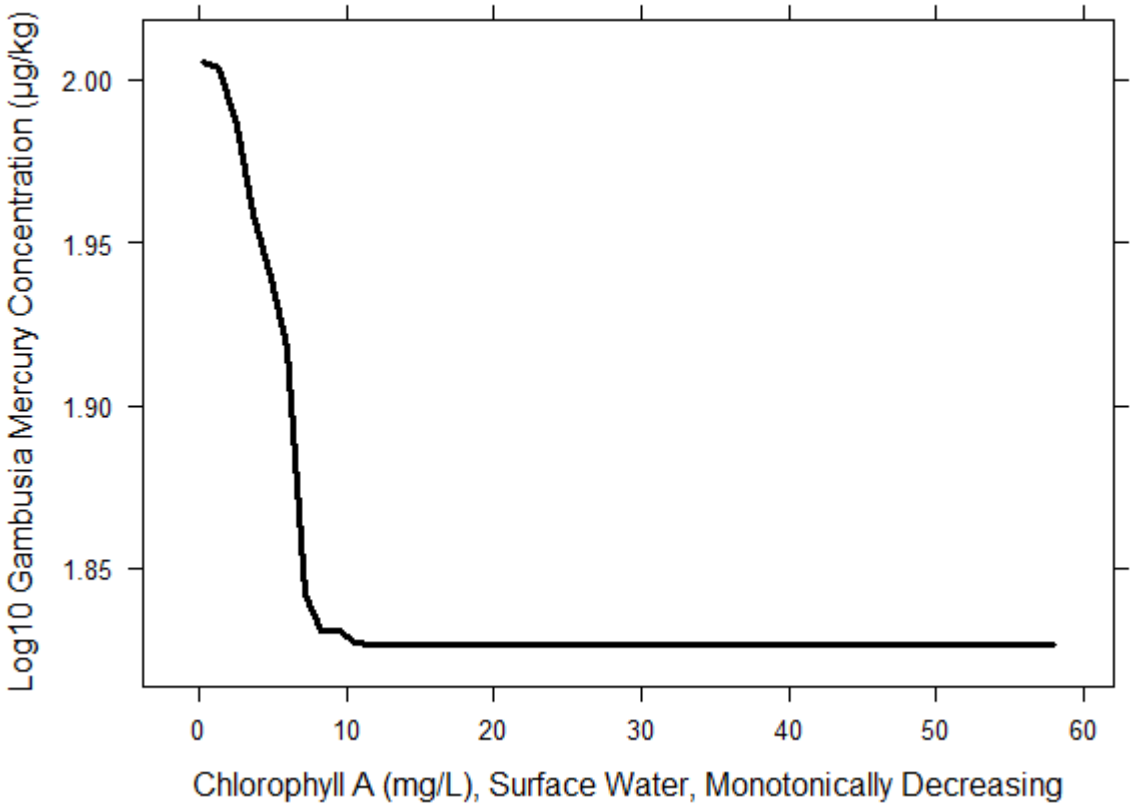

Figure 4h. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on chlorophyll-*a* (ug L^{-1})in surface water. The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

