Hierarchical Agglomerative Cluster Analysis Applied to WIBS 5-Dimensional Bioaerosol Data Sets

M.W. Gallagher[1], N.H.Robinson[1,4] , P. H. Kaye[2], and V.E. Foot [3]

[1]Centre for Atmospheric Science, University of Manchester, Simon Bldg, Oxford Rd, Manchester, M13 9PL, UK

[2] Centre for Atmospheric & Instrumentation Research, STRI, University of Hertfordshire, Hatfield, AL10 9AB, UK

[3] DSTL, Porton Down, Salisbury, Wiltshire, SP4 0JQ, UK

[4] *Now with UK Meteorological, Office, Fitzroy Road, Exeter, Devon, EX1 3PB, UK*

Introduction

Primary biological aerosol particle (PBAP) classification requires discrimination of particles various diverse sources which may have wide reaching effects in the atmosphere. In order to predict these effects under future emissions scenarios it is useful to be able to identify ambient PBAP concentration. To date, this has largely been achieved by the use of off-line techniques, which, whilst allowing accurate identification of different aerosols, are labour intensive and have poor time resolution.

To improve on this we have investigated the use of hierarchical agglomerative (HA) cluster analysis applied to single-particle multi-spatial (5-D) datasets comprising optical diameter, particle asymmetry and three induced fluorescence waveband measurements, from two commonly used dual Waveband Integrated Bioaerosol Spectrometers (WIBS), (Kaye et al., 2005). We show that HA cluster analysis, without the need for any *a-priori* assumptions concerning the expected aerosol types, can reduce the level of subjectivity compared to the more standard analysis approaches for multi-parameter aerosol measurements.

Methods

We use two WIBS– a model 3 and a model 4 (Gabey *et al.*, 2011). In both models the single particle elastic scattering intensity (at 633 nm) is used to infer particle optical-equivalent diameter, D_O. A quadrant PMT measures the variation in azimuthal scattering and hence provides a particle asymmetry factor, A_F. This measurement triggers pulses from filtered xenon flash-lamps at 280 nm and 370 nm, designed to excite tryptophan and NAD(P)H molecules within the particle. Fluorescence is measured in two wavelength regimes, FL1 & FL2 providing three fluorescence channels; FL1 & FL2 following the 280 nm excitation and FL2 following the 370 nm excitation. The FL1 and FL2 fluorescence detection regimes overlap spectrally in the WIBS3, but have been separated in the WIBS4.

A software tool (WIBS Analysis Program, WASP) was developed that applies the average-linkage HA-cluster analysis algorithm (Everitt, 1993, Robinson et al. 2012) to WIBS data. Average-linkage defines the two most similar clusters as those with the smallest distance across an *n*-dimensional space, where *n* is the number of particle diagnostics. The distance between two clusters is defined as the average squared Euclidian distance between all possible pairs of particles, or

$$L_{A,B} = \frac{1}{pq} \sum_{i=1}^{p} \sum_{j=1}^{q} \left\| A_i - B_j \right\|^2$$

where $L_{A,B}$ is the cluster distance, A the coordinate vector of cluster A containing p members, and B the vector for cluster B with q members. Physically realistic cluster solutions and particle apportionment is achieved with d-metrics and z-score analysis. This was applied to laboratory test data using NIST non-fluorescent and fluorescence particles. It was then applied to ambient measurements collected during the BEACHON-RoMBAS experiment at an elevated pine forest (Manitou National Forest, Colorado USA).

Conclusions

We successfully demonstrate application of the HA-cluster analysis approach to real-time induced fluorescence bioaerosol data sets. Different diehl cycles for the various bioaerosol classes, associated with fungal spores, bacteria, dust etc., are presented from a forest ecosystem. The different influences of rainfall on different bioaerosol classes is highlighted as an example of the power of the technique. Finally micrometeorological fluxes of different bio-aerosol classes are inferred from multi-height measurements.

Everitt, B. S. (1993). Cluster Analysis (3rd ed.). *New York: John Wiley & Sons.*

Gabey, A. M., Gallagher, M. W., *et al.* (2010). *Atmospheric Chemistry and Physics*, 10(10), 4453-4466.

Kaye, P. H., *et al.* (2005). *Opt. Express*, 13, 3583. 6389.

Robinson, N.H., *et al.*. (2012). *Atmos. Meas. Tech. Discuss.*, 5, 6387–6422.