Kriging with Atmospheric Mercury Concentrations

Hugh McCreery, Johnathan Castillo

April 25, 2020

Abstract

This paper attempted to examine the issue of mercury pollution by using spatial methods to predict the levels of three different types of pollutants at locations primarily in the Eastern United States. Using data from the National Atmospheric Deposition Program, we attempted to use both empirical variograms as well as maximum likelihood methods to estimate the covariance structure for each type of pollutant. After maximum likelihood analysis, we decided to use the estimated covariance structure from one type of pollutant for all pollutants as the other structures did not have logical physical interpretations. We then used simple kriging to generate predictions at locations without values. The usefulness of these predictions were hampered by the lack of input data and the input data being potentially non-stationary, and as a result we weren't able to confidently draw conclusions.

Contents

1 Introduction

In this paper, we worked with a data set from the National Atmospheric Deposition Program describing mercury pollution levels primarily centered around the eastern United States. We decided to use this data as the rising levels of pollution is a critical issue facing the world today. This mercury pollution, created primarily from coal mining and small-scale gold mining, can cause irreversible toxic effects to both the health of humans and the ecosystems around us (Ankrah). These potential harms of mercury and the resulting desire to monitor the levels of their pollution lead us to attempt to find potential sources of pollution. We hope to use the observed data of the mercury concentrations at different atmospheric locations to estimate the mercury concentration at locations without known values across the eastern United States. If we are able to create meaningful predictions of the pollution levels, we also wanted to consider if these predictions could be used to determine the possible sources of the pollution.

The data set contains information about three different types of mercury pollution - Particulate Bound Mercury (PBM), Gaseous Oxidized Mercury (GOM), and Gaseous Elemental Mercury (GEM). The data spans from 2008 to 2018, but as we didn't want to consider time in our analysis, we took the average amounts of each pollution for each station from 2015-2017 and used those as the values to be considered. In addition, we didn't consider stations that were very far away from the rest of our data (such as one in Taiwan) and any data marked as "invalid" in the metadata. This left us with 21 stations with valid data for each of the three types of pollution, as shown in the plots below.
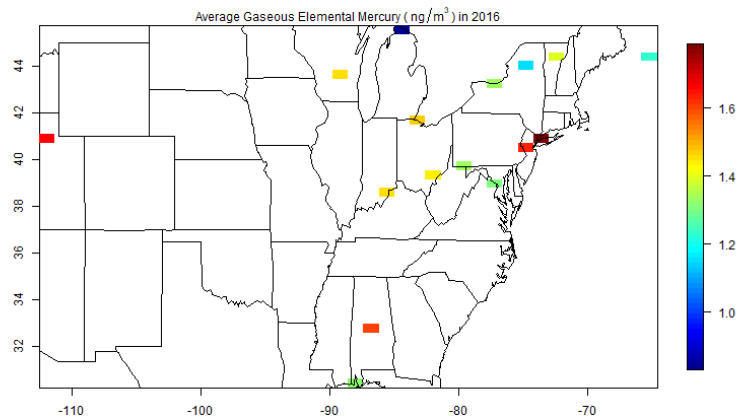

