Deep learning based classification of biological aerosols

Hao Zhang[1], David Topping[1], Ian Crawford[1], Martin Gallagher[1], Man Nin Chan[2], Hing Bun martin Lee[2], Sinan Xing[2], Tsin Hung Ng[2], Amos Tai[2]

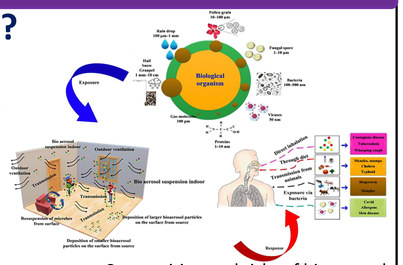
[1]*The University of Manchester, Centre for Atmospheric Science, Manchester, United Kingdom*

[2]*Faculty of Science, The Chinese University of Hong Kong, Hong Kong, China*

1. Research motivation & objectives

Why distinguish different bioaerosols?

- Bioaerosols, derived from biological sources, are a subset of atmospheric aerosols, consisting mainly of viruses, fungal spores, and pollens.
- Different bioaerosols have various effects on human health, agriculture, climate, and other aspects of the human and earth systems.



Composition and risks of bioaerosol
(Gondwana Research 2021, 99: 178-203.)

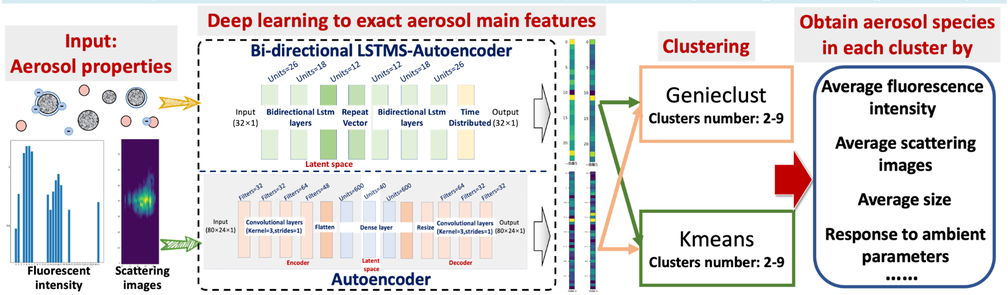
Research Objectives

- Clustering aerosol by deep learning and clustering methods.
- Obtain the species of bioaerosol in different clusters.
- Evaluate the performance of different combinations of two deep learning and two clustering methods in bioaerosol classification.

2. Methodology and data description

Methodology and workflow

- Two deep learning methods (Autoencoder/Bilstms-Autoencoder) were used to get aerosol features, then Kmeans/Genieclust clustered aerosols via aerosol features.
- Clusters species were then identified via their properties (average FL/images, etc.).



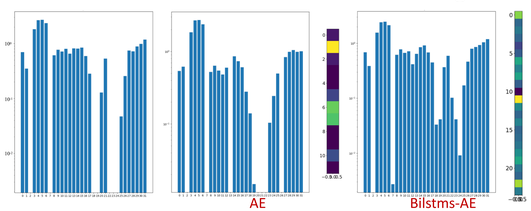
Dataset——by Rapid-E

- Rapid-E was deployed at HKUST supersite from June to November 2018 to monitor bioaerosols.
- Rapid-E could obtain a variety of aerosol information to help distinguish aerosols.

Fluorescent intensity: FL (32 channels)	Scattering image (80×24×1)
Aerosol Size	Aerosol number (17116266)

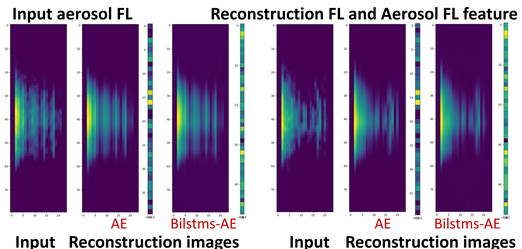
3. Aerosol features from deep learning

- The reconstruction accuracy of Bilstms-Autoencoder (Bilstms-AE) was higher than that of Autoencoder (AE).
- Feature dimension of Bilstms-AE was greater than AE.
- The FL features obtained by deep learning served as the foundation for the aerosols clustering.



