REVERSE MODELLING FOR THE DETERMINATION OF FUGITIVE SOURCES OF PM10

Guido Cosemans and Clemens Mensink
VITO – Flemish Institute for Technological Research, Boeretang 200, B-2400 Mol, Belgium

INTRODUCTION

A compost facility Cf and a sand trader St have their facilities at 160 m to 370 m east of the ambient air quality monitoring site M705 in Kortrijk (VMM, 2006). The pollutant rose on figure 1 shows the impact of these facilities –and other local sources- upon the PM10 concentrations measured during 2002. That impact of the local sources on the year average in 2002 is 8.4 µg/m³, 48% of which is due to the two facilities (Table 1).

The local source impact has been determined as follows. The year average of the original time series of ½hourly PM10 concentrations is 38.8 µg/m³. The the PM10 monitoring sites of the Flemish Environmental Agency VMM in 2002 in Flanders are indicated on Figure 2. There are a number of background monitoring sites located at places remote from local sources. The ½hourly PM10 concentrations measured there vary very synchronously in time and have quite comparable values. For each ½hour, the median value of these background concentrations is taken to construct a PM10-background time series. By subtracting this background time series from the original ½hourly PM10 data at M705, a time series with the impact of local sources is obtained which can be used for making pollutant roses and for analyzing the local source impact with respect to time- and wind speed variability.

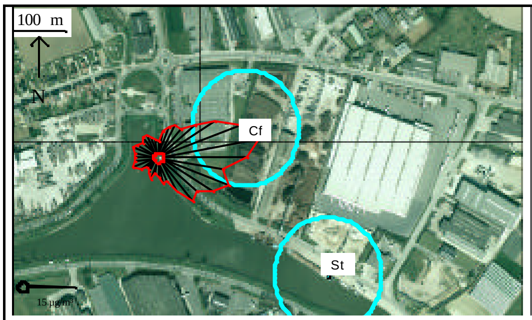

Figure 1

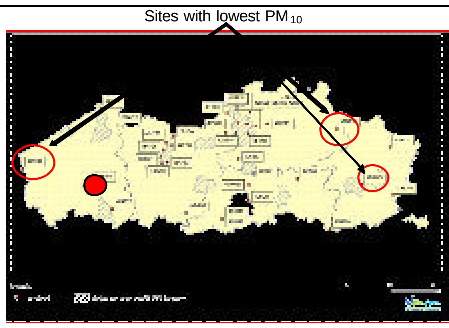

Figure 2. PM10 monitoring sites in Flanders, 2002. The red dot is M705

Table 1

Wind sector	average concentration (µg PM10/m³)	% of time	% of total concentration
50°-150°	21.5	19	48
other	5.3	81	52
all	8.4	100	100

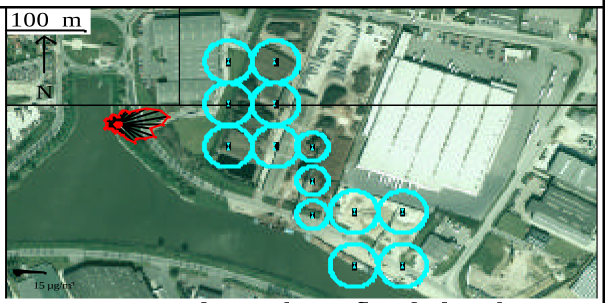

Figure 3. Industrial site flooded with unit sources

A simple reverse modelling exercise, using two unit sources of 1 ton PM10/year, one source at the midpoint of the Cf facility and the other one at the midpoint of the Sf facilities, gave an emission of 4.8 ton/year for Cf and 11.6 ton/year for Sf. This emission leads to an IFDM-computed impact of 4.7 µg/m³ on the year average. These results have been reported to the environmental authorities who ordered this study (Mensink, 2007).

Wee shall use the above data set to explore some possibilities, limits and pitfalls of using least squares regression for reverse modelling, next give a more elaborated example on how regression and reverse modelling helps to understand a time series of observed PM10 data.

What you should know about least squares regression

Averaging time. Regression is done on a time series of observations and a set of time series of computed impacts for unit sources. At each potential fugitive source location, an unit source is placed. The time series can consist of ½hourly concentrations of day-averaged concentrations. For industrial sites with pollutants as heavy metals or PAH's, only the latter will be routinely available, so it is useful to explore the potential of regression for this kind of data as well. One reason to do regression on day averages instead of ½hourly data, is the uncertainty on the transport wind. The wind direction used for the M705 example is onsite, measured at 30 m above the ground and above the PM10-monitor. The fugitive PM10 emissions are transported by the wind near ground-level. Its direction will certainly be affected by the buildings and sand and compost piles located upwind. We do not know the direction of the transport wind, as we know only the measured wind. This uncertainty on transport wind affects regression using day averages only in so for the day wind roses of measured and transport wind are different, while for regression using ½hourly data, each ½hour with a significant difference between both wind direction leads to an observation that can not be explained by regression. This probably explains why the correlation coefficient between observed and predicted time series with day-averages is 2 to 3 times higher than for regression with ½hourly time series.

