IL NUOVO CIMENTO VOL. 21 C, N. 1 Gennaio-Febbraio 1998

Probability density functions of photochemicals over a coastal area of Northern Italy

T. GEORGIADIS (1), F. FORTEZZA (2), L. ALBERTI (2), V. STROCCHI (2)
A. MARANI (3) and G. DAL BO' (3)

(1) *CNR-FISBAT - Via Gobetti 101, I-40129 Bologna, Italy*
(2) *ARPA, Sez. Ravenna - Via Alberoni 17, I-48100 Ravenna, Italy*
(3) *Dipartimento di Scienze Ambientali, Università di Venezia*
D.D. 2137, I-30123 Venezia, Italy

(ricevuto il 13 Marzo 1997; approvato il 28 Novembre 1997)

Summary. — The present paper surveys the findings of experimental studies and analyses of statistical probability density functions (PDFs) applied to air pollutant concentrations to provide an interpretation of the ground-level distributions of photochemical oxidants in the coastal area of Ravenna (Italy). The atmospheric-pollution data set was collected from the local environmental monitoring network for the period 1978-1989. Results suggest that the statistical distribution of surface ozone, once normalised over the solar radiation PDF for the whole measurement period, follows a log-normal law as found for other pollutants. Although the Weibull distribution also offers a good fit of the experimental data, the area's meteorological features seem to favour the former distribution once the statistical index estimates have been analysed. Local transport phenomena are discussed to explain the data tail trends.

PACS 92.60.Sz – Air quality and air pollution.

1. – Introduction

Several studies have been conducted over the last few years in coastal areas to determine the evolution and distribution of pollutants and, prevalently, photochemical oxidants [1-5]. The interest in monitoring these areas is justified by the fact that coastal regions, in particular in the Mediterranean Basin, are highly industrialised and densely populated.

The provincial district of Ravenna, situated in Northern-Italy's Po Valley, is bordered on the East by the upper Adriatic Sea and features an extensive seaboard industrial belt with petrochemical and power plants that produce large amounts of pollutant emissions. The morning land breeze drives the primary pollutants offshore and after midday the sea-breeze transports the oxidants produced by photoreactions inshore.

In order to control and to improve the air quality of Ravenna, the local government installed an environmental monitoring network in the area in the 1970s. Meteorological parameters and concentrations of air pollutants such as ozone, nitrogen oxides, sulphur dioxide and hydrocarbons are continuously monitored by ten permanent stations. This network has made it possible to characterise the air quality features of the area, along with the determination of physico-chemical patterns related to emission, transformation and transport of oxidant precursors and photochemical-reaction products [6-9]. The particular role of the local circulation (land-sea breeze system) in determining conditions for high-efficiency conversion of primary pollutants in oxidants during the summer periods has been highlighted [10, 11].

Despite the great amount of data collected by monitoring the time-space evolution of primary and secondary pollutants at these coastal sites, which provide a reliable framework when applying dynamic or photochemical models to describe the environmental impact of such compounds, uncertainties still remain as to the determination of the statistical distribution of surface pollutant concentrations. Furthermore, the determination of probability density functions (PDFs) can lead to the expected number of days in which the observed pollutant concentrations are above the national environmental air quality standards. The data collected from these measurements constitute in any case an information set-up that is difficult to handle because of the sheer amount of data involved.

These data sets require automatic procedures geared to the main issues involved: i) the quality control of data [12] and ii) the interpretation of the data. The first step is to determine the probability distribution of the given values. This is not easy because along with the determination of the distribution parameters, it is necessary to define the mathematical structure. The literature is rich in proposals mainly based on empirical formulations derived without taking into account the processes by which the distribution is generated, although there are exceptions that do consider the influences of the physical processes involved [13-15].

The aim of the present study is to test some PDFs in determining the best representation of photochemical ozone concentration distribution and to try an interpretation of the results by considering the well-known time-space evolution of the oxidants in the provincial district of Ravenna. The operational procedure followed here, due to Holland and Fitz-Simons [16], produces some uncertainties in the results [17].

2. – PDFs **and statistics of goodness of fit**

The application of PDFs to air quality data can take into account some features of the local environment. Air pollutant concentrations are often auto-correlated and exhibit fluctuations due to changes in weather patterns.

