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

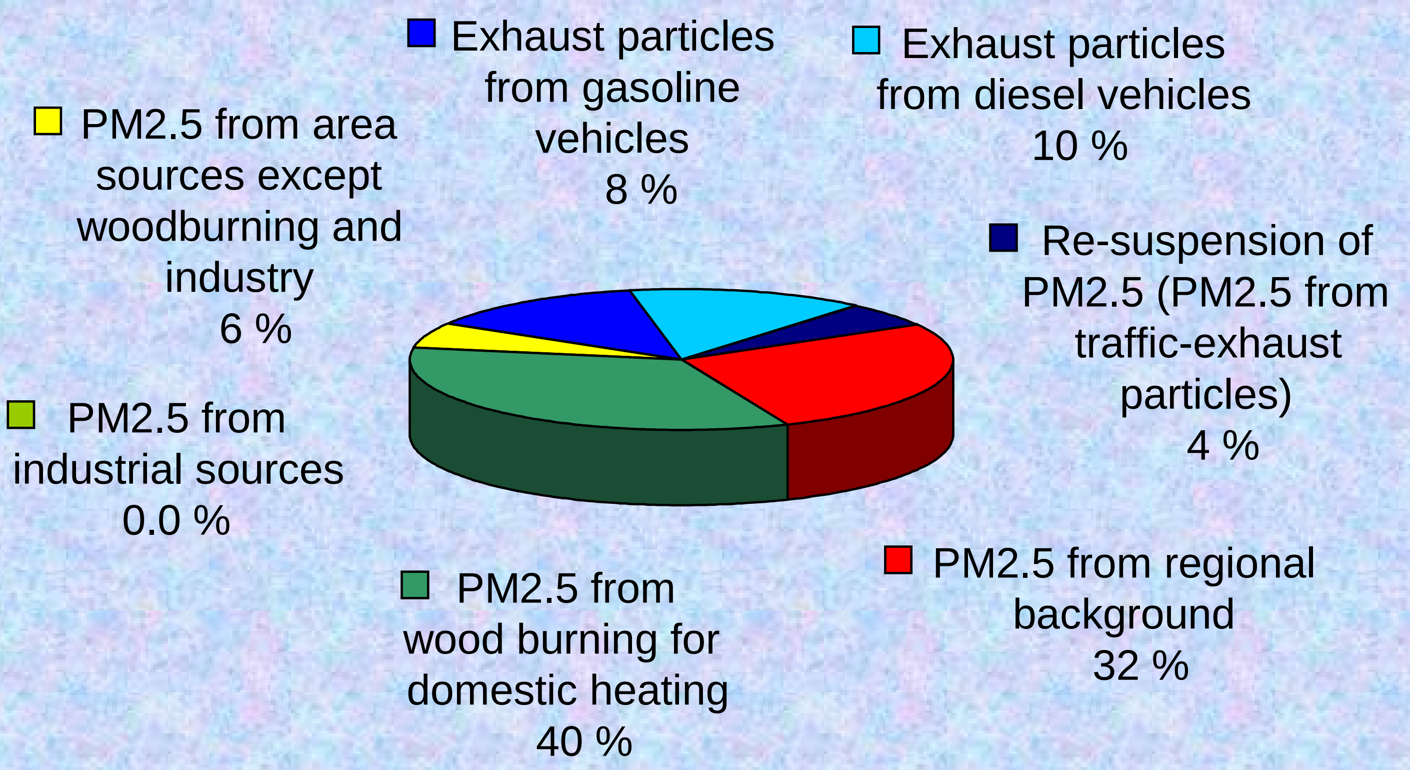
norden
Nordic Council of Ministers

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**

Contributions to $PM_{2.5}$ in Oslo

- **Modelled source contributions at a traffic station in Oslo based on the current emissions inventory**