Regression constant. Regression can be done with or without the addition of a regression constant. If there is pollution coming from other wind sectors then those occupied by the unit sources, the use of a constant is recommended. For the example of the introduction (Table 1), regression should give the constant a value of 4.4 (µg PM10/m³), namely that part of the yearly average that can not be explained by the (unit) sources used. This value is obtained when doing regression with day-averaged concentrations. Regression of half-hourly data produces a constant of –depending upon the source configuration entered- 5.2 to 7.6 µg/m³, the lowest (and most correct) value being obtained for unit sources that reflect time- and wind speed dependency as derived from observations.

In the example of Figure 1, regression without constant typically gives source strengths that are 50% to 100% higher than given by regression with a constant, and this for as well regression on ½hourly data as on day-averaged data. There are reasons for this difference. Figure 4a shows that this is to be expected, as one imposes a steeper slope on the regression line when requesting that the regression constant be zero. Figure 4b shows the same reasoning applied to time series. If the peak value is 6 µg/m³, and the regression constant is 4 µg/m³ (dashed line on Figure 4b), than the unit source must only produce 2 µg/m³ to match the peak value. If the regression constant is zero, then the unit source must only produce all 6 µg/m³ to match the peak value.

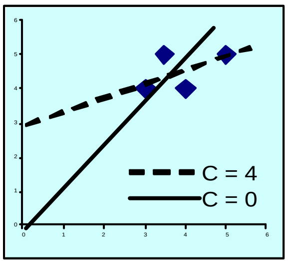

Figure 4a. Regression constant and slope of fitting line

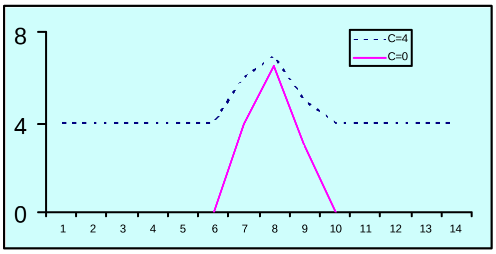

Figure 4b. Regression constant and predicted time series

Noise fitting. Regression does not give proportionality factors $\boldsymbol{Q_i}$ to the unit sources $\boldsymbol{i}$, but it produces values $x_i = \boldsymbol{Q_i} \pm D_i$, where the D_i can be as large as –or even larger- than $\boldsymbol{Q_i}$.. Regression does so to obtain a smaller sum of squares of differences between observed and predicted concentrations. Figure 3 shows a unit source configuration that will certainly lead to noise fitting. The Cf- and St-site terrains are covered with 13 small unit sources. Regression assigns to these sources strengths that vary from -100 tons to +150 tons PM10/year. One can remove sources to which a negative source strength was given from the regression, until all remaining sources receive a source strength greater than zero. In this case, only two sources remain, located quite closely to the two fugitive sources on Figure 1.

Correlation coefficient R. Successful noise fitting increases the correlation only slightly. A noise fitting variable can be recognized by that the corresponding regression coefficient has a large standard deviation. In the subsequent example, all variables with a standard deviation on the regression coefficient larger than 30% are removed from the regression.

Understanding the time series

The time series of local-source PM10 impact at the M705 site for the year 2002 is given Figure 5. Shown are the central moving 14-day averages computed from the original ½hourly data. In an attempt to understand the sources that cause up and downs in this time series, a ring of unit area sources has been laid around the measuring point. (Figure 6. This figure gives also the pollutant rose of local-source PM10 impact.) Each circle on Figure 6 represents a unit source with constant emission, and a 'unit source' whose strength depends upon wind speed in the following way. The source is active at wind speeds of 2 m/s or above, has a 'unit emission' of {1 times $\max(0,\min(8,(\boldsymbol{u}-2)))^3$} tons/year, where $\boldsymbol{u}$ is the wind speed at 10 m above the ground. This strength of this 'unit' area source is proportional to the 3rd power of the wind speed minus 2 m/s, up to a wind speed of 10 m/s, above which wind speed the source term is at maximum strength of 8^3 or 512 tons/year.

