Interactive comment on "Urban aerosol size distributions over the Mediterranean city of Barcelona, NE Spain" by M. Dall'Osto et al.

Reply in Italic

Anonymous Referee #1

General Comments

The manuscript is focused on an analysis using K-means clustering and PMF of the size distribution measured in Barcelona. The different aspects are discussed in details and interesting results from the comparison of PMF and K-means clutering are obtained. However, there are a few points that need further discussion (see my specific comments) and some technical corrections are needed. Therefore I believe that the paper merit publication after a moderate revision that take into account the following comments.

We thank reviewer one for considering this manuscript for ACP and for asking only for minor revision.

Specific Comments

Page 16465. Lines 12-14. It is discussed the possibility the enhancement in concentrations in the afternoon, shown in Figure 2 could be due to a photochemical nucleation event. However, this enhancement is visible only on relatively large particles (larger than 20-30 nm). If it was a nucleation I would expect the increase to start at lower size and be visible from the first stage of the DMPS (13 nm). Could traffic (or another anthropic source) be the cause for this enhancement?

Yes, we agree with this comment on the fact that the nucleation events should be more visible on the lower sizes bins (13nm). By using particle number counters (CPCs), Reche et al. (2011) and Dall'Osto et al (2012, this issue) have shown that the increase of particle number concentrations in the afternoon is mainly due to nucleation events. The reason why this is not well seen in Figure 2 is the long time averages (yearly in Figure 2b). However - relative to the other seasons - a shift in the lower size bins can be seen during summer time (Figure 2a). Additionally, this is later on seen in the clustering classification. The reason why we do not see a lower mode in Figure 2b is the low time resolution used in the figure. When using the clustering techniques, we clearly see a maximum in the smallest sampled size bins (see clusters 6 and less extent cluster 8). Figure 2 is indeed a limitation of presenting a year of data, that is why clustering techniques throw light in explaining better the aerosol sampled population.

Page 16468. Lines 23-25. The sentence "The position of the clusters to the bottom left of Fig. 6 is consistent with this origin, although the relatively low values of N13−50 are surprising, but consistent with the low frequency of nucleation events." It is not clear to me how the frequency of these events is actually related to their position in the graphs of

Figure 6. Even if the frequency is low the increase of particles in the size range 13-50 nm should be relevant. Could the authors discuss in more detail this aspects?

We discuss this in more details and we incorporate the text in the manuscript. In other words, this cluster group is associated with the lowest pollutants (NOx) values, supporting clean conditions (low NOx, low PM and low condensation sink favouring nucleation events). We believe there are two contrasting aspects here, first the higher wind speed favouring dispersion (hence low particle number concentrations) second the low condensation sink (hence favouring nucleation events) resulting overall in lower size particle diameters although with overall not large particle number concentrations.

Page 16472. Lines 1-3. It is reported significant relationships between factor 1 and factor 2 and between factor 4 and factor 5. Probably it would be better to quantify the meaning of significant relationships indicating the probability level given that the determination coefficients are relatively low especially between factor 4 and factor 5. On table 2 only the concentration in the size range 13-50 nm is reported. If I understand well this is only for display but all the available range (i.e. 13-800 nm) is used in the clustering. I was wondering why showing only this range and not three ranges associated with typical nucleation, Aitken and accumulation modes?

We answer to the two points raised in this paragraph:

1. We refer to statistical correlation of the temporal trends of the different PMF factors, and they are valid at (p-value<0.05). Text included.

2. A main objective of the study was to apportion the UF particle concentrations, as described in the introduction "Ultrafine particles (UF) are defined as those with diameters smaller than 0.1 μm (100 nm), they make the highest contribution to the total particle number concentrations". We modified Figure 6 and also added the ranges of Ultrafine and total particle number concentrations in table 2 (N<100, N>100 and Ntot).

On Supporting Information material. Page 2. In the discussion relative to Cluster 2 three modes are reported, however, in Figure SI1b only two modes are included in the fit. Authors should modify the text or include all the modes in the Figure.

We thank reviewer one for well spotting the mistake in the text, modified accordingly right values in Figure SI1b.

On Supporting Information material. Page 6. At the end of the discussion relative to Cluster 8 it is reported ". . .with growth events as discussed in the next section." However, it is not clear what section are referring because the only thing after this is the discussion of another cluster

Sentence deleted. Cluster 6 and 8 are attributed to growth of nucleation events.