Observations of $PM_{2.5}$ in Oslo

- **$PM_{2.5}$ observational network during winter 2004**

Traffic stations

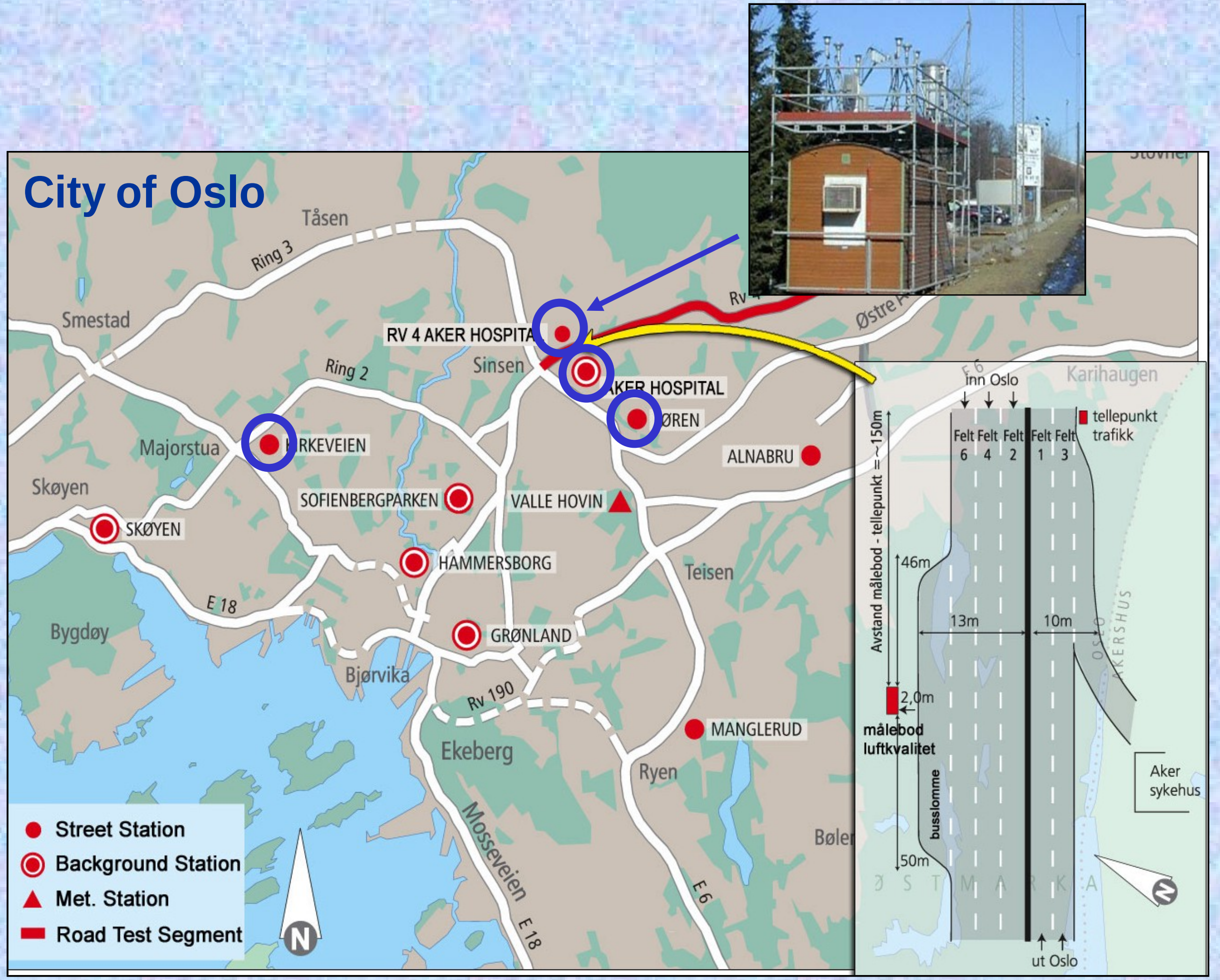
- **RV4**
- **Kirkeveien**
- **Løren**

Urban background

- **Aker Hospital**

Filter samples

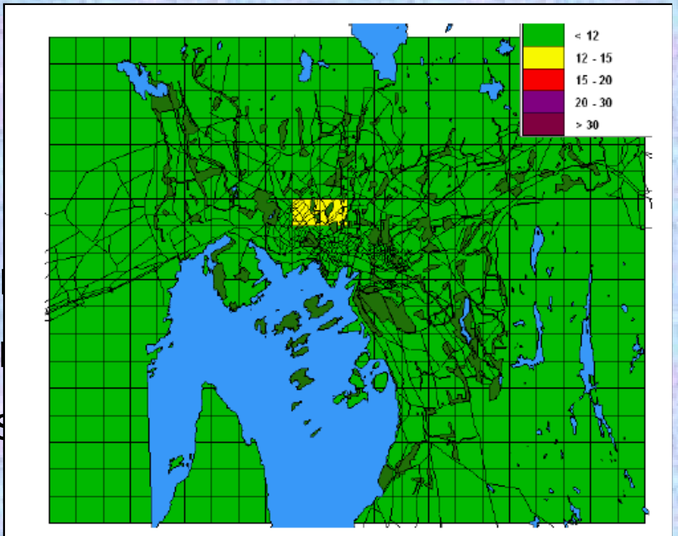
- **RV4**
- 38 twelve hour samples



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-**
  - Area sources, e.g. wood burning, use
  - Traffic sources use Gaussian line sou
  - Industrial sources use a Gaussian poi
  - Meteorology using meteorological mas field model (MATHEW)