Figure 1: Levels of Gaseous Elemental Mercury at each of the 21 valid measuing stations
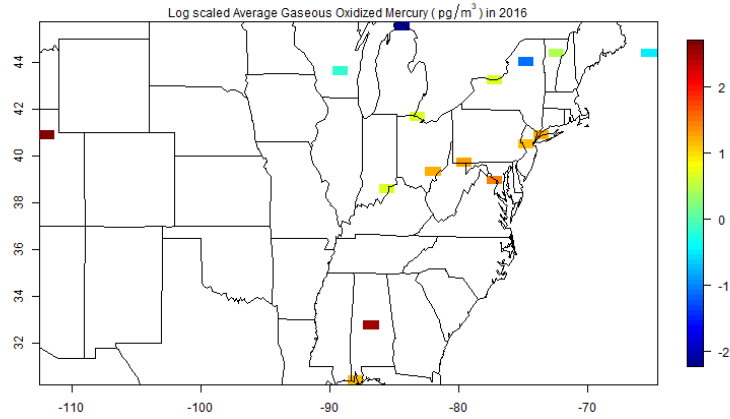
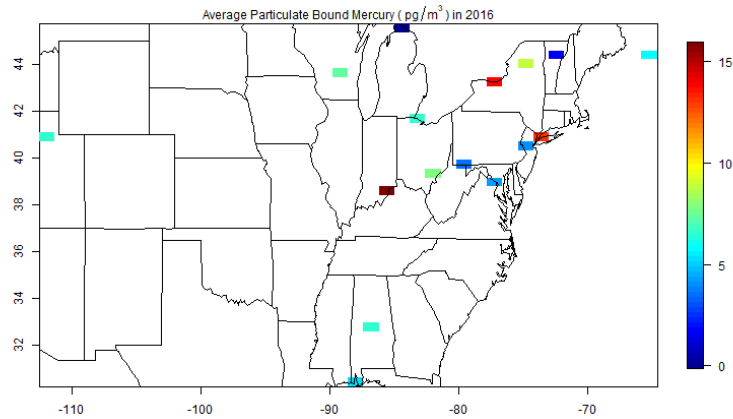Figure 2: Levels of log-scaled Gaseous Oxidized Mercury at each of the 21 valid measuing stations

Figure 3: Levels of Particulate Bound Mercury at each of the 21 valid measuring stations

2 Methods and Results

To start our analysis, we first obtained the mercury data from the NADP and imported the data into R ("AMNet..."). Then, using the provided information about each site, we matched each station from the initial data file with its geographic location. Invalid data was then removed based on the metadata descriptions.

After plotting the mercury data for each station, the data was partitioned into 3 areas with roughly the same number of stations. The means of the mercury concentrations in each partition were calculated and compared to each other to test for stationarity. It was found that the GEM concentrations appeared to be relatively stationary, while the GOM and PBM concentrations appeared to be potentially non-stationary. To simplify further analysis, we've decided to assume stationarity for each of the mercury concentrations despite the potentially non-stationary nature of the data.

Empirical variograms were also calculated for each mercury concentration using the **variog** function from the **geoR** package in R and have been included below.
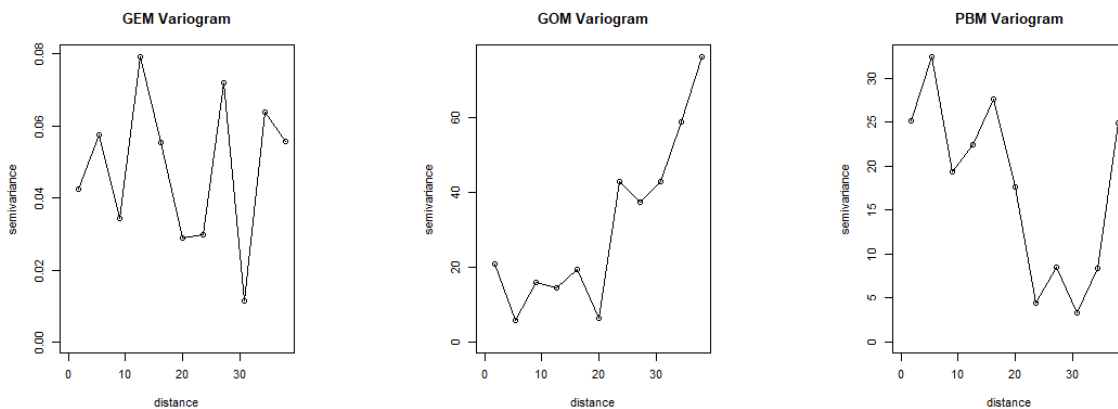

Figure 4: Empirical Variograms for GEM, GOM, and PBM concentrations based on the observed data.