Technical corrections

Page 16468. Line 13. In the sentence ". . ..solar radiation associated with of these two clusters" please remove "of".

Modified

Page 16471. Line 15. Given the meaning of the sentence it would probably be better to substitute the symbol N<13-800 with N<13.
Correct, modified

Anonymous Referee #2

This work discusses about the source and time variability of atmospheric aerosol particle number concentrations in Barcelona. Thanks to the extended dataset of hourly averaged size distributions two different techniques (K-means clustering and Positive Matrix Factorization) have been applied, providing interesting piece of information. The paper is clearly written and properly supported by summary Tables and Figures; supplementary material provided online is useful for a complete understanding of the results, as well. However, there are a few points that need some explanation, Overall I believe that the paper is worth publication after minor revisions that take into account the following comments.

Page 16462, line 7: "a software programmed at the JRC". I presume that JRC is the EU Joint Research Centre: maybe it'd better to avoid the acronym. Page 16462, lines 7-8: "scans of seven minutes each" Does it mean that you have a complete size distribution every 7 minutes?

1. *Yes, it is the Joint Research Centre – modified*
2. *Yes, modified*

Page 16463, line 6: "with DIGITEL PM1025 inlets". I suggest using "PM10 and PM2.5 inlets"

Modified

Page 16463, Section 2.4. If I correctly understand PM mass concentrations were derived as follows: first, 1-hour PM mass readings from Grimm 1108 instrument have been averaged to obtain daily mean values; then, these daily values have been compared with concurrent gravimetric data from Digitel instrument; finally, the C-factor in the Grimm instrument was determined based on best data fitting. I guess that this procedure (if I got the point, of course!) was separately repeated for PM10 and PM2.5; the question is how did you calibrate PM1 without PM1 gravimetric data?

Modified – we used PM_1 mass readings too, text modified accordingly. We apologise for missing the PM_1 number mass.

Page 16464, lines 8-9: For the reader's ease it would be useful to provide some information (and reference) on the meaning of the 0.20 values obtained for the Hopkins Index.

We modified the text and added additional references, where we already reported detailed information on K-mean clustering, which we would only repeat here (Beddows et al., 2009; Dall'Osto et al., 2011). The hourly DMPS size distributions were subsequently normalised by their vector-length and cluster analysed (Beddows et al., 2009). The choice of K-means clustering was made from a selection of the partitional cluster packages (Beddows et al., 2009). The mathematical details of the method presented are available elsewhere (Beddows et al., 2009; Dall'Osto et al., 2011b). Briefly, K-means method aims to minimize the sum of squared distances between all points and the cluster centre. In order to choose the optimum number of clusters the Dunn-Index (DI) was used, which aims to identify dense and well-separated clusters. DI is defined as the ratio between the minimal intercluster distance to maximal intra-cluster distance. In other words, for Dunn's index we wanted to find the clustering which maximizes this index. The use of cluster analysis was justified in this work using a Cluster Tendency test, providing a calculated a Hopkins Index of 0.20 and implying the presence of structures in the form of cluster in a dataset (Beddows et al., 2009). The Dunn-Index for the results of the K-means analysis for different cluster numbers showed a clear maximum for 14 clusters. By carefully looking at the clusters, these were reduced to 9 as the difference among some of them was minimal. It is important to note that the different aerosol size distribution clusters were merged not only upon their similar size distributions among each other but also by considering strong correlations with other physical and chemical parameters obtained with other instruments.

Page 16464, Section 2.7.1: "were subsequently normalised by their vector-length". Does it mean that size resolved concentration data have been normalised for the total number concentration? It is not clear whether for clustering purpose only DMPS data have been used or also meteorological and criteria pollutant data have been used altogether.

Only DMPS data were used in the K-means clustering analysis, added in the expanded section. It does not mean that size resolved concentration data have been normalised for the total number concentration. Only the SMPS data were normalised and clustered. Normalisation by vector-length means that for each SMPS spectrum each dN/dLogDp value is divided by the root sum of the square of all the dN/dLogDp values. This means that when you calculate the vector length of each normalised SMPS spectrum, it is equal to 1.

eg for a vector [x,y] it becomes [x/sqrt(xx +yy),y/sqrt(xx +yy)] when normalised by the vector length. Its length then becomes
*sqrt(x/sqrt(xx +yy)*x/sqrt(xx +yy) + y/sqrt(xx +yy)]*y/sqrt(xx +yy)]) =*
sqrt(xx/(xx +yy) + yy/(xx +yy)]) = sqrt((xx + yy)/(xx + yy)) = 1

Page 16464, lines 10-13: “The Dunn-Index for the results of the K-means analysis for different cluster numbers showed a clear maximum for 14 clusters. By carefully looking at the clusters, these were reduced to 9 as the difference among some of them was minimal.” The 9-cluster solution the authors present in the result section looks reasonable; however, I find a some mismatch in the abovementioned sentence while stating “clear maximum” and “minimal difference”. What’s the Dunn-Index for a 9-cluster solution?

