A statistical index for the evaluation of typical period trends

Tiziano Tirabassi

Institute ISAC of CNR
e-mail: t.tirabassi@isac.cnr.it

Abstract

The work presents a daily trend of pollutant concentrations, referred to as the "representative day". The methodology can be considered as an objective method for identifying critical and anomalous situations.

Resumo

O trabalho apresenta uma tendência diária das concentrações de poluentes, conhecido como o "dia representativo". A metodologia pode ser considerada como um método objectivo para a identificação de situações críticas e anômalas.

Introduction

Currently, hourly concentration means are the most widespread method employed and represents the maximum desegregation with which pollution data are generally collected. It is often adopted in applications and allows the evaluation of average concentrations over long periods, such as a day, a season or a year. In addition hourly concentrations are used to define specific "typical" periods that may be of particular interest in the study of pollutant diffusion (for example, a typical working day, a typical holiday, a typical seasonal day etc.) (Tirabassi and Nassetti, 1999; Tirabassi et al., 2005)).

The work presents a daily trend of pollutant concentrations, referred to as the "representative day", i.e. the day for which the overall sum of the square differences between its concentration, averaged within each hour, and the concentrations for all other days at the same hour, is a minimum.

The representative day

What we call a “representative day” is a 24 hour data set, actually recorded at a field station, which is characterised by the least differences with respect to all the 24 hour measurements of that station’s temporal series: that is to say, the daily series whose sum of the squared differences over 24 hours turns out to be the smallest compared to all the other days of the period under consideration.

In mathematical notation:

$$A_{ij} = \sum_{k=1}^{24} (c_{ki} - c_{kj})^2 \quad i,j = 1,2...N \qquad (1)$$

where N is the number of days for which the representative day is calculated and c_{ki} is the pollutant concentration of the i-th day at the k-th hour.

We adopt A_i to indicate the sum of all the squared residuals of the i-th line (or column, A_{ij} being a symmetrical matrix with all zeros in the diagonal):

$$A_i = \left(\sum_{j=1}^{N} A_{ij} \right) \qquad (2)$$

The representative day (RD) is the one with the lowest sum, i.e. the i-th day where A_i is the smallest of the quantities obtained:

$$\min(A_i) \Rightarrow \text{RD} \qquad (3)$$

The approach also allows the identification of the “least representative day” (LRD), that is the daily series that maximises the mean sum of squared residuals:

$$\max(A_i) \Rightarrow \text{LRD} \qquad (4)$$

The least representative day identifies an anomalous situation of pollutant dispersion, often (although not always) characterised by maximum ground concentrations.

The purpose of such typifying is that of outlining characteristic scenarios for a given period under investigation and then mathematical models make it possible to attempt simulations of a typical period trend, without the need to simulate all the days of the time interval covered by the typical period.

To compare the degree of representativety of the most or least representative days with that obtained for other time periods and/or in other measurement stations (also in complete different areas), a normalisation is required in order to make the day independent of the length of the measurement series, sampling period and characteristics of the area under study.

The "representativety" of a representative day can be quantified by introducing the following adimensional index:

$$DI = \sqrt{\sum_{i=1}^{N}\sum_{k=1}^{24}\left(\Gamma_k - c_{ik}\right)^2} \bigg/ \sqrt{\sum_{i=1}^{N}\sum_{k=1}^{24}\left(\overline{c_k} - c_{ik}\right)^2} \qquad (5)$$

where $\overline{c_k}$ is the mean hourly concentration at the k-th hour calculated over the period under consideration and Γ_k is the mean hourly concentration of the representative day at the k-th hour.

DI is an adimensional quantity (DI ≥ 1), which is closer to unit the more the day RD is representative of the period in consideration.

The least representative day can also be normalised in the same way: one simply substitutes in equation (5) the hourly concentration of RD (Γ_k) with the least representative ones. In this case the value of DI will be greater than one 1, providing an indication of the low degree of representativety of the day obtained; the more DI is greater than 1, the more the least representative day is "anomalous", compared to the trend of RD.

The normalisation procedure described above presents two important features: i) independence from the size of the measured concentrations (if the concentrations are multiplied by a factor, the results of the normalisation do not change); ii) independence from the number of days (N) included in the time period considered (the quantities of the numerator and denominator of equation (5) are calculated over all the N days of the period under consideration).

Conclusions

The special advantage of the representative day method derives from the fact that, being an actual day; it allows the identification of the date on which the representative trend occurred and, thus, a knowledge of the meteorological and emission parameters that characterised it. It can therefore be more easily and less expensive simulated with diffusion models.

The same approach also allows the identification of the least representative day, that is, the day on which an anomalous, nearly always critical situation occurred, compared to the average trend recorded for that period. So the methodology presented can be considered as an objective method for identifying critical and anomalous situations.

References

Tirabassi T. and Nassetti S.. The representative day. Atmos. Environ., 33, pp. 2427-2434, 1999.

Tirabassi T., Cocci G. R. and Passerini G. "The Representative Day Technique in the Study of Photochemical Smog Pollution in Falconara Industrial Area" Water, Air, and Soil Pollution, 165, pp. 65-81, 2005.