Reconstruction Accuracy of FL

78.7% (AE)	
81.9% (Bilstms-AE)	

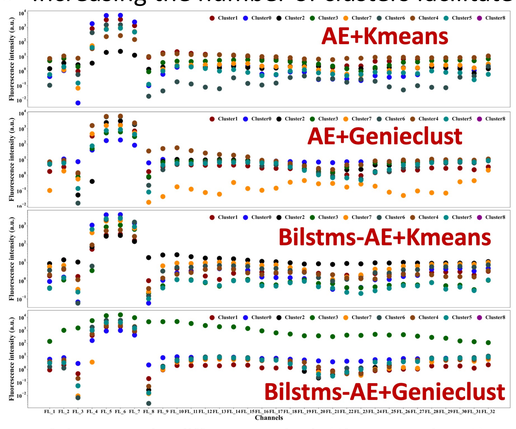


Feature Dimension

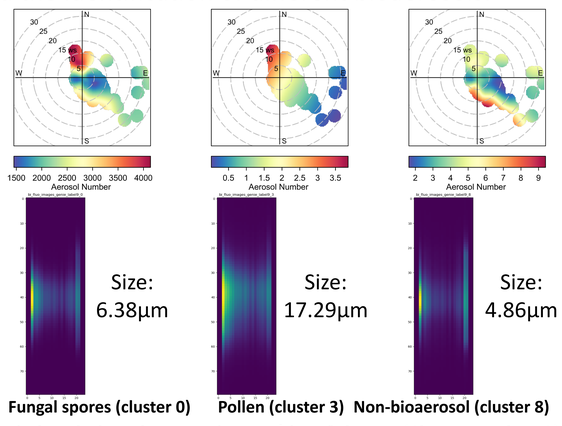
12—FL 40—images (AE)	
24—FL 48—images (Bilstms-AE)	

4. Determine bioaerosol species in each cluster

- Cluster 8 should be non-bioaerosol aerosols (e.g. marine aerosols) as they didn't have FL response and mainly came from the south direction (ocean).
- The Bilstms-AE and Genieclust combination successfully separated rare pollen (cluster 3-its FL pattern was consistent with the experimental data), while no pollen was identified in the other combinations. Other clusters were suspected to be fungal spores.
- Increasing the number of clusters facilitated the identification of different bioaerosols.



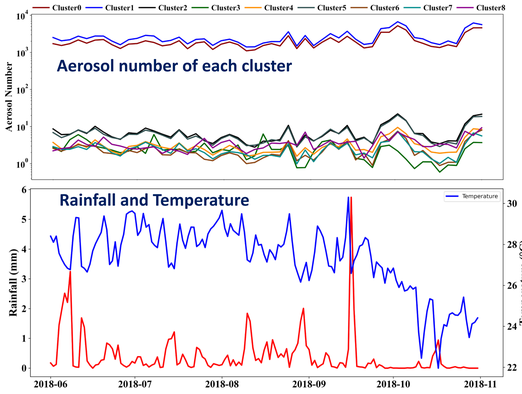
FL of clusters under different methods (Cluster number=9)



Wind analysis and average images/size of clusters (Cluster number=9)

5. Time series variation of bioaerosols

- Fungal spores were the main species.
- Fungal spores increased after rainfall.
- The peaks of pollens and fungal spores didn't exactly coincide.



Time series change of ambient data and aerosol number

- Bilstms-AE + Genieclust demonstrated strong performance in bioaerosol classification.
- Increasing the cluster number was beneficial for bioaerosol classification.
- The bioaerosols in Hong Kong were mainly fungal spores and exhibited sensitivity to temperature and rainfall.