Journal of Mathematical Chemistry 8(1991)283–289

ANALYSIS OF OZONE MONITORING DATA

Jelena JEFTIĆ and Tomislav CVITAŠ
Ruder Bošković Institute, P.O. Box 1016, 41001 Zagreb, Yugoslavia

Abstract

The concentration data obtained by monitoring pollutants do not follow a symmetric or even regular, for example log-normal, distribution so that statistical methods which are insensitive to the type of distribution are required. Such a robust statistical method has been applied for the analysis of data obtained by monitoring ozone in different environments. Since ozone is formed at ground level through photochemical reactions, its volume fractions follow a typical diurnal behaviour with a maximum at about noon and a minimum during the night. A convenient way of representing ozone monitoring data is described. The distribution is represented by percentiles, and for each hour of a day averaged over the whole period. Five most characteric percentiles are represented by so-called "boxes and whiskers".

1. Introduction

Ozone has recently reached the attention of media and laymen worldwide because it is the subject of two concerns. One is the problem of its depletion in the stratosphere where it has a protective role, preventing harmful UV radiation reaching the Earth's ground [1–4]. The other is its rising concentration in the lower troposphere, which is adversely affecting many plants, forests, animals, humans and various materials [5,6]. Tropospheric ozone is therefore monitored at many stations and data are collected by individual scientists, sometimes by certain nationwide services and recently also on an international level. An example is the EUREKA environmental project EUROTRAC (**Euro**pean experiment on **tra**nsport and transformation of environmentally relevant trace **c**onstituents in the troposphere over Europe) and its subproject TOR (**t**ropospheric **o**zone **r**esearch).

Ozone is a natural constituent of the troposhpere as a result of stratospheric intrusions. Its volume fractions are around 30 ppb (30×10^{-9}). It is, however, also formed as a result of photochemical and chemical reactions of primary pollutants emitted into the atmosphere, primarily in densely populated areas.

The first step in ozone formation is the photolysis of nitrogen dioxide, which is formed by oxidation of nitrogen oxide in the atmosphere. Nitrogen oxide is a primary pollutant in car exhaust and from other combustion processes.

$$NO_2 \xrightarrow{h\nu} NO + O. \qquad (1)$$

The rate coefficient k_1 of this reaction is a function of solar irradiance (the radiation power received at the Earth's surface). The subsequent reaction of atomic and molecular oxygen yields ozone:

$$O + O_2 + M \rightarrow O_3 + M, \tag{2}$$

where M is a collision partner taking up surplus energy. Ozone reacts with nitrogen oxide, another primary pollutant from combustion processes:

$$O_3 + NO \rightarrow O_2 + NO_2. \tag{3}$$

Thus, the steady-state concentration of ozone is determined by the concentration ratio of the two nitrogen oxides

$$[O_3] = \frac{k_1}{k_3} \frac{[NO_2]}{[NO]} . \tag{4}$$

This ratio, however, and consequently the ozone concentration, is highly sensitive to the presence of other pollutants such as hydrocarbons in the ambient air. Hydrocarbons are attacked by OH radicals from the photolysis of ozone in the presence of water vapour and react with oxygen to form peroxy radicals, which on the other hand oxidize NO into NO_2, thus strongly affecting the steady-state concentration of ozone, eq. (4). The concentration of ozone in a polluted area is therefore a function of the amounts of hydrocarbons and nitrogen oxides emitted as primary pollutants and of solar irradiance.

When its volume fraction rises above 80 ppb (as hourly average), first effects on humans can be noticed and values of 100 ppb and greater are typical of photochemical smog.

Since ozone is a secondary pollutant, it often reaches highest concentration at a considerable distance away from primary pollution sources. It is also fairly stable in relatively clean atmospheres and can therefore be transported over long distances of several hundred kilometers, causing harmful effects to vegetation far away from the source of pollution [7]. The study of its concentration in different areas and the relation to meteorological conditions is therefore of considerable importance. Some countries (e.g. The Netherlands, Germany, USA, . . .) have extensive monitoring networks, while other monitor air pollution at only very few sites (e.g. Italy, France, . . .). In quite a number of countries, ozone is not regularly monitored at all. This applies also to Yugoslavia and the whole East Mediterranean region.