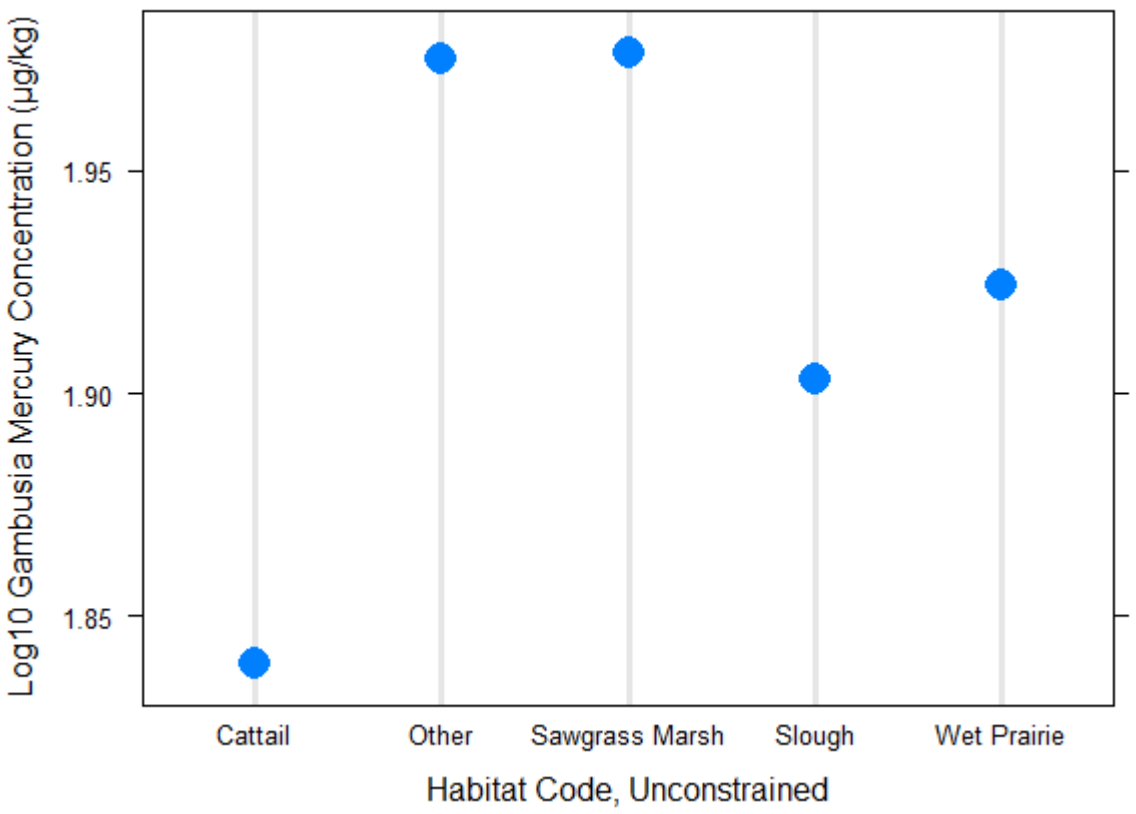

Figure 4i. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on habitat code.