It is often considered that the statistical distribution of a pollutant follows a log-normal law when a long period of observations is taken into account [18]. Yet this distribution neither fits the tail concentration values correctly [19], nor interprets correctly the skewness and kurtosis of the experimental distributions [20].

Of the other PDFs found useful in describing pollutant concentrations, the gamma and the Weibull distribution are the most widely employed ones in air quality modelling. Once the experimental data are fitted to the chosen distributions, statistical indices can be used as descriptors of the goodness of fits.

Our statistical analysis of air quality data takes into account many factors related to the local concentration values as well as weather and circulation patterns like transport phenomena. The persistence of high values overnight due to an oxidant recirculation driven by the sea breeze with an average wind speed of about 3-4 m/s during the spring-summer periods, tends to increase the statistical appearance of extreme values.

Although photochemical processes are highly non-linear and the ozone evolution is characterised by non-stationarity [21], it is possible to apply quasi–steady-state (QSS) photochemical model for measurements performed in the vicinity of the emissions (*i.e.* few kilometres) [22, 23]. Near isolated sources, some estimations showed [24] that non-homogeneity in the diffusion processes can cause deviations from QSS. In the case of measurements performed within an area where continuous-emission points are widespread, conventional statistics applied for the calculation of probability density functions can appropriately predict average surface concentrations albeit it is not possible to give a statistical description of the instantaneous patterns because of the smoothing produced by averaging scalars.

A computer software product, MAXFIT, was used to provide conventional estimations of PDFs and statistical indices. The MAXFIT estimate and the related algorithms are fully treated by Holland and Fitz-Simons [16].

3. – Results and discussion

Three of the six PDFs supplied by MAXFIT were used to fit O_3 half-hourly and daily concentration averages, NO_x daily averages (μg/m^3) and solar radiation daily intensity (W/m^2). The log-normal, gamma and Weibull distribution results were related to the following statistical indices:

- absolute deviations (AD), which is related to the existing deviation between theoretical and observed frequencies;
- weighted absolute deviations (WAD), the same as AD but weighted over the theoretical probability of data in the considered interval;
- Kolmogorov-Smirnov (KS), which is the maximum distance between the empirical distribution of the function and the observed cumulative distribution frequencies;
- Cramer-von Mises-Smirnov (CVS), corresponding to the sum of the squares differences between observed and theoretical cumulative probabilities weighted for the data theoretical probability;
- likelihood function (SSQ), the same as CVS not weighted for the data theoretical probability.

The outputs of MAXFIT are presented in this paper as PDF tables *vs.* ranking indices. The numbers in the cells indicate the goodness of the fit from the best (1) to the worse (3) case.

The whole data set of O_3 half-hourly concentration averages used for the analysis, covering the period 1978-1989, consists of 86118 validated data. Table I shows the results of MAXFIT's application to these data. The findings show that it is very difficult to assign greater statistical significance to one or another distribution.

The distribution analysis of the O_3 daily average leads to the attribution of the log-normal behaviour to this pattern (all tests indicate the log-normal as the best

TABLE I. – *Results of* PDF *analysis vs. statistical indices* (SIs) *for half-hourly ozone concentration data.*

	Log-normal	Weibull	Gamma
KS	3	1	2
CVS	1	2	3
SSQ	3	1	2
AD	2	3	1
WAD	1	3	2

TABLE II. – *Validity percentages of the couple-coherence test for* SIs.

	AD	WAD	KS	CVS	SSQ
AD	100				
WAD	82.7	100			
KS	83.1	75.3	100		
CVS	93.8	88.9	84.0	100	
SSQ	58.0	51.9	55.6	94.0	100

TABLE III. – *Results of* PDFs *vs.* SIs *for daily ozone concentration averages under conditions of solar-radiation intensity* $\geqslant 24.5\ \mathrm{mW/cm^2}$ *and wind speed* $\leqslant 1.5$ m/s.

	Log-normal	Weibull	Gamma
KS	1	2	3
CVS	1	2	3
SSQ	2	1	3
AD	2	1	3
WAD	1	3	1

distribution). The contrast of this result with the attribution uncertainties in table I leads to the conclusion that the effects of the temporal scales on ozone measurements are marked and the half-hourly and daily values are related to two different processes. The daily values appear to indicate the presence of a dilution mechanism, as that proposed by Ott [15], which would support the diurnal trends based on the analysis of air-parcel trajectories [11].

