POLLUTION ROSES FOR 24 H AVERAGED POLLUTANT CONCENTRATIONS BY REGRESSION

Guido Cosemans and Jan Kretzschmar
Flemish Institute for Technological Research (VITO) , Centre for Remote Sensing and Atmospheric Processes (TAP), B2400 Mol, BELGIUM

INTRODUCTION

Pollution roses are polar diagrams that show how air pollution depends on wind direction. If an ambient air quality monitoring station is markedly influenced by a source of the pollutant measured, the pollution rose shows a peak towards the local source. The construction of a pollution rose is simple when both wind direction and pollutant concentrations are averages over one hour. The measurements are divided into 36 partitions, each partition corresponding to a 10° wind direction sector. For the pollutant concentrations in each partition, one can calculate the arithmetic average, and also determine the 95-th or 98-th percentile if needed. Such pollution roses of averages and percentiles can be used for model validation, as we did in *Cosemans, G. et al.* (1983).

More sophisticated, *Henry, R.C. et al.* (2002), determine a locally optimal wind direction bin width $\Delta\Theta$ and compute average concentration with a confidence interval for each bin using nonparametric regression; finally, they use interpolation in order to find the peak locations with greater precision than 1°.

The construction of pollution roses for 24 h averaged concentrations is of great practical interest, although we found no papers on this subject in the leading journals in the field of air pollution. For a number of pollutants, such as heavy-metals, day averaged concentrations are routinely monitored at many locations, for instance in Belgium since 1972. Ambient air quality standards do not justify the cost of monitoring such pollutants with greater time resolution. Yet one is of course interested in the wind direction dependency of these pollutants. In *Kretzschmar, J.G. and G. Cosemans* (1979), pollution roses superimposed on geographic maps were used to illustrate the source areas of the heavy-metal concentrations measured. These pollution roses were computed according to:

$$C_{dd} = \sum_{j=1,n} p_j f_{dd,j} \alpha_j \bigg/ \sum_{j=1,n} f_{dd,j}\, \alpha_j \qquad (1)$$

where C_{dd} is the average concentration for wind sector dd, n is the number of days in the period for which the rose is constructed, p_j is the measured concentration on day j, $f_{dd,j}$ is the frequency that the wind came from sector dd and α_j is some weight function based on the persistency of the wind vector during day j (*Thiessen, L. and Y. Lenelle,* 1991). In this paper, α_j is set to the inverse of the number of wind direction bins with non-zero frequency. Equation (1) basically assumes that, on a given day, all hourly concentrations are equal.

AN EXAMPLE

We construct a simple example. The wind direction is classified in three bins. During the hours that the wind direction comes from bin 1, the air carries a pollutant concentration of 10 µg/m³. Otherwise, the pollutant concentration is 0 µg/m³. In Table 1, the wind direction frequencies f_{dd} for 4 days are listed, together with the resulting sum of hourly concentrations and 24 h averaged pollutant concentration.

Table 1. Data for a simple example.

Real concentration $C_{R,dd}$ of pollutant in the air from wind direction bin dd_i					
	dd_1	dd_2	dd_3		
	10 µg/m³	0 µg/m³	0 µg/m³		
day (j)	number of hours (f_{dd}) that the wind is in direction bin dd_i			Sum of hourly concentrations[1]	24 h averaged concentration (p_j)
1	5	8	11	50 µg/m³	50/24 µg/m³
2	12	6	8	120 µg/m³	120/24 µg/m³
3	9	12	3	90 µg/m³	90/24 µg/m³
4	9	12	3	70 µg/m³ [²]	70/24 µg/m³

[1]: The sum of the hourly concentrations for day j is $\Sigma C_{dd} * f_{dd}$

[2]: For the 4-th day, we lowered the sum of hourly concentrations from 90 µg/m³ to 70 µg/m³ to mimic the fact that real pollution data do not follow the simple arithmetic rules we used for days 1-2-3.

