

Method	Summary	No. of clusters
HAC	<ul style="list-style-type: none"> <li>- Does not rely on training data.</li> <li>- The conclusion we make when using the CH index may be incorrect when a large proportion of the particles are from one broad class.</li> <li>- How the data were prepared greatly impacted upon performance.</li> <li>- Particles from different categories were sometimes clustered together, e.g. pollen with fungal.</li> </ul>	Determined using the maximum value of the CH index produced for clusterings containing between 1 and 10 clusters.
DBSCAN	<ul style="list-style-type: none"> <li>- Produced a clustering which contained three distinct clusters each containing primarily one broad class of bioaerosol in the case of one of the data sets.</li> <li>- Data preparation greatly impacted upon performance.</li> <li>- It is not clear at this point whether the values of epsilon and the minimum number of points would be applicable to ambient data.</li> </ul>	Naturally determined by setting epsilon and the minimum number of points required for a neighbourhood.
Gradient boosting	<ul style="list-style-type: none"> <li>- Performance was consistently good irregardless of data preparation provided that a threshold, either 3 or 9 standard deviations, was applied to the fluorescence measurements</li> <li>- Relies on adequate training data being collected and it is not clear at this point whether the data collected will be sufficient.</li> </ul>	Always the same as the number of groups in the training data.