Unfortunately, the small sample size of the data resulted in only a few points at most for each distance (lag), and so our empirical variogram calculations were not useful. This meant we needed to use the maximum likelihood estimation approach to estimate the covariance structures for the mercury concentrations.

For maximum likelihood estimation, we first assumed that each of the concentrations would have a exponential covariance structure. We felt confident making this assumption because exponential covariance structures are common in climate modeling. A possible nugget term was also included, giving us a total of 3 spatial parameters to estimate: τ^2 (nugget variance), σ^2 (process variance), and λ (range parameter of exponential covariance). Additionally, since it appeared the concentrations did not have a mean centered at 0, the data sets were detrended.

Maximum likelihood estimation of $(\tau, \sigma^2, \lambda)$ using the negative log likelihood was done using

both a 3-dimensional grid search and R's built in **optim** function. Using the grid search method, it was found that for the GOM concentrations the MLE estimates were:

$$(\hat{\tau}, \hat{\sigma}^2, \hat{\lambda}) = (0,\ 0.4081,\ 234.7)$$

with similar results using the **optim** function.

An important note is that the station located in Utah seemed too far away from the rest of the observations and thus was not used in the rest of our analysis. The station being very far from the rest of the points was distorting the parameter values. Initially, we believed that this point was close enough spatially to the rest of our data but got better results not using it. The results of our grid search have been visualized below.
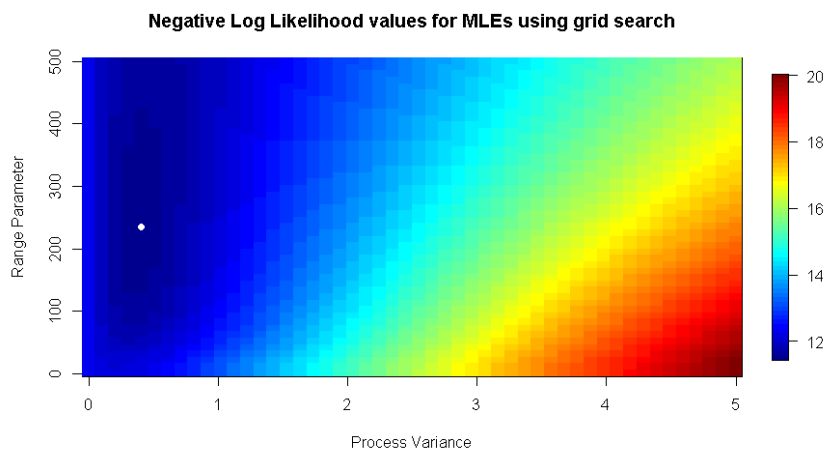

Figure 5: Negative log-likelihood values for MLE estimation of GOM covariance parameters $\hat{\sigma}^2$ and $\hat{\lambda}$ ($\hat{\tau}$ not shown since estimate was 0). The white dot indicates where the negative log-likelihood is minimized.

Having determined the covariance structure for the GOM data, we then attempted to find the MLEs for the covariance structure of the PBM and GEM using the same process. However, we ran into issues where the resulting estimates had interpretations that weren't logical. For the GEM data, the MLE for σ^2 was 0, indicating the data had no spatial correlation. Similarly, when we created MLEs for the PBM data, we found that the estimate for the range parameter was extremely small, which lead the kriged values to equal the mean at almost every point. As neither of these estimates gave useful results, we chose to assume that each of the types of mercury pollution would have similar covariance structures and used the parameter estimates from the GOM data to krige the PBM and GEM data.

