H14-83

ON THE RELATION BETWEEN DAILY EXCEEDANCES AND YEARLY AVERAGES FOR PM10

Joost Wesseling[1], Ferd Sauter[1] and Sander Jonker[2]

[1] National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands
[2] TNO Built Environment and Geosciences, Utrecht, The Netherlands

Abstract: The daily limit value for particulate matter (PM10), which limits the number of days with PM10 concentration levels exceeding 50 µg/m3, is in practice much more restricting than the yearly average limit value. The Dutch system for assessing urban air quality primarily relies on a simple model that calculates yearly average concentrations only. It is therefore important to understand the relation between yearly average concentrations and the accompanying number of exceedance days. Here, we present a simple analytical expression to describe the relation between the yearly averaged PM10 concentration and the number of days with average concentrations exceeding 50 µg/m3. This analytical expression, which comprises a two-parameter error function, is based on the assumption of a three-parameter log-normal distribution of the PM10 concentration data. It has been tested in depth using daily averaged PM10 concentrations from measuring stations in the Netherlands and, in less detail, using yearly averaged PM10 concentrations from European countries, as available in the Airbase database. Although the fitted parameters in the relation vary significantly between countries, the resulting values for the so-called ‘critical’ PM10 concentration, i.e. the yearly averaged concentration that corresponds with the EU-norm of 35 daily exceedances per year of a threshold of 50 µg/m3, are not dissimilar among these countries. In fact, the 95% confidence interval for almost all countries for which there were sufficient data contains a European average of the critical PM10 concentration of 30 µg/m3.

***Key words:** particulate matter (PM10) exceedance days, log-normal distribution.*

INTRODUCTION

The Dutch system for assessment of urban air quality relies mainly on a simple model, denoted Calculation of Air pollution from Road traffic (CAR), which has been derived from wind tunnel measurements. This model can estimate yearly averaged concentrations in an urban environment from a few simple parameters, such as the amount of traffic, the type of street and information on background concentrations. Over the years the model has been calibrated extensively (Wesseling and Sauter, 2007), and it is updated each year by the Dutch ministry of housing, spatial planning and the environment (VROM). The number of PM10 exceedance days has to be estimated from the yearly average concentration as it is not calculated directly by the model. It is therefore important to understand the relation between yearly average concentrations and the accompanying number of exceedance days. The aim of this article is to investigate the nature of this relation.

PM10 MEASUREMENTS BY THE RIVM

The Dutch National Institute for Public Health and the Environment (RIVM) is a centre of expertise in the fields of health, nutrition and environmental protection. It manages a network of measurement stations in the Netherlands, known as the National Air Quality Monitoring Network (NAQMN). In this network, PM10 is measured at some 40 different locations using fully automatic beta-ray absorption analysers. An equivalence study, based on the guideline of the EC Working Group (2005), with some modifications, has been conducted to determine the equivalence between the beta-ray analysers in the NAQMN and the European reference method (Beijk et al., 2008). The measuring locations are distributed throughout the Netherlands, although most locations are in the densely populated western part of the country. During the last few years, the number of locations where PM10 is measured has almost doubled. The measurement locations can be categorized into three basic types: regional, urban background and street. Regional locations are usually situated in rural areas of the Netherlands where there are no industrial sources or roads in the neighbourhood. Urban background locations are located in cities, at least a few blocks distant from main urban streets. Street locations are, as the name implies, located in streets with a substantial amount of traffic. Where possible, the measurement stations are located on the sidewalk, close to the road. The (preliminary) hourly concentrations measured at each location are reported directly on the internet (see: www.lml.rivm.nl). After the data have been checked and validated, they are officially reported to Airbase, the public air quality database system of the European Environmental Agency (EEA), and released to the public.

THE LOG-NORMAL CHARACTER OF PM10 CONCENTRATION DISTRIBUTIONS

The log-normal distribution is used for a variety of applications in all fields of science, as described by Limpert et al. (2001) who lists a large number of diverse applications of the log-normal distribution, including the distribution of particles in the environment, the length of spoken words in phone conversations, latency periods of diseases, species abundance, the age that women marry in Denmark and hydrological data and farm size in England and Wales. Field data from a wide variety of scientific areas are standardly fit to log-normal distributions.