A more detailed analysis of the test indications is attempted in table II, which reports the couple-coherence percentage to determine the best distribution. The AD and CVS proved to be the statistics that can indicate the best distribution over the considered data set.

An additional analysis is applied by considering only data in which atmospheric patterns favourable to ozone formation were present. Data were chosen for solar-radiation intensity greater than $245\ \mathrm{W/m^2}$ and wind speed less than 1.5 m/s. Table III indicates that for SSQ and AD the Weibull is the best distribution and for CVS, WAD

TABLE IV. – *The same as table* III *but with wind speed* $\leqslant 1$ m/s.

	Log-normal	Weibull	Gamma
KS	2	1	3
CVS	1	3	2
SSQ	2	1	3
AD	2	1	3
WAD	1	3	2

TABLE V. – *Results of* PDFs *vs.* SIs *for the* O_3*/solar-radiation intensity ratio.*

	Log-normal	Weibull	Gamma
KS	1	2	3
CVS	1	2	3
SSQ	1	2	3
AD	1	3	2
WAD	3	2	1

TABLE VI. – *Results of* PDFs *vs.* SIs *for daily nitrogen oxides concentration averages.*

	Log-normal	Weibull	Gamma
KS	2	3	1
CVS	1	3	2
SSQ	1	3	2
AD	2	3	1
WAD	1	3	2

and KS the best is the log-normal one. An overall analysis of the statistics also indicates the Weibull as the better distribution because, when it did not reach the goodness-of-fit value 1, the numerical difference with the value of the best of the test is very small. If we also consider as a classification criterion the accordance between the CVS, AD and the other statistics, the functions representing the best fits of the experimental data are the log-normal and the Weibull.

This latter result tends to confirm the findings of Georgopoulos and Seinfeld [25], in which the Weibull distribution proved to be the best descriptor of the experimental data consisting of daily ozone concentration averages collected in Los Angeles. It is in effect conceivable that the differences found between Ravenna and Los Angeles (the predominance of the log-normal distribution for the Ravenna's data set) can be ascribed to the different meteo-climatic and diffusive conditions of the sites. The Los Angeles area is characterised by strong atmospheric stability for long periods of the year, with high solar radiation and light winds [26-28]. Ravenna, on the other hand, is subjected to more marked periods of unstable weather and higher wind speeds. These differences tend to diminish if we look only at Ravenna data involving lower wind speed, *i.e.* corresponding to anticyclonic conditions, as reported in table IV.

The comparison of the results for the two different wind velocities evinces the influence of this factor on the probability distribution. This sensibility suggests that great care be taken in the analysis of ozone concentration values because, as in this case, data are often utilised without considering the seasonal effects influencing ozone, as, for example, the solar-radiation intensity.

The role played by solar-radiation intensity on daily ozone concentration averages has been investigated. Given that ozone formation in the atmosphere is regulated by solar radiation and its concentration is time-dependent on the variation of radiation, a linear relation between solar-radiation intensity and ozone production has been found [29]. PDF analysis was applied to the values as the ratio of daily ozone concentration averages to daily solar-radiation intensity averages. This analysis was performed to find a distribution linked only to the meteo-diffusive characteristics of the atmosphere.

The results in table V clearly indicate the log-normal distribution as the best descriptor and an increased statistical confidence in the goodness of the fit between