Using these MLE estimates for the covariance structure, the concentrations at unobserved locations were predicted using simple kriging methods.
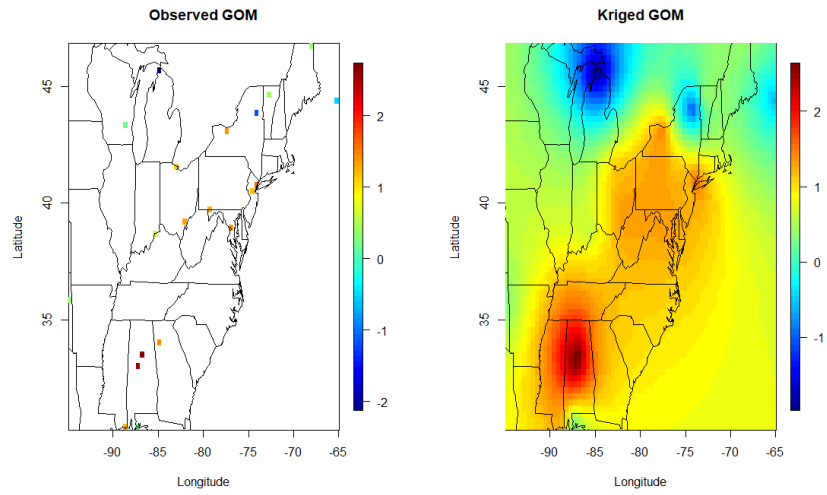
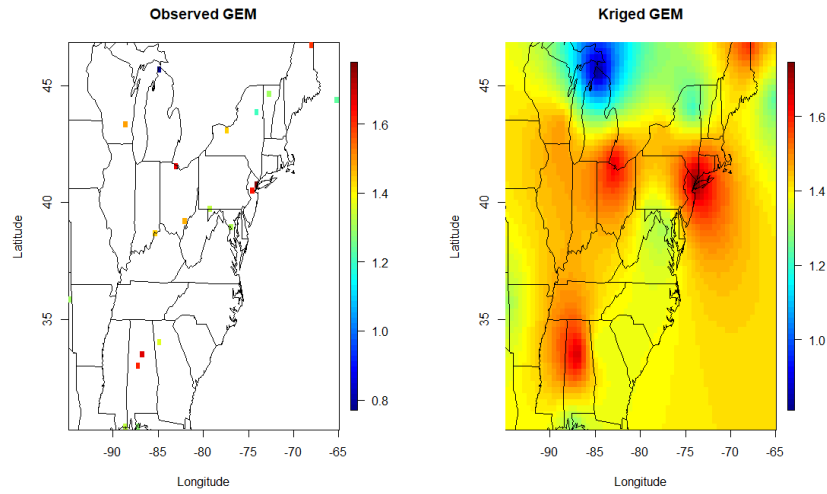Figure 6: Observed GOM concentrations and Kriged GOM concentrations

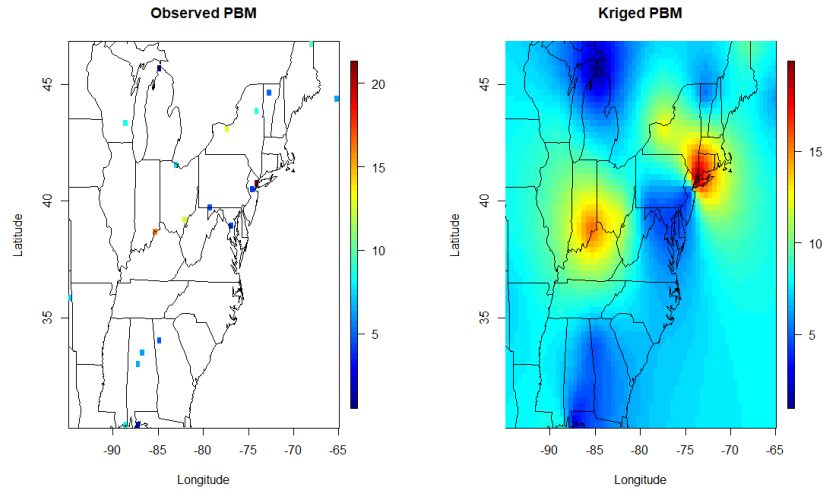Figure 7: Observed GEM concentrations and Kriged GEM concentrations

