7.09 POLLUTANT ROSES FOR 24 H AVERAGED POLLUTANT CONCENTRATIONS BY RESPECTIVELY LEAST SQUARES REGRESSION AND WEIGHTED SUMS

Guido Cosemans and Jan Kretzschmar
Flemish Institute for Technological Research (Vito), Mol, Belgium

INTRODUCTION

Pollutant 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 pollutant rose shows a peak towards the local source. When both wind direction data and pollutant concentration are measured as (½)-hourly averages, the pollutant rose is mathematically well defined and the computation is simple.

When the pollutant data are averages over 24 h, as is the case for heavy metals or dioxin levels or in many cases PM10-levels in ambient air, the pollutant rose is mathematically well defined, but the computational scheme is not obvious.

In this paper, two practical methods to maximize the information content of pollutant roses based on 24 h pollutant concentrations are presented. These methods are applied to time series of 24 h SO_2 concentrations, derived from the ½-hourly SO_2 concentrations measured in the Antwerp harbour, industrial, urban and rural regions by the Telemetric Air Quality Monitoring Network of the Flemish Environmental Agency (VMM). The pollutant roses computed from the ½-hourly SO_2 concentrations constitute reference or control-roses to evaluate the representativeness or truthfulness of pollutant roses obtained by the presented methods. The presented methodology is very useful in model validations that have to be based on measured daily averaged concentrations as only available real ambient levels.

While the methods give good pollutant roses in general, this paper especially deals with the case of pollutant roses with ‘false’ peaks.

SYMBOLS USED

Indices

i : index of wind direction sector. Range: 1-36 for wind direction sectors of 10°;

j : day index for the period under investigation. Range: 1-365 for a one year period.

Other symbols

n : a number of times:

$n_{i,j}$: the number of hours that the wind came from wind direction sector ***i*** on day ***j***;

n_i : the number of hours that the wind came from wind direction sector ***i*** during the period;

C : a pollutant concentration;

$\overline{c}_i$: the average pollutant concentration over the hours that the wind came from sector ***i*** during the period of interest;

$\overline{c}_{i,j}$: the average pollutant concentration on day ***j*** when the wind came from sector ***i***

c_j : 24 h pollutant concentration on day ***j***; c_j can have been measured as such, or it can have been computed from hourly data;

r_i : ratio for the wind direction sector ***i***.

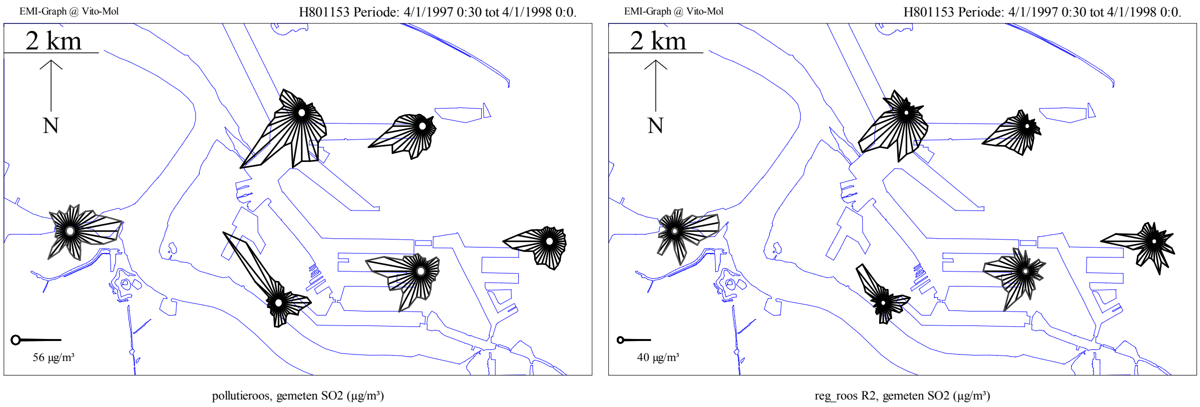

Figure 1. Pollution roses for the Antwerp harbour, 1997, using representative wind data. Left: calculated from half hourly SO_2-concentrations.
Right: calculated from 24 h SO_2-concentrations using least squares regression.

