Reply to referee #1 on behalf of all co-authors

General Comments

The MS mainly deals with the evaluation of SMPS data sets for 5 measurement sites by k-means clustering analysis. The idea of the study is good, its goals are relevant, timely and of interest for the international scientific community in the field. Unfortunately, the work has not been performed on a scientific level that merits the high standard of the ACP, and it contains several confusions and mistakes.

Response: We thank the referee for believing that the idea of the study is timely and scientifically relevant, and we regret that he feels the manuscript does not meet ACP standards. We have worked a lot to prepare a completely revised manuscript and we believed that we took into account all the specific issues raised by the referee. Please see details below.

1. Selection of the measurement sites and corresponding time intervals is not justified. Brisbane does not belong to the Mediterranean climate zone as shown in Fig. 1, which is in contradiction with the title. (The source of Fig. 1 is not given and the type of the climate classification system is not mentioned.) Los Angeles can not be accepted as well since a 3-month long data set is only available for it, which A) does not cover one full year, and therefore the seasonality in the nucleation frequency (which is obvious in many areas) is disregarded, and B) it is also much shorter than for the other sites, thus 1 or 2 years. Furthermore, the measurements in Rome were performed at a regional background site (p. 6), which is again in contrast with the title, and questions the representative character of the conclusions for urban areas.

R: We agree that the title was not completely in accordance with the selection of the measurement sites and their characteristics. Therefore, the title has been changed to "Frequency of nucleation events in high insolation urban environments". We modified completely the manuscript to focus mainly in the cities of Barcelona, Madrid and Brisbane, which are all located in urban environments with high insolation. Data from Rome and Los Angeles are used only to complement the results and give supportive examples that these regions may also undergo the same UFP processes than the core cities, but are not longer the main focus of the paper in accordance with the comments from the referee. Although the data for Los Angeles are scarce in quantity, the photochemically driven nucleation processes have been previously documented (Hudda et al., 2010) in the study area, being very intense during the warmer months. We believe it is important to show this phenomenon is occurring in such urban areas. Although Rome is considered a regional background site according to EMEP, it is regularly impacted by the pollution plumes from the city centre of Rome, and the detection of grown particles that nucleated downwind from that area and were advected to the sampling site has already been reported by Costabile et al. (2010). But, as suggested, we took out these sites from the core interpretations of the paper and are used only to exemplify processes.

Fig. 1 has been changed, its source acknowledged and the climatic classification system (Köppen) has been mentioned in the text.

2. It is generally and well accepted that frequency of the new particle formation event is determined on a daily basis, and that it represents the number of nucleation days with respect to all/relevant days on a certain time scale (week, month or year). The title of the paper is misleading not only because the concept of the frequency is completely different here, but - more importantly - since it can not be related to the nucleation event itself (see also comment 3). Instead, it expresses the time share of the particle growth process. At the same time, it is the end of the growth process that is difficult to determine in urban environments due to, for instance, substantial emissions, and therefore, the frequency concept suggested here is doubted.

R: The reviewer is right pointing out that frequency is generally attributed to the percentage of days on which nucleation events occur. Given that our measurements have an hourly resolution, we provide the frequency accordingly. There are several papers reporting the daily occurrence of nucleation events (Yoon et al., 2006 and 2007; Salma et al., 2014), hourly occurrence (Dall'Osto et al., 2011, 2012, 2013) or periods longer than 24 hours (O'Dowd et al., 2010). Moreover, in previous studies that have applied the k-Means clustering technique it is common to report the percentage of time (hours, usually) each cluster represents (see Dall'Osto et al., 2012; Sabaliauskas et al., 2013; Brines et al., 2014; Beddows et al., 2014; Salimi et al., 2014).

It is true that many processes and sources may affect urban areas, which might be difficult to separate. Therefore, simultaneous gaseous pollutants concentrations and meteorological parameters were recorded in order to attempt a realistic interpretation of the results. Moreover, the k-Means analysis has been reported to be a very strong statistical tool to apply on size distribution data, which highly simplifies its analysis. This technique has been compared to other statistical techniques and has been found to be the most adequate for such analysis (see Beddows et al., 2009; Salimi et al., 2014). It has been successfully applied to large data matrices containing large data sets and from different sites (Beddows et al., 2014). The method itself is quite robust in separating the most different clusters while keeping the cluster number to the minimum.

To account for the suggestions of the referee and complement our study, the percentage of nucleation days in each city has been calculated and added to the discussion. Moreover, the nucleation days were classified regarding the uninterrupted number of hours the Nucleation cluster prevailed at each of the main cities. The following table was included and discussed in the manuscript:

Table 4: Percentage of days with nucleation events at the main cities BCN, MAD and BNE, and the uninterrupted time prevalence of these events.