7

Figure 8: Observed PBM concentrations and Kriged PBM concentrations

Kriging uncertainties were also calculated for the kriging estimates. Since each type of concentration were assumed to have the same covariance model, the kriging uncertainties for each concentration are the same.
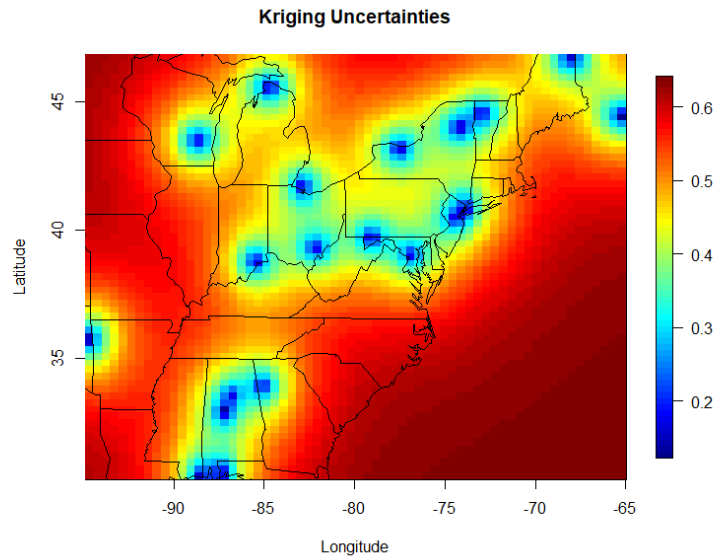

Figure 9: Simple Kriging uncertainties based on location. Due to the small sample size, our uncertainties are quite high in many areas of the map.

3   Implications and Analysis

Overall, our lack of data hampered our ability to draw any strong conclusions from our analysis and generate the accurate predictions we desired. The empirical variograms we calculated didn't lead us to believe that we would be able to determine a spatial correlation through that method and the estimates of parameters using MLEs had large variances. Additionally, our MLE estimates relied on assuming all three types of pollution had the same covariance structure, which is likely incorrect. Therefore, while we did get some results, they were hindered by their large uncertainties and as a result didn't give very useful information. This also resulted in us not being able to use our analysis to predict any outside sources of pollution. Our kriged estimates mainly capture the hot and cold spots of the observed locations which limits our ability to identify any potentially high pollution areas outside of observations.

Finally, our assumption that the data is stationary for each of the types of pollution may not necessarily hold. When we tested for stationarity, our results indicated that the GOM and PBM concentrations may not be stationary. While we performed our analysis under the assumption of stationarity anyway, it represents a significant weakness in our analysis that we weren't certain that the processes were stationary.

4   Further Research

If we were to continue research on this topic, the most important improvement we would make would be to find a more complete data set. The fact that we had very few data points limited our ability to draw any strong conclusions. Using the same analysis that we completed earlier on a larger data set would greatly improve many of our issues determining stationarity and the large uncertainties associated with the MLEs and kriged values.

Additionally, we could attempt to use non-stationary techniques to analyze our data. As noted earlier, while it appeared the GEM data was stationary, the GOM and PBM data looked to be potentially non-stationary. If we don't use this assumption of stationarity, our potential results may be more accurate.

Finally, we might attempt find possible covariates for our data to help improve our predictions. As noted earlier, the major causes of mercury pollution come from coal and gold mining, so finding data relating to where these industries operate could be useful as covariates in future analysis. We may also find other types of pollution (such as air pollution from coal burning) useful as well.

5   References

"AMNet Data Retrieval" *National Atmospheric Deposition Program.* 2018.
http://nadp.slh.wisc.edu/data/AMNet/. 8 Dec 2018.

Ankrah, Rodges. "Mercury Emissions: The Global Context." *United States Environmental Protection Agency.* 21 Nov 2018. https://www.epa.gov/international-cooperation/mercury-emissions-global-context. 8 Dec 2018.