Project 2011DC129B has resulted in no reported publications as of FY2011.

Report Follows
Pollution Source Identification in Washington DC Stormwater using Bayesian Chemical Mass Balance Modeling

Progress Report

Submitted to the
DISTRICT OF COLUMBIA WATER RESOURCES RESEARCH INSTITUTE

By:

Arash Massoudieh, Ph.D.
Catholic University of America

Tolessa Deksissa, Ph.D.
University of the District of Columbia
Progress Report: Pollution Source Identification in Washington DC storm-water using Bayesian Chemical Mass Balance modeling

Arash Massoudieh, Tolessa Deksisa

1 Executive Summary

To prioritize the reduction of pollutions and nutrients in the stormwater and also in order to determine the most effective approaches in the reduction of the pollutants entering streams as a result of stormwater discharge, it is important to identify the major contributors of specific pollutants in the stormwater. Particularly in highly urbanized areas this is an important and at the same time a challenging task due to the large number of potential contributors to stormwater pollution. The goal of the proposed research is to identify the source major pollutants and nutrients at a highly urbanized area at the vicinity of the Anacostia River in North East Washington DC. This is to test/demonstrate the ability of the method in finding the sources of contaminants with an acceptable confidence. For this purpose, an innovative method herein referred to as Bayesian Chemical Mass Balance modeling will be used. This method uses the elemental profiles of potential sources as well as the stormwater runoff samples to infer the contribution of each sources. Traffic and non-traffic related sources (e.g. street dust, wet deposition, and roof runoff) are identified and multiple samples of each is collected and analyzed for their elemental profiles. The elemental profiles of the source samples and the discharged water will be analyzed using mass spectrometry technique. Then, the Bayesian CMB method (Massoudieh et. al, in review) will be utilized to infer the contribution of various sources into the stormwater runoff. We have already submitted proposals to EPA or NSF proposing to apply the method at a larger scale by applying it to a larger number of discharge points in the city of Washington to the stormwater being released into the Anacostia and Potomac rivers.

2 Introduction

Receptor modeling has been widely used for source apportionment of air pollutants (Miller, Hidy et al. 1972; Friedlander 1973; Gordon 1988). Although they have been also used in some disciplines related to water resources, including surface water (Su, Christensen et al. 1998; Kelley and Nater 2000), groundwater (Olmez and Hayes 1990), wastewater (Soonthornnonda and Christensen 2008) and sediments source apportionment (Collins and Walling 2002; Collins and Walling 2004), the use of such models in studies of water pollution is still in its infancy (Kelley and Nater 2000; Su, Li et al. 2010). One of the main reasons for the rare use of receptor models in apportioning water pollutants is that diffuse sources often play an important role and source signatures can change during the transport from source to receptor. However, it has been suggested that information from several receptor sites can be used to incorporate retardation and chemical transformation into multivariate receptor models (Grimvall and Stalnacke 1996).
One of the main assumptions in source apportionment is that the elemental composition of the phases being studied (e.g., particulate or aqueous phases) will not change during the transport as a result of mass exchange with the surrounding medium of reactions. This assumption holds in most cases for air pollutants as the mass exchange between the aerosols and the surrounding air media is very small. In aquatic systems however, there can be a significant mass exchange between the particulate phase and the aqueous phase for at least some of the elements. Furthermore, the facts that a) the particulate matter and dissolved chemicals can have different transport behaviors and b) the dissolved phase elements can undergo adsorption and desorption to the solid surfaces such as pavement or stormwater channel surface during their transport, makes the use of receptor modeling in stormwater source apportionment more challenging.

Furthermore, there are uncertainties in the estimated pollutant source contributions due to the uncertainties in measured concentrations of sources, and runoff caused by both measurement errors and spatial and temporal heterogeneities. Thus, it is important to determine the reliability of the outcomes of such models by evaluating the confidence intervals of the estimated source contributions. Bayesian approach has been used in conjunction with various receptor modeling techniques including CMB (Billheimer 2001; Fox and Papanicolaou 2008) and positive factor analysis (Park, Oh et al. 2000). In none of the works done in the past, however, the impact of the non-persistence of the elemental compositions of the sources have been evaluated or have been incorporated into the Bayesian approach. The method proposed in this research can identify the impact of solid-water mass exchange on the elemental fraction of different elements used for receptor modeling. Using the Bayesian approach, the variabilities of the fraction of each element will be implicitly incorporated as a weighting factor into the receptor modeling. The outcome of this study will also reveal which elements comprise the most useful signature for pollutant sources.