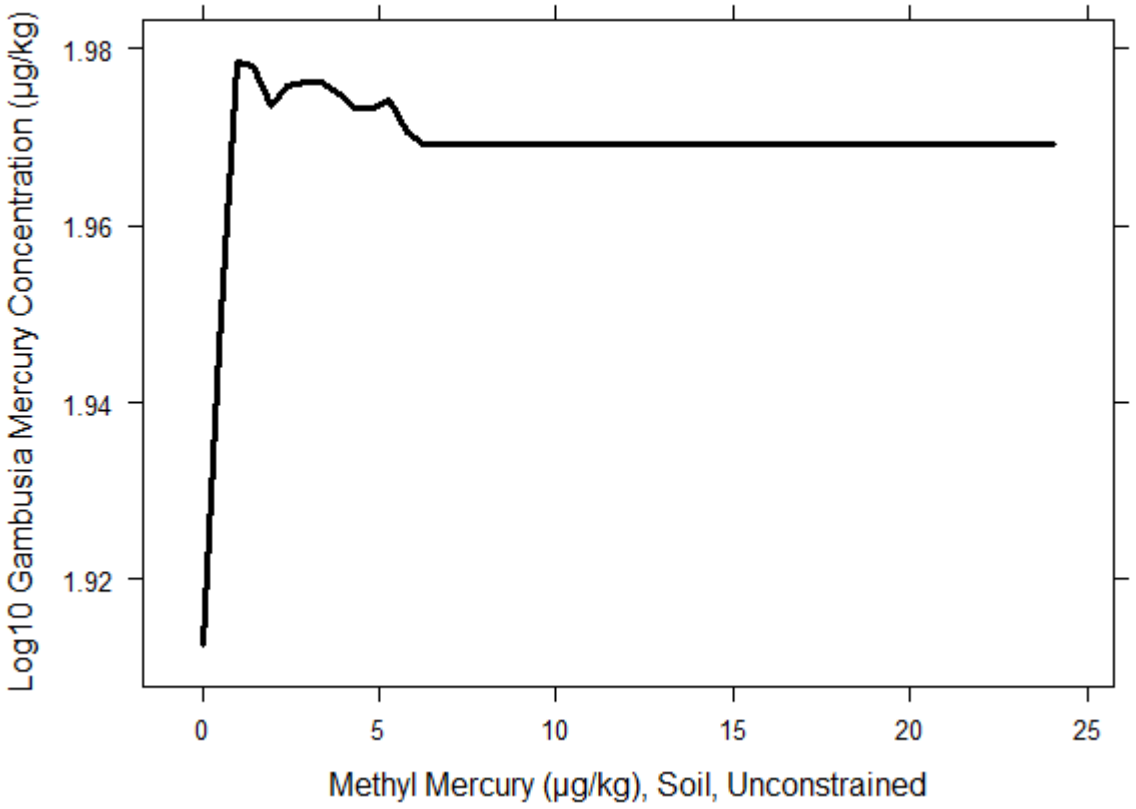

Figure 4j. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on methyl mercury (µg kg^{-1}) in soil. The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

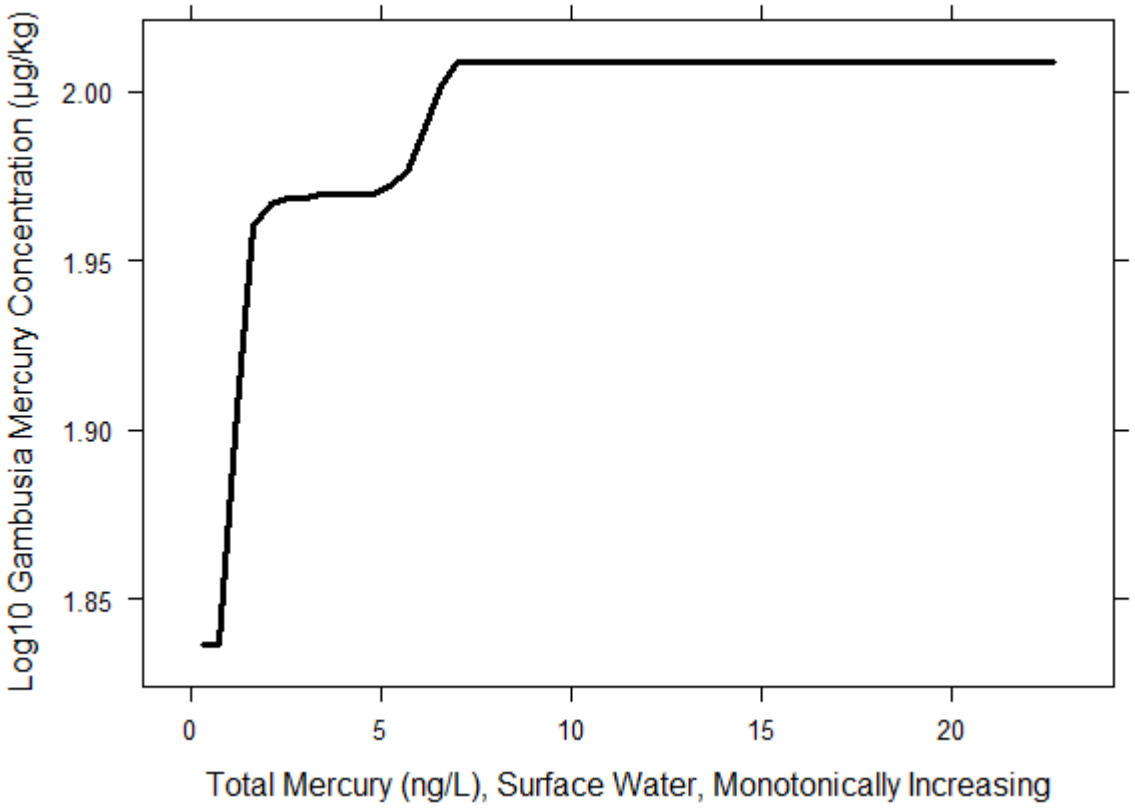

Figure 4k. PDP of mosquitofish Log_{10} mercury on total mercury in surface water (ng/L). The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