Figur 14: Årsmiddelverdier for PM_{25} 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 (ε) where the scaling factor (a_i) is the weight

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

Multiple linear regression

- We wish to minimise the error (ε)
- 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.

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

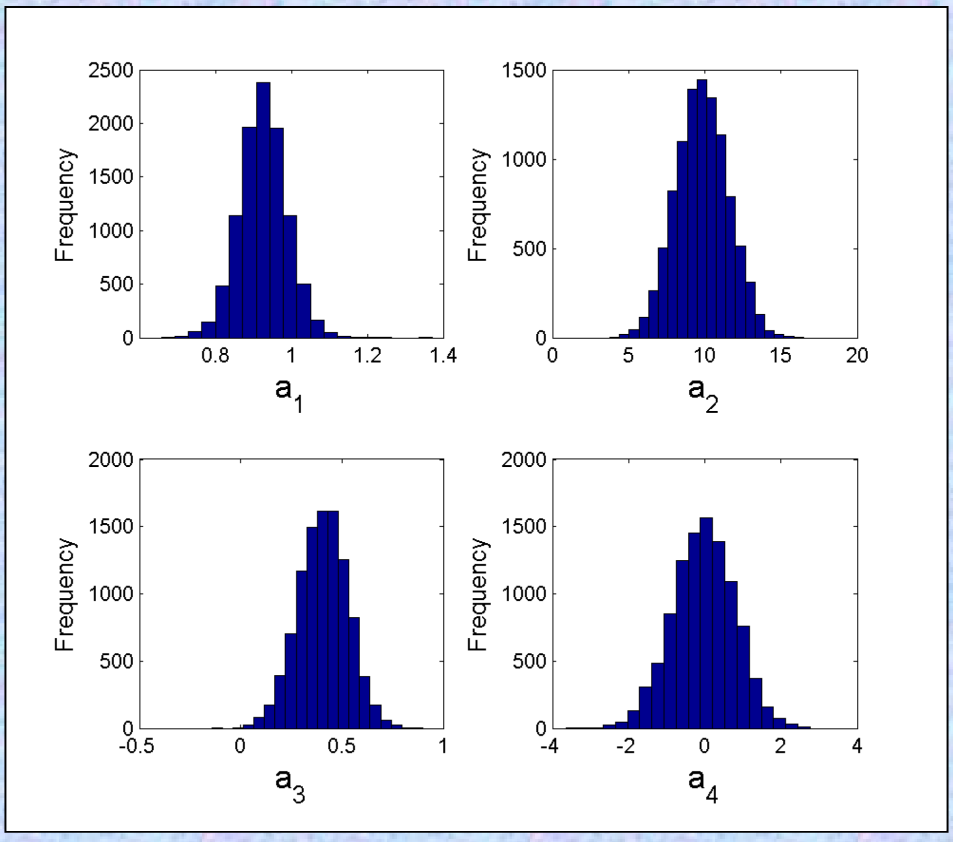
- When can MLR be applied?
    1. When the different source contributions are not well correlated
    2. When two more more sources are of a similar order of magnitude
    3. There are no significant missing sources
    4. Linearity is applicable