The dry weather accumulation of pollutants is spatially variable due to the different nature of various sources of pollutants and different behavior of different types of land surfaces. For example, the wet and dry atmospheric deposition can be to a large extent considered uniform over a small watershed. On the other hand, the traffic related pollutants are highly heterogeneous even over small scales. This variability can have a highly important implication when designing BMPs or water quality control strategies. Stormwater model calibration often times cannot reveal the spatial variability of the buildup model parameters, due to the fact that when incorporating different accumulation rates into the model, they become over parameterized and obtaining a unique parameter set to represent the observed concentrations becomes impossible. Therefore, it is especially useful to link the accumulation rates of pollutants with the sources generating them, and consequently, linking the accumulation of pollutants from various sources to land use/climate and anthropogenic factors, such as population density and traffic. This research is an attempt to enhance the stormwater quality models by separately incorporating the dynamics of pollutants from various sources into them.
3 Methods

3.1 Bayesian Chemical Mass Balance Receptor Modeling

The chemical mass balance method used in this study is based on Massoudieh et al., (under review). The method is briefly described here. The mass balance method can be written in Matrix form as:

\[ C = YX \]  

(1)

where \( C = [c_i]_{mx1} \) is a vector representing the true elemental composition of the fluvial sample normalized by the sum of the measured elemental concentration recipient elemental profile vector henceforth referred to as the fluvial sample elemental composition vector. \( Y = [y_{ij}]_{mrxn} \) is the true source elemental composition matrix, each of its columns representing the normalized elemental composition of sources by the sum of the measured elements' concentrations in each source. \( X = [x_{jn}]_{nx1} \) is the source contribution vector, containing the true values of fractional contributions of each source into the fluvial sample, \( m \) is the number of elements measured and \( n \) is the number of sources considered. Since \( X \) is defined as the fractional contribution of sources the sum of its elements should be unity:

\[ \sum_{j=1}^{n} x_{j} = 1 \]  

(2)

It should be noted that due to measurement errors and heterogeneities in the elemental composition of both the fluvial and source samples, one can never know the exact values of the true fluvial sample elemental composition vector, \( C \), and the true source elemental composition matrix \( Y \) and therefore one can never know the true source contribution vector \( X \). The goal here is to infer the posterior probability distribution of \( X \) based on observed elemental compositions of sources and fluvial samples henceforth referred to as \( \tilde{Y} \) and \( \tilde{C} \) using Bayesian inference. Considering that the prior distributions of \( X \) and \( Y \) are independent, based on Bayes’ theorem the posterior distribution of \( X \) and, given the observed source and fluvial sample elemental composition can be expressed as:

\[ p(Y, X | \tilde{C}, \tilde{Y}) \propto p(\tilde{C} | Y, X) \cdot p(\tilde{Y} | Y) \cdot p(Y) \cdot p(X) \]  

(3)

In Eq. (3) \( p(\tilde{C} | Y, X) \cdot p(\tilde{Y} | Y) \) is the likelihood function, and \( p(Y) \) and \( p(X) \) are the prior distributions for \( Y \) and \( X \) respectively. Assuming the recipient elemental profiles \( \tilde{C} \) is log-normally distributed with known variances with its elements independent of each other, the first component of the likelihood function can be expressed as:
\[ p(\tilde{C} \mid Y, X) \propto \frac{1}{\prod_{i=1}^m \tilde{c}_i \delta_{c,i}} e^{-\sum_{i=1}^n \left[ \ln(\tilde{c}_i) - \ln(\sum_{i=1}^n x_i) \right]^2} \tag{4} \]