The distribution of PM10 concentrations in the Netherlands has been found to be log-normally distributed (Elzakker and Buijsman, 1999; Kruize et al., 2000). In a study of the statistical nature of PM10 concentrations in Taiwan, Hsin-Chung Lu (2002) used the log-normal distribution, a Weibull distribution and a type V Pearson distribution to fit measured concentration distributions for several stations. The results showed that concentration distributions were best fitted using the log-normal distribution, although in one case the concentration distributions were bi-modal, probably due to discrepancies in local meteorological conditions in different seasons. Therefore, a combination of two distributions was used. The ratio between measured and calculated number of days exceeding 125 µg/m3 varied from 44% to 192%. Hsin-Chung Lu and Guor-Cheng Fang (2003) subsequently extended the work of Hsin-Chung Lu by using parent distributions with different distributions at higher concentration values in which one approach was based on extreme value theory and a second approach was based on a two-parameter exponential distribution. The log-normal distribution was found to be the most appropriate parent distribution, although at high concentrations the two-parameter exponential distribution better matched the data. More recently, Mijić (2009) investigated the applicability of the log-normal distribution, a Weibull distribution and a type V Pearson distribution to describe measured PM10 distributions in Belgrade during the period 2003–2005. The type V Pearson distribution gave the best result, with the measured number of days exceeding 125 μg/m3 overestimated by 17%.

The primary objective of the recent studies was to determine the distribution that best describes the shape of measured PM10 concentration distributions, also at high concentrations. In the analysis reported here, we limited ourselves to the log-normal distribution, as this not only fits the measured concentration distributions in the Netherlands, but it can also produce a simple analytical relation between measured yearly average concentrations and the accompanying number of exceedance days. An aim of our study was to describe measured concentration distributions and the accompanying number of exceedance days for a large number of stations, using concentration levels that may be above or below limit values.

For our analysis of PM10 distributions we initially assumed a two-parameter log-normal distribution for the daily average PM10 concentrations. The probability p(c) of observing days with a daily average PM10 concentration of c is then given by:

$$p(c) = \frac{e^{-\frac{(\ln(c)-\mu)^2}{2\sigma^2}}}{c\,\sigma\,\sqrt{2\pi}} \tag{1}$$

The parameters μ and σ determine the shape of the distribution. However, in a relatively large number of cases, a shift (δ) of the distribution (to the right) is found to be necessary in order to satisfactorily describe the measured concentration distributions, as there is always a certain background level present. For concentrations above δ the modified three-parameter log-normal distribution can be written as:

$$p(c) = \frac{e^{-\frac{(\ln(c-\delta)-\mu)^2}{2\sigma^2}}}{(c-\delta)\,\sigma\,\sqrt{2\pi}} \tag{2}$$

For concentrations below or equal to δ the modified three-parameter log-normal distribution is not defined, as the argument to the natural logarithm in the relation becomes zero or less. We then define a value of zero for the distribution. In statistics, the parameter σ is often called the 'shape parameter', δ the 'location parameter' and m = exp(μ) the 'scale parameter' (NIST, 2008). The yearly average concentration ($\overline{C}$) of a log-normal concentration distribution with parameters σ, δ and μ can be written as:

$$\overline{C} = e^{(\mu+\frac{1}{2}\sigma^2)} + \delta \tag{3}$$

The shape of the distributions varies substantially when the parameters are changes, even though the yearly average concentrations may remain the same. More detailed study shows that the parameter μ has the largest relative influence on the yearly averaged concentration. An example of an individual concentration distribution is shown in figure 1. In this case, it is the distribution of measured daily average PM10 concentrations in 2006 at rural station 633 in the Netherlands. This station is located near the village of Zegveld, in the province Utrecht.