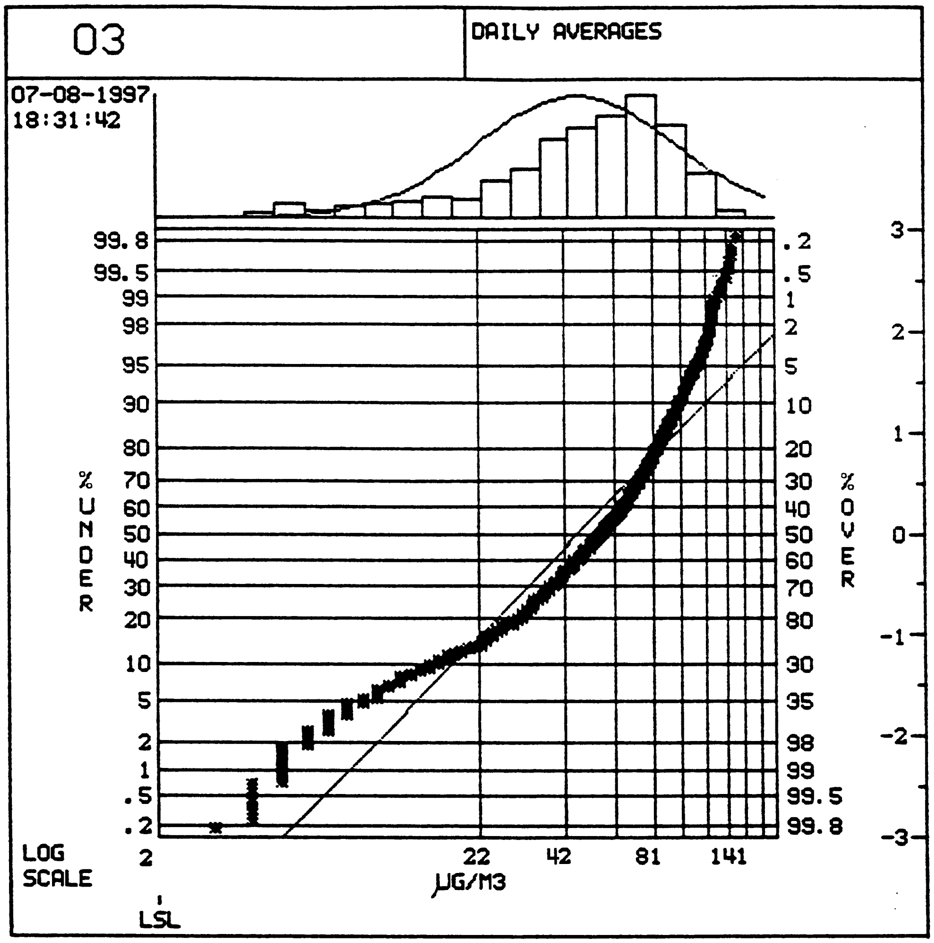

Fig. 1. – Plot of the daily ozone concentration averages on a log-normal scale.

