SOURCE APPORTIONMENT OF PM$_{2.5}$ IN URBAN AREAS USING MULTIPLE LINEAR REGRESSION AS AN INVERSE MODELLING TECHNIQUE

Bruce Denby and Herdis Laupsa
bde@nilu.no

Norwegian Institute for Air Research (NILU), Norway
Overview

- Sources of PM$_{2.5}$ in Oslo
- Observations
- Modelling (AirQUIS)
- Multiple linear regression
- Uncertainty assessment

Results
- All data
- Filter days at RV4 (validation)

Conclusions
Modelled source contributions at a traffic station in Oslo based on the current emissions inventory:

- PM2.5 from area sources except woodburning and industry: 6%
- PM2.5 from industrial sources: 0.0%
- PM2.5 from wood burning for domestic heating: 40%
- Exhaust particles from gasoline vehicles: 8%
- Exhaust particles from diesel vehicles: 10%
- Re-suspension of PM2.5 (PM2.5 from traffic-exhaust particles): 4%
- PM2.5 from regional background: 32%
Observations of PM$_{2.5}$ in Oslo

- PM$_{2.5}$ observational network during winter 2004

Traffic stations
- RV4
- Kirkeveien
- Løren

Urban background
- Aker Hospital

Filter samples
- RV4
- 38 twelve hour samples
Modelling of PM$_{2.5}$ in Oslo

• **PM$_{2.5}$ emissions**
  - Wood burning based on questionnaires and emission factors (climatological temperature dependence)
  - Traffic exhaust is a bottom up inventory
  - Resuspension related to exhaust emissions, studded tyre percentage and surface conditions (precipitation and temperature)
  - Number of other combustion sources, e.g. shipping.

- **Dispersion modelling (AirQUIS-EPISODE)**
  - Area sources, e.g. wood burning, use a Eulerian CTM
  - Traffic sources use Gaussian line source model (HIWAY-2)
  - Industrial sources use a Gaussian point source model
  - Meteorology using meteorological mast and diagnostic wind field model (MATHEW)

*Figure 14: Årsmiddedverk for PM$_{2.5}$ for 2003.*
Inverse modelling

- The aim is to provide an assessment of the average contributions from the different source sectors to the total observed PM$_{2.5}$ mass concentration.
- Consider the total concentration ($C$) to be the sum of the individual source contributions ($c_i$)

$$C(x, y, t) = \sum_{i=1}^{n} c_i(x, y, t)$$

- The observed concentration is the weighted sum of the model source contributions ($c_{mod\ i}$) plus an error ($\varepsilon$) where the scaling factor ($a_i$) is the weight

$$C_{obs}(x, y, t) = \sum_{i=1}^{n} a_i c_{mod\ i}(x, y, t) + \varepsilon_i(x, y, t)$$
Multiple linear regression

• We wish to minimise the error ($\varepsilon$)
• In this case we minimise the mean square error (MSE)
• This is equivalent to multiple linear regression when forcing the intercept to pass through 0.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i(x, y, t)^2$$

When can MLR be applied?
1. When the different source contributions are not well correlated
2. When two or more sources are of a similar order of magnitude
3. There are no significant missing sources
4. Linearity is applicable
**Uncertainty in the factors ($a_i$)**

- **Boot strapping** methods are applied
  
  The random selection, with replacement, of the data

- 10,000 realisations are made and the standard deviation of the source correction factors ($a_i$) are determined

- Provides an uncertainty in the scaling factors based on the limited sample representation
Two sets of data used:

- **All data**: 103 daily mean modelled and observed PM$_{2.5}$ concentrations from 4 stations

- **Filter days RV4**: 38 twelve hourly mean modelled and observed PM$_{2.5}$ concentrations corresponding to the filter samples at the RV4 site (for validation)

---

**Results**

All data: 103 days at RV4 site
Results: all data (1)