Uncertainty in the factors (a_i)

- **Boot strapping** methods are applied

  The random selection, with replacement, of the data

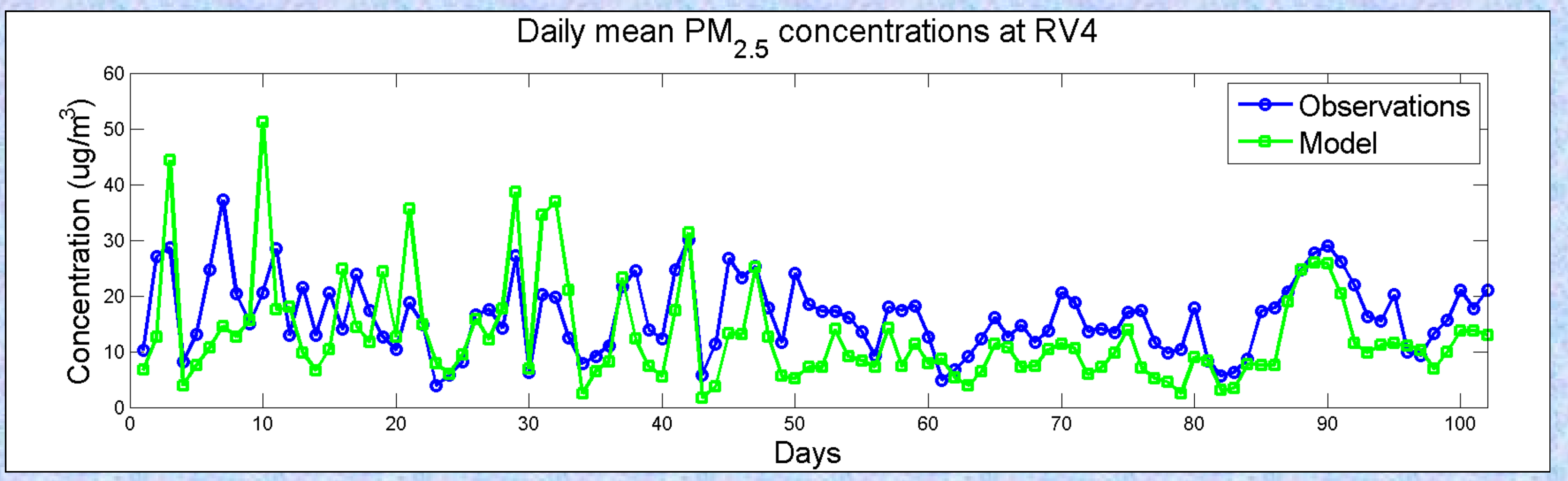
- 10 000 realisations are made and the standard deviation of the source correction factors (a_i) are determined
- Provides an uncertainty in the scaling factors based on the limitted sample representation



Results

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)



All data: 103 days at RV4 site

Results: all data (1)

