Green-Duwamish River Watershed PCB Congener Study Phase 2 Summary

Factor analysis results

Lisa A. Rodenburg
Department of Environmental Sciences
Rutgers, the State University of New Jersey
My research specializes in analyzing large data sets on PCBs and other pollutants:

- **New York/New Jersey Harbor**
  - Water column, dischargers, sediment
- **Delaware River PCB TMDL data**
  - Water column, sediment, dischargers, air
- **Integrated Atmospheric Deposition Network - Chicago**
- **San Francisco Bay**
  - BDEs in sediment, PCBs in water
- **Portland Harbor Superfund Site**
  - Water column and sediment, biota
- **Duwamish/Green River, Washington**
  - Water, sediment, biota, air
Introduction

- Green-Duwamish River Watershed PCB Congener Study: Phase 1
  - Hundreds of environmental samples collected and analyzed for 209 PCB congeners by method 1668
  - This data already underwent QA/QC
Phase 2 - evaluation data from phase 1

Task 1 – can the data be used for factor analysis?

Criteria for data to be used in factor analysis:

- **Enough data:** at least as many samples as congeners/peaks
- **Enough data above detection**
- **Surrogate recoveries or duplicates needed to estimate uncertainty**
  - duplicates should be assimilated with regular samples so statistical independence between samples is preserved for the analysis
- **Detection limits are needed**—must be estimated if not available
Purpose of Phase 2, Task 3

- Identification of PCB chemical signatures (fingerprints)
- Determination of the relative contribution of these source signatures
- Identification of potentially known/unknown sources of and/or pathways for PCBs in the LDW
- A recommended set of PCBs (individual congeners and/or suites of congeners) to be included in modeling for the Green/Duwamish watershed PLA
- Recommendations for data collection and/or analysis approaches for future PCB congener data collection
Additional considerations for fingerprinting analysis

- GC column and co-elution pattern
  - SPB-octyl is the main column used for 1668
  - SGE-HT8 and DB-5 or equivalent are alternatives
  - Very different co-elution patterns

- Blanks

- Obvious trends
  - Non-Aroclor PCBs
  - Aroclors
Factor Analysis Equation

Applies to Principal Components Analysis, PMF etc.

View the PCB signal as a mixture of mixtures

Some of those mixtures are Aroclors …some are not.

Use this equation to predict concentration of each congener, based on number, fingerprint and concentration of sources.

You do NOT need any information about the sources, such as their fingerprints, or even how many there are!

\[ X = G F + E \]

\((m \times n)\) \hspace{1cm} \((m \times p)\) \hspace{1cm} \((p \times n)\)

\(X = \) input data matrix
\(G = \) matrix of conc of each factor in each sample generated by model
\(F = \) matrix of fingerprint of each factor \((p)\) generated by model
\(E = \) leftover or residual
\(n = \) number of analytes
\(m = \) number of samples
\(p = \) number of factors (sources)

Note: in all forms of factor analysis, the user has to decide what is the ‘correct’ number of sources based on model output.
Advantages of Positive Matrix Factorization over other models, for example Principal Components Analysis

- Positive correlations only – mass balance model
  - Great for concentrations of contaminants
- Assign a point-by-point uncertainty estimate
- Missing and below detection limit values can be included by assigning them a high uncertainty
- “Robust” mode can be used so that outlier values will not skew the factor profiles
  - Data can span many orders of magnitude
- PMF provides the quantitative contribution estimate from each factor for each sample.
The Soda Analogy

• Several different soft drinks to choose from
• Sometimes kids like to mix these...

• Say we have 100 kids who made mixed drinks from the same soda fountain...
Analytes

- Sugar = most non-diet sodas
- Aspartame = some diet sodas
- Carmel coloring = most colas, root beer, etc.
- Citric acid = Sprite, 7-Up, some fruity drinks such as Cherry Coke, etc.
- Cola flavoring = most colas
- Caffeine = most colas
## Data matrix

<table>
<thead>
<tr>
<th></th>
<th>Caramel color</th>
<th>sugar</th>
<th>aspartame</th>
<th>citric acid</th>
<th>cola flavoring</th>
<th>caffeine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anna</td>
<td>0.50</td>
<td>0.62</td>
<td>0.41</td>
<td>0.58</td>
<td>0.99</td>
<td>0.87</td>
</tr>
<tr>
<td>Bruce</td>
<td>0.58</td>
<td>0.86</td>
<td>0.25</td>
<td>0.78</td>
<td>0.35</td>
<td>0.14</td>
</tr>
<tr>
<td>Carlos</td>
<td>0.65</td>
<td>0.06</td>
<td>0.68</td>
<td>0.75</td>
<td>0.50</td>
<td>0.06</td>
</tr>
<tr>
<td>Donna</td>
<td>0.33</td>
<td>1.00</td>
<td>0.98</td>
<td>0.39</td>
<td>0.63</td>
<td>0.92</td>
</tr>
<tr>
<td>Emily</td>
<td>0.38</td>
<td>0.10</td>
<td>0.40</td>
<td>0.14</td>
<td>0.11</td>
<td>0.06</td>
</tr>
<tr>
<td>Francis</td>
<td>0.67</td>
<td>0.60</td>
<td>0.44</td>
<td>0.60</td>
<td>0.50</td>
<td>0.10</td>
</tr>
<tr>
<td>George</td>
<td>0.07</td>
<td>0.23</td>
<td>0.65</td>
<td>0.37</td>
<td>0.82</td>
<td>0.54</td>
</tr>
<tr>
<td>Harriet</td>
<td>0.95</td>
<td>0.53</td>
<td>0.02</td>
<td>0.25</td>
<td>0.51</td>
<td>0.86</td>
</tr>
<tr>
<td>Inga</td>
<td>0.46</td>
<td>0.67</td>
<td>0.19</td>
<td>0.92</td>
<td>0.23</td>
<td>0.45</td>
</tr>
<tr>
<td>John</td>
<td>0.32</td>
<td>0.97</td>
<td>0.79</td>
<td>0.19</td>
<td>0.88</td>
<td>0.21</td>
</tr>
<tr>
<td>Karl</td>
<td>0.81</td>
<td>0.42</td>
<td>0.68</td>
<td>0.70</td>
<td>0.15</td>
<td>0.08</td>
</tr>
<tr>
<td>Lisa</td>
<td>0.22</td>
<td>0.62</td>
<td>0.47</td>
<td>0.94</td>
<td>0.52</td>
<td>0.75</td>
</tr>
<tr>
<td>Michael</td>
<td>0.00</td>
<td>0.95</td>
<td>0.98</td>
<td>0.19</td>
<td>0.45</td>
<td>0.88</td>
</tr>
<tr>
<td>Nick</td>
<td>0.49</td>
<td>0.46</td>
<td>0.25</td>
<td>0.02</td>
<td>0.97</td>
<td>0.02</td>
</tr>
<tr>
<td>Olga</td>
<td>0.36</td>
<td>0.49</td>
<td>0.55</td>
<td>0.62</td>
<td>0.94</td>
<td>0.07</td>
</tr>
</tbody>
</table>

### Concentrations (mg/L)
PMF results

PMF can tell you:
- How many sources (fingerprints, factors)
- Their fingerprints (F matrix)
- How abundant each fingerprint is in each sample (G matrix)

"F matrix"
PMF results - F matrix
Fingerprints

PMF can’t tell you:
• What it all means

• YOU have to interpret this information
PMF Results – G matrix

- **G matrix:** abundance of each factor in each sample

- **Helps with questions like:**
  - Older people prefer diet soda?
  - Women prefer non-caffeinated drinks?
  - More caffeine consumed later at night?