The few measurements that have been carried out in this area relied on a small number of instruments, so the data are far from being sufficient to understand what is happening.

It is our aim here to describe a way of analyzing hourly average ozone monitoring data in order to obtain a deeper insight into the phenomena characterizing the status of air pollution at a particular site.

2. Results and discussion

Since ozone volume fractions in a polluted area show a behaviour which more or less parallels the solar irradiance, it is much more meaningful to determine the typical diurnal variation of the volume fraction than just to report an average volume fraction.

The typical behaviour can best be estimated by taking the average value for individual hours of the day. If the data are assumed to follow a normal distribution, bars can be used to indicate the standard deviations. Narrow ranges of values then indicate a fairly stable meteorological situation during the analyzed period of time [8]. However, the assumption of a normal Gaussian distribution is usually not justified for data obtained by monitoring a pollutant. In general, measured environmental data in long time series are believed to represent samples from a log–normal probability distribution [9]. However, a log–normal analysis is not useful for the analysis of surface ozone data because of the asymmetric distribution of the diurnal ozone concentrations due to chemistry and meteorology [8]. The distribution is usually asymmetric, and for ozone it varies considerably in shape from site to site, and even at a particlar site it varies for different periods of time. Thus, a better picture of the typical diurnal variation can be obtained by plotting selected quantities of the distribution [10] rather than averages with misleading standard deviation bars.

We have found that a good picture can be obtained by indicating and plotting the quantiles of order 0.10, 0.25, 0.50, 0.75 and 0.90, as well as the minimum and maximum values. This can be done by the so-called "box and whiskers" representation. A quantiles of order *p* is the value (say ozone volume fraction) below which there is the fraction *p* of all the data. Thus, the quantile of order 0.5 (the 50th percentile) is the median. In fig. 1, a representation by percentiles is compared with the representation by the average (μ) with standard deviations of (σ). For a normal

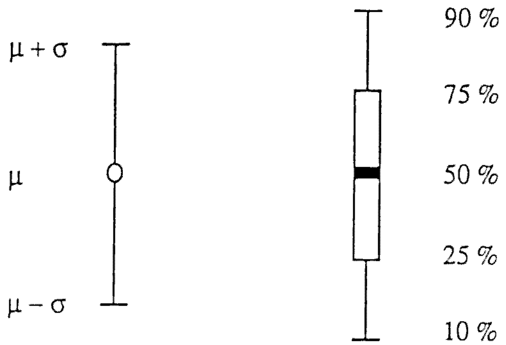

Fig. 1. Representation of a normal distribution by the average and standard deviation bars (on the left-hand side) and by the "box and whiskers" (on the right-hand side).

distribution, the two representations are of course perfectly equivalent. However, for asymmetric distributions the normal distribution law is not applicable and the adequate "box and whiskers" representation of percentiles gives a better picture.

An example for real data is shown in fig. 2. The figure shows the diurnal variation of ozone volume fractions as determined by monitoring ozone in the suburb of Šibenik during the summer of 1986. The total number of hours is 744.

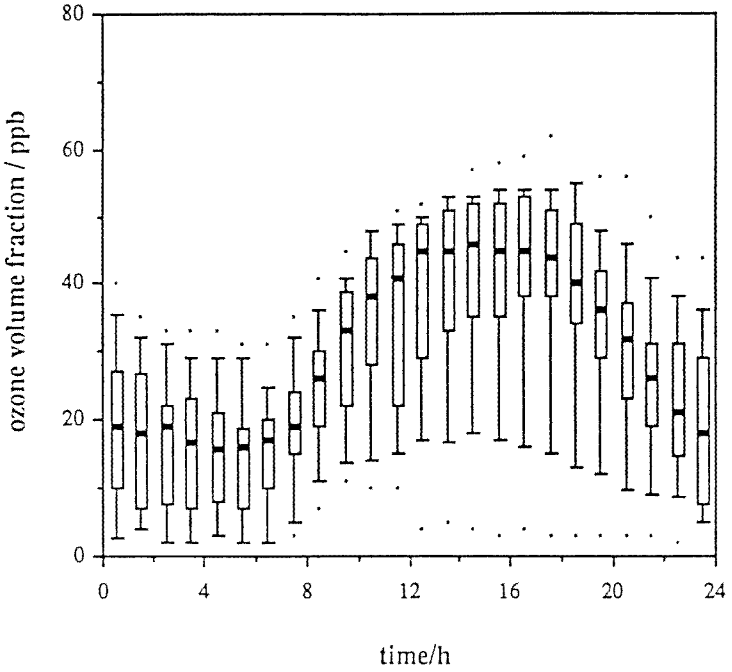

Fig. 2. Diurnal variation of ozone volume fraction in Šibenik, July 1986. The "box and whiskers" are explained in fig. 1 and the dots indicate the minimum and maximum values.