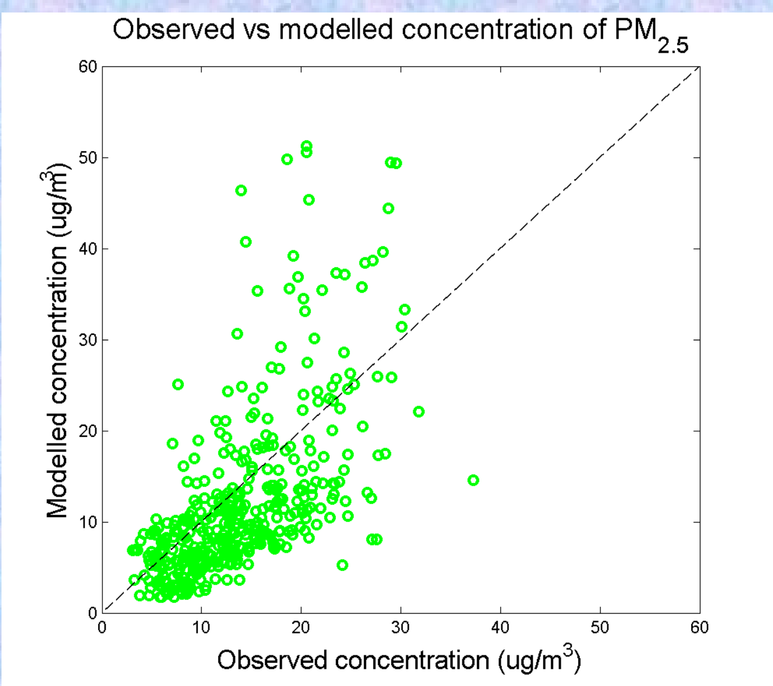
- Model source contributions and correlation (r^2) matrix

Model source	(%)	1.	2.	3.	4.	5.	6.
1. Background	32	1	0	0	0.01	0.01	0.01
2. Exhaust	18	0	1	0.86	0.17	0.38	0.35
3. Suspension	4	0	0.86	1	0.06	0.33	0.24
4. Wood burning	40	0.01	0.17	0.06	1	0.25	0.22
5. Area sources	6	0.01	0.38	0.33	0.25	1	0.79
6. Industrial	0	0.01	0.35	0.24	0.22	0.79	1

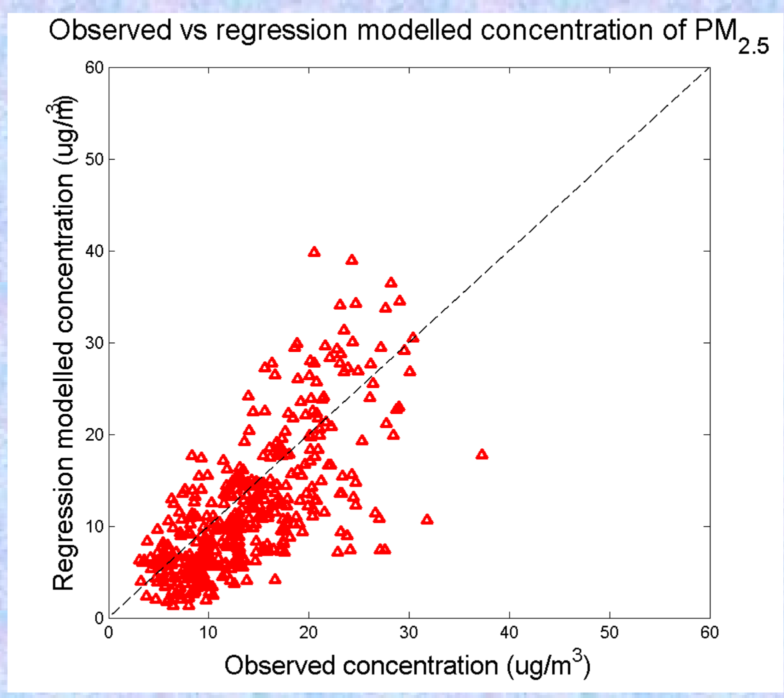
Modelled source contribution to total $PM_{2.5}$ mass

Four sources used in the multiple linear regression

Results: all data (2)



