Machine Learning Approaches for Classification of Single Particle Aerosol Mass Spectra

Maria A. Zawadowicz, Brookhaven National Laboratory

with contributions from: José Perez-Chavez (Howard U) and Dan Cziczo (Purdue)

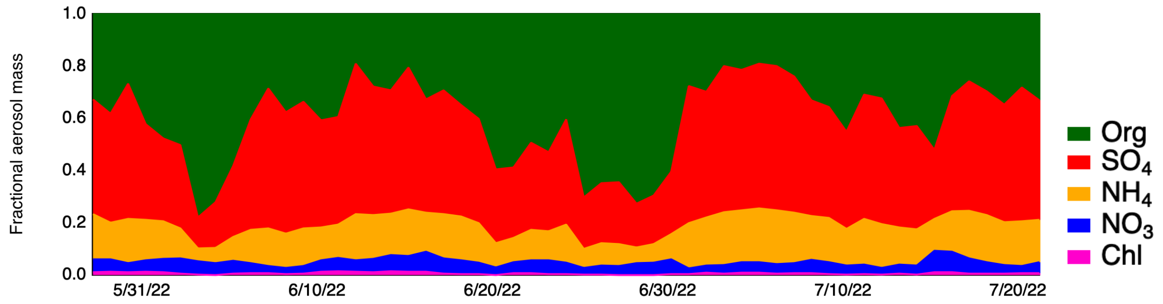
ARM ASR PI Meeting, August 8, 2023

Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

@BrookhavenLab

Aerosol chemistry measurements are an essential aerosol property



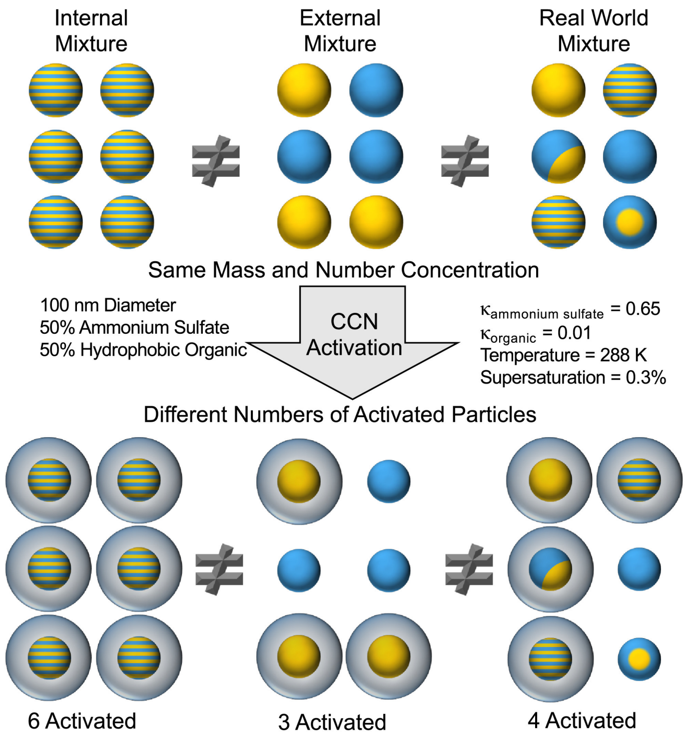
Aerosol composition is a frequently requested ARM capability.

The currently used technique, the ACSM only provides bulk non-refractory submicron aerosol composition