Using α_j = 1, the pollution rose for the 24 h averaged concentrations in Table 1 according to formula (1) is:

$$C_{dd1} = 3.73\ \mu g/m^3, \qquad C_{dd2} = 3.33\ \mu g/m^3, \qquad C_{dd3} = 3.25\ \mu g/m^3 \tag{2}$$

From Table 1, one can also derive the system of linear equations (3) which states that the sum of the products f_{dd} times C_{dd} is equal to the sum of the hourly concentrations per day.

$$\begin{aligned} 5\,C_{dd1} + 8\,C_{dd2} + 11\,C_{dd3} &= 50 \\ 12\,C_{dd1} + 6\,C_{dd2} + 8\,C_{dd3} &= 120 \\ 9\,C_{dd1} + 12\,C_{dd2} + 3\,C_{dd3} &= 90 \\ 9\,C_{dd1} + 12\,C_{dd2} + 3\,C_{dd3} &= 70 \end{aligned} \tag{3}$$

The system of linear equations (3) can be solved by a Least Squares method. We obtain:

$$C_{dd1} = 10.0\ \mu g/m^3, \qquad C_{dd2} = -1.0\ \mu g/m^3, \qquad C_{dd3} = 0.7\ \mu g/m^3 \tag{4}$$

Solution (4) is clearly a better estimate of the real concentrations $C_{R,dd}$ than is solution (2). Using the symbols of equation(1), the system of equations (3) is:

$$\sum_{dd=1,ndd} c_{dd} f_{dd,j} = p_j \sum_{dd=1,ndd} f_{dd,j} \tag{5}$$

where $\Sigma f_{dd,j}$, the sum of frequencies of the wind direction over the *ndd* different bins during day j, is 24 for hourly wind direction data and 48 for half hourly wind data.

A PRACTICAL SYSTEM OF EQUATIONS FOR THE REGRESSION

The Least Squares Algorithm (LSA), applied to the system of equations (5b), usually returns values for the concentrations C_{dd} that, for successive wind direction bins, oscillate between large positive values and large negative values. In order to prevent the LSA to produce large negative values, we add the solution of equation (1) to the system (5). In doing so, the system of equations (3) in our example becomes system (6) below, where a new weighting function β is introduced.

$$
\begin{array}{l}
5\,C_{dd1} + 8\,C_{dd2} + 11\,C_{dd3} = 50 \\
12\,C_{dd1} + 6\,C_{dd2} + 8\,C_{dd3} = 120 \\
9\,C_{dd1} + 12\,C_{dd2} + 3\,C_{dd3} = 90 \\
9\,C_{dd1} + 12\,C_{dd2} + 3\,C_{dd3} = 70 \\
\beta\ C_{dd1} = \beta\ 3.73 \\
\beta\ C_{dd2} = \beta\ 3.33 \\
\beta\ C_{dd3} = \beta\ 3.25
\end{array}
\tag{6}
$$

The system of equations (6) is solved iteratively with varying values for β, where we look for a value of β below which, the LSA returns negative values in the solution vector, and above which the solution vector has only elements greater than or equal to zero. For half hourly meteorological data, a good starting value for β is 0.03*n*48/ndd, where n is the number of days and ndd the number of wind direction bins.

Remark 1. For large values of β, the solution of the system of equations (6) becomes equal to the solution of equation (1). Therefore, if there still are negative values in the solution vector after four iterations, we apply a smoothing operator over the solution vector such as a moving average [$C_{n,new} = (C_{n-1} + aC_n + C_{n+1})/(2+a)$] or a polynomial interpolation. The smoothing must be done over the doses (the product of C_{dd} with f_{dd}) and over the frequencies f_{dd}.
Remark 2. This smoothing operator also removes the needle-like peaks in the pollution rose, that are created by the LSA-algorithm when it replaces near equal concentrations C for three (or more) adjacent wind direction bins with a sequence (zero, 3*C, zero). This is illustrated in Figure 2.
Remark 3. The detection limit of the analytical method, used to determine the pollutant concentrations, can degrade the efficiency of the algorithm. We advise that, prior to the construction of the system of equations, all concentrations lower than the 60-percentile over the monitoring period are changed into zero.