FORMULAS FOR POLLUTANT ROSES

Reference pollutant rose (for hourly pollutant data)

Equation (1) gives the components $\overline{c}_i$ of a pollutant rose when a time series of measured hourly pollutant concentrations is available. For equation (1), the quantity $\overline{c}_{i,j}$ is to be calculated from the individual hourly measured pollutant concentrations. Also, the quantities $n_{i,j}$ and n_i have to be determined from the measured hourly wind direction values.

$$\overline{c}_i = \sum_j n_{i,j} \overline{c}_{i,j} \Big/ n_i \qquad (1)$$

From a time series of measured hourly pollutant concentrations, one can derive a time series of 24 h averaged concentrations $c_j = \sum_i n_{i,j} \overline{c}_{i,j} \Big/ \sum_i n_{i,j}$. For pollutant roses computed from such a time series of computed c_j, equation (1) provides some kind of a reference rose.

Pollutant rose by least squares regression (for 24 h pollutant concentrations)

Equation (2) gives the system of linear equations that is to be solved to obtain a pollutant rose from measured 24 h pollutant concentrations c_j. There is one equation per day (index ***j***). The pollutant rose components $\overline{c}_i$ in the right hand side of the equation are the variables the system is to be solved for.

$$c_j \sum_i n_{i,j} = \sum_i n_{i,j} \overline{c}_i \qquad (2)$$

For practical work, constraints must be added to this system in order to limit the occurrence of large negative values in the solution (*Cosemans, G. and J. Kretzschmar,* 2002). For the roses in Figure 1, the constraints are: $0 = k\,\beta \overline{c}_i$ with β = 15, an initial weight of the added constraints, and k =2, because, for this data set, 2 was the smallest multiple of β for which the system yields a solution without negative values. A smoothing operator (x(i-1)+6x(i)+x(i+1))/8 has been applied to remove most of the serrated appearance of the regression rose. The pollutant roses in Figure 1 show that most of the information in the ‘reference rose’, based on hourly pollutant concentrations, is extracted from the 24 h pollutant concentrations by least squares regression.

False peaks due to non representative wind direction data
Experience has learned that pollutant roses obtained by least squares regression do sometimes show ‘false’ peaks. A ‘false’ peak is a peak that points to a wind sector where, in reality, are no pollutant sources. A ‘false’ peak can happen if the meteorological data used is not perfectly representative for the wind flow at the air quality monitoring site and its nearby pollutant sources. Figure 2 shows such a pollutant rose with a ‘false’ peak to the south-east. On some days, at the meteorological site used, local wind is south-east, while the wind flow over the broader region is south. This phenomenon has no impact on the reference rose, but it produces extra peaks in the least squares regression rose.

When only a time series of daily pollutant concentrations is available, one is not able to construct a reference rose and, consequently, one has no way to know whether a peak in the regression pollutant rose does point to a source region or whether it is only a regression artefact.

We will present an other method to derive pollutant roses from 24 h pollutant concentrations and investigate whether this method also produces ‘false’ peaks if an not-entirely-representative meteorological data set is used.

Pollutant rose by weighted sums (for 24 h pollutant concentrations)
Consider the ratio r_i of the sum of weighted 24 h pollutant concentrations $w_{i,j}c_j$ versus the sum of the weights $w_{i,j}$ (equation (3)).

$$r_i = \sum_j w_{i,j}c_j \Big/ \sum_j w_{i,j} \qquad (3)$$

The weighting factor $w_{i,j}$ could be chosen to be proportional to $n_{i,j}$, the number of hours that the wind came from wind direction sector ***i*** during day ***j***. If the ratio r_i turns out to be high for those wind direction sectors ***i*** where sources of the pollutant are located, and r_i is low for other sectors, then r_i satisfies the requirements of a pollutant rose. A simple weighting scheme is to set $w_{i,j}$ equal to some integer m power of $n_{i,j}$:

$$w_{i,j} = (n_{i,j})^m, m > 0 \qquad (4)$$

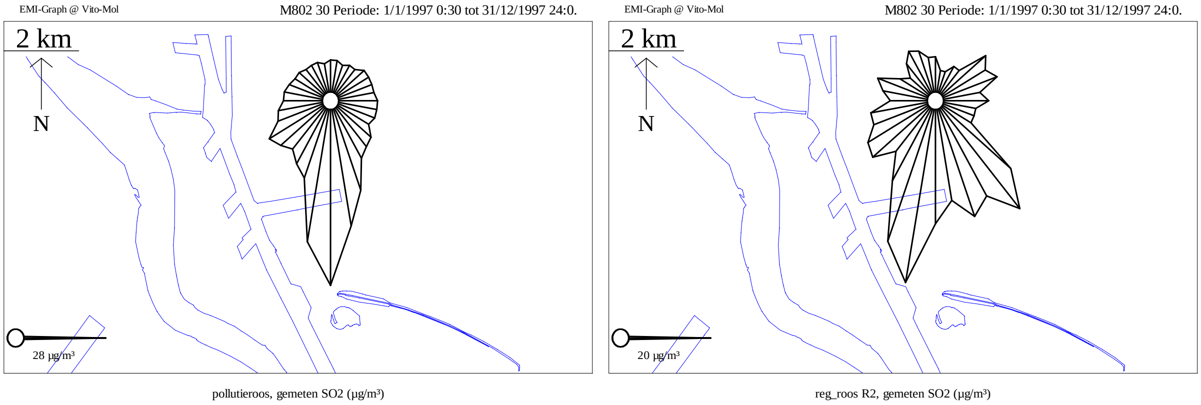

Figure 2. Pollution rose for R831, using wind data that is sometimes not representative
Left: 'reference rose', calculated from half hourly SO_2-concentrations.
Right: calculated by regression from 24 h SO_2-concentrations with 'false' peaks.

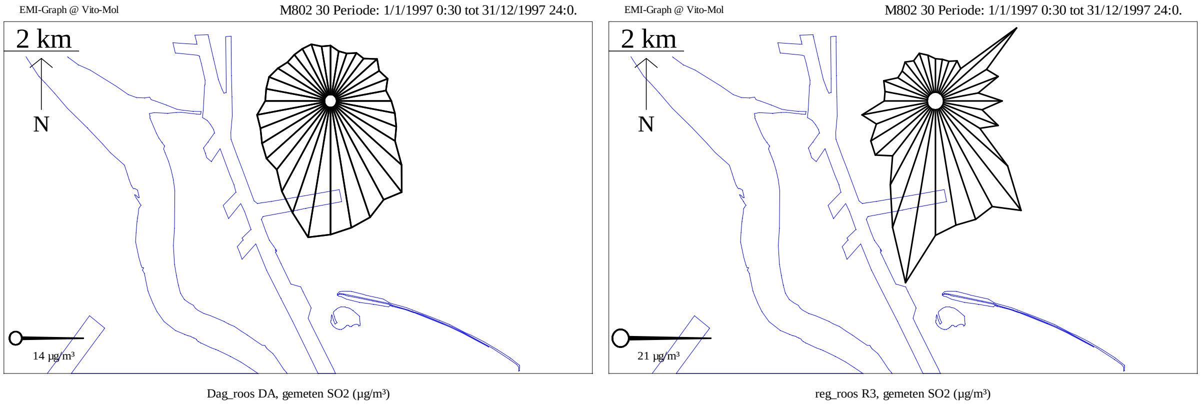

Figure 3. Pollutant roses for R831, using wind data that is sometimes not representative.
Left: calculated from24 h SO_2-concentrations by weighted sums, $m = 1$.
Right: calculated from 24 h SO_2-concentrations by weighted sums, $m = 4$.

