Atmos. Chem. Phys. Discuss., 2, S694–S697, 2002
www.atmos-chem-phys.org/acpd/2/S694/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Long term measurements of submicrometer urban aerosols: statistical analysis for correlations with meteorological conditions and trace gases" *by* B. Wehner and A. Wiedensohler

Anonymous Referee #2

Received and published: 29 November 2002

General comments

There is no doubt in my mind that long-term aerosol particle size distribution data sets such as the one presented in this paper will prove to be extremely valuable in the near future. The data set is unique considering its duration (4 years), consistency (nearly complete data coverage), size range (down to 3 nm), and high time resolution (15 min). I can only hope that the researchers persevere for many more years to come. I also recommend that they establish contact with epidemiologists studying the effects of particulate matter on human health. Any such effect should be more evident when mortality and morbidity in a large population (Leipzig) is correlated with a long-term data set. Furthermore, the detailed size distributions presented are likely to give vital clues as to the effects of certain size ranges, for instance ultrafine particles. Since one urban site is probably not fully representative of the population exposure, it would also be very exciting if other sites in Leipzig and its surroundings were included in future studies.

When given such an extensive data set, the question naturally arises how best to present it. The authors have chosen time series analysis and PCA. One might suggest additional ways.

Normally, size distributions can be split into (log-normal) modes, for instance the nucleation, Aitken, and accumulation modes. The mode parameters (number concentration in each mode, count median diameter, geometric standard deviation) can then be presented to give a condensed overview of the data. This parameterisation requires fitting all size distributions, which is a huge task if not performed in an automated way. The mode parameters might then be used in for instance PCA. Simple statistics - presented in "boring" tables - is often very useful to the scientific community, and should not be underrated.

Trend analysis might prove useful, since the data set covers 4 years.

The data could be used to test whether the size distribution at any given time can actually be predicted with adequate accuracy as a function of simple variables. These explaining variables would then include time of day, weekday, temperature, relative humidity, wind speed and direction, and global radiation, perhaps also trace gases. The hypothesis is then that these simple variables can account for the major processes affecting the size distribution. Traffic, the dominating source, is largely described by the time of day and weekday, defining the driving pattern. Nucleation is largely described by global radiation etc. Essentially, the authors have already described these covariations in their PCA, but they have not attempted to build a model that can predict size distributions. There are several statistical techniques to do this, for instance PLS

ACPD

2, S694–S697, 2002

Interactive Comment

(Partial Least-Squares regression) and PMF (Positive Matrix Factorization). It would be asking too much to include this in the present paper, but given the long-term data set, it would be very interesting to try it. If it works, simple variables might be used to estimate size distributions. That would enable people that want to make use of the data set for other purposes (e.g. epidemiologists) to extend the data some years forwards and backwards in time, and to fill in missing data.

To summarize, the paper is very well written, and the outcome of the data analyses performed are clearly presented. I very much enjoyed reading it. I strongly recommend it for publication in ACP.

Specific comments

Introduction, line 4: "cardiopulmonary", should probably be "cardiovascular".

Introduction, 2nd paragraph, line 8: should probably be "$>$10-20 nm".

Page 3. The hypothesis on prerequisites for the formation of new, ultrafine particles is plausible, however difficult to prove using only this data set.

Fig 4. The morning rush hour peak is much more pronounced in winter than in summer. This might be due to more intense particle formation immediately after the exhaust gases leave the tailpipe during the cold winter conditions. Lower temperatures lead to higher supersaturation ratios for the emitted vapours, and consequently higher nucleation rates.

Section 3.2.3: From a statistical point of view, PCA and factor analysis (FA) are two different methods. Their underlying philosophies differ: PCA attempts to account for maximum covariance, while FA separates residuals (unique factors) that are uncorrelated with respect to the common factors. PCA and FA are often confused. If the authors used PCA, then they should not use terms such as factor loadings, but rather PC loadings.

Section 3.2.3, Page 4-5: The section describing interdependent variables is somewhat

ACPD

2, S694–S697, 2002

unclear. There is no reason why highly correlated variables should not be included in the PCA. They would then turn up with high loadings on the same PC. Including also temperature and relative humidity might give clues to the nucleation phenomenon.

Section 3.2.4, Page 5: Since the CO/NOX ratio is considerably lower for diesel vehicles compared to gasoline driven vehicles, it would probably have been wise also to include CO in the analysis.

Section 3.3, Page 5: What is meant by the "relative peaks" describing the time lags of ¡ day etc.?

Section 3.5, Table 3, Page 6: Was the original PC solution subjected to an orthogonal Varimax rotation?

Section 3.5.2, Page 7, last paragraph: I fully agree with the authors that we still lack a good explanation for the frequently observed particle nucleation events. The good news about this is that it still leaves us with more exciting science to be done in the future.

Section 3.5.3, Page 8: The authors assume that the PCs 4-14 contain noise only. It should be kept in mind that the variance in PCs 4-14 could also be due to processes not accounted for in the PC model. This could be either a PC loading pattern that could not be interpreted properly by the authors, or a process that is not manifested in the variables included. For instance, a process related to temperature, which was excluded from the PCA, would not be easily visible in the PCA solution.

Interactive comment on Atmos. Chem. Phys. Discuss., 2, 1699, 2002.