Brookhaven National Laboratory

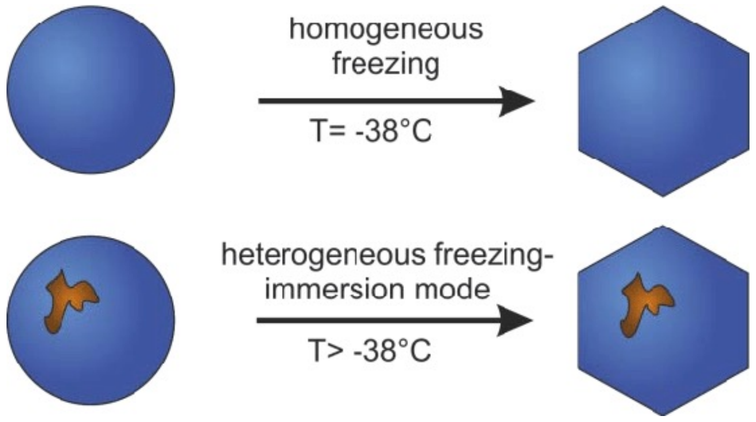
Why single particle?

CCN activation



Riemer, et al. (2019)

Heterogeneous ice nucleation



Many atmospheric processes, especially within the realm of aerosol-cloud interactions, depend strongly on properties of individual aerosol particles

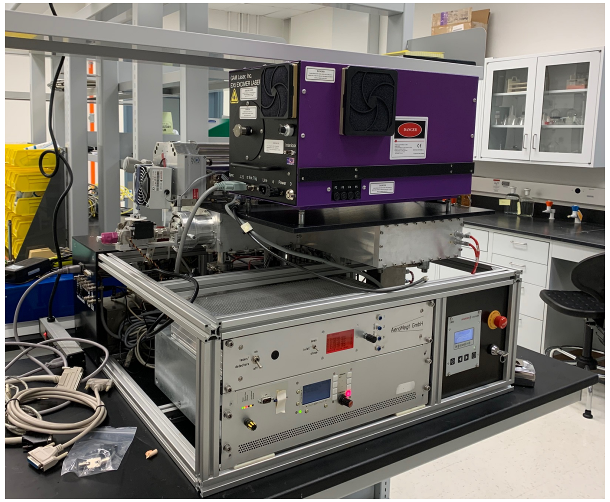
<u>"Bulk parametrizations" can result in model inaccuracies</u>

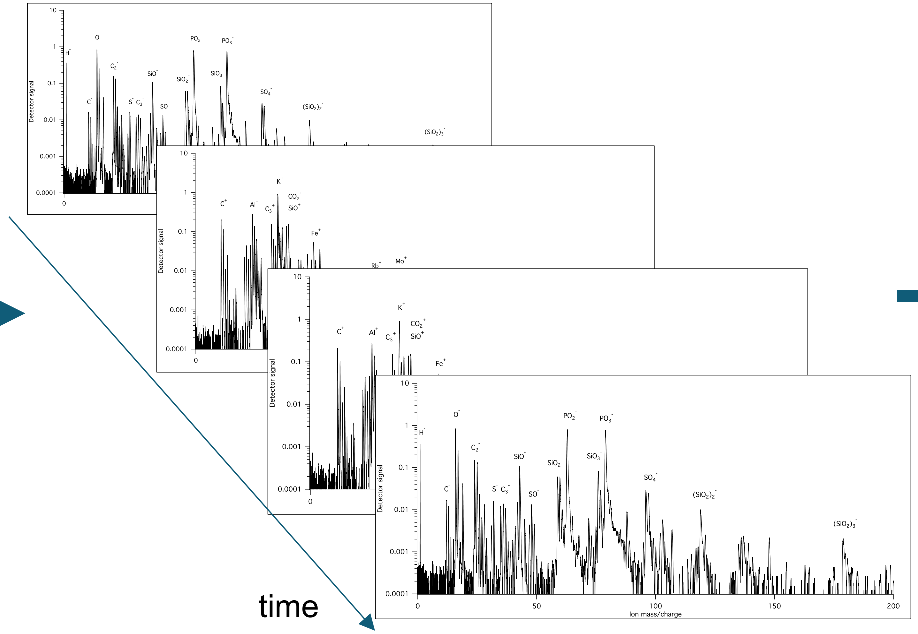
For example, Zaveri, et al. (2010) found that averaging the CCN composition resulted in overestimates of 40% in CCN efficiency compared to a particle-resolving model.

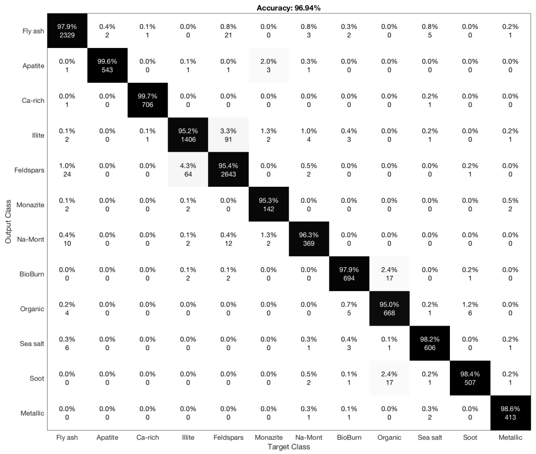
Brookhaven National Laboratory

Single-particle mass spectrometry at BNL

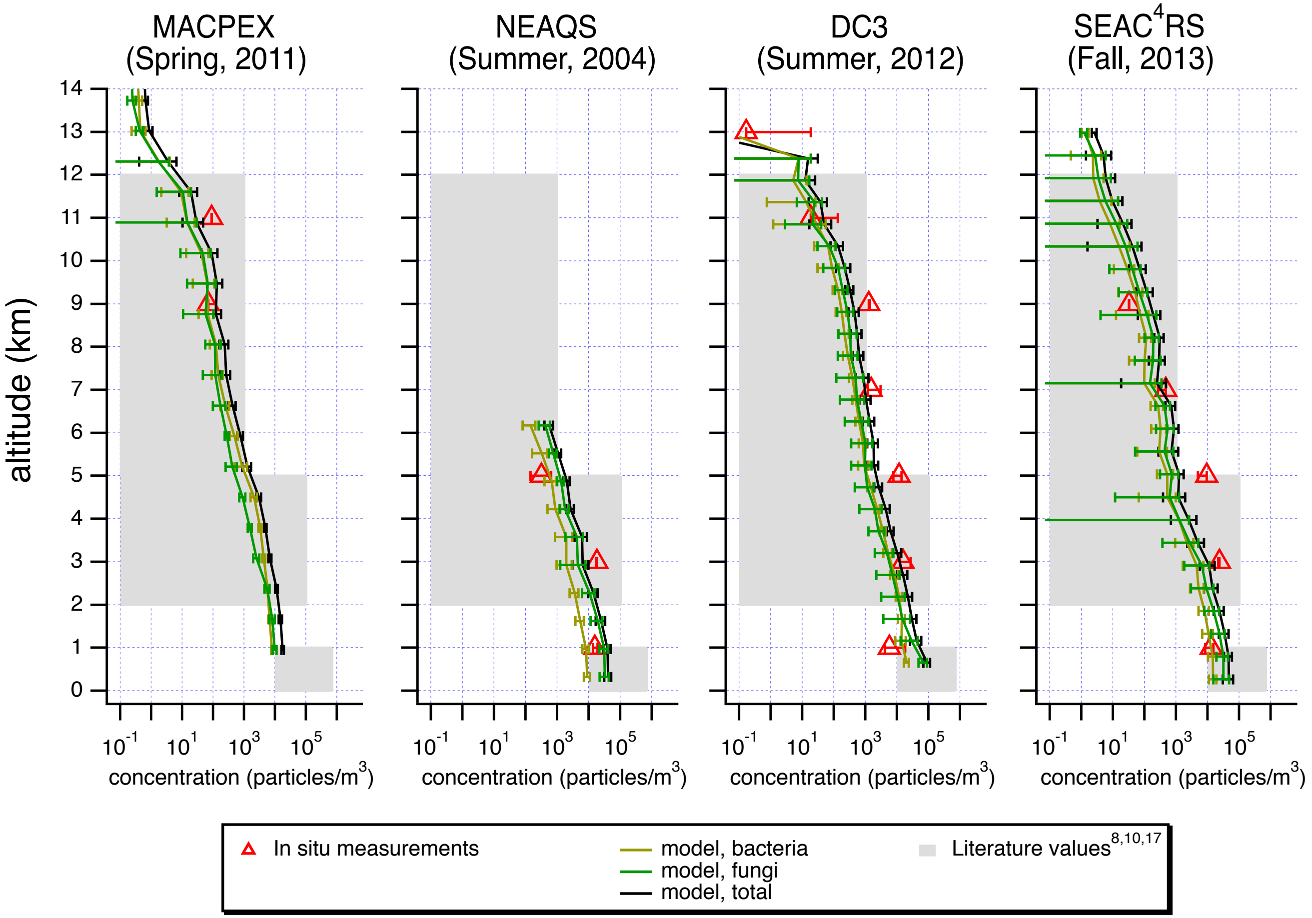
- Aims for the most comprehensive characterization of single particle composition to date through both hardware improvements and computational advances