The typical behaviour at this site can be seen from the black rectangles in fig. 2, representing the median. The boxes and whiskers, on the other hand, indicate that the distribution of values is far from being symmetrical. Deviations below the expected value given by the median are sometimes very large. This can be understood when considering the location of the monitoring site. It was located approximately 1 km to the south-west of the city centre of Šibenik (a Mediterranean town with 80,000 inhabitants), across the bay. There was also a factory producing iron alloys located about 700 m to the north of the monitoring site. When the wind blew the factory emissions to the monitoring site, all the ozone was destroyed and therefore very low values were recorded.

Another example showing opposite deviations is shown in fig. 3, which is based on the hourly average ozone volume fractions measured in Zagreb at the

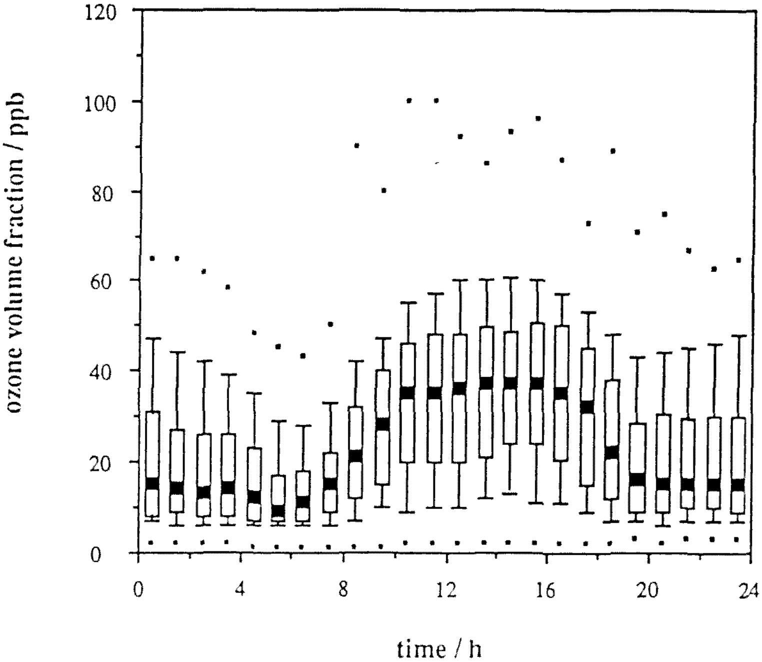

Fig. 3. Diurnal variation of ozone volume fraction in Zagreb, summer 1989. The "box and whiskers" are explained in fig. 1 and the dots indicate the minimum and maximum values.

Ruder Bošković Institute during the summer of 1989 (a total of 4392 hours). One can again immediately observe the typical diurnal behaviour for a city but, in addition, there are also rather rare occasions when the ozone levels exceed the critical values.

The computation of quantiles is also very useful when computing monitoring data obtained at different sites [11]. An illustrative example is, for instance, the comparison of values obtained in a polluted city with those obtained in a cleaner environment. A polluted environment is characterized by ozone volume fraction peaks around noon, which frequently exceed the air quality standards, and also by very low volume fractions at night and especially at highest nitrogen oxide emission rates during morning and evening rush hours. In a cleaner environment, the average volume fraction may be equal or even higher, yet there will be no extreme peak values, unless caused by long-range transport and, even more characteristic, there will be no extremely low values.

