

Supplement

A Statistical Analysis of North East Atlantic (submicron) Aerosol Size Distributions

M. Dall'Osto^{1*}, C. Monahan¹, R. Greaney¹, D.C.S. Beddows², R. M. Harrison², D. Ceburnis¹ and C. D. O'Dowd¹

(1) School of Physics and Centre for Climate and Air Pollution Studies,
Ryan Institute, National University of Ireland Galway, University Road,
Galway, Ireland

(2) National Centre for Atmospheric Science, School of Geography, Earth &
Environmental Sciences, Division of Environmental Health & Risk Management,
University of Birmingham Edgbaston, Birmingham, B15 2TT, United Kingdom

*now at: Institute for Environmental Assessment and Water Research (IDÉA-CSIC),
Barcelona, Spain (manuel.dalosto@gmail.com)

S1. Calculation reference (Beddows et al. 2009)

K-Means Clustering

In the cluster analysis, a set of H hourly SMPS measurements are presented to the algorithm as a matrix of particle counts $smps_{h,m}$, where row h depicts the hourly measurement and column m depicts one of the M particle size bin mobility diameter. Given that S_k represents the set of spectra in the k^{th} cluster and $\overline{smps_{k,m}}$ represents the mean for the variable m over cluster k , the K-Means method partitions the data into G groups such that the ‘within-cluster sum of squares’ function is minimised using an iterative cycle (function (1)).

$$\sum_{k=1}^G \sum_{h \in S_k} \left\| smps_{h,m} - \overline{smps_{k,m}} \right\|_2^2 \quad (1)$$

Known as the Hartigan and Wong method, this starts with a random initial partition and continually reassigning the SMPS spectra to the clusters based on the similarity between the SMPS spectra and the cluster centres until a convergence criterion is met, i.e. there is not reassignment of any patterns from one cluster to another, or the ‘within-cluster sum of squares’ ceases to decrease significantly after some number of iterations.

The Dunn Index, D

Cluster validation is an evolving field of research which provides various indices, derived from a number of distance based cluster statistics, used to describe and compare cluster results. The distance based measurements include: cluster size (i.e. number of SMPS spectra within each cluster); cluster diameter (maximum separation of the SMPS spectra within the cluster); cluster separation (i.e. minimum distance between SMPS spectra of neighbouring clusters); the average distances within clusters; and the average distances between clusters. In this work the Dunn index (equation 2) was found to be the most useful index (Beddows et al. 2009) which is a function of the ratio of the minimum cluster separation to the maximum cluster diameter defined more precisely as,

$$D = \min_{i=1\dots G} \left\{ \min_{j=i+1\dots G} \left[\frac{d(C_i, C_j)}{\max_{k=1\dots G} \Phi(C_k)} \right] \right\} \quad (2)$$

where

$$d(C_i, C_j) = \min_{h_1 \in C_i, h_2 \in C_j} d(h_1, h_2)$$

is the minimum distance function between two cluster centres C_i & C_j and $\Phi(C_k)$ is the diameter of a cluster, which is defined as,

$$\Phi(C_k) = \max_{h_1, h_2 \in C_k} d(h_1, h_2)$$

From this it can be deduced that the larger the Dunn index the more compact and well separated are the clusters within the space.

S2. Peculiar feature of each K-means SMPS cluster

Aerosol size distribution clusters presented unique features within the same aerosol category, therefore each individual cluster is presented in this S1 section, whereas a discussion on each category is instead presented in section 4 (Discussion).