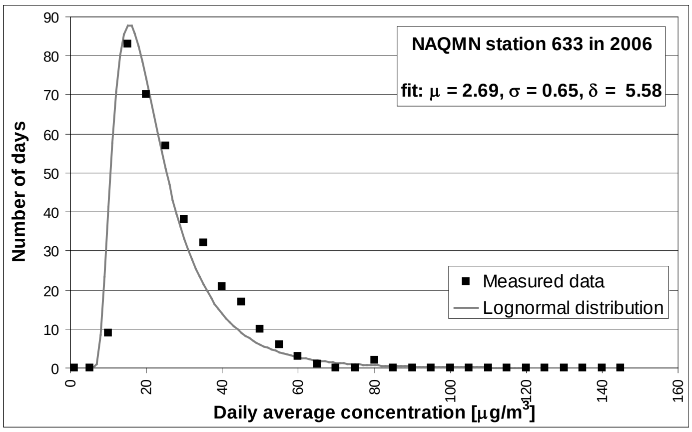

Figure 1: An example of an individual concentration distribution

For all Dutch stations, hourly measured PM10 concentrations were combined to provide daily average concentrations using a procedure that complied with relevant EU rules regarding data quality and availability. To each measured distribution, a three-parameter log-normal distribution, as shown in relation (2), was fitted using a Nelder–Mead simplex method (MATLAB, version 7.1; see Lagarias et al., 1998). The log-normality of the measured distribution was verified using a Jarque–Bera test (Jarque and Bera, 1980) to test whether the sample skewness and kurtosis are unusually different from their expected values, as measured by the chi-square statistic at the 5% significance level. Interestingly, the percentage of rejected fits is quite small for all years, with the exception of 2001. In fact, for most of the years in the analysis, the percentage of rejected fits varies between 0% and 10%. In 2001, the percentage of rejected fits reaches a maximum at just over 50%, due to very high peaks in a few days in January, which lead to skewed distributions.

EXCEEDANCE DAYS

Using the assumed log-normal distribution discussed earlier, the fraction of days (ν) having concentrations above a specific threshold concentration G can be calculated by integrating the distribution. Relation (3) can be used to link the yearly average concentration to the shape parameters. The result is:

$$\nu = \tfrac{1}{2}(1 + erf\left(\tfrac{1}{2}\sqrt{2}\left(\frac{\ln(\overline{C}-\delta) - \tfrac{1}{2}\sigma^2 - \ln(G-\delta)}{\sigma}\right)\right) \tag{4}$$

Here erf() is the well-known error-function. In the case of PM10, the relevant value of G is 50 μg/m3. The PM10 data with yearly averages above 30 μg/m^3 can be described quite well using a linear function that follows quite naturally from our relation (4). A second-order series expansion of (4) around a specific concentration "M" yields the following linear relation between the yearly average concentration and the fraction of exceedances:

$$\nu = \tfrac{1}{2}(1 + erf\left(\tfrac{1}{2}\sqrt{2}\left(\frac{\ln(M-\delta) - \tfrac{1}{2}\sigma^2 - \ln(G-\delta)}{\sigma}\right)\right) \quad + \quad \frac{e^{-\tfrac{1}{2}\left(\frac{\ln(M-\delta) - \tfrac{1}{2}\sigma^2 - \ln(G-\delta)}{\sigma^2}\right)^2}}{\sqrt{2\pi}(M-\delta)\sigma}(C-M) \tag{5}$$

The parameters in (5) can be fitted from de observed exceedance data or they can be fitted to all underlying individual log-normal distributions of the PM10 concentrations. In figure 2 the result of relation (4) is compared to measured numbers yearly exceedance days at all Dutch stations and for all available years.

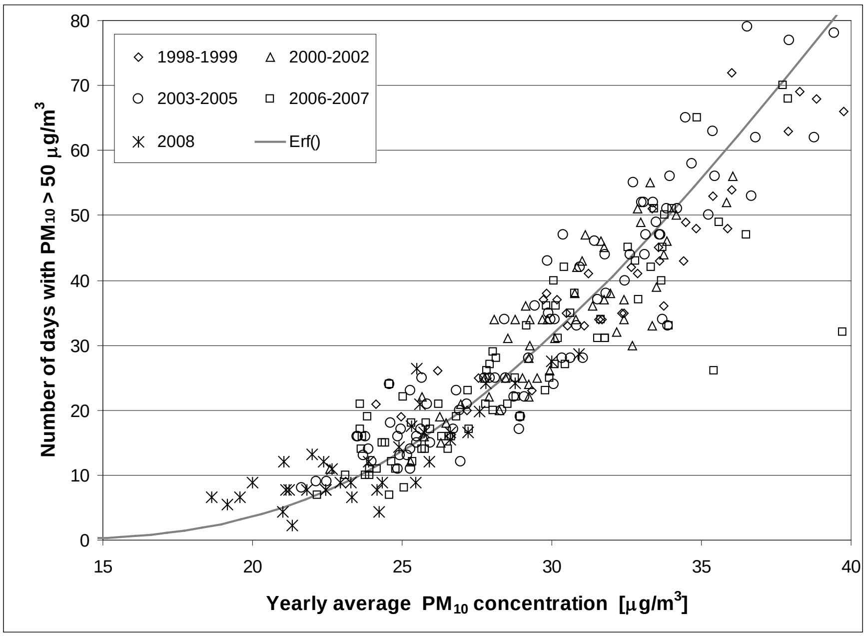

Figure 2. A plot of the relation between the measured yearly average PM10 concentration and the number of days having average concentrations above 50 μg/m^3. The curve labeled 'Erf()' has been calculated using relation (4) with fixed values δ=6 μg/m^3 and σ=0.56.