where \( \delta_{c,i} \) is the variance of the logarithm of elemental composition and \( \sum_{j=1}^m y_{ij} \cdot x_j \) is the “true” recipient elemental composition for element \( j \). In order to make sure all source elemental fractions \( \tilde{y}_{ij} \) vary between zero and one, the transformation \( \tilde{y}_{ij} / (1 - \tilde{y}_{ij}) \) of observed source elemental compositions are assumed to be log-normally distributed and therefore we can express the second component of the likelihood function as:

\[ p(\tilde{Y} \mid Y) \propto \frac{1}{\prod_{j=1}^n \prod_{i=1}^m \tilde{y}_{ij} \left(1 - \tilde{y}_{ij}\right) \delta_{y,ij}} e^{-\sum_{j=1}^n \sum_{i=1}^m \left[ \ln\left( \frac{\tilde{y}_{ij}}{1 - \tilde{y}_{ij}} \right) - \ln\left( \frac{y_{ij}}{1 - y_{ij}} \right) \right]^2} \tag{5} \]

where \( \delta_{y,ij} \) is the variance of element \( j \) measured in source \( i \) calculated from multiple observations of elemental fractions for each source:

\[ \delta_{y,ij} = \text{STD} \left[ \ln\left( \frac{\tilde{y}_{ij}}{1 - \tilde{y}_{ij}} \right) \right] \tag{6} \]

The prior distribution for \( X \) was considered a Dirichlet distribution with parameters all equal to one and therefore satisfying constraint in Eq. (2):

\[
\begin{align*}
\begin{cases}
p(x_1, \ldots, x_n) = n! & \text{for } 0 \leq \sum_{j=1}^{n-1} x_j \leq 1 \\
p(x_1, \ldots, x_n) = 0 & \text{otherwise} \\
x_n = 1 - \sum_{j=1}^{n-1} x_j
\end{cases}
\end{align*}
\tag{7}
\]

Since no additional information about the source elemental profiles other than the measured elemental concentrations is available, a uniform PDF between 0 and 1 was considered for the prior distribution of \( Y \). Substituting Eqs. (5) and (4) into Eq. (3), the following relationship for the posterior probability is obtained:

\[
p(X, Y \mid \tilde{Y}, \tilde{C}) \propto \frac{1}{\prod_{i=1}^m c_i \delta_{c,i}} e^{-\sum_{i=1}^n \left[ \ln(c_i) - \ln(\sum_{i=1}^n x_i) \right]^2} \frac{1}{\prod_{j=1}^n \prod_{i=1}^m \tilde{y}_{ij} \left(1 - \tilde{y}_{ij}\right) \delta_{y,ij}} e^{-\sum_{j=1}^n \sum_{i=1}^m \left[ \ln\left( \frac{\tilde{y}_{ij}}{1 - \tilde{y}_{ij}} \right) - \ln\left( \frac{y_{ij}}{1 - y_{ij}} \right) \right]^2} p(Y)p(X) \tag{8}
\]
Eq. (8) can be used to calculate various moments of the posterior probability distributions of $X$ and $Y$ through integration. Due to the large number of dimensions, evaluating the integral in equation (8) using conventional methods is prohibitive. Therefore a Markov Chain Monte Carlo (MCMC) approach (Gamerman and Hedibert 2006) is used to generate random samples according to the posterior distribution of $X$ and $Y$. Specifically in this research the Metropolis-Hasting Algorithm (Metropolis, Rosenbluth et al. 1953) is used to obtain a sequence of random numbers from the posterior probability distribution presented in Eq. (8).