- Cluster 1 (8.9%): Figure 3a shows size distribution of cluster 1, revealing a strong nucleation mode in the smallest detectable SMPS size bin at about 4-5nm, as well as an Aitken mode (about 60nm) and accumulation mode (at about 200nm). WS and WD were $7.0 \pm 3 \text{ m s}^{-1}$ and $236 \pm 98^\circ$, respectively (Figure 4a-b); temperature of $13.3 \pm 3 \text{ }^\circ\text{C}$ and RH of $75 \pm 12 \%$, respectively (Figure 4c-d); with the highest average temperature of all clusters. This cluster presented the second highest average particle number concentration ($N_{D>3}$ $23,352 \text{ cm}^{-3}$) and also the largest difference between the two CPCs ($N_{D>10} - N_{D>3}$, Table 3).
- Cluster 2 (3.8%): Figure 3a presents a similar distribution to cluster 1, but the nucleation mode is shifted towards larger diameters (6-8nm) whilst both the Aitken and accumulation modes are found similar to cluster 1. WS and WD were $5.7 \pm 3 \text{ m s}^{-1}$ and $219 \pm 96^\circ$ respectively; the temperature was $11.4 \pm 4 \text{ }^\circ\text{C}$ and the RH was $75 \pm 13 \%$. $N_{D>3}$ revealed the highest particle number concentration ($45,574 \text{ cm}^{-3}$) as well as the largest difference $N_{D>3} - N_{D>10}$ ($42,675 \text{ cm}^{-3}$).
- Cluster 3 (8.6%): Figure 3a shows the third clear nucleation size distribution, with a larger nucleation mode at about 9-11nm relative to cluster 1 and cluster 2. WS and WD were $5.7 \pm 3 \text{ m s}^{-1}$ and $193 \pm 96^\circ$ (respectively), the temperature was $11.5 \pm 3 \text{ }^\circ\text{C}$ and the RH was $78 \pm 13 \%$ (respectively) representing similar pattern of cluster 2. This cluster presented similar high particle number concentrations to cluster 1 and cluster 2 and therefore it was merged in the subgroup of particle size distributions attributed to coastal nucleation events (Figure 3a).
- Cluster 4 (12.6%): This cluster (Figure 3b) showed a nucleation mode at about 15nm, significantly larger than the previous 3 clusters attributed to coastal nucleation events. An accumulation mode at about 200nm is also seen. WS and WD were $6.38 \pm 3 \text{ m s}^{-1}$ and $205 \pm 86^\circ$, the temperature was $11.4 \pm 3 \text{ }^\circ\text{C}$ and the RH was $81 \pm 12 \%$. Furthermore, the particle number concentration was found to be

- high ($N_{D>3}$ 13,733 cm^{-3} and $N_{D>10}$ 2,249 cm^{-3}) but the difference is not as high as the ones of clusters 1 to 3.
- Cluster 5 (11.0%): Figure 3b shows a broad mode at about 30nm for this cluster, as well as a smaller accumulation mode at about 200nm. WS and WD were 6.6 ± 3 m s^{-1} and $216\pm 86^\circ$ (respectively), the temperature was 11.6 ± 3 $^\circ\text{C}$ and the RH was 83 ± 12 % (respectively). This cluster was found to be similar to cluster 4 and therefore merged into the same subgroup shown in Figure 3b. Merged with cluster 4, it showed the lowest average $\text{PM}_{2.5}$ concentration (7.8 ± 3 $\mu\text{g m}^{-3}$, see Table 3 column d).
 - Cluster 6 (12.5%): This cluster (Figure 3c) presents a broad Aitken mode at about 40nm and an accumulation mode shifted towards bigger diameter at about 250nm. WS and WD were 6.0 ± 3 m s^{-1} and $184\pm 105^\circ$, the temperature was 9.3 ± 3 $^\circ\text{C}$ and the RH was 84 ± 12 %. Given the broad particle size distribution spanning across the Aitken and the accumulation mode, this cluster is different from the ones previously described and therefore it is presented in Figure 3c.
 - Cluster 7 (15.2%): Figure 3d shows much smaller particle concentration relative to previous clusters, with both the Aitken and accumulation mode shifted towards smaller diameter (30nm and 150nm, respectively). WS and WD were 9.0 ± 4 m s^{-1} and $240\pm 63^\circ$, temperature of 9.1 ± 3 $^\circ\text{C}$ and RH of 84 ± 10 %, with the lowest average temperature among all clusters. It was associated with the most westerly wind detected (Figure 4b, 241 ± 30 $^\circ$) and second strongest WS among all clusters (Figure 4a, 9 ± 4 m s^{-1}). The BC concentration was found the third lowest among all clusters (Table 3 column e). This cluster was also peculiar as presented the lowest $N_{D>10}$ concentration (282 ± 40 cm^{-3}) among all 12 classifications.
 - Cluster 8 (9.0%): This cluster (Figure 3b) was found similar to cluster 5, with modes slightly bigger at about 50nm and 200nm (Aitken and accumulation, respectively). WS and WD were 6.7 ± 3 m s^{-1} and $202\pm 90^\circ$, temperature of 11.5 ± 3 $^\circ\text{C}$ and RH of 86 ± 9 %, also reflecting similar averages of cluster 4 and 5.
 - Cluster 9 (4.9%): Figure 3c shows a broad Aitken mode for this cluster type at about 60nm, and also a broad mode at higher diameter for the accumulation mode