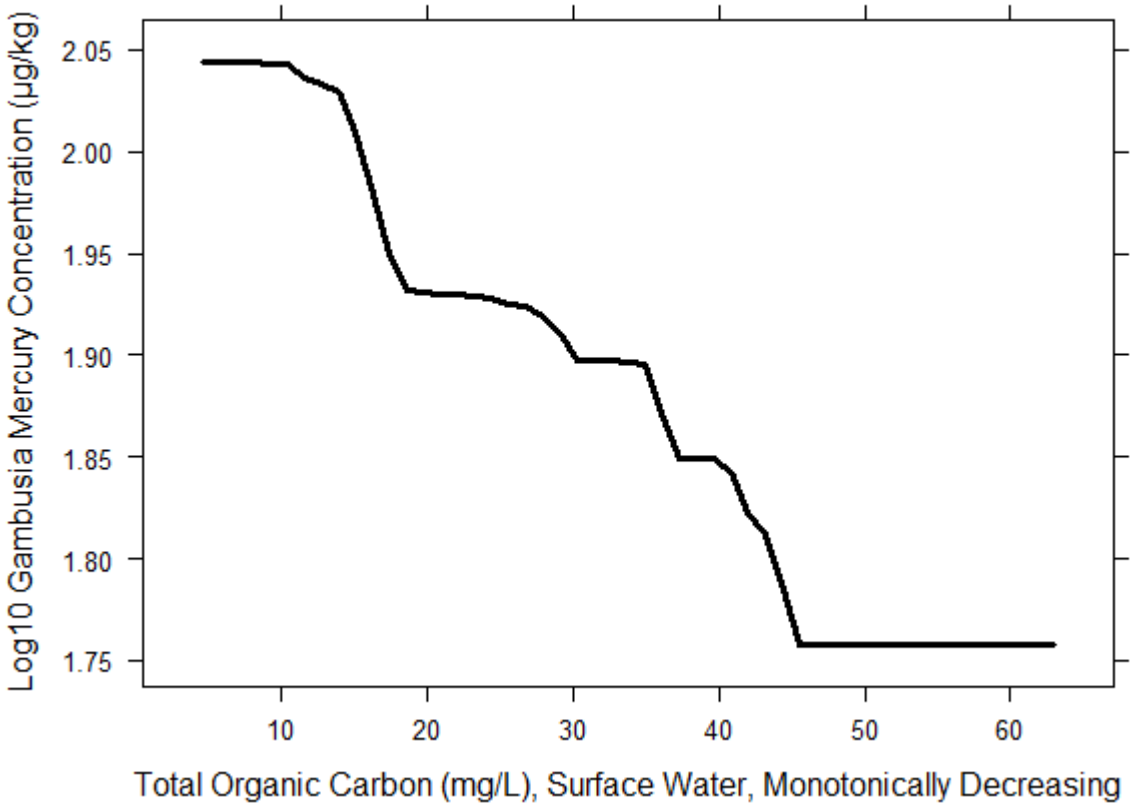

Figure 4l. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on total organic carbon (mg L^{-1}) in surface water. The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