HALF HOURLY AND 24 H AVERAGE SO_2 ROSES FOR ANTWERPEN
Half hourly SO_2 concentrations near five oil refineries are measured in the Antwerp harbour, as well as half hourly wind. We use these half hourly data, year 1997, to calculate pollution roses with 36 wind direction bins of 10° each. Next we calculate 24 h averaged SO_2 concentrations, consecutive in time, which are used to construct a system of equations analogue to system (6), which is solved with the LSA method. In Figure 1, the pollution roses on the left are calculated straight from the half hourly data, the roses of the right-hand are obtained by regression from the 24-h averaged data (4 iterations, no smoothing). In general, the agreement between the roses in Figure 1 is excellent, only one regression rose has a 'false' peak (2nd rose from the right below).

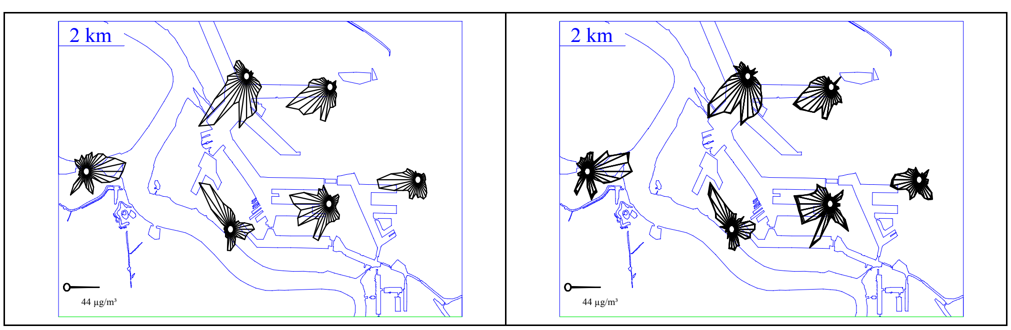

Figure 1. Pollution roses for the Antwerp harbour, 1997. Left: calculated from half hourly SO_2-concentrations. Right: Calculated from 24-hour averaged SO_2-concentrations.

WIND SPEED DEPENDENCY

We calculate the 33-percentile and 66-percentile of the half hourly wind speed measurements over one year to define three wind speed classes: Low, Medium and High, and extend equations (1) and (5) to handle bins defined by wind direction sector and wind speed class. For each of these wind speed classes, we calculate straight pollution roses using the measured half hourly SO_2 concentrations. These roses, for a monitoring station situated in the middle of five oil refineries in the Antwerp harbour, are shown at the left side of Figure 2 above the label '30 minutes/(reference)'. The highest SO_2-concentrations are linked to strong wind from the northwest. Next to these roses is the outcome of formula (1). Peaks in these roses (above the label '(1)') are less pronounced than in the reference roses. The rose for the Medium wind speed class is missing the northwest peak. Above the label 'regression' are the pollution roses obtained after 4 LSA-iterations for β. The Medium northwest peak is reproduced in part. The very tall but needle-thin peak in the 'High wind speed' rose is a typical regression artifact. Applying a smoothing operator produces the pollution roses at the right end of Figure 2.

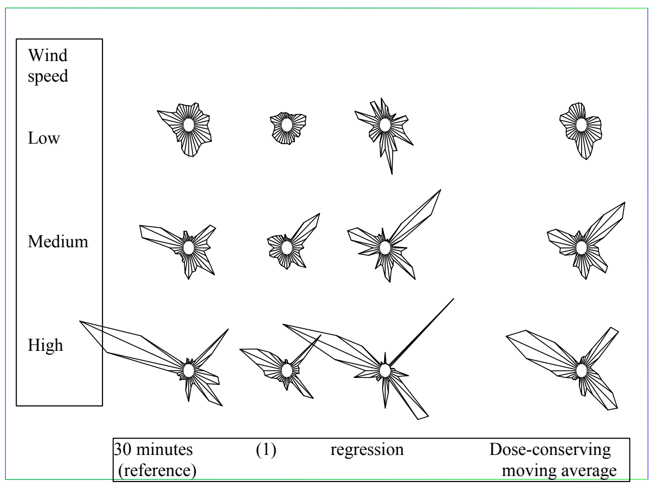

Figure 2. Pollution roses for wind speed dependency. See text for further details.