City	1 h or more	2 h or more	3 h or more	4 h or more
Barcelona	67%	54%	43%	28%
Madrid	69%	58%	41%	30%
Brisbane	67%	53%	37%	27%

3. The lower diameter measurement limits (between 10.2 and 17.5 nm) and the corresponding measurement diameter interval make the evaluation of the atmospheric nucleation events rather difficult in particular in cities since the most valuable diameter range, namely the interval below 10 nm is completely missing. As a consequence, the authors show a contour plot in Fig. 5 for Rome as a nucleation event although there is no indication of the nucleation mode (below 10-20 nm), and the elevated concentrations only appear above 20 nm, which is typical for emissions. This all questions if there was atmospheric nucleation at all that day. Such an unusual atmospheric event can not be classified or regarded as nucleation without firm and detailed explanations and evidence. Thus, the conclusions draw at a later stage are also not plausible. Let me also mention here that the heading of Fig. 5 "Daily average SMPS size distributions on a nucleation day" seems to be obscure similar to many other formulations (p. 2: collected size distributions, p. 6: data were sampled, title of section 2.2.1, etc.) in the text, which may indicate that the MS was not elaborated carefully and by all co-authors.

R: It is well accepted that nucleation clusters form at 1-3 nm, however very few research groups in the world have access to technologies required to measure those clusters. Many research papers have reported the growth of nucleated particles with instruments having a low size range of 10 nm. We accept that it might have not been clearly stated in the text that we were

measuring grown nucleated particles, and we have amended it by adding this explanation to the text in the methodology.
We also added to the text that nucleation events were also evaluated visually by inspecting the trends of the SMPS size distributions (i.e., the “banana” or nucleation burst events). Furthermore – i.e. Figure 5 - shows the trends of NOx and the frequency of the nucleation cluster- to check whether the ultrafine plumes were of primary or secondary origin.

In general, NOx concentrations were 30-65% lower during nucleation events than usual (as stated in Fig 5 legend). Namely, in the case of Rome, it has been demonstrated that the air masses transported with the sea-breeze while passing over the city centre of Rome become progressively enriched in photochemical oxidants, and that under high pressure conditions the maximum photochemical production in the Tiber valley occurs between the city limits of Rome and the suburban areas located 15 km from the city centre (Ciccioli et al., 1999). Indeed, the dominant wind direction for the nucleation cluster is SW (morning sea breeze) therefore indicating the transport of nucleated particles downwind of Rome towards the sampling site. It must also be taken into account that this cluster occurs in the afternoon, and it is entirely plausible that the nucleated particles downwind of Rome have grown in size while being transported towards the sampling site by the sea-breeze. This concept is reported in great detail in Dall’Osto et al. (2013), where simultaneous measurements of a growing nucleation event is reported from the urban city centre of Barcelona, growing while transported outside the city, in the afternoon, with minimum amount of BC and NOx. Moreover, Costabile et al. (2010) found a PCA factor (PC2) attributed to an aged nucleation mode with a size peak comprehended between 20.2-33.4 nm, which is in agreement with the size distribution of the Rome Nucleation cluster (23±1 nm). In any case, Rome and Los Angeles data sets are no longer the main focus of the manuscript and the text was modified accordingly.

The title of section 2.2.1 has been changed to “Particle number size distributions”.
The heading of Figure 5 has been changed to “Mean SMPS size distributions on a nucleation day at each selected city...”.
We improved the English usage in the revised version that has been validated again by all authors.

4. It is not described at all how the number of representative clusters between 7 and 15 was reduced “after a careful consideration” (p. 8) to 4-7, which could be a critical issue, and lacks objectivity in its present form.
R: The number of clusters was conservatively chosen using the Dunn Index and the Silhouette Width. The larger the Dunn Index and Silhouette Width, the more compact, well separated and similar were the elements within each cluster (Beddows et al., 2009). Preference was given to a solution with a higher cluster number to reduce the likelihood that any one of the clusters grouped together spectra reflects more than one source. Although we reduce the possibility of losing information by 'over-clustering', it is likely that when comparing the average size distributions - together with the corresponding gaseous pollutants, meteorological parameters, and various temporal trends (daily, weekday-weekend, monthly) - that more than one size distribution may (or even may not) originate from a similar process/source. More often than not, when considering the average size distributions and auxiliary measurements from over-clustered data (e.g. similarly low NO concentrations among the clusters, similar daily trends…), one or more clusters are combined together thus reducing the number of clusters in the final solution. This technique has been applied in several works (Beddows et al., 2009; Dall’Osto et al., 2012; Brines et al., 2014). An explanatory text has been added in the supporting information to clarify this issue.