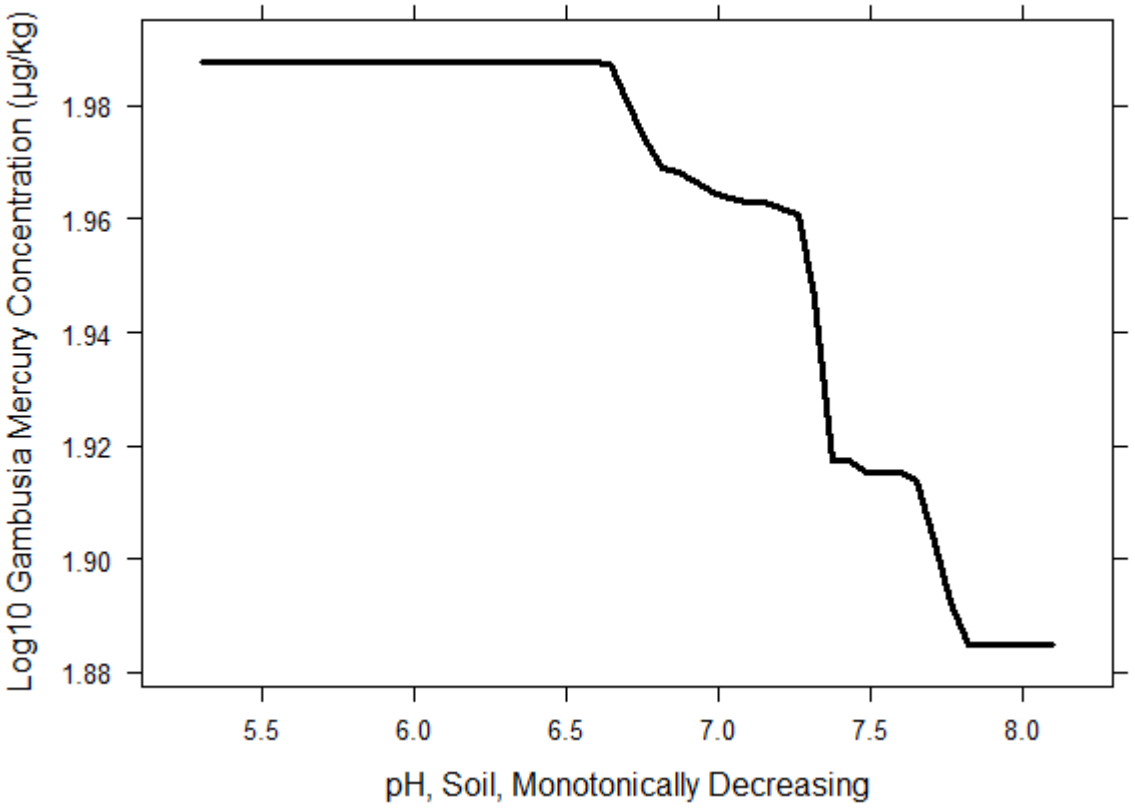

Figure 4m. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on soil pH (S.U.). The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