We modified the text accordingly. The DI for the 9 cluster solution was slightly lower than the 14 clusters solution (about 15%). We looked at the 14 clusters solution (highest DI index) and we merged some of the clusters (down to 9 in total) as some of the differences among them were minimal.

Page 16464, line 25: “. . .only 5000 can be cluster analysed at a time..” Please check this sentence. Probably, I did not get the point. I was expecting that the PMF was applied to each single size distribution, regardless for the results of the cluster analysis.

As stated in Harrison et al. (2011), “Using a PC running Windows XP with 4GB of RAM available, only 5000 can be cluster analysed at a time and many datasets exceed this limit.” To address this problem, 5000 spectra are randomly selected from the dataset and used in the cluster analysis. Out of the 5345 spectra used in this dataset (as reported at beginning of section 2.7), we used only 5000 (randomly selected – 94%).

Page 16465, lines 13-15: “A less distinct enhancement is also seen in the afternoon, likely to be due to the photochemical nucleation events occurring during summer time”. To better support this latter sentence I suggest to provide Figures for the temporal profiles on seasonal basis (cold and warm season). Furthermore, an enhancement apparently occurs for particles in the 30-40 nm range too; have you any clue for this behaviour?

Pey et al. (2008, 2009) have already reported the temporal profiles of the different size-resolved particle number concentrations of this dataset, which we feel it is not necessary to report again in this study. The objective of this study is to report unique aerosol size distributions and relative unique temporal profiles, not only size-resolved particle number concentrations. We modified the text and added such references.

Page 16465, line 20: “both high volume (V) and particle number (N) distributions”.
Actually, table 2 reports particle mass concentration data and not particle volume data. I suggest replacing “volume” with mass”

Modified

Page 16467, line 20: “They lie towards the top right of the N15−50 vs. NOx plot”. It should be N13−50

Modified. We also used the better classification N13-100 as reported in the introction for ultrafine particles (<100nm).

Page 16470, line 10: “observing the onset of much larger increases of Q.” Q is not defined: I suggest avoiding the symbol and explaining what Q stands for.

OK, removed and reference added with the mathematical details (Harrison et al., 2011)

Page 16472, lines 23-27: In table 5 the authors compare the relative contributions to the three modes obtained by PMF and by lognormal fitting of k-mean clusters size distributions. Actually, for these clusters the authors could also provide the measured relative contributions for the three size intervals.

In Figure SI 1 we show the log normal fitting of the 9 aerosol size distributions obtained by K-mean clustering. We report the peak maximum of the nucleation, aitken and accumulation mode in Table 4, whose ranges are 18-36 nm, 41-90 nm and 103-315 nm, respectively. However, it should be kept in mind that these are the peak maximum and not the actual size interval, as of course as seen in Figure SI1 there are overlaps in the shapes of the fittings. Text added.

Page 16472-16473: I suggest inserting a comment concerning the results of this study concerning the sources affecting particle number concentrations with those obtained by the cited work of Pey et al., 2009, where 7 emission sources were recognized to have influence on size-selected particle number concentrations.

Sentence and text added in line with the study of Pey et al., 2009.

Page 16482, Table 2: Why do you present N13-50 only? Particles below 100 nm are usually considered as UF, why not presenting more data for the particle number concentration? (for instance: N13-50, N50-100, N100-800 or N13-50, N13-100, N13-800)

Modified. We modified Table 2 and we added the N13-100, N100-800 and N13-800 as also suggested by reviewer 1. We also modified Figure 6 in the better context of both the introduction and the discussion sections when referring to ultrafine particles (<100nm).

Page 16492, Figure 7: I suggest to keep the same scale for the Score axis. For the reader’s ease, I suggest to explain what the wind roses report.

We tried to do so but the current scale better visually explain the wind roses.