Least squares regression is done on the time series of all ½hourly data in the summer of 2002. (Using the ring of sources of Figure 6, regression with day-averaged data would lead to unwanted and hard to avoid noise fitting effects.) Regression withholds 22 sources, totaling 19 tons PM10/year. Most area sources received a constant emission smaller than 0.5 ton PM10/year. Sources on the Cf-St terrain received 12 tons/year, of which 8 ton wind speed depended. (The summer period was chosen because regression on the entire year yielded results that weren't very useful.) The emissions, derived from the summer PM10-measurements, reproduces fairly well the base line of the pollution measured (Figure 7).

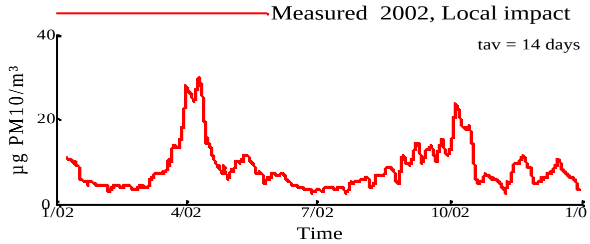

Figure 5. Time series observed impact of local PM10 sources

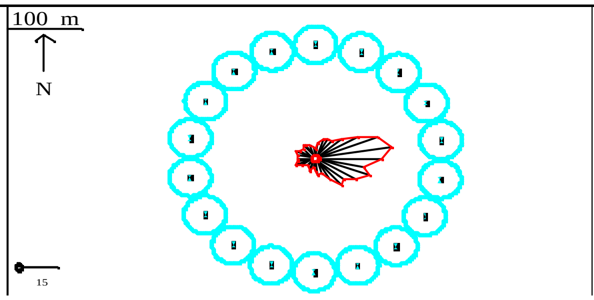

Figure 6. A ring of area sources and pollutant source of local-source PM10 impact.

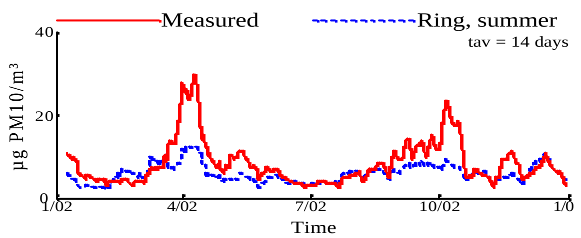

Figure 7. Base line of pollution reproduced by Ring of sources (summer), 22 sources, 19 ton PM10/year, 6.0 µg/m³ (R=0.59, R²=0.35, for day averages)

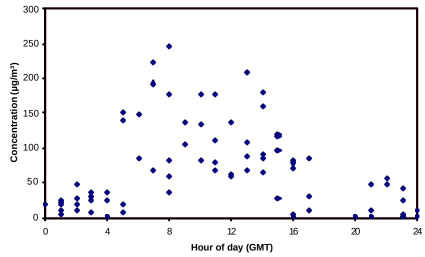

Figure 8. Time dependency of observed concentrations, 2-6 April 2002

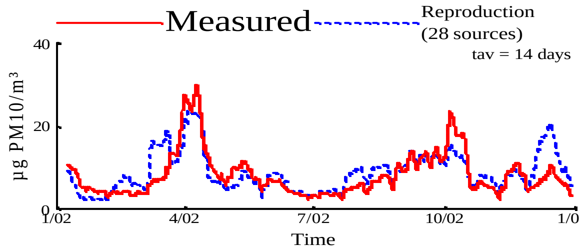

Figure 9a. Observed and reproduced time series 2002. (regression summer + regression October) (R=0.61, R²=0.38, for day averages)

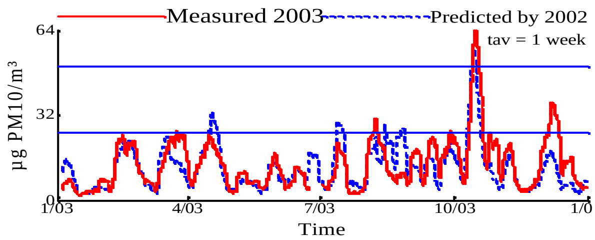

Figure 9b. Observed and predicted time series 2003. Emissions estimated from the 2002 data. (R=0.67, R²=0.45, for day averages)

An analysis of the ½hourly concentrations of 2-6 April 2002 shows that elevated concentrations occur between 6 h and 15 h GMT, that is, between 8 and 17 hours local time (Figure 8). We now subtract the computed base line pollution from the original ½hourly local-impact PM10 time series, and try regression on the time series for the period 29 September – 12 October 2002, a period with higher observed concentrations than reproduced by the 22 sources regression has already given.