Model vs observations



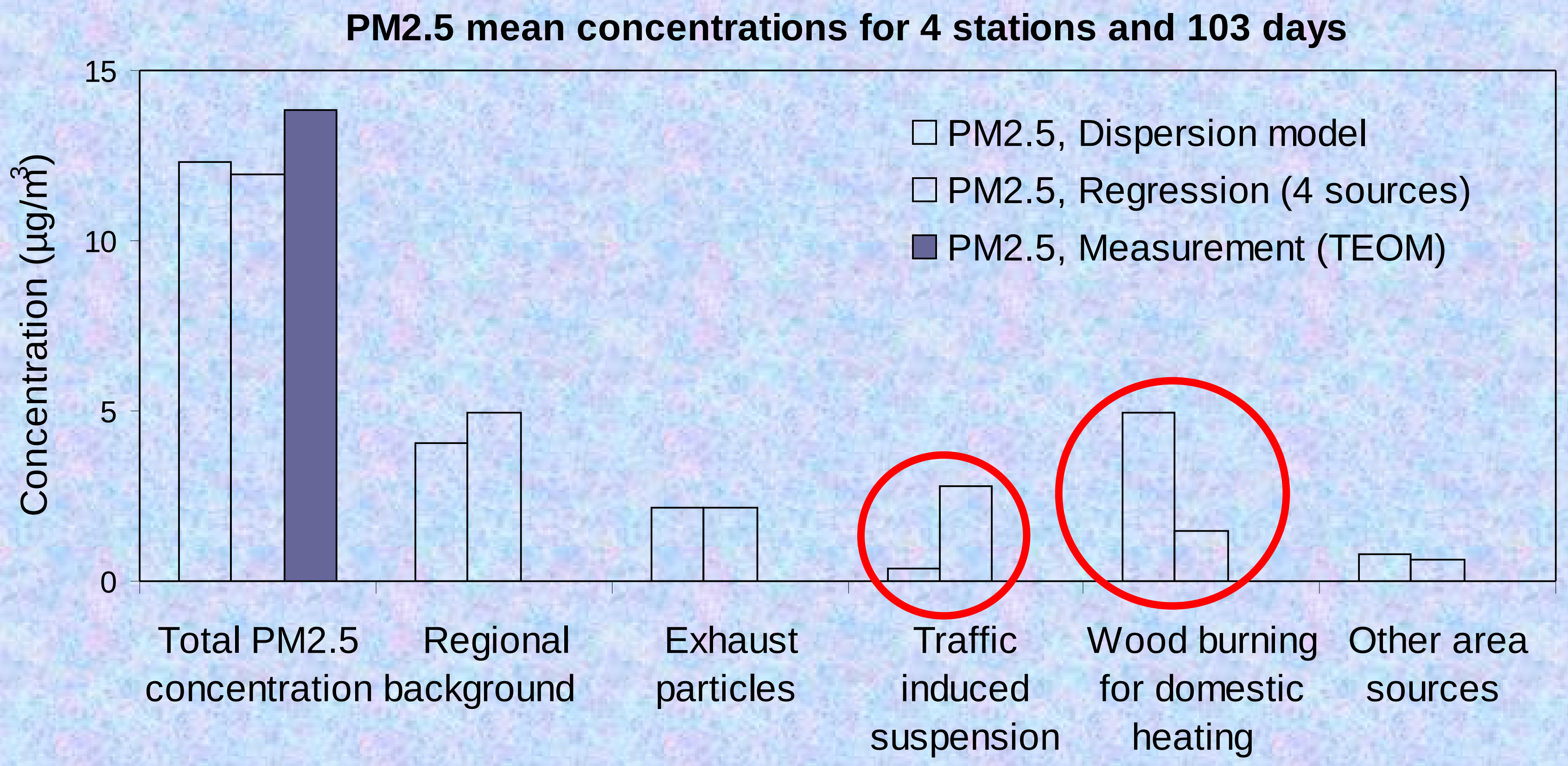
Regression model vs observations

Multiple Linear Regression

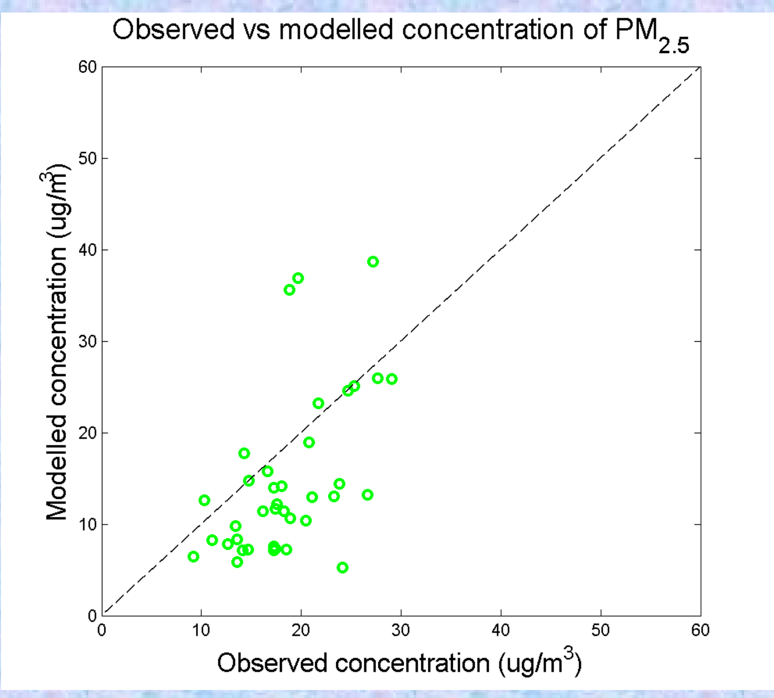
Model source	Scaling factor (a_i)
1. Regional background	1.22 ± 0.07
3. Traffic induced suspension	7.6 ± 1.0