A PRACTICAL APPLICATION

By means of calculated 24-h averaged SO_2 concentrations, derived from measured half hourly SO_2 concentrations, we have shown that regression can be used to derive wind direction and wind speed dependent information from 24 h pollutant concentrations. For heavy metals, such as lead, we only have 24 h averaged concentrations. In *Cosemans, G. and E. Roekens* (2001), it was shown in a rather laborious way that the fugitive emissions of a lead works increased with the wind speed. Similar information shows up immediately from the pollution rose in Figure 3. Note that the roses in Figure 3 are constructed for wind sectors of 20°. The rose for 10° wind direction bins had too many regression artifacts.

CONCLUSIONS

We illustrated that Least Squares Regression (LSR) can be used to construct pollution roses from 24 h averaged pollutant concentrations. The correctness of these roses is verified by reference roses calculated straight from measured half hourly SO_2 concentrations. These LSR-roses convey more information than the pollution roses obtained by the more traditional methods we know about.

The straigth LSR solution vector for this problem usually contains features that render the LSR solution useless for practical use. We counteract the tendency of LSR to place large negative values in the solution vector by adding equations that put a large penalty on these negative components of the solution. Further, we remove artificial peaks in the LSR solution, due to the tendency of LSR to replace sequences (...,C,C,C,..) by a sequence (...,0, 3C, 0,...), by applying a smoothing operation over the solution vector.

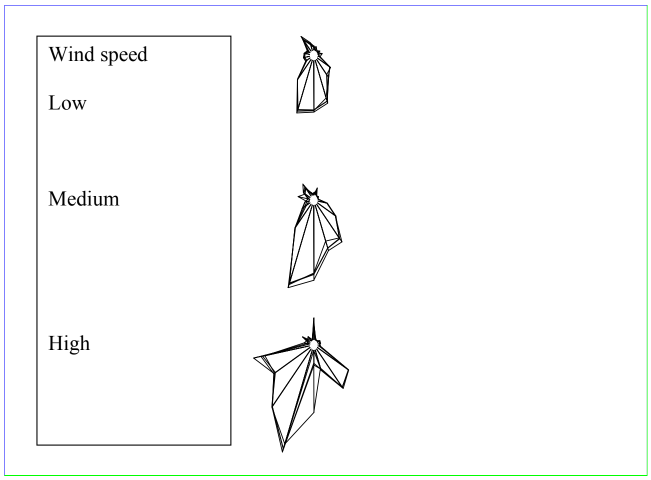

Figure 3. Pollution roses for three wind speed classes. Pollutant Pb, measured near a lead smelter, years 1996-1997.

ACKNOWLEDGEMENT
We thank E. Roekens of the Flemish Environmental Agency (VMM) for having made available the measured daily Pb concentrations, the half hourly SO_2 concentrations and the meteorological data used for the illustrations in this paper.

REFERENCES
Cosemans, G., Kretzschmar, J., De Baere, G. and J. Vandervee, 1983: Large scale validation of a bi-Gaussian dispersion model in a multiple source urban and industrial area, in: Air Pollution Modelling and Its Application II, Ed. C.De Wispelaere, Plenun Press, 709-727.

G. Cosemans, G. and E. Roekens, 2001: Air quality monitoring and modelling near a lead works, Proc. 7th Int.Conf. On Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Ed. C.Cuvelier et al.,EC JRC Ispra, 225-227.

Henry, R.C., Chang, Y.-S., and C.H. Spiegelman, 2002: Locating nearby sources of air pollution by nonparametric regression of atmospheric concentrations on wind direction. *Atmos. Environ,* **36,** 2237-2244.

Kretzschmar, J.G. and G. Cosemans, 1979: 5 Year survey of some heavy-metal levels in air at the Belgian North-Sea coast. *Atmos. Environ.,* **13**, 267-277.

Thiessen, L. and Y. Lenelle, 1991: Evaluatie van de gehalten aan zware Metalen in de omgevingslucht in België, 10th annual report. Instituut voor Hygiëne en Epidemiologie, Brussel.