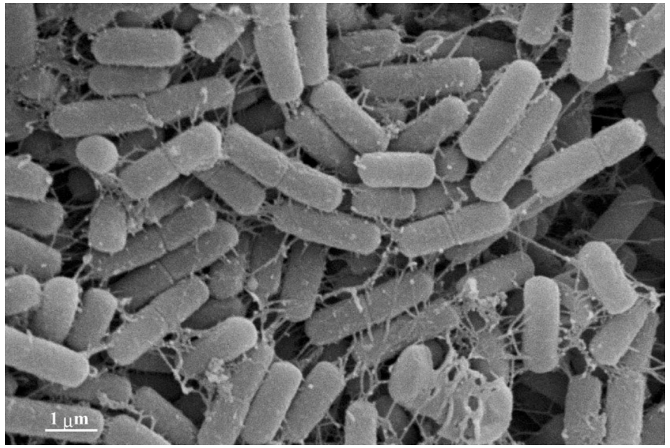




Why mass spectrometry?



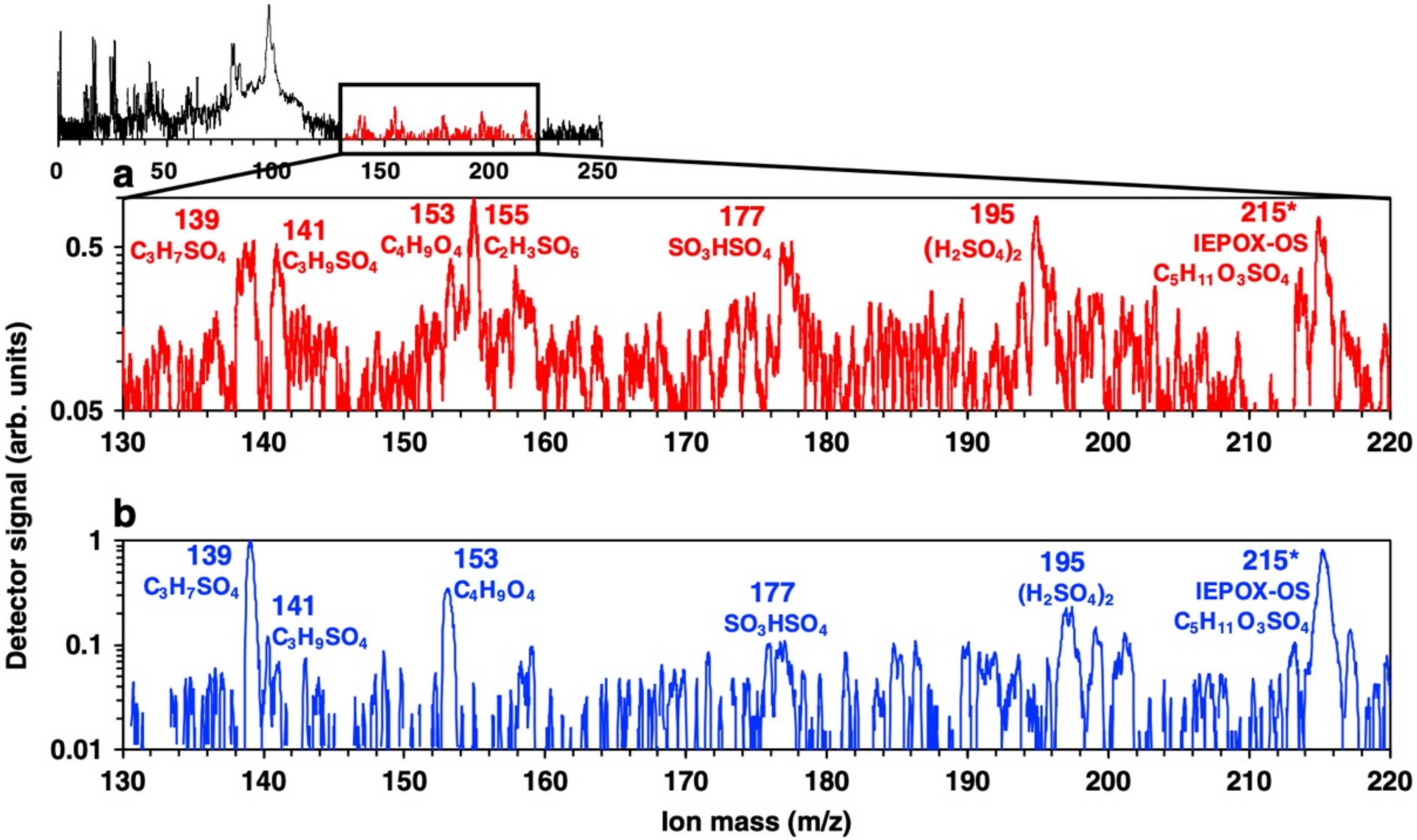
Because of its mode of ionization and single particle resolution, SPMS is sensitive to compositions that other aerosol measurements often miss, such as bioaerosol