In fig. 4, the quantiles of the ozone volume fractions measured during the summer of 1989 in the city of Zagreb and at the 1000 m high and 4100 m distant site on the mountain of Medvednica are compared. It can be seen that the frequency of low volume fractions is much greater in the city, about 50% being below 20 ppb, while only 20% of the values are below 20 ppb at the elevated site. The similarity

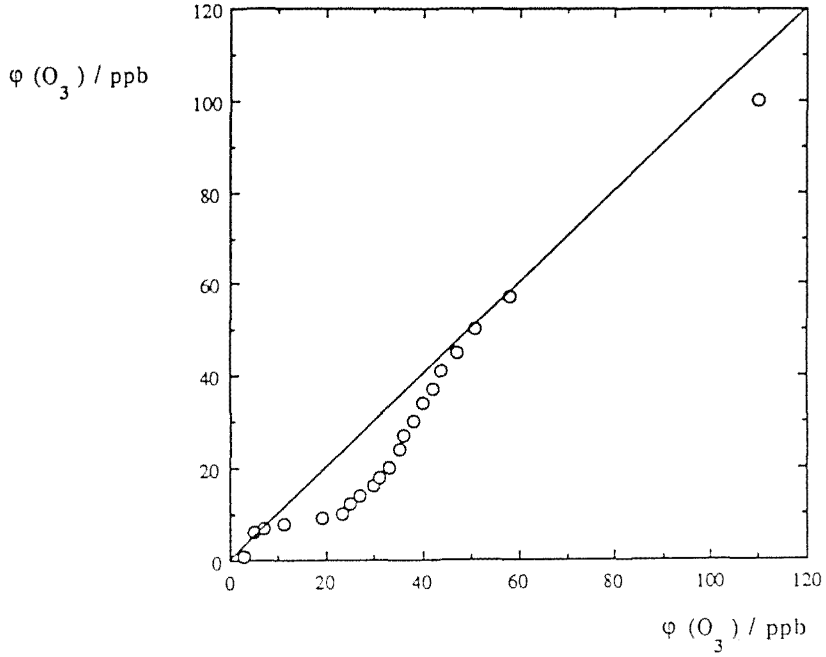

Fig. 4. A quantile–quantile plot comparing [9] the distributions of hourly average ozone volume fractions in Zagreb (site "R. Bošković" Institute, 180 m a.s.l.) and on the Medvednica mountain ridge (site Puntijarka, 1000 m a.s.l.). The dots represent every fifth percentile.

in peak values indicates that there must be transport effects at the elevated site which cause similar results as the photochemistry in the more polluted city. A more drastic example for two sites in Athens, a city with frequent photosmog episodes, has been given by Cvitaš et al. [11].

3. Conclusion

Examples were given to illustrate the advantages of the analysis of quantiles characterizing the distribution of ozone volume fraction values and the graphical representation by the so-called "box and whiskers" as compared to the more usual average values plot. An analysis of the deviations from the typical behaviour at a particular site yields more detailed information about the causes of the deviations. Details of such an analysis will be given elsewhere.

References

[1] J.C. Farman, B.G. Gardiner and J.D. Shanklin, Nature 315(1985)207.
[2] R.S. Stolarski and M.R. Schoeberl, Geophys. Res. Lett. 13(1986)1210.

[3] M.R. Schoeberl and A.J. Krueger, Geophys. Res. Lett. 13(1986)1217.

[4] R.S. Stolarski, Sci. Amer. 258(1988)20.

[5] B.J. Finlayson-Pitts and J.N. Pitts, Jr., *Atmosphere Chemistry: Fundamentals and Experimental Techniques* (Wiley, New York, 1986).

[6] H. Güsten, Formation, transport and control of photochemical smog, in: *The Handbook of Environmental Chemistry*, vol. 4/part A, ed. O. Hutzinger (Springer, Berlin, 1986), p. 53.

[7] Ø. Hov, E. Hesstvedt and I.S.A. Isaksen, Nature 273(1978)341.

[8] Z. Božičević, V. Butković, T. Cvitaš and L. Klasinc, Kem. Ind. 27(1978)177.

[9] R.O. Gilbert, *Statistical Methods for Environmental Pollution Monitoring* (Van Nostrand Reinhold, New York, 1987).

[10] M.B. Wilk and R. Gnanadeskian, Biometrika 55(1968)1.

[11] T. Cvitaš, H. Güsten, G. Heinrich, L. Klasinc, D.P. Lalas and M. Petrakis, Staub-Reinhalt. Luft 45(1985)297.