4. Wood burning	0.30 ± 0.06
5. Other area sources	0.75 ± 0.42

Results: all data (3)

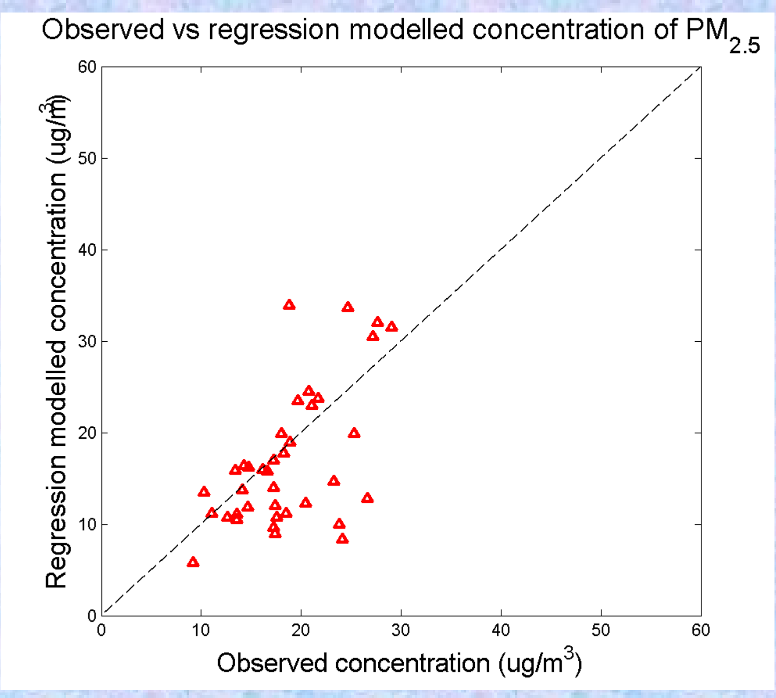
- Correlation (r^2) increases from 0.36 to 0.50
- RMSE decreases from 7.9 µg/m^3 to 5.7 µg/m^3



Results: validation at RV4 (1)