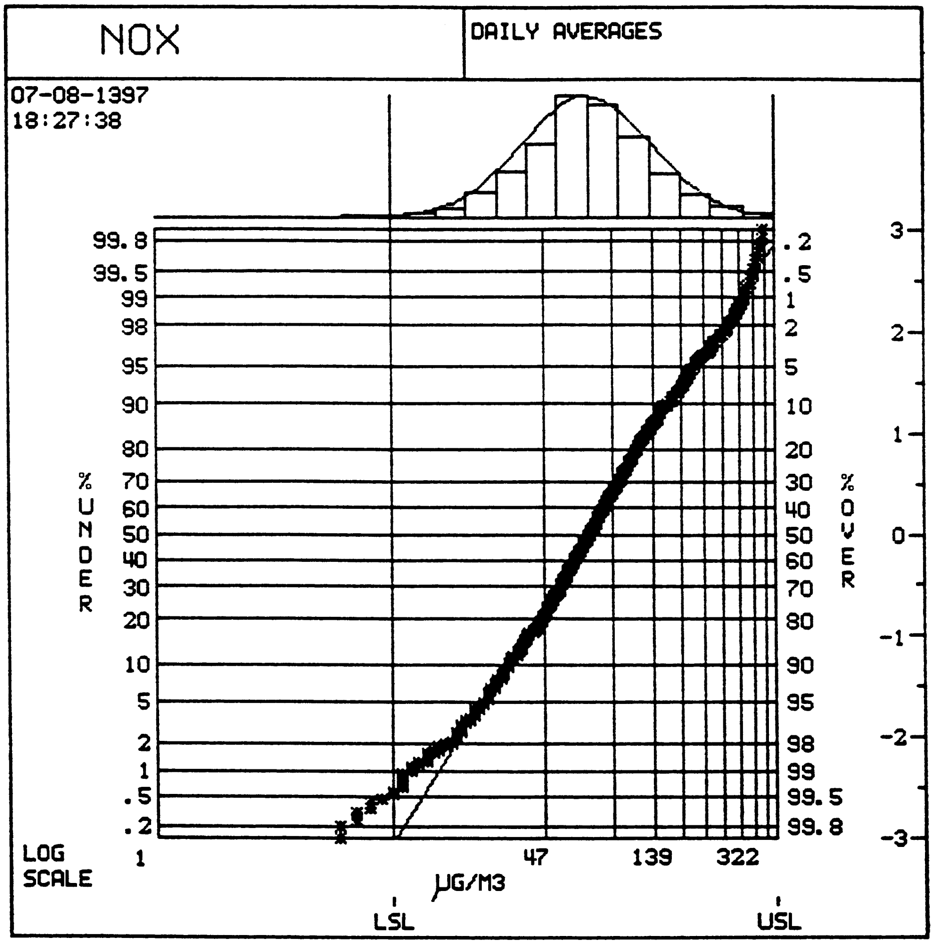

Fig. 2. – Plot of the daily nitrogen oxides concentration averages on a log-normal scale.

theoretical and experimental distributions. The analysis of daily nitrogen oxide concentration averages (table VI) indicate, again, that the log-normal distribution is an effective descriptor.

Figure 1 shows the plot of the log-normal PDF as a function of the daily ozone concentration average. The two tails of experimental data show a large variation in comparison to the log-normal line. Lower deviation can be attributed to the experimental uncertainties associated with instrumental accuracy. In fact, the concentration levels are less than 20 μg/m^3, which represents the analyser threshold. Upper deviations indicate that 20% of the data are poorly fitted by the distribution. The analogous plot of daily nitrogen oxide concentration average (fig. 2) shows a good agreement between theoretical and experimental distribution, the only exception being the lower tail where the limit of instrumental precision was reached.

Figure 3 has the plot of the O_3/solar radiation ratio. The physical meaning of the ratio (μg/Wm) is not to be considered as an observed quantity, but it represent only a relationship between correlated quantities. Clearly evident is the increased statistical

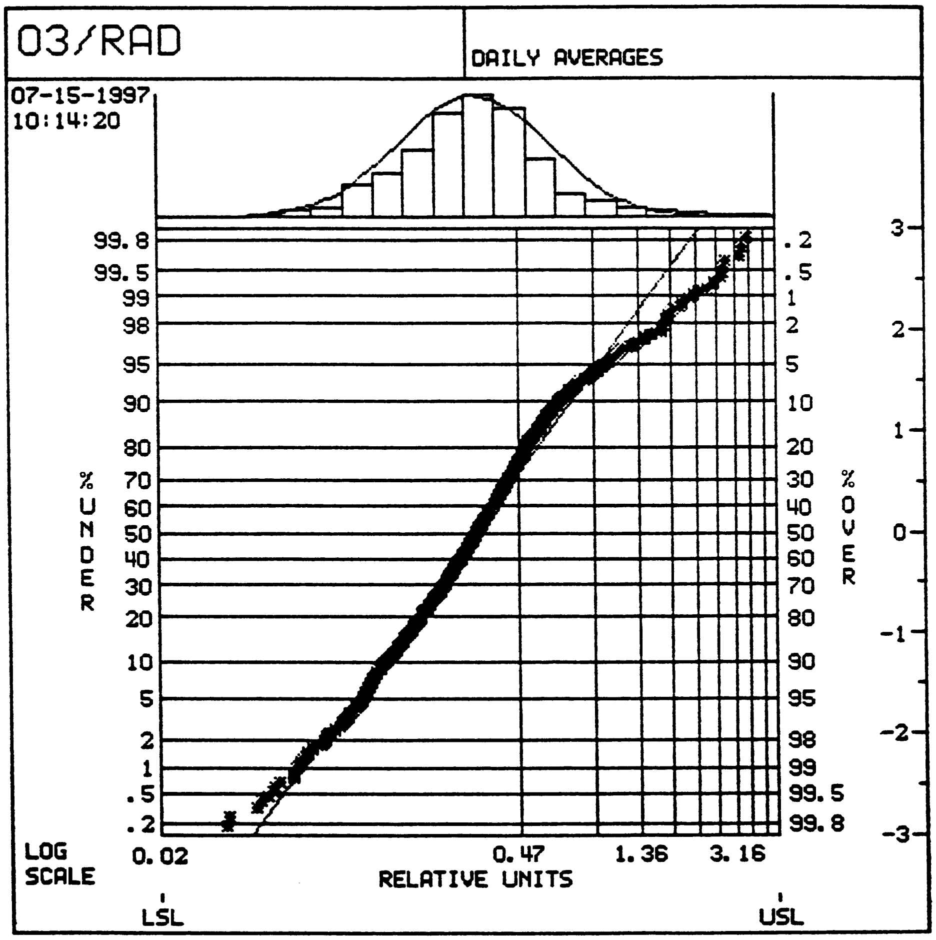

Fig. 3. – Plot of the daily O_3/solar radiation ratio on a log-normal scale. Relative units correspond to the ratio μg/Wm.