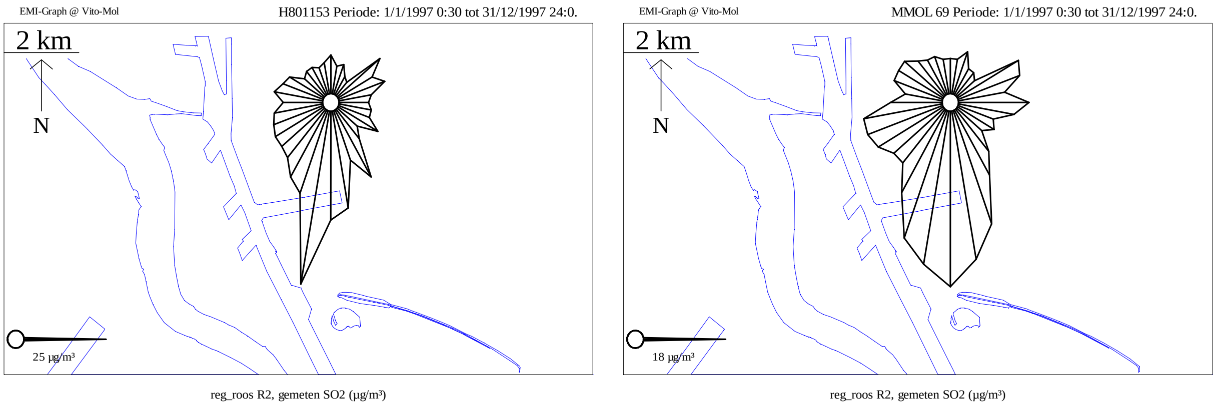

Figure 4. Pollutant roses for R831, calculated from 24 h SO_2-concentrations by least squares regression.
Left: meteorological data from Antwerp harbour, site H801, at 12 km from R831.
Right: meteorological data from Mol, at 59 km from R831.

For m = 1, the ratio's r_i form a quite blunt, nearly oval pollutant rose (Figure 3, Left); for m = 4, the ratio's r_i form a peaked pollutant rose that is very similar tot the pollutant rose obtained by least squares regression, including the false peak towards the south-east.

Pollutant roses with meteorological data from a more distant place

The quality of a pollutant rose, derived from hourly pollutant concentrations and wind direction data, is based on the fact that hours with an high value of the pollutant concentration coincide exactly with the hours that the wind blows from the pollutant source region to the air quality monitoring site. For this reason, one prefers to make pollutant roses with wind data from the nearest meteorological site, because this give the highest probability on this synchronism between high concentrations and wind coming from over a pollutant source. The algorithms for pollutant roses from 24 h pollutant concentrations are based on the frequencies $n_{i,j}$ that the wind came from sector ***i*** on day ***j***. So meteorological data from a more remote place, where the wind is still governed by the same systems of high and low pressure (and land-sea breeze and so on), might be better than the data from the nearest meteorological site if the wind there is some times distorted by local phenomena that are not representative for the air quality monitoring site.

Figure 4 shows pollutant roses for 24 h pollutant data at R831 made with two other meteorological data sets. The rose at the figure on the left is made with the wind directions measured in the Antwerp harbour at 11 km from the SO_2 monitoring site, the rose on the figure at the right is made wind measured at Mol at 60 km east of the SO_2 monitoring site. The largest peak is within 10° (one wind direction sector) accurate. It is clear that the minor peaks in such roses should be considered as 'noise', and not as indicators towards important sources.

CONCLUSIONS

Pollutant roses from 24 h pollutant concentrations can be made. These roses often give valuable information on the location of the most important pollutant sources that affect the monitoring site. Occasionally, such a pollutant rose can contain an important peak that points to a region without pollutant sources. This can be due to the fact that the wind direction data used is not perfectly representative for the monitoring site. One can verify whether this is the case by using meteorological data from a more distant site. Finally, one should not give attention to the smaller peaks in such pollutant roses.

REFERENCES

Cosemans, G. and J. Kretzschmar, 2002: Pollution roses for 24h averaged pollutant concentrations by regression, in: Eighth International. Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Sofia, Bulgaria, 14-17 Oct. 2002, E. Batchvarova and D. Syrakov (Eds.), ISBN 954-9526-12-7, pp. 414 - 418. 7.10