3.2 Site Description and Sampling

The sample collection is already conducted during the rain event on 22nd of May 2012. Four student researchers participated in the sampling campaign. Figures 1 and 2 show the map of the study site at the shore of Anacostia River in North East Washington DC. The location of all the sampling stations including source samples and runoff samples are shown in Figures 3 and 4. The stations and number of water samples collected at each station is listed in Table 1 along with the designated codes to the samples. Samples were collected totally at 12 locations at each of them at least three water samples were collected to represent the temporal variations of the elemental profiles of the samples. The runoff samples were collected at two outfalls into the Anacostia River. Each sample were filtered onsite using 0.45μm Whatman Autovial 5 Syringeless Filters (Fischer Scientific). The portion passing through the filters were considered. The portion passing through the filter was considered dissolved and the unfiltered part was considered as the total. The particulate elemental composition will be obtained by subtracting the dissolved from the total elemental compositions. At least 5mL of each fraction (i.e. dissolved and total) of each sample were collected. The sample were put in fridge and were shipped to UC Davis for ICP-MS analysis at Peter Green’s lab.
Figure 1: The location of the study site in the District of Columbia
Figure 2: Detailed location of the study site at the shore of Anacostia River
Figure 3: Detailed Sampling locations at the study site

Figure 4: Source and runoff sampling locations in the study site
3.3 Future Steps

Due to lack of significant rain the project was delayed and we asked for a no-cost extension until September which was granted. We expect the ICP-MS analysis to be finished by the end of May. As soon as we receive the results of the ICP-MS analysis the Bayesian CMB will be performed separately on the dissolved and particulate fractions. We plan to identify the contribution of the potential sources including (grass, street, and roof runoff) into the Phosphorus, Pb, Cu, Zn and Cd into the Anacostia River.

Table 1: List and location of samples collected for source identification and the codes attributed to each sample

<table>
<thead>
<tr>
<th>source</th>
<th>Site</th>
<th>codes</th>
<th>Dissolved</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>10 min</td>
<td>20 min</td>
</tr>
<tr>
<td>Grass</td>
<td>Near Basketball field</td>
<td>G1</td>
<td>G1-10-D</td>
<td>G1-20-D</td>
</tr>
<tr>
<td>Out-fall</td>
<td>Clay St. (1)</td>
<td>O1</td>
<td>01-10-D</td>
<td>01-20-D</td>
</tr>
<tr>
<td></td>
<td>Blaine St. (2)</td>
<td>O2</td>
<td>02-10-D</td>
<td>02-20-D</td>
</tr>
<tr>
<td>Street</td>
<td>Dix St.NE</td>
<td>S1</td>
<td>S1-10-D</td>
<td>S1-20-D</td>
</tr>
<tr>
<td></td>
<td>Blaine St.</td>
<td>S2</td>
<td>S2-10-D</td>
<td>S2-20-D</td>
</tr>
<tr>
<td></td>
<td>Clay St.</td>
<td>S3</td>
<td>S3-10-D</td>
<td>S3-20-D</td>
</tr>
<tr>
<td></td>
<td>Anacostia Ave</td>
<td>S4</td>
<td></td>
<td>S4-30-D</td>
</tr>
<tr>
<td></td>
<td>Anacostia Ave</td>
<td>S5</td>
<td></td>
<td>S5-30-D</td>
</tr>
<tr>
<td>Bridge</td>
<td>East Capitol St. NE (South)</td>
<td>B1</td>
<td>B1-10-D</td>
<td>B1-20-D</td>
</tr>
<tr>
<td>Parking lot</td>
<td>Enterprise Rent car Pump of gas</td>
<td>P1</td>
<td>P1-10-D</td>
<td>P1-20-D</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2</td>
<td>P2-10-D</td>
<td>P2-20-D</td>
</tr>
<tr>
<td>Roof</td>
<td>Dix St.NE</td>
<td>R1</td>
<td>R1-10-D</td>
<td>R1-20-D</td>
</tr>
<tr>
<td></td>
<td>Clay St.</td>
<td>R2</td>
<td>R2-10-D</td>
<td>R2-20-D</td>
</tr>
<tr>
<td></td>
<td>Enterprise Rent car</td>
<td>R3</td>
<td>R3-10-D</td>
<td>R3-20-D</td>
</tr>
<tr>
<td></td>
<td>First Alley (near Bridge)</td>
<td>R4</td>
<td>R4-10-D</td>
<td>R4-20-D</td>
</tr>
<tr>
<td>Rain</td>
<td>Blaine St.</td>
<td>Rain1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blaine St.</td>
<td>Rain2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blaine St.</td>
<td>Rain3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4 References: