Atmospheric Environment 42 (2008) 6982–6991

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Pollutant roses for daily averaged ambient air pollutant concentrations

Guido Cosemans*, Jan Kretzschmar, Clemens Mensink

Flemish Institute for Technological Research (VITO NV), Boeretang 200, BE-2400 Mol, Belgium

ARTICLE INFO

Article history:
Received 6 June 2007
Received in revised form 31 March 2008
Accepted 7 April 2008

Keywords:
Pollutant roses
Averaging time
Outliers
Weighted averages
Least squares regression
Lognormal distribution

ABSTRACT

Pollutant roses are indispensable tools to identify unknown (fugitive) sources of heavy metals at industrial sites whose current impact exceeds the target values imposed for the year 2012 by the European Air Quality Daughter Directive 2004/207/EC. As most of the measured concentrations of heavy metals in ambient air are daily averaged values, a method to obtain high quality pollutant roses from such data is of practical interest for cost-effective air quality management.

A computational scheme is presented to obtain, from daily averaged concentrations, 10° angular resolution pollutant roses, called PRP roses, that are in many aspects comparable to pollutant roses made with half-hourly concentrations.

The computational scheme is a ridge regression, based on three building blocks:

1. ordinary least squares regression;
2. outlier handling by weighting based on expected values of the higher percentiles in a lognormal distribution;
3. weighted averages whereby observed values, raised to a power m, and daily wind rose frequencies are used as weights. Distance measures are used to find the optimal value for m.

The performance of the computational scheme is illustrated by comparing the pollutant roses, constructed with measured half-hourly SO_2 data for 10 monitoring sites in the Antwerp harbour, with the PRP roses made with the corresponding daily averaged SO_2 concentrations. A miniature dataset, made up of 7 daily concentrations and of half-hourly wind directions assigned to 4 wind sectors, is used to illustrate the formulas and their results.

1. Introduction

1.1. Subject

Pollutant roses are indispensable tools to identify unknown (fugitive) sources of heavy metals at industrial sites whose current impact exceeds the target values imposed for the year 2012 by the European Air Quality Daughter Directive 2004/207/EC. As most of the measured concentrations of heavy metals in ambient air are *daily* averages, a method to obtain high quality pollutant roses from such data is of great practical interest for cost-effective air quality management.

Given, for some ambient air quality monitoring site, a time series of k simultaneously measured *half-hourly* pollutant concentrations c_k and wind directions dd_k over a period of one year, the pollutant rose $P=(p_i, i=1\ldots N_{dir})$ is obtained by dividing the wind direction into N_{dir} sectors – for instance 36 sectors of 10° each – and by taking p_i equal to the arithmetic average of the concentrations c_k during the half hours that the wind came from wind sector i (Gräfe, 1972; Henry et al., 2002).

* Corresponding author. Tel.: +32 14 33 59 61; fax: +32 14 32 11 85.
E-mail address: guido.cosemans@vito.be (G. Cosemans).

1352-2310/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2008.04.025

Pollutant roses, when drawn as polar diagrams (see Fig. 1), show peaks towards the source area(s) with a high impact on the concentrations measured at the monitoring site.

It is, however, not obvious how to make high angular resolution pollutant roses from *daily* averaged concentrations; consequently, an investigation of techniques to do so is justified.

1.2. Organization of the paper

Section 2 describes the SO_2 dataset of Antwerp, 1977. Pollutant roses made with the half-hourly concentrations measured at ten monitoring sites are a benchmark (or reference roses) to evaluate methods producing pollutant roses from daily averaged concentrations.

Section 3 introduces the symbolic notation to deal with time series and pollutant roses in formal notation.

Formulas and results are illustrated on the miniature dataset introduced in Section 4.

Section 5 discusses current practice: pollutant roses made with a daily average wind direction.

In Sections 6–8, the three building blocks of the computational scheme are presented.

In Section 6, the problem of finding a pollutant rose for daily concentrations is formulated as a system of equations solved with Least squares regression (LSR). The LSR solution is very different from the desired reference roses; we indicate the sources of these differences and explain the rationale behind the LSR results.

Section 7 deals with outlier handling based on samplings from a lognormal population.

Section 8 describes a weighted averaging scheme giving fairly good pollutant roses.

The computational scheme is described in Section 9. This approach gives an improvement of pollutant rose accuracy from Section 8, so that $90 \pm 5\%$ of the variance of the reference roses is explained.

Section 10 discusses some qualitative and quantitative aspects of the roses obtained.

2. Datasets used and reference roses

2.1. Benchmark datasets

To develop the computational scheme, about a hundred datasets of one-year half-hourly pollutant data (SO_2, PM_{10}), measured in the northern part of Belgium, West Europe, during the period 1996–2003, have been investigated. These data were converted into time series of daily concentrations and the pollutant roses obtained with the computational scheme were compared with the pollutant roses based on the original half-hourly data.

Fig. 1 shows such half-hourly based pollutant roses, year 1997, for eight SO_2 monitoring sites located in the Antwerp industrial harbour. This industrial site is one of the largest clusters of petrochemical industry in the world. In 1997, about 93 000 tons of SO_2 were emitted from more than 100 industrial stacks (Anoniem, 2007).

The peaks of the roses clearly indicate the sectors where SO_2 sources with a high impact on measured SO_2 levels are located. These roses have one or several peaks in both the prevailing and in the non-prevailing wind directions, hence they provide a complex real-world benchmark for methods that produce pollutant roses from daily averaged concentrations.

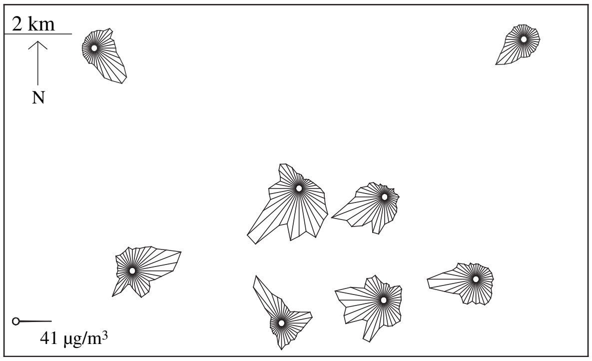

Fig. 1. Reference pollutant roses.

2.2. Requirements for a good pollutant rose

A ‘good’ pollutant rose, derived from daily concentrations, gives the same visual information as does a half-hourly pollutant rose. Therefore, it must satisfy at least the following criteria:

- the daily rose must have peaks in the same direction(s) as the half-hourly data based rose;
- the peak values in both roses must be comparable in magnitude;
- the daily rose may not have peaks absent in the half-hourly rose.

3. Formal notation

The computational scheme starts with a time series of N_{DAYS} daily pollutant concentrations C_j, $j = 1\ldots N_{DAYS}$, and for the same period, simultaneously measured half-hourly wind direction data. The wind direction is divided into 36 sectors of 10° each.

The half-hourly wind direction data for day j are converted into a daily wind rose $\{a_{ij}, i = 1\ldots36\}$ where a_{ij} is the number of half hours that the wind came from wind direction i on day j. The set of observations, that is, the daily concentration C_j and the corresponding daily wind rose $\{a_{ij}, i = 1\ldots36\}$ for day j, is represented by

$$(C_j, a_{1j}, \ldots, a_{36j}). \quad (1)$$

4. The miniature dataset

4.1. A random sample

Table 1 defines a miniature dataset with 4 wind sectors, 90° each, and with 7 daily averaged concentration values C_j.

The dataset can be regarded as a random sample of 7 days drawn from a one-year dataset. Reflected in the dataset are: the presence of two prevailing wind directions