<table>
<thead>
<tr>
<th></th>
<th>Cherry Coke</th>
<th>Coke</th>
<th>Sprite</th>
<th>Diet Coke</th>
<th>Mt Dew</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anna</td>
<td>16%</td>
<td>20%</td>
<td>13%</td>
<td>19%</td>
<td>32%</td>
</tr>
<tr>
<td>Bruce</td>
<td>20%</td>
<td>30%</td>
<td>9%</td>
<td>28%</td>
<td>13%</td>
</tr>
<tr>
<td>Carlos</td>
<td>25%</td>
<td>2%</td>
<td>26%</td>
<td>29%</td>
<td>19%</td>
</tr>
<tr>
<td>Donna</td>
<td>10%</td>
<td>30%</td>
<td>29%</td>
<td>12%</td>
<td>19%</td>
</tr>
<tr>
<td>Emily</td>
<td>34%</td>
<td>9%</td>
<td>35%</td>
<td>12%</td>
<td>10%</td>
</tr>
<tr>
<td>Francis</td>
<td>24%</td>
<td>21%</td>
<td>16%</td>
<td>21%</td>
<td>18%</td>
</tr>
<tr>
<td>George</td>
<td>3%</td>
<td>11%</td>
<td>30%</td>
<td>17%</td>
<td>38%</td>
</tr>
<tr>
<td>Harriet</td>
<td>42%</td>
<td>23%</td>
<td>1%</td>
<td>11%</td>
<td>23%</td>
</tr>
<tr>
<td>Inga</td>
<td>19%</td>
<td>27%</td>
<td>8%</td>
<td>37%</td>
<td>9%</td>
</tr>
<tr>
<td>John</td>
<td>10%</td>
<td>31%</td>
<td>25%</td>
<td>6%</td>
<td>28%</td>
</tr>
<tr>
<td>Karl</td>
<td>29%</td>
<td>15%</td>
<td>25%</td>
<td>25%</td>
<td>5%</td>
</tr>
<tr>
<td>Lisa</td>
<td>8%</td>
<td>22%</td>
<td>17%</td>
<td>34%</td>
<td>19%</td>
</tr>
<tr>
<td>Michael</td>
<td>0%</td>
<td>37%</td>
<td>38%</td>
<td>7%</td>
<td>18%</td>
</tr>
<tr>
<td>Nick</td>
<td>22%</td>
<td>21%</td>
<td>11%</td>
<td>1%</td>
<td>44%</td>
</tr>
<tr>
<td>Olga</td>
<td>12%</td>
<td>16%</td>
<td>19%</td>
<td>21%</td>
<td>32%</td>
</tr>
</tbody>
</table>

Rows sum to 100% → Need ancillary info, such as age, gender, time of day etc.
## Versions of PMF

<table>
<thead>
<tr>
<th></th>
<th>PMF2</th>
<th>PMF 5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pros</strong></td>
<td>--Always converges on a solution</td>
<td>--Bootstrapping and displacement routines help identify optimal number of factors</td>
</tr>
<tr>
<td></td>
<td>--Robust mode</td>
<td>--Runs in Windows</td>
</tr>
<tr>
<td></td>
<td>--Generated factors resemble Aroclors</td>
<td>--EPA continues to develop</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--GUI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--Some pre- and post-processing of data done by the program</td>
</tr>
<tr>
<td><strong>Cons</strong></td>
<td>--Older, runs in DOS</td>
<td>--Often doesn't converge on a solution</td>
</tr>
<tr>
<td></td>
<td>--No pre- or post-processing of the data</td>
<td>--Often have to arbitrarily add uncertainty to get convergence</td>
</tr>
<tr>
<td></td>
<td>--No longer supported</td>
<td>--No robust mode</td>
</tr>
<tr>
<td></td>
<td>--No GUI</td>
<td>--Generated factors less likely to resemble Aroclors</td>
</tr>
</tbody>
</table>
PMF 5.0
Main PCB sources in most watersheds

• AROCLORS!

• Non-Aroclor congeners from pigments

• Reductive dechlorination of Aroclors by bacteria
Aroclor fingerprints have been measured.

Aroclor 1242 (51%)  
Similar to Aroclor 1016 (13%)

Aroclor 1248 (7%)

Aroclor 1254 (16%)

Aroclor 1260 (11%)
Known inadvertent non-Aroclor PCB sources

- Organic pigments, especially diarylide yellow, contains primarily PCB 11, among others (like 12?, 13?, 35, 77, 52 etc)

- Titanium dioxide (white pigment) may contain PCBs 206, 208, and 209

- Silicone rubber tubing produces PCBs 68, 44 and 45, etc. (Perdih and Jan Chemosphere 1994)
  - Don’t sample using silicone rubber tubing!
Microbial dechlorination of PCBs