The agreement is quite satisfactory, although the calculated exceedance days, on average, underestimate the measured values by 5%. Where it matters most, around 35 exceedance days per year, 5% amounts to an average underestimation of 1.75 days. In figure 2 the values of δ (6.0 μg/m^3) and σ (0.57) in relation (5) were fitted directly to the measured yearly averages and exceedance days. In performing such a fit, it is assumed that δ and σ do not vary significantly over the years. This assumption is adequate for the Dutch data. Fitting δ and σ directly is like fitting the tails of the concentration distributions only, given the known value of the number of exceedance days, while neglecting the shapes of the remainder of the concentrations. The critical concentration, i.e. the yearly average concentration at which according to relation (4) exactly 35 exceedances occur, is 31.3 μg/m3.

OTHER COUNTRIES IN EUROPE

Using data taken from Airbase, we have extended our analysis to data from other countries within Europe. Data for yearly average PM10 concentrations and corresponding days with average concentrations over 50 μg/m3 were available in Airbase for the period 1991 up to and including 2007. The number of exceedance days as a function of the yearly average PM10 concentrations is shown for several countries in Figure 10. Both the number of years with data for each station and the number of measuring stations vary from country to country. The data for Italy, Poland and Czechoslovakia include quite large yearly average PM10 concentrations and corresponding exceedance days, whereas all concentration levels are quite low for Finland and Ireland.

The relation between yearly average PM10 concentrations and exceedance days is clearly quite similar in each country and consistent with relation (5). For a more quantitative analysis we have fitted the data for every country using relation (5). In fitting the relations between yearly average PM10 concentrations and corresponding exceedance days directly, rather than fitting the underlying log-normal distributions, we assume that the values of δ and σ do not change much in time, as was indeed the case for the Dutch data. The results are presented in Figure 3 as curves. The abbreviations of the country names are taken from Airbase. From the analysis we observe that, although fitted values of δ and σ vary significantly, the resulting values for Ccrit (the 'critical' PM10 concentration that corresponds with the EU-norm of 35 daily exceedances per year of a threshold of 50 µg/m3) are not so different for the different countries. In fact, the 95% confidence interval for almost all countries with sufficient data contains a European average of Ccrit = 30 µg/m3.