Then, the part of the ring of sources covering the Cf-St facilities was given day-time-unit emissions, both constant and wind speed dependent. Regression returned 6 sources, four constant and two wind speed dependent. With these sources, both the April and October peak could be reproduced. However, the ½hourly time series for the entire year 2002, computed with this source data, showed a few ½hourly concentrations that were 3 to 4 times the greatest ½hourly-value ever measured (1200 µg/m³ versus 350µg/m³). Analyzing these small set of cases (some 40 ½hours over one year) suggested to reduce the two wind speed dependent source terms by a factor of 4, resulting in a less well reproduced October peak, but giving a better agreement for the rest of the year.

The time series of the computed impact of these 22+6 = 28 sources is shown in Figure 9a. As measured PM10 data for 2003 was available, these sources were used to 'predict' the 2003 measurements. This turned out to be satisfactory (Figure 9b).

Total emissions assigned to the Cf and St facilities for 2002 are 27 tons/year: 8tons of constant sources, 19 tons to wind speed driven sources. Constant sources added 4 µg/m³ to the year average, wind speed driven sources 1.8 µg/m³. (There is some double counting in this example on how to use regression.) Taken apart, a constant source causes a maximum ½hourly concentration of 10-20 µg/m³, a wind driven one 100-200 µg/m³. It are these sources that cause the high PM10 concentrations at the M705 air quality monitoring site.

The analysis up to now used only the time series as source of information, but pollutant roses and the cumulative frequency distribution (Figure 10) also provide information to assess the quality of a solution given by least squares regression.

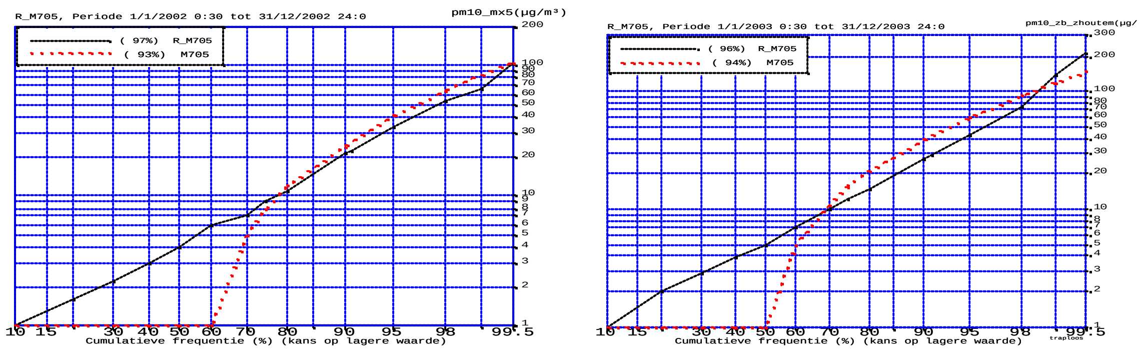

Figure 10. Cumulative frequency distribution of observed and computed local-source PM10 impact, 2002 (left) and 2003 (right).

CONCLUSIONS

Reverse modelling using least squares regression can reveal some interesting features of the pollutant sources in the vicinity of an ambient air quality monitoring site. The method can be applied to as well ½hourly as to day-averaged data; one must however be aware of the dangers of noise fitting. Advantages of the method are that no emission factors or a priori knowledge on the sources is required.

REFERENCES

Cosemans, G., Kretzschmar, J.G. and Maes, G. 1992:, The Belgian Immission Frequency Distribution Model IFDM, in Olesen, H. and Mikkelsen, T. (Editors) Proceedings of the DCAR Workshop on Objectives for Next Generation of Practical Short-Range Atmospheric Dispersion Models, NERI, Roskilde, Denmark.

Horst, A, 1977: Surface depletion model for deposition from a Gaussian plume. Atm.Env. **11**, pp. 41-46

Mensink, C. et al., 2007: Qantification of diffuse and fugitive sources of pm10 by means of the integrated "hot-spot" method, DustConf 2007, 23-24 April, Maastricht, The Netherlands , http://www.dustconf.org/

VMM, 2006: Luchtkwaliteit in het Vlaamse Gewest – 2005. Vlaamse Milieumaatschappij, Aalst. 166 pp. + bijlagen (yearly reports)

Wampler R.H., 1979: Algorithm 544: L2A and L2B, Weighted Least Squares Solutions by Modified Gram-Schmidt with Iterative Refinement. ACM TOMS, **5**, 494 - 499