Atmos. Chem. Phys. Discuss., 9, C3217–C3220, 2009
www.atmos-chem-phys-discuss.net/9/C3217/2009/
© Author(s) 2009. This work is distributed under
the Creative Commons Attribute 3.0 License.

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Origin of aerosol particles in the mid latitude and subtropical upper troposphere and lowermost stratosphere from cluster analysis of CARIBIC data" *by* M. Köppe et al.

Anonymous Referee #1

Received and published: 23 July 2009

Review of "Origin of aerosol particles in the mid latitude . . ..." By Koppe et al (acp –2009-320)

1. General Comment The study presented here uses CARIBIC data coupled with, primarily, cluster analysis to derive information on the origin of nucleation and Aitken mode particles in the upper troposphere. The data themselves are quite interesting given the scarcity of such information for the upper troposphere and I found the analysis of the particle origin also interesting though somewhat more ambiguous than I think the authors portray it. For example, it is never made clear precisely where the

particles are actually formed (nucleated). Additionally, I feel that the justification for the extreme winnowing of the data (only 26-38% of the data points were used) needs a bit more work. Nevertheless, I am sure that the study will prove valuable and recommend publication after the few moderately important issues I raise below are addressed.

2. Specific Comments 2.1 Page 13528, lines 15-25. The issue of the rather extreme paring down of the data set arises here. Of course one always expects to have to discard some data during quality assurance screening but to discard two-thirds to three-quarters of the data is very unusual and does, I feel, require more discussion. The authors do point out that the missing values are evenly distributed geographically and thus assume that their analysis is sound. However, this merely means that a geographic bias is unlikely, not that other biases cannot have arisen. For example, they should state what percentage of the measurements was rejected due to calibration as opposed to instrument malfunction. I would imagine that mostly it would be due to malfunction. If so, then what sort of malfunctions are we talking about here? The instrument package must necessarily be autonomous and if something malfunctions it must somehow repair or correct itself (since if it remained "dead" until the flight ended there WOULD be a geographic bias). This much malfunction and self-correction must be explained to assure the reader that the data that WERE retained are in fact credible. And there are other considerations that should be discussed. Were there one or two instruments that accounted for most of the data lose or was the lose associated approximately equally over all the instruments? If the culprits were few in number, were they operating close to their range limits when the malfunctions occurred or were the problems uniformly distributed over the instrument ranges? If the former, then a clear bias arises. Did the malfunctions occur at specific regions in the parameter space such as for high RH, high altitude etc.? In any case, I think that my point is clear without further belaboring the issue. More discussion and justification is necessary here.

2.2 Page 13532, lines 4-24 I found the discussion of the statistical analysis presented here at best confusing. Normally distributed variables will certainly yield more accurate results but strict normality is NOT a necessary prerequisite for any statistical analysis of which I am aware. (Given that the authors are taking 10 second averages ab initio – which will tend to normalize the resultant variables – I would not think that the distributions were all that skewed.) The citations given to support this very surprising assertion are all in German, which I unfortunately do not read, and perhaps I am simply misinterpreting the authors here. Certainly one would want a population distribution that was not TOO far off normal or the analysis tests usually employed would be inaccurate (for example, differentiation of cluster means).BUT, strict normality is NOT necessary. For a relevant example, there have been extant for some time now, cluster algorithms that do not require normally distributed variables at all (e.g., Sugar and Gareth, J. Amer. Statist. Assoc., 98, 750-763, 2003). Possibly the SPSS algorithm does require this (I am not familiar with it) but if this is the case, say this, not that the technique in general requires this. The next bit of confusion concerns the transformation of the variables. As the text reads, it appears as if some variables were transformed to render them more normally distributed and some were not. If so, then a very definitely inhomogeneous data set has been created and the variance structure has been distorted. The so-called centering and re-scaling the authors invoke then becomes quite important. There is, let me hasten to add, nothing at all wrong with doing these things in any case, it is simply, despite the claim, not essential in the general case. For example, rescaling variables to an internal normalization parameter does get rid of biasing due to different unit scales or even widely different concentration means but there can be instances when this is actually beneficial (it is like weighting). In short, while I see nothing obviously wrong with what the authors have done, I did have to read over the text a number of times to come to this conclusion. The authors should re-write this section to more clearly differentiate between what they did and what is actually necessary. They are not the same thing.

2.3 Page 13533, lines 8-9 PCA factor extraction CAN be done to minimize correlations between factors – e.g., the Varimax rotation – but, once again, this is not the only option. Other matrix rotations such as Quartimax (which sacrifices orthogonality in favor

of minimizing the number of factors need to explain the variables) are also possible. Once again, the authors should clarify what they have done as opposed to what could be done.

2.4 Page 13541, lines 7-16 The Ukraine (and much of the area to the east as well) does indeed commonly have a burning maximum in the late summer/fall but the main maximum is usually in the spring. This can in fact be seen in the MODIS fire product to which the authors refer for the period in question. In 2006, while there was a maximum in August it was appreciably less than that in April/early May. In 2007, the spring and late summer maxima were about equal and there was a maximum in the spring of 2008 as well. More to the point, do the CH3CN concentrations reflect this temporal trend? Are the BL cluster points for these seasons significantly higher than those for the winter and mid summer seasons? As presented, the analysis seems incomplete.

2.5 Page 13542, lines 4-19 There is some ambiguity as to what the authors are actually saying about where some of the observed particles are being formed. The authors (p. 13533) are supposedly talking about “particle origins” which to me means where they are formed – or nucleated. They point out here that the nucleation mode (or at least the N4-12 concentration, which is indicative of recent new particle formation but certainly does not directly correspond to new particles) is highest in the boundary layer cluster but does this mean that these particles are actually nucleated in the BL and then transported to the upper troposphere? This would be consistent with the geographic location of the BL clusters, which occur preferentially in areas where strong convection or vertical transport might be expected (see figure 5, for example) and in fact these regions have been so identified by the authors. . On the other hand, the vertical transport could simply be moving new particle precursors into the upper troposphere where the new particle formation actually takes place as per, for example, Clarke and Kapustin (J. Atmos. Sci., 59, 363-382, 1992). So which is it? This should at least be discussed.

Interactive comment on Atmos. Chem. Phys. Discuss., 9, 13523, 2009.