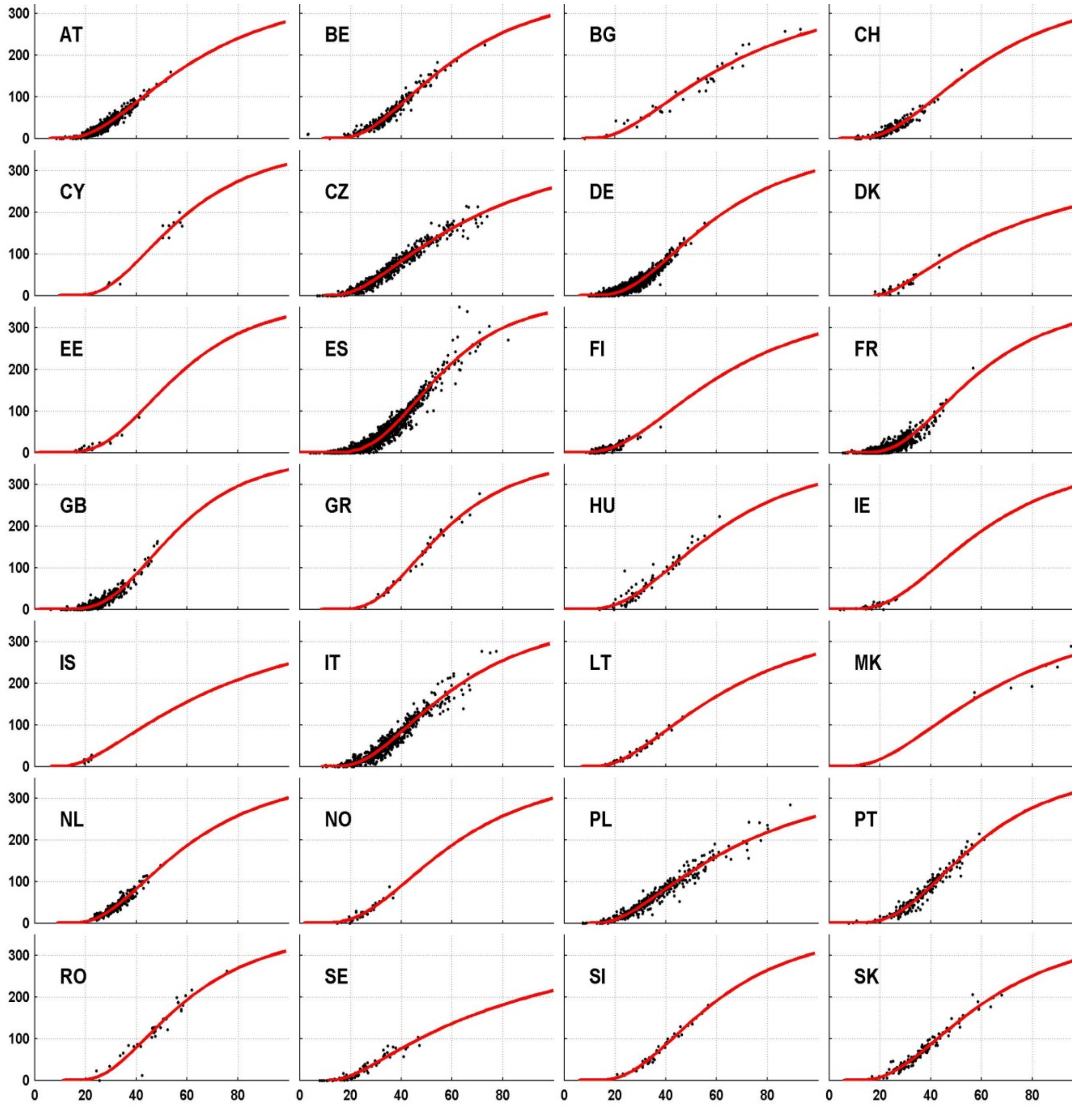

Figure 3. The number of exceedance days (y-axis) vs. the yearly average concentrations (x-axis in $\mu g/m^3$) for several European countries. The lines are the fitted error functions according to our analysis.

CONCLUSIONS

In this paper we have presented a simple analytical expression to describe the relation between the yearly averaged PM10 concentration and the number of days with average concentrations exceeding 50 µg/m3. This expression, which is based on the assumption of a three-parameter log-normal distribution of the PM10 concentration data, has been tested in detail using daily averaged PM10 concentrations from measuring stations in the Netherlands and, in less detail, by using yearly averaged PM10 concentrations from European countries, as available in the Airbase database.

By fitting the relation between the concentrations and the exceedance days, we estimated a so-called ‘critical’ PM10 concentration (Ccrit), i.e. the yearly averaged concentration that corresponds with the EU-norm of 35 daily exceedances per year of a threshold of 50 µg/m3, of 31.3 µg/m3. A similar analysis was perform for most EU-countries. The 95% confidence interval for almost all countries with sufficient data contains a European average of Ccrit = 30 µg/m3.

REFERENCES

Elzakker BG van, Buijsman E, 1999, "The Dutch National Air Quality Monitoring Network: monitoring program in 1999", in Dutch, RIVM Rapport 723101032

Hsin-Chung Lu, 2002, “The statistical character of PM10 concentration in Taiwan Area", Atm. Env. 36 (2002) 491-502.

Hsin-Chung Lu and Guor-Cheng Fang, 2003, "Predicting the exceedances of a critical PM10 next term concentration-a case study in Taiwan", Atmospheric Environment, Volume 37, Issue 25, August Pages 3491-3499

Jarque, Carlos M., Anil K. Bera, 1980, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals". Economics Letters 6 (3), 255–259.

Lagarias Jeffrey C., James A. Reeds, Margaret H. Wright, Paul E. Wright, 1998, "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions", SIAM Journal of Optimization, 9(1): p.112-147.

Limpert, E. Stahel, W.A. and Abbt, 2001, “Log normal distributions across the sciences: Keys and clues”, BioScience, may 2001, Vol. 5, No. 5, 341.

Mijic, Z, et al, "The statistical character of PM10 in Belgrade area", Atmospheric Research (2009), doi:10.1016/j.atmosres 2009.01.002.

Wesseling J.P. and Sauter F.J., 2007, "Calibration of the program CAR II using measurements of the national measuring network of the RIVM", RIVM report 680705004 (in Dutch).