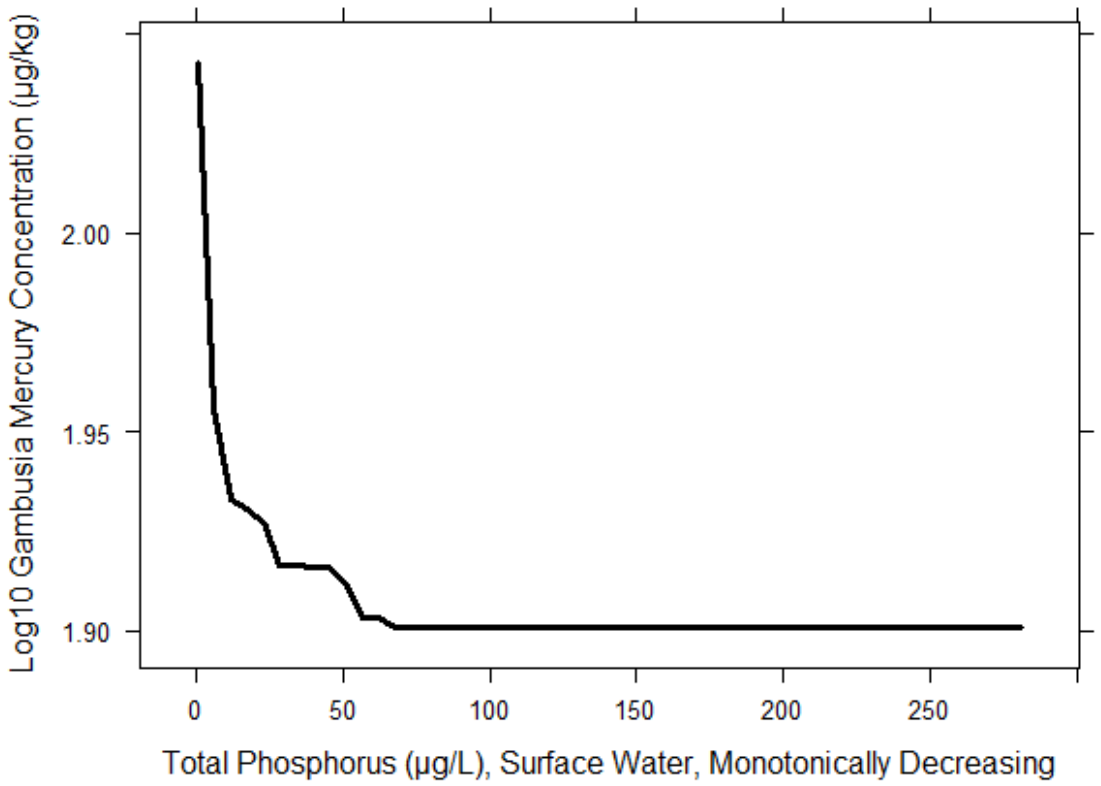

Figure 4n. PDP of mosquitofish Log_{10} mercury (ug/kg) on total phosphorus in surface water (ug/L). The x-axis label indicates whether or not the covariate was constrained using the "var.monotone" parameter in the GBM model statement.

Discussion

The R^2 Testing values in Table 1 indicate a model that accounts for about one-half of the environmental influence on variation in mosquitofish mercury (for example, Overall model R-square testing = 0.59). This outcome is robust for a large, disturbed ecosystem such as the Everglades, given its seasonal, annual, and spatial differences.

The PDP for surface water sulfate (Figure 4g) shows an increase in fish Hg between 20 and 40 mg/l, which contradicts previous reports that the highest mercury methylation occurs in a sulfate range of 1-20 mg/L, or even 2-10 mg/L,

with no increase above 20 mg/L (reviewed in Orem et al. 2011). However, the PDPs of covariates in a GBM are influenced by the other covariates in a model, as their effect is "integrated out" in order to produce the PDP for a specific covariate. In a simple GBM model with only Surface Water Sulfate and Surface Water Temperature, the PDP for Surface Water Sulfate (Figure 5) is closer to what would be expected based on theory. Nevertheless, our model does suggest that the influence of surface water sulfate on fish Hg extends further up the range of sulfate, beyond 20 mg/L as reported by Orem et al. 2011.

The PDP for sulfate in soil (Figure 4e) has a spike at the bottom of the range, consistent with a unimodal relationship between sulfur and mercury. This relationship could be explained by the activity of sulfate-reducing bacteria, which methylate mercury (into its bioavailable form), until inhibition by moderate to high levels of sulfide occurs (Orem et al. 2011). Since methylation in the Everglades can occur in the oxygen-depleted environment of the soil-water interface, soil sulfate can be expected to influence fish Hg.