- Model source contributions and correlation ($r^2$) matrix

<table>
<thead>
<tr>
<th>Model source</th>
<th>(%)</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Background</td>
<td>32</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>2. Exhaust</td>
<td>18</td>
<td>0</td>
<td>1</td>
<td>0.86</td>
<td>0.17</td>
<td>0.38</td>
<td>0.35</td>
</tr>
<tr>
<td>3. Suspension</td>
<td>4</td>
<td>0</td>
<td>0.86</td>
<td>1</td>
<td>0.06</td>
<td>0.33</td>
<td>0.24</td>
</tr>
<tr>
<td>4. Wood burning</td>
<td>40</td>
<td>0.01</td>
<td>0.17</td>
<td>0.06</td>
<td>1</td>
<td>0.25</td>
<td>0.22</td>
</tr>
<tr>
<td>5. Area sources</td>
<td>6</td>
<td>0.01</td>
<td>0.38</td>
<td>0.33</td>
<td>0.25</td>
<td>1</td>
<td>0.79</td>
</tr>
<tr>
<td>6. Industrial</td>
<td>0</td>
<td>0.01</td>
<td>0.35</td>
<td>0.24</td>
<td>0.22</td>
<td>0.79</td>
<td>1</td>
</tr>
</tbody>
</table>

Modelled source contribution to total PM$_{2.5}$ mass

Four sources used in the multiple linear regression
Results: all data (2)

Model vs observations

Regression model vs observations

<table>
<thead>
<tr>
<th>Model source</th>
<th>Scaling factor ($a_i$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Regional background</td>
<td>1.22 ± 0.07</td>
</tr>
<tr>
<td>3. Traffic induced suspension</td>
<td>7.6 ± 1.0</td>
</tr>
<tr>
<td>4. Wood burning</td>
<td>0.30 ± 0.06</td>
</tr>
<tr>
<td>5. Other area sources</td>
<td>0.75 ± 0.42</td>
</tr>
</tbody>
</table>
Results: all data (3)

- Correlation ($r^2$) increases from 0.36 to 0.50
- RMSE decreases from 7.9 µg/m$^3$ to 5.7 µg/m$^3$

PM2.5 mean concentrations for 4 stations and 103 days

- Total PM2.5 concentration
- Regional background
- Exhaust particles
- Traffic induced suspension
- Wood burning for domestic heating
- Other area sources

Concentration (µg/m$^3$)
Results: validation at RV4 (1)

- Model vs observations
- Regression model vs observations

<table>
<thead>
<tr>
<th>Model source</th>
<th>Scaling factor ($a_i$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Regional background</td>
<td>-</td>
</tr>
<tr>
<td>3. Traffic induced suspension</td>
<td>$10.6 \pm 1.6$</td>
</tr>
<tr>
<td>4. Wood burning</td>
<td>$0.34 \pm 0.22$</td>
</tr>
<tr>
<td>5. Other area sources</td>
<td>-</td>
</tr>
</tbody>
</table>
Comparison of regression model with receptor modelling for the filter days at RV4

PM2.5 mean concentrations for the RV4 station and the 38 filter days

- Total PM2.5 concentration
- Regional background
- Exhaust + other combustion area sources
- Traffic induced suspension
- Wood burning for domestic heating

Concentration (µg/m³)

PM2.5, Dispersion model
PM2.5, Regression (2 sources)
PM2.5, Receptor model (PMF-2)
PM2.5, Measurement (Kleinfiltersgerät)
Conclusions

- The inverse modelling indicates a significant discrepancy in the dispersion model source contribution for wood burning and traffic suspension.

- This deviation has been quantitatively confirmed by comparison with independent source apportionment studies using receptor modelling.

- For wood burning this deviation could be due to either emissions or to model formulation. The dispersion model is sensitive to emission height and wind speed.

- For traffic induced suspension this deviation is due to emissions.

- Combination with receptor modelling results is important for interpretation.