- Previously, seen *only* in aquatic sediments, but we found it in:
  - *Sewers* (esp. combined)
  - *Landfills*
  - *Groundwater* at contaminated sites

- Mediated by *chloroflexi*
  - Use organochlorine compounds as electron acceptors

- Usually removes chlorines at *meta* and *para* but not *ortho* positions
  - Several pathways identified
  - Main products are PCB 4 and 19

---

[Chemical structures of PCB 4 and PCB 19: PCB 4 (2,2’), PCB 19 (2,2’,6)]
Data sets analyzed by PMF

<table>
<thead>
<tr>
<th>compartment -&gt;</th>
<th>air</th>
<th>sediment</th>
<th>surface water</th>
<th>tissue</th>
<th>storm drain</th>
</tr>
</thead>
<tbody>
<tr>
<td>columns</td>
<td>SPB-octyl</td>
<td>SPB-octyl &amp; SGE-HT8</td>
<td>SPB-octyl</td>
<td>SPB-octyl</td>
<td>SPB-octyl &amp; DB-5</td>
</tr>
<tr>
<td>samples</td>
<td>64</td>
<td>146</td>
<td>209</td>
<td>128</td>
<td>74</td>
</tr>
<tr>
<td>peaks</td>
<td>64</td>
<td>80</td>
<td>42</td>
<td>90</td>
<td>73</td>
</tr>
<tr>
<td>congeners</td>
<td>100</td>
<td>154</td>
<td>69</td>
<td>135</td>
<td>142</td>
</tr>
<tr>
<td>% of mass</td>
<td>88%</td>
<td>94%</td>
<td>60%</td>
<td>96%</td>
<td>92%</td>
</tr>
<tr>
<td>% data points Below Detection Limit</td>
<td>18%</td>
<td>9%</td>
<td>30%</td>
<td>1.4%</td>
<td>15%</td>
</tr>
</tbody>
</table>

- % of mass = % of the total mass contained in all the data that was included in the PMF analysis
- Air and storm drain congener lists limited by number of samples
- Water congener list limited by large numbers of Below Detection Limit (BDL) values

Includes duplicates
Uncertainty

The PMF model and results are highly reproducible. This does not necessarily imply low uncertainty.

Uncertainty arises from:

- Insufficient data: not enough samples or detected analytes
  - Esp. for water compartment

- Different models may give different results for the same data
  - Tried PMF2 and PMF 5.0 – very different results

- Various permutations of the same data set may give different model results, even when the same model is used
  - We ran many permutations and got essentially the same results giving us higher confidence

- Choosing a sub-optimal number of factors
  - # of factors was relatively obvious for most compartments, less so for water

- Factors may be misinterpreted
  - Similarity between Aroclors?
Results

Types of sources:
- Across all five compartments, Aroclors are the dominant PCB sources
  - $1260 > 1254 >> 1248 > 1016/1242$
- Non-Aroclor PCB sources are minor
- No dechlorination

Spatial trends in sources:
- Spatial trends are consistent across water, sediment, biota
Air (atmospheric deposition)

- 6 factors found:
- 4 surprisingly similar to Aroclors
- 1016 > 1260 > 1248 > 1254
- Lower MW formulations more abundant in the atm dep
- Air4 (5% of mass) does not resemble any Aroclor – composition is variable
Air – spatial trends

Higher PCB flux → more urban/industrial

- More 1260 in the more urban/industrial areas?
- No ‘urban fractionation effect’ – local sources?
Sediment

5 factors found:
4 similar to Aroclors
1260>>1254>1248>1016

Sed4 not similar to Aroclors, contains a lot of PCB 11

Wastewater/stormwater/CSO?
Or atmospheric deposition?
Sediment – spatial trends

- Sed5 (Aroclor 1260) dominates near river mouth
- Sed4 (PCB 11) more important upstream
Surface Water

- Four factors
- All resemble Aroclors
- Non-Aroclor congeners excluded
Surface water – spatial trends

- Mass-weighted average contribution to PCBs at each RM location
- Aroclor 1260 dominates nearer to river mouth
- PCB11, PCBs 206+208+209 were not included in the PMF model
Surface Water – non-Aroclor congeners