5. It is unusual to use “traffic-related nucleation mode” (e.g. on p. 9) because the particles which are formed within the source, plume or exhaust are considered as primary particles contained in the Aitken mode in contrast to the nucleated particles contained in the nucleation mode. The present reviewer admits that this can be somewhat more complex (see Robinson et al., Science 315, 1259-1262, 2007) but the usage of such expression without further specific explanations is not tolerable.

R: Vehicular exhausts gases emitted into the atmosphere are cooled and diluted after leaving the tailpipe, leading either to nucleation and new particle formation (in the large nucleation and early Aitken mode, 10-30 nm) or condensation onto pre-existing particles (Aitken and accumulation mode) according to Charron and Harrison (2003). The volatile components of the particles can later evaporate and condensate onto other existent particles, according to Robinson et al. (2007). Therefore the study of the processes affecting traffic particles is rather complex and beyond the main objectives of the paper. But as suggested we clarified and changed the nomenclature to avoid the size mode and secondary/origin process links for the exhaust emissions.

6. Fig. 2 shows particle number size distributions that resulted from k-means clustering. After a detailed examination of many curves, the readers can wonder if resolving the distributions of atmospheric aerosol particles into two modes is indeed realistic, or in other words, whether the clusters T1, T2 and T3 containing 2 modes each at 1)20-40 nm and 70-130 nm, 2) 20-40 nm and 60-90 nm, and 3) 10-20 nm and 50-80 nm are indeed different.

R: The size modes for each curve at each site were obtained by the log-normal fitting method. Moreover the complementary gaseous pollutants concentrations averages, meteorological parameters and temporal trends pointed to some differences that did not enable to merge the traffic clusters. Namely, traffic T1 was related to fresh traffic emissions, thus containing high concentrations of smaller particles in a range of 20-30 nm. T2, on the other hand, was observed in the evening and night, reflecting the possible traffic particle growth due to condensation of volatile gaseous compounds on existing particles and coagulation processes. Regarding T3, the reduction in size of the 20-40 nm of the T1 and T2 clusters may indicate the occurrence of some evaporation processes, as it is detected during daytime. The biggest difference in the spectra could be observed between T3 and the other Traffic clusters. The sources and processes reflected in the 3 Traffic clusters are in accordance with the complex scenario described by Robinson et al. (2007) and merging them into one cluster would lead to a loss of useful information. This same classification and a detailed analysis on the link between T1-3 can be found in Brines et al. (2014). Also, different traffic clusters had been previously reported in Dall’Osto et al. (2011) in a different environment (London, UK).

7. In relation to comment 6, a sensitivity analysis or arguments should have been added on the uncertainty of some results. Without these, it can be questioned whether the frequencies of 6% (section 3.1.3) or 7% (section 3.1.4) are significant or just within the uncertainly limits.