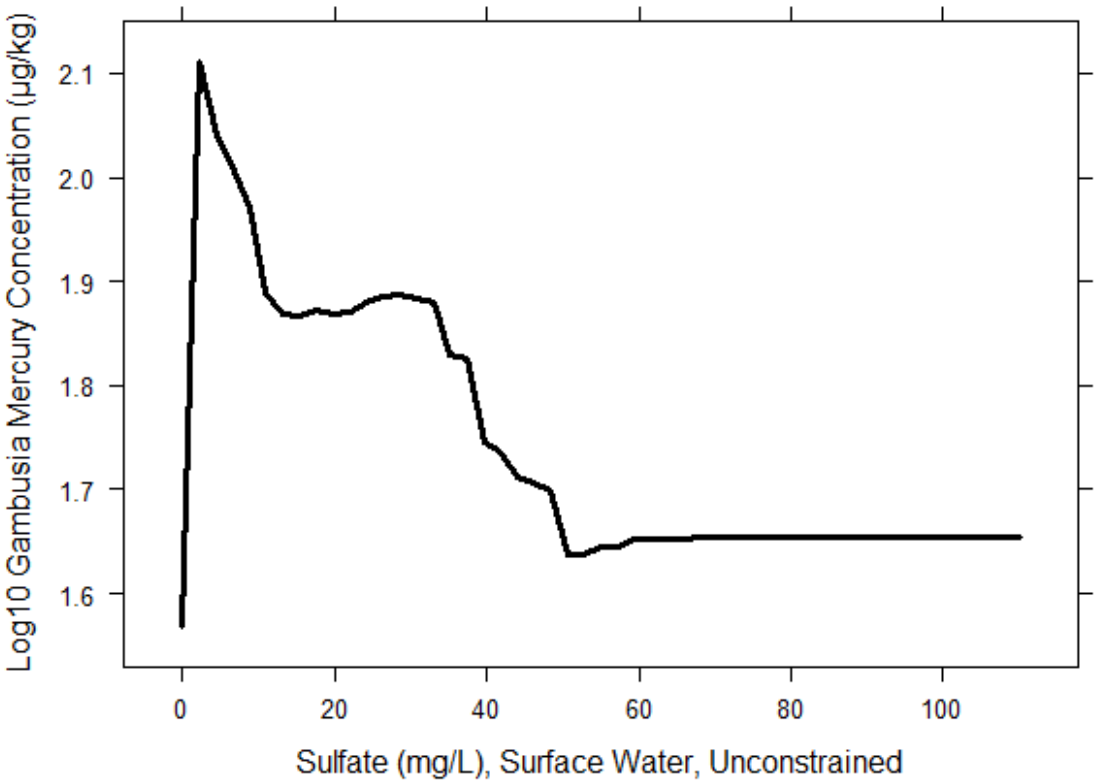

Figure 5. PDP of mosquitofish Log_{10} mercury (µg kg^{-1}) on surface water sulfate (mg L^{-1}) in a model fit to the the Overall dataset, with surface water temperature as the only other covariate in the model.

Other Partial Dependence Plots

The curves for mercury (methyl mercury in periphyton Figure 4c, methyl mercury in surface water Figure 4d, methyl mercury in soil Figure 4j, total mercury in surface water Figure 4k) show a largely straightforward relationship with mosquitofish. Inorganic mercury enters the water from the atmosphere, gets methylated, and then is taken up by biota, including periphyton, that form the food web leading to fish. The curve for soil shows a slight drop after the rise probably because the parts of the system with the most methyl mercury in the soil (Stober et al. 2001) are the same places where the food web has been degraded. Biomagnification is less in these areas (Scheidt and Kalla 2007). The

methyl mercury that cannot be efficiently incorporated into the food web is instead sequestered in the soil.

The curve for conductivity (Figure 4f) resembles the curves for sulfate in Figures 4e and 5, in that it is unimodal. The peak may be spatially associated with optimal levels of other constituents needed for efficient methylation, namely those of organic carbon and sulfur.

The curve for soil pH (Figure 4m) drops off at about 6.5. The portion between 6.0 and 6.5 may have been influenced by the historically high fish Hg concentrations (Stober et al. 2001) found in southwestern Water Conservation Area 3. However, all soil pH values less than 6.0 occur in the Refuge, which has low mosquitofsh mercury (Scheidt and Kalla 2007).

The curve for water temperature (Figure 4b), though erratic, has the general form of lower fish Hg levels at higher temperatures. This result may be explained by sunlight. Higher temperatures tend to be found in wet prairies and sloughs during daylight because the water is less shaded than in sawgrass marshes and cattails. Photodegradation has been proposed as the predominant mechanism of mercury demethylation in surface waters (reviewed in Tai et al. 2014). Figure 4i provides corroboration for this explanation by showing that mosquitofish in wet prairies and sloughs have less mercury than those in sawgrass marshes.

There are three covariates in surface water, total phosphorus (Figure 4n), alkaline phosphatase activity (Figure 4a), and chlorophyll-*a* (Figure 4h), that can be associated with trophic state. These three covariates, along with total organic carbon (Figure 4l), show less biomagnification of mercury at higher concentrations of the covariate. The alkaline phosphatase curve is the inverse of the phosphorus curve because there is greater activity in low-phosphorus environments. In addition to sulfur and organic carbon, agricultural runoff that enters the Everglades contains phosphorus at levels above the very low background concentrations found in the pristine oligotrophic parts of the system (reviewed in Scheidt and Kalla 2007). Over time, this input has led to eutrophication in some areas, which in turn degraded the original habitat, producing depauperate food webs and short food chains, without periphyton, which are less efficient at biomagnification (reviewed in Kalla and Scheidt 2017).

The best example of this effect is cattail marshes, which have the least fish Hg of any habitat (Figure 4i).

