Atmos. Chem. Phys. Discuss., 6, S5686–S5692, 2006
www.atmos-chem-phys-discuss.net/6/S5686/2006/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Application of absolute principal component analysis to size distribution data: identification of particle origins" *by* T. W. Chan and M. Mozurkewich

T. W. Chan and M. Mozurkewich

Received and published: 23 December 2006

Main Points:

P10496 regarding discussion of PCA vs. PMF. We devote just over one page in the ACPD paper to this issue. Since we are neither claiming superiority of one method over the other nor seeking to provide a guide to choosing one method or the other, we do not believe that this is the place for a more detailed discussion of both methods. Each method has its advantages and disadvantages. What is appropriate here is to give our rationale (largely simplicity of implementation) for the choice we made, that is what we have done. We will try to clarify our discussion, keeping the reviewer's comments in mind.

P10501 and P10516 regarding inclusion of wind speed in the APCA analysis. We see the referee's point: if the components are regarded as "sources" then there is a conceptual problem with including wind speed (or solar radiation, in the case of the Egbert data). But there is a different way of looking at the components: that they are merely statistical groupings of correlated (or anti-correlated) variables. In this view, which we adopt, there is no difficulty with including such non-material variables. A group of material variables that have a common source might be part of such a component, but need not be all of it. We do need to include a brief discussion of this point in the revised paper and will do so. Leaving wind speed out of the analysis has very little effect on either the loadings or the scores and would not alter the conclusions. Including wind speed does provide additional support for our identification of the "boundary layer dynamics" component, which is the only one with a substantial dependence on wind speed.

P10502/L14 regarding particle nucleation. The points made by the referee are essentially the same as given in the paper, but our text is overly terse and not very clear. We will expand and clarify it.

P10501-10506 regarding clarifying the presentation. We will carefully consider ways of doing this and will, at a minimum, include either a table of all components observed at each site or a paragraph in section 4.1 to the same effect.

P10501, sec. 4 regarding mixed components. The term 'mixed components' is only applied to component loadings prior to the Varimax rotation. After the Varimax rotation, the variance explained by different factors is re-distributed so that the rotated components are linear combinations of all the components included in the rotation. Therefore, once rotated, we can not differentiate the signal components from the mixed components.

P10501 sec. 4 regarding the average size distribution for each component. As discussed in the MS, we first apply APCA to all the size distribution data to decompose the size distribution into different numbers of independent aerosol components. Then

ACPD
6, S5686–S5692, 2006

EGU

the scores for these aerosol components are combined with the trace gas measurements and meteorological data in a traditional PCA analysis. The average sizes of all aerosol components that were used in each study were summarized in Table 3 in this MS; their shapes were shown in Fig 1 in the companion paper. An average size distribution for each component could be obtained by a linear combination of the corresponding aerosol components. The resulting distributions look just like one would expect from the bar graphs given in Fig. 2.

P10501/Fig 3 regarding flat component scores when there are no clouds. Figure 3 is a plot of the component scores as a function of time. The major and minor tick marks on the x-axis represent mid-night and noon, respectively. There were no cloud observations from 21:00 to either 2:00 or 4:00, depending on the night. We will revise Fig. 3 to clearly indicate these intervals. The reason that the scores were flat during those night-time periods is because there was no photochemistry. Also, since the cloud coverage data were hourly averages, they can not always represent short term variations in cloud coverage. We will modify the text on the top of page 10502 and the figure caption to clarify this point.

Detailed points:

General question regarding the units of the loading and scores. In this case, since the input data was normalized prior to the PCA analysis, both the loadings and scores are dimensionless. We will clarify this in the revised paper.

General question regarding how the scores of the factors strongly associated with the particle sizes compare to the scores of the particle sizes in the APCA. The scores obtained from PCA analysis of the mixed data can be viewed as a weighted average (by the loading amount) of the number concentration for the individual sizes in the original size distribution except that each of the size bin in the original data should be divided by the standard deviation in the corresponding size bin.

P10494/L13-14 regarding particle loadings in the boundary layer dynamics component.

ACPD
6, S5686–S5692, 2006

EGU

As shown in Fig. 3, the boundary layer dynamics component does not have strong loadings on the particle components. This does not mean that there are no changes in particle size distributions associated with that component, only that the associated variations are small in comparison with the contributions from the other components.

P10495/L1-2 regarding suggested citation of the Zhang et al (2005) paper. We don't cite that work because it is concerned with the analysis of aerosol mass spectra, not size distributions.

P10496/L20 regarding clarification of arbitrary positive/negative loadings. We address this issue in our reply to the referee's comments on the companion paper.

P10497/L11-19 regarding doing the analysis in one step. In the first MS, we emphasized the importance of weighting in applying PCA to the size distribution data. As described in section 3.2, we found that the standard scaling to unit variances was most appropriate. We did try doing the analysis one step and it was not successful; the reason seems to be the need to use two different types of scaling. We will revise section 3.2 to make this clearer.