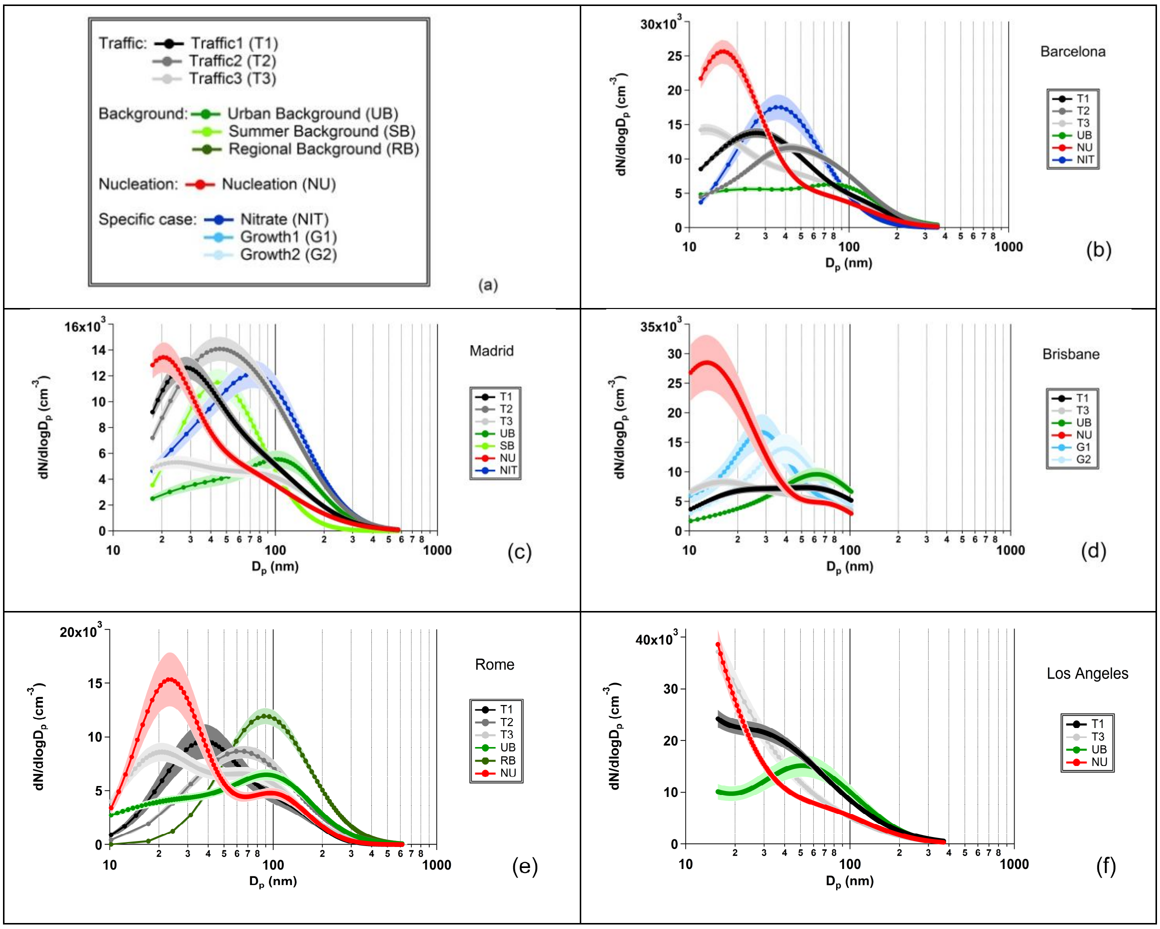
R: The nucleation clusters are unique, showing a very distinctive particle size distribution with very high particle concentration in the nucleation mode, high solar radiation, high ozone concentrations, low black carbon/NOx concentrations, etc. Therefore we believe the nucleation clusters resulting from the k-Means are accurate. The Nitrate cluster reported in Barcelona is site-specific, and although it might contain other particle sources, its temporal and seasonal trends are quite revealing (higher occurrence at night during cold months). Moreover, it has already been reported by Brines et al. (2014) for the city of Barcelona.

We have calculated the 99.99% uncertainty for each cluster size distribution at each city using the confidence limits μ:

$$\mu = mean\ (x) \pm t \frac{\sigma}{\sqrt{n}}$$

where x are the size bin values $dN/dlogD_p$, n is the number of values used in the average, σ is the standard deviation, t is the Student t-value. We approximated the degrees of freedom to ∞, due to the high number of hours contributing to each cluster - in the range of hundreds to thousands. We considered 99.9% of confidence level, obtaining a t-value of 3.291 according to http://www.webassign.net/harrischem/4-02tab.gif. An explanatory text has been added to the manuscript and a more detailed explanation can be found in the supporting information to address this issue.

The uncertainty bands plotted for each cluster show that there is a 99.99% chance than any of the elements within each cluster are miss-classified by the analysis. As can be observed in the modified figures below, the highest uncertainty can be found in the size peaks, although no spectra overlapping is detected at any of the sites. Therefore, the k-Means clustering method is proven to be very robust.



Several comments listed above represent excluding criteria or arguments for rejection, and it is thought that the MS needs such an extensive improvement which can only be realised within the frame of a new submission.

R: As the referee will see, the revised version has been completely modified to account for all suggestions and comments raised. We have put a considerable effort into this revision and we believe that now the revised version has improved a lot the quality of the presentation of the results.

References:

Beddows, D.C.S., Dall'Osto, M. and Harrison, R.M.: Cluster analysis of rural, urban and curbside atmospheric particle size data, Environ. Sci. Technol., 43, 4694-4700, 2009.