- PCB11, PCBs 206+208+209 not very abundant in the water column

Affected by one outlier sample
Tissue

Five factors:
All resemble single Aroclor or mix
1260 > 1254 > mix = 1248

Tissue4 and Tissue5 both resemble 1260, one more weathered (ADME processes)
Species vary in their ability to metabolize PCBs
Storm drain

Six factors:
Four resemble Aroclors

Storm3 = PCB 11

Storm6 = 206+208+209
Storm drain water vs. storm drain solids

- Storm water shifted toward lower MW Aroclors
Summary

Match \( R^2 \) between Aroclors and factors for each compartment:

<table>
<thead>
<tr>
<th>compartment</th>
<th>1016</th>
<th>1248</th>
<th>1254</th>
<th>1260</th>
</tr>
</thead>
<tbody>
<tr>
<td>closer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>storm drain</td>
<td>0.96</td>
<td>0.86</td>
<td>0.86</td>
<td>0.98</td>
</tr>
<tr>
<td>↑</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sediment</td>
<td>0.42</td>
<td>0.84</td>
<td>0.94</td>
<td>0.99</td>
</tr>
<tr>
<td>sources</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>surface water</td>
<td>0.73</td>
<td>0.44</td>
<td>0.84</td>
<td>0.91</td>
</tr>
<tr>
<td>↓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>air</td>
<td>0.81</td>
<td>0.57</td>
<td>0.85</td>
<td>0.88</td>
</tr>
<tr>
<td>further</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tissue</td>
<td>NA</td>
<td>0.43</td>
<td>0.7</td>
<td>0.84</td>
</tr>
</tbody>
</table>
Results

Types of sources:

• Across all five compartments, Aroclors are the dominant PCB sources
  – 1260 > 1254 >> 1248 > 1016/1242
• Non-Aroclor PCB sources are minor
• No dechlorination – probably due to salinity

Spatial trends in sources:

• Spatial trends are consistent across water, sediment, biota
Comparisons to other watersheds

Other systems have more ‘other’, more non-Aroclor, and often more dechlorination

In Delaware and Portland Harbor, Aroclor 1260 was associated with shipyards

Often difficult for the model to discern between Aroclors 1242 and 1248
Implications for modeling

• **What is your endpoint?**

• **Model homologs or total PCBs?**
  
  • When 1668 data is available, many systems have modeled PCB homologs, for example:
    – New York/New Jersey Harbor
    – Delaware River
  
  • When 1668 data is not available, systems model either total PCBs or a sub-set
    – Upper Hudson River models: total and Tri+ and a few congeners
    – Green Bay model: total and 5 congeners
Recommendations for homolog modeling

- Get input from the fate modeling team **before** sampling, not after

- Always measure PCBs using 1668 with SPB-octyl column
  - How to incorporate Aroclor data?

- Pick a short model calibration period of about one year
  - Opposite of monitoring

- Characterize loads
  - Head of tide, ocean/sound boundary, point sources (WWTPs, CSOs, other dischargers), non-point sources

- Hundreds of samples of water, sediment, and biota needed for the calibration of the model across full range of flow conditions
If you do homologs...

- Probably don’t need to model all 10

- Could potentially ignore homologs 1, 2, 9, and 10
  - Not very abundant in water, sediment, biota
  - Difficult to model due to non-Aroclor sources
  - 1 & 2 are subject to aerobic degradation which is hard to model
Acronyms

- BDL – Below Detection Limit
- CSO – combined sewer overflow
- LDW – Lower Duwamish Waterway
- MW – molecular weight
- PCB – Polychlorinated biphenyl
- PLA – pollutant loading assessment
- PMF – Positive Matrix Factorization
- QA/QC – quality assurance/quality control
- TMDL – total maximum daily load
Acknowledgements

- Work performed under sub-contract to Leidos and for WA Department of Ecology.

- EPA – esp. Gary Norris
- Brittney Blackburne, undergraduate intern, wrote code to automate much of the PMF2 analysis
Further information

Lisa Rodenburg - Rutgers University
Lisa.Rodenburg@Rutgers.edu  848-932-5774

Iris Winstanley - Leidos
iris.winstanley@leidos.com  425-205-5189

Rick Thomas, and Rachel McCrea