Relative Percent Influence of Covariates

Of the six most influential covariates (those meeting the threshold in all seven models in Table 2), three were mercury in various forms and media, one can be a trophic indicator (alkaline phosphatase), one can be a marker of stormwater transport (conductivity), and one can be an enabler of methylation (sulfate). While these covariates had an average individual influence ranging from 3.7% to 9.6%, together they accounted for almost 40% of the influence in each model, on average. Sulfate, the enabler, is transported to and through the Everglades by canals (reviewed in Scheidt and Kalla 2007), along with phosphorus and organic carbon.

According to our model, mercury's presence in the food web is affected by sulfur (4.9%). Moving water that has high sulfur into parts of the Everglades where food webs are complex and food chains are long may result in greater biomagnification of mercury.

One noteworthy contrast in the results is the larger number of influential (>/= 2%) covariates in 2005 and 2014 compared to the 1990s. The covariates below the dotted line for 2005 and 2014 may have had more influence in those years simply because mosquitofish mercury concentrations were lower in 2005 and even lower in 2014, as compared to the 1990s (Scheidt and Kalla 2007, Kalla and Scheidt 2017). Sulfate also had lower concentrations in those years (Scheidt and Kalla 2007, Kalla and Scheidt 2017), which could have allowed other covariates to rise in influence.

Comparing seasons overall, there are more influential covariates in the wet season than in the dry season. Perhaps as rewetting of the marsh proceeds during the summer the system becomes more varied and complex both ecologically and hydrodynamically, with more covariates coming into play.

Citations

Friedman, J.H. (2001). "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics 29(5):1189-1232

García-Laencina, P., Sancho-Gómez, J., & A. Figueiras-Vidal. (2010). Neural Computing and Applications 19: 263. https://doi.org/10.1007/s00521-009-0295-6

Greenwell, B., Boehmke, B., & J. Cunningham. (2019). gbm: Generalized Boosted Regression Models. R package version 2.1.5. https://CRAN.R-project.org/package=gbm

Kalla, P.I., and D.J. Scheidt. 2017. Everglades ecosystem assessment – Phase IV, 2014: Data reduction and initial synthesis. U.S. EPA, Science and Ecosystem Support Division, Athens, GA. SESD 14-0380. 58 pp. https://www.epa.gov/everglades/everglades-ecosystem-assessment-phase-iv-2014-report

Orem, W., C. Gilmour, D. Axelrad, D. Krabbenhoft, D. Scheidt, P. Kalla, P. McCormick, M. Gabriel, and G. Aiken. 2011. Sulfur in the South Florida ecosystem: distribution, sources, biogeochemistry, impacts, and management for restoration. Critical Reviews in Environmental Science and Technology 41(sup1):249-288. https://doi.org/10.1080/10643389.2010.531201

Pollman, C.D. 2014. Mercury cycling in aquatic ecosystems and trophic state-related variables – implications from structural equation modeling. Science of the Total Environment 499:62-73. https://doi.org/10.1016/j.scitotenv.2014.08.036

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Scheidt, D.J., and P.I. Kalla. 2007. Everglades ecosystem assessment: water management and quality, eutrophication, mercury contamination, soils and habitat: monitoring for adaptive management: a R-EMAP status report. USEPA Region 4, Athens, GA. EPA 904-R-07-001. 98 pp.

https://www.epa.gov/everglades/everglades-ecosystem-assessment-water-management-and-quality-eutrophication-mercury

Stober, Q.J., K. Thornton, R. Jones, J. Richards, C. Ivey, R. Welch, M Madden, J. Trexler, E. Gaiser, D. Scheidt, and S. Rathbun. 2001. South Florida ecosystem assessment: Phase I/II – Everglades stressor interactions: hydropatterns, eutrophication, habitat alteration, and mercury contamination. EPA 904-R-01-002.

Tai, C., Y. Li, Y. Yin, L.J. Scinto, G. Jiang, and Y. Cai. 2014. Methylmercury photodegradation in surface water of the Florida Everglades: Importance of dissolved organic matter-methylmercury complexation. Environ. Sci. Technol. 48:7333-7340. https://dx.doi.org/10.1021/es500316d

End of Report