Beddows, D. C. S., Dall'Osto, M., Harrison, R. M., Kulmala, M., Asmi, A., Wiedensohler, A., Laj, P., Fjaeraa, A.M., Sellegri, K., Birmili, W., Bukowiecki, N., Weingartner, E., Baltensperger, U., Zdimal, V., Zikova, N., Putaud, J.-P., Marinoni, A., Tunved, P., Hansson, H.-C., Fiebig, M., Kivekäs, N., Swietlicki, E., Lihavainen, H., Asmi, E., Ulevicius, V., Aalto, P. P., Mihalopoulos, N., Kalivitis, N., Kalapov, I., Kiss, G., de Leeuw, G., Henzing, B., O'Dowd, C., Jennings, S. G., Flentje, H., Meinhardt, F., Ries, L., Denier van der Gon, H. A. C., and Visschedijk, A. J. H.: Variations in tropospheric submicron particle size distributions across the European continent 2008–2009, Atmos. Chem. Phys., 14, 4327-4348, doi:10.5194/acp-14-4327-2014, 2014.

Brines, M., Dall'Osto, M., Beddows, D.C.S., Harrison, R. M., and Querol, X.: Simplifying aerosol size distributions modes simultaneously detected at four monitoring sites during SAPUSS, Atmos. Chem. Phys., 14, 2973-2986, doi:10.5194/acp-14-2973-2014, 2014.

Charron, A. and Harrison, R. M.: Primary particle formation from vehicle emissions during exhaust dilution in the roadside atmosphere, Atmos. Environ., 37, 4109–4119, 2003.

Ciccioli, P., Brancaleoni, E., Frattoni, M.: Reactive hydrocarbons in the atmosphere at urban and regional scale. In: Hewitt, N.C. (Ed.), Reactive Hydrocarbons in the Atmosphere. Academic Press, pp. 159-207, 1999.

Costabile, F., Amoroso, A. and Wang, F.: Sub-µm particle size distributions in a suburban Mediterranean area. Aerosol populations and their possible relationship with HONO mixing ratios, Atmos. Environ., 44, 5258-5268, 2010.

Dall'Osto, M., Thorpe, A., Beddows, D. C. S., Harrison, R. M., Barlow, J. F., Dunbar, T., Williams, P. I., and Coe, H.: Remarkable dynamics of nanoparticles in the urban atmosphere, Atmos. Chem. Phys., 11, 6623–6637, 2011.

Dall'Osto, M., Beddows, D.C.S., Pey, J., Rodriguez, S., Alastuey, A., Harrison, R. M. and X. Querol: Urban aerosol size distributions over the Mediterranean city of Barcelona, NE Spain, Atmos. Chem. Phys., 12, 10693-10707, doi:10.5194/acp-12-10693-2012, 2012.

Dall'Osto, M., Querol, X., Alastuey, A., O'Dowd, C., Harrison, R.M., Wenger, J. and Gómez Moreno, F.J.: On the spatial distribution and evolution of ultrafine particles in Barcelona, Atmos. Chem. Phys., 13, 741-759, doi:10.5194/acp-13-741-2013, 2013.

Hudda, N., Cheung, K., Moore, K.F. and Sioutas, C.: Inter-community variability in total particle number concentrations in the easter Los Angeles air basin, Atmos. Chem. Phys., 10, 11385-11399, 2010.

O'Dowd, C., C. Monahan and M. Dall'Osto: On the occurrence of open ocean particle production and growth events, Geophys. Res. Lett., 37, L19805, doi:10.1029/2010GL044679, 2010.

Sabaliauskas, K., Jeong, C.-H., Yao, X., Jun, Y.-S. and Evans, G.: Cluster analysis of roadside ultrafine particle size distributions, Atmos. Environ., 70, 64-74, 2013.

Salimi, F., Ristovski, Z., Mazaheri, M., Laiman, R., Crilley, L. R., He, C., Clifford, S., and Morawska, L.: Assessment and application of clustering techniques to atmospheric particle number size distribution for the purpose of source apportionment, Atmos. Chem. Phys., 14, 11883-11892, doi:10.5194/acp-14-11883-2014, 2014.

Salma, I., Borsós, T., Németh, Z., Weidinger, T., Aalto, P., and Kulmala, M.: Comparative study of ultrafine atmospheric aerosol within a city, Atmos. Environ., 92, 154-161, 2014.

Yoon, Y. J., C. D. O'Dowd, S. G. Jennings, and S. H. Lee: Statistical characteristics and predictability of particle formation events at Mace Head, J. Geophys. Res., 111, D13204, doi:10.1029/2005JD006284, 2006.

Yoon, Y. J., Beburnis, D., Cavalli, F., Jourdan, O., Putaud, J.P., Facchini, M.C., Decesari, S., Fuzzi, S., Sellegri, K., Jennings, S.G. and O'Dowd, C.D.: Seasonal characteristics of the physicochemical properties of North Atlantic marine atmospheric aerosols, J. Geophys. Res., 112, D04206, doi:10.1029/2005JD007044, 2007.