goodness of the fit between experimental data and the log-normal distribution with respect to fig. 1. Only 5% of the data are poorly fitted *vs.* high values of the ratio. In this case it is not surprising that the experimental data were underestimated by the log-normal distribution. The main reason resides in the fact that transport played a key role in increasing higher values throughout the measurement period. The upper slope of the experimental data must be attributed to the presence of high value concentrations due to the re-entry of ozone-rich air masses transported by the sea breeze. This phenomenon, which persists for two seasons a year, largely affects the ozone concentration maxima.

4. – Conclusions

The determination of the best PDF descriptor for a certain atmospheric pollutant, particularly a reactive pollutant such as ozone, must take into account some specific patterns of the site in which measurements are performed. In particular, when treating photochemical pollutants evolution the non-stationariry and non-linearity of the processes should be taken into account. In the investigated case this problem was overcome because of the peculiar distribution of the emission sources in the territory which can lead, in first approximation, to a quasi-steady state.

Under the assumption of QSS, the conventional statistical analysis here applied can furnish a description, even though partial, of pollutants statistical distributions. It should be considered also that most models currently applied are prevalently empirical and so it is impossible to attribute to them a universal validity under the varying environmental circumstances.

Ravenna's peculiar atmospheric patterns tend to ascribe to the log-normal distribution a major role in the description of surface ozone concentration, once the latter is normalised over solar-radiation intensity. Weibull and gamma distributions can also be applied, but the result is less useful to describe ozone concentrations.

Some important indications emerge by comparing the studies conducted in Los Angeles to those in Ravenna. The role played by wind speed seems to be critical in order to find the apparent discrepancies between the Weibull and log-normal distribution, in that the lower the wind speed, the greater the statistical confidence in the Weibull. This case confirms the findings of Georgopoulos and Seinfeld [25].

REFERENCES

[1] VIEZZE W., JOHNSON W. B. and SINGH H. B., *Stratospheric ozone in the lower troposphere. - II. Assessment of downward flux and ground-level impact, Atmos. Environ.*, **17** (1983) 1979.

[2] KITADA T., CARMICHAEL G.R. and PETERS L. K., *Numerical simulation of the transport of chemically reactive species under land- and sea-breeze circulations, J. Clim. Appl. Meteorol.*, **23** (1984) 1153.

[3] GUSTEN H., HEINRICH G., CVITAS T., KLASVIC L., RUSCIC B., LALAS D. P. and PETRAKIS M., *Photochemical formations and transport of ozone in Athens, Greece, Atmos. Environ.*, **22** (1988) 1855.

[4] CHANG Y. S., CARMICHAEL G. R., KURITH H. and HUEDA H., *The transport and formation of photochemical oxidants in central Japan, Atmos. Environ.*, **23** (1989) 363.

[5] MANTIS H. T., REPAPIS C. C., ZEREFOS C. S. and ZIOMAS J. C., *Assessment of the potential for photochemical air pollution in Athens: a comparison of emissions and air-pollutant levels in Athens with those in Los Angeles, J. Appl. Meteorol.*, **31** (1992) 1467.

[6] GIOVANELLI G., GEORGIADIS T., FORTEZZA F. and STROCCHI V., *Transport of photochemical ozone along the western Adriatic shore near a petrochemical plant, Nuovo Cimento C*, **8** (1985) 727.

[7] GEORGIADIS T., ALBERTI L., BONASONI P., FORTEZZA F., GIOVANELLI G. and STROCCHI V., *Seasonal budget of ozone and oxidant precursors in an industrial coastal area of northern Italy. Quadrennial Ozone Symposium, June 4-13, Charlottesville Va., USA, NASA Conf. Publ. 3266* (1992) pp. 48-50.

[8] GEORGIADIS T., BONASONI P., GIOVANELLI G., RAVEGNANI F., FORTEZZA F., ALBERTI L. and STROCCHI V., *Ozone transport and deposition at an Adriatic coastal site, 6th European Symposium on Physico-Chemical Behaviour of Atmospheric Pollutants, October 18-22, Villa Ponti, Varese, Italy*, edited by G. ANGELETTI and G. RESTELLI, *European Commission Publ.,* **1** (1993) 537.