- (250nm). WS and WD were $6.3 \pm 3 \text{ m s}^{-1}$ and $186 \pm 90^\circ$, the temperature was $11.6 \pm 3^\circ$ and the RH was $87 \pm 11 \%$. This cluster presented the second highest BC concentration ($207 \pm 218 \text{ ng m}^{-3}$, Table 3 column e).
- Cluster 10 (2.8%): This cluster shows the highest loading of accumulation mode particle concentration, with two modes peaking at about 100nm and 300nm, respectively (Figure 3c). WS and WD were $5.6 \pm 3 \text{ m s}^{-1}$ and $155 \pm 82^\circ$, respectively whilst the average temperature was $10.4 \pm 3^\circ \text{C}$ and the average RH was $86 \pm 11 \%$. Furthermore, it presented the highest BC concentration (almost two-fold increase over all cluster of $573 \pm 540 \text{ ng m}^{-3}$ - Table 3 column e), and also highest $\text{PM}_{2.5}$ concentration ($16.5 \pm 8 \text{ } \mu\text{g m}^{-3}$ - Table 3 column d).
 - Cluster 11 (3.1%): Figure 3d shows that this cluster exhibited very low particle average particle number size distributions, with two clear modes at about 60nm and 200nm, respectively, showing the typical clean marine bimodal size distribution. WS and WD were $8.4 \pm 3 \text{ m s}^{-1}$ and $229 \pm 53^\circ$ (respectively), whereas the average temperature was found to be $10.9 \pm 3^\circ \text{C}$ and the average RH $86 \pm 10 \%$. The $N_{D>3}$ and $N_{D>10}$ average concentrations were found the second lowest of all clusters ($764 \pm 80 \text{ cm}^{-3}$ and $321 \pm 40 \text{ cm}^{-3}$, respectively). Whilst the CPC was the second lowest, the $\text{PM}_{2.5}$ concentration ($11.5 \pm 9 \text{ } \mu\text{g m}^{-3}$) was found the third highest and likely to be due to coarse sea salt particle contribution (Table 3).
 - Cluster 12 (7.8%): This cluster (Figure 3d) shows a broad Aitken mode merged with the accumulation mode at about 100nm. A second shoulder at about 350nm is also seen, not seen in any of the other 12 clusters. The average WS and WD were $11.4 \pm 3 \text{ m s}^{-1}$ and $237 \pm 53^\circ$, the temperature was $9.0 \pm 3^\circ \text{C}$ and the RH was $83 \pm 10 \%$. Average particle number concentrations were found to be the second ($N_{D>3}$) and third ($N_{D>10}$) lowest of all clusters ($773 \pm 80 \text{ cm}^{-3}$ and $379 \pm 40 \text{ cm}^{-3}$, respectively), whilst $\text{PM}_{2.5}$ concentration ($13.3 \pm 9 \text{ } \mu\text{g m}^{-3}$) was the second higher, reflecting the same feature of cluster 11 with CPC and PM values being anti correlated. In other words, bimodal size distributions associated with the lowest particle number concentrations were found to have the highest values of particle $\text{PM}_{2.5}$ mass.

REFERENCE SUPPORTING INFORMATION

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).