Zawadowicz et al. (2019), ACP

Brookhaven National Laboratory

Applications: ice and droplet residual composition

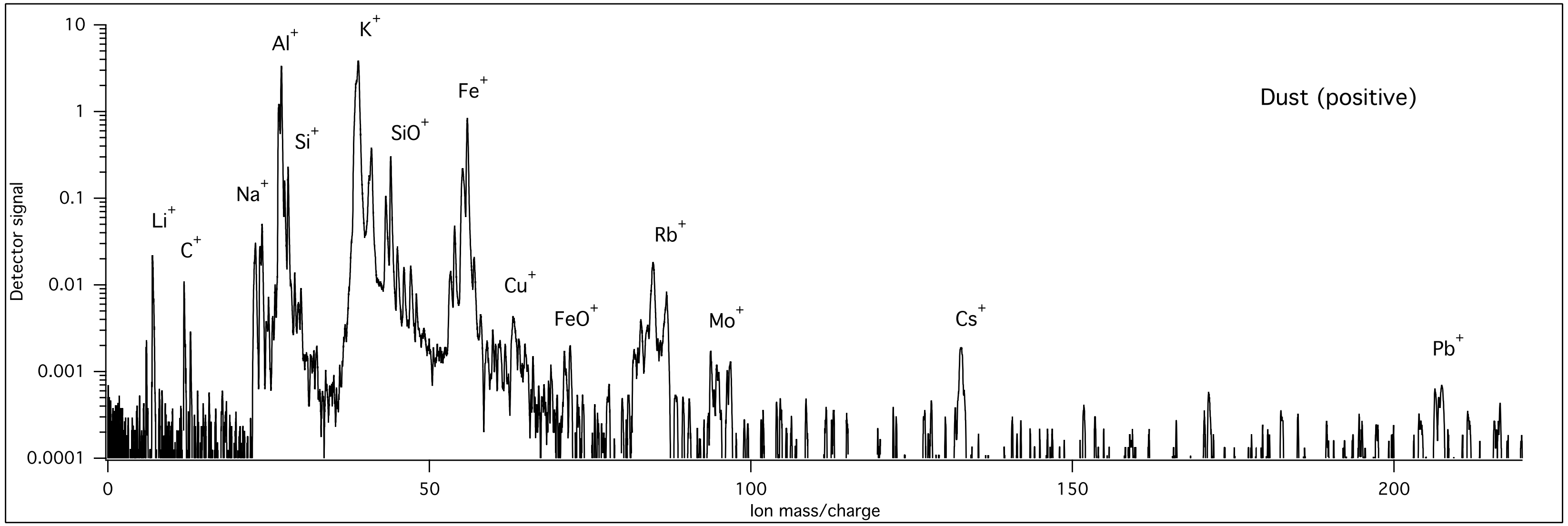


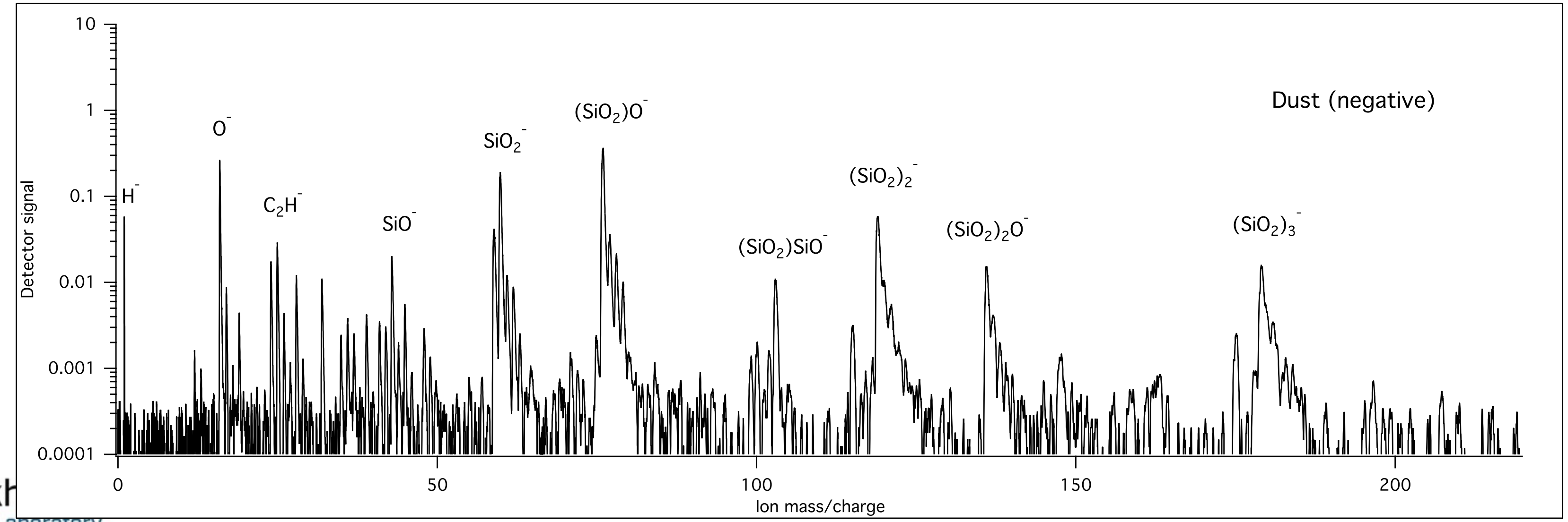
SPMS allows to probe the composition of single, evaporated ice crystals (ice residuals)

This study found a fingerprint associated with a secondary organic product of isoprene chemistry in ice residuals

Wolf et al. (2020), Nat. Commun.