[9] Georgiadis T., Giovanelli G. and Fortezza F., *Vertical layering of photochemical ozone during land-sea breeze transport, Nuovo Cimento C*, **17** (1994) 371.

[10] Fortezza F., Strocchi V., Bonasoni P., Georgiadis T. and Giovanelli G., *Time space model of summer ozone distribution around Ravenna, J. Geophys. Res.*, **98** (1993) 10615.

[11] Fortezza F., Strocchi V., Giovanelli G., Bonasoni P. and Georgiadis T., *Transport of photochemical oxidants along the northwestern Adriatic coast, Atmos. Environ.*, **15** (1993) 2393.

[12] EPA, *Quality Assurance Handbook for Air Pollution Measurements Systems*, Vol. I: *A Field Guide to Environmental Quality Assurance*, Office of Research and Development, Washington DC 20460, EPA/600/R-94/038a (1994).

[13] Aitchison J. and Brown J. A. C., *The Lognormal Distribution* (Cambridge Univeristy Press, London) 1973, pp. 173.

[14] Marani A. and Bendoricchio G., *Models of statistical distributions for NPSP concentrations*, in *Agricultural Nonpoint Source Pollution: Model Selection and Application*, edited by A. Giorgini and F. Zingales (Elsevier, Amsterdam) 1986, pp. 363-382.

[15] Ott W. R., *A physical explanation of the lognormality of pollutant concentrations, J. Air Waste Manag. Assoc.*, **40** (1990) 1378.

[16] Holland D. M. and Fitz-Simons T., *Fitting statistical distributions to air quality data by the maximum likelihood method, Atmos. Environ.*, **16** (1982) 1071.

[17] Marani A., Lavagnini I. and Buttazzoni C., *Statistical study of air pollutant concentrations via generalized Gamma distributions, J. Air Pollut. Control Ass.*, **36** (1986) 1250.

[18] Larsen R. I., *A new mathematical model of air pollutant concentration averaging time and frequency, J. Air Pollut. Control Ass.*, **19** (1969) 24.

[19] Nieuwstadt F. T. M., *Prediction of air pollution frequency distribution. Part II. The Gaussian plume model, Atmos. Environ.*, **14** (1980) 259.

[20] Marani A., *On mathematical models of natural systems, Proceedings of the VI ISEM International Conference, Venice, June*, in *Advances in Environmental Modelling*, edited by A. Marani (Elsevier, Amsterdam) 1988, pp. 5-26.

[21] Kewley, D. J., *The effect upon the pohotostationary state relationship when clean air and photochemical smog mix, Atmos. Environ.*, **14** (1980) 1445.

[22] Bange P., *Hidden photostationary equilibrium: a case study on the effect of monitor averaging on the calculated oxidation rate of NO to NO_2 in the plume of a power plant, Atmos. Environ.*, **27** (1993) 573.

[23] Hess, G. D. and Cope, M. E., *A note on subgrid-scale processes in photochemical modelling, Atmos. Environ.*, **23** (1989) 2857.

[24] Lamb, R. G., *Note on the apllication of K-theory to diffusion problems involving nonlinear chemical reactions, Atmos. Environ.*, **7** (1973) 257.

[25] Georgopoulos P. G. and Seinfeld J. H., *Statistical distributions of air pollutant concentrations, Environ. Sci. Technol.*, **16** (1982) 401A.

[26] Mak M. K. and Walsh J. E., *On the relative intensities of sea and land breezes, J. Atmos. Sci.*, **33** (1976) 242.

[27] McElroy J. L. and Smith T. B., *Creation and fate of ozone layers aloft in southern California, Atmos. Environ. A*, **27** (1993) 1917.

[28] Miller P. R., De Lourdes- De Bauer M., Nolasco A. Q. and Tejeda T. H., *Comparison of ozone exposure characteristics in forested regions near Mexico City and Los Angeles, Atmos. Environ.*, **28** (1994) 141.

[29] Weston K. J., Kay P. J. A., Fowler D., Martin A. and Bower J. S., *Mass budget studies of photochemical ozone production over the U.K., Atmos. Environ.*, **23** (1989) 1349.