Model vs observations



Regression model vs observations

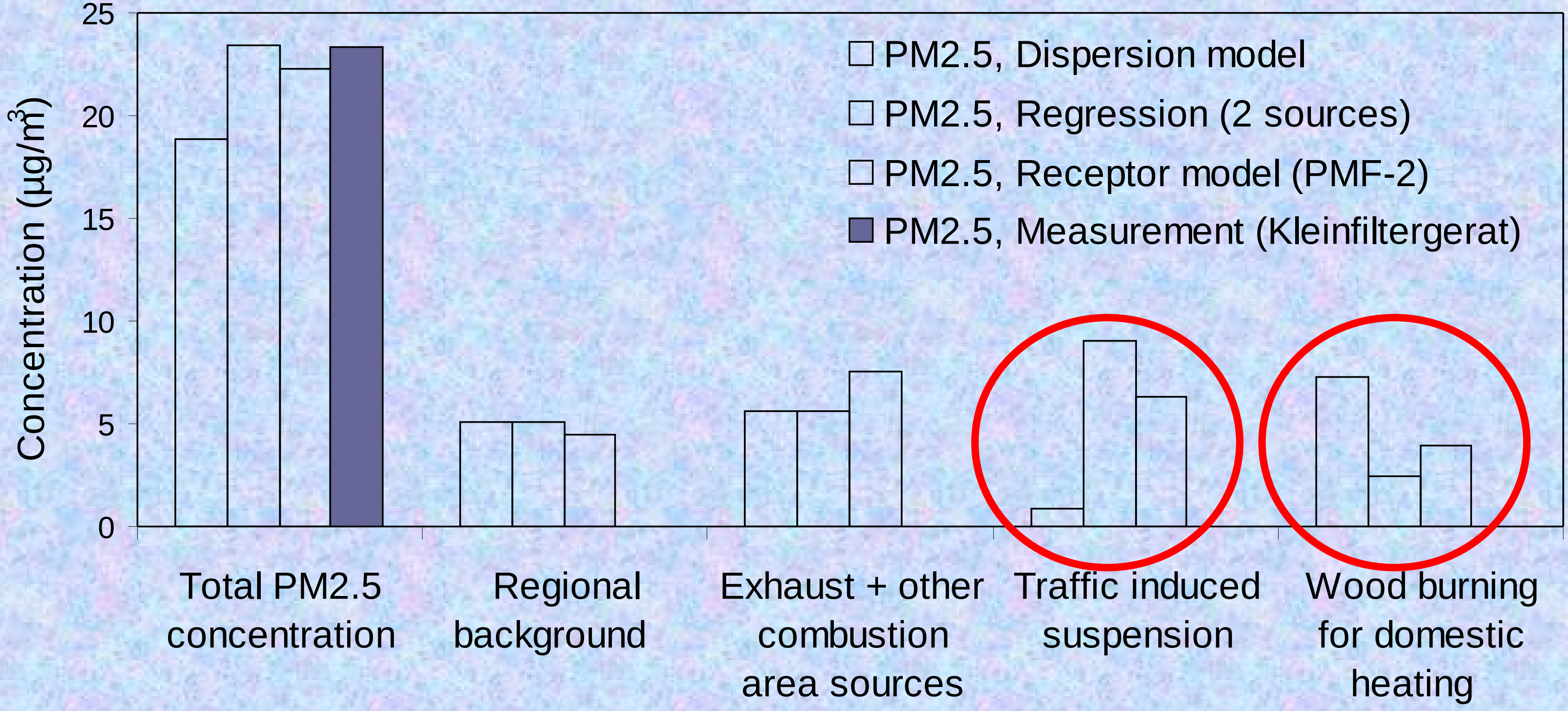
Multiple Linear Regression

Model source	Scaling factor (a_i)
1. Regional background	-
3. Traffic induced suspension	10.6 ± 1.6
4. Wood burning	0.34 ± 0.22
5. Other area sources	-

Results: validation at RV4 (2)

- 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



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