P10499/L20 regarding number of aerosol components included in the Hamilton 2000 analysis and why Pacific 2001 was not included. In the first paper (P10479/L1-6), we indicated the minimum and maximum numbers of aerosol components to retain for different field studies. In this paper, we generally used the maximum number of aerosol components in order to preserve maximum information. That maximum number ranged from 5 to 8, depending on the study. In the case of Hamilton 2000, although the maximum number was 5 components, we chose to use 4 aerosol components. This is because when we compare the average size distributions observed during Hamilton 1999 and 2000 studies (Fig. 4), we found that the observed accumulation mode particles during the two studies were very similar. When we compared the modal diameters of the aerosol components between the Hamilton 2000 and Hamilton 1999 studies, we found that the 4 aerosol components in Hamilton 2000 provide a better comparison

ACPD
6, S5686–S5692, 2006

EGU

with the 6 components used in Hamilton 1999 study, which is consistent with our observation from the size distributions. As a result, we decided that using 4 aerosol components instead of 5 in the mixed data for the Hamilton 2000 study would enable us to better determine the common sources present in the two Hamilton data sets. We will explain this in the revised MS. The Pacific 2001 data were not included since a main objective was to determine if we could get consistent results from a number of sites in the same region. The Pacific 2001 data would not have contributed to this and would only have made the paper longer and more difficult to understand.

P10499/L24 regarding the word “mixed”. The mixed data set in the second MS has no ‘special’ meaning and has no connection with the “mixed components” defined in the first MS. The mixed data set in the second MS simply refers to a data set that contains more than one type of data. As explained above with respect to the main comment re P10501, sec. 4, the mixed components have no relevance here.

P10500/L16 regarding the ‘unsatisfactory results’. The problem was that the results were dominated by whatever variables had the greatest variance with respect to the measurement uncertainty. The reducing the measurement uncertainty gives a variable greater significance in the analyis even if the experimental error is an insignificant portion of the variance in that variable. This is not physically reasonable. We will clarify this in the text.

P10501/L2 regarding the modified scree plots. As mentioned in the MS, the modified scree plot works best for the size distribution data and was not as useful for the mixed data set. We only used the results as a starting point and that is often not the final solution. We do not believe it will add extra information by including them in the MS.

P10501/L8 regarding ‘reasonable physical interpretation’. This is discussed at length in section 4.

P10501/L24, regarding why some scores (e.g., Fig. 3 & 8) are negative. These appear to be an artifact, but we do not have a fully satisfactory explanation for their origin.

ACPD

6, S5686–S5692, 2006

They may be connected with the orthogonality requirement or with the fitting of the data. Since these negative values are always small in magnitude compared to the positive values and occur only a small fraction of the time, we have not been overly concerned with them.

P10503/L15 regarding regional vs. local SO2 variations. The variations at Hamilton and Egbert were much larger than at the other sites, implying that the difference is local. For example, at the Hamilton site the SO2 concentration generally fluctuates between 5-8 ppbv except when the wind is from the direction of the steel mills; then the SO2 concentration is as high as 50 ppbv.

P1004/L5 regarding back trajectories vs. local wind direction. We did not find anything that alters our conclusion in this MS. The local factors identified by the wind directions are very short range.

P10505/L6 regarding the boundary layer dynamics in Egbert. The presence of the boundary layer dynamics factor in this study mainly involves loadings on NOx, Ox, and wind speed. In the case of Egbert, none of these were available in the data set, so there is no reason that this should have any effect on the other factors. Note that NOy was available, but this behaves very differently than NOx and is mainly associated with the regional pollutants.

P10505/L22 regarding the "weak" justification of the processed nucleation mode particles in Simcoe. We disagree.

P10506/L18 regarding reference for the origin of the transported particles. This was based on evidence presented in this paper.

P10513/Table 2 regarding size distributions with different resolutions. Although these data sets have different measured size ranges, they were measured with the same resolution, with 16 bins per decade. This is reasonably compatible with the lower sheath air to aerosol flow ratios used. In any case, the PCA representation used for the input

ACPD
6, S5686–S5692, 2006

EGU

data further lowers the effective resolution, but with minimal loss of information; this is because atmospheric size distribution data usually do not show the sort of features that really require very high resolution. Therefore, resolution is not an issue here.

P10521/Fig 7 regarding the data source of the polar plot. This represents only Hamilton 1999 data. Hamilton 2000 data show similar results.

P10521/Fig 7 regarding the value of the polar plot. Fig 7 is a simple polar plot, not a conditional probability function plot. The values on the plot represent the values of the component scores.

P10522/Fig 8 regarding additional graphs. Overlaying the graphs would not be very clear, but an image plot of limited time periods clearly show the growth and could be included.

We will correct the typos and grammatical errors.

Interactive comment on Atmos. Chem. Phys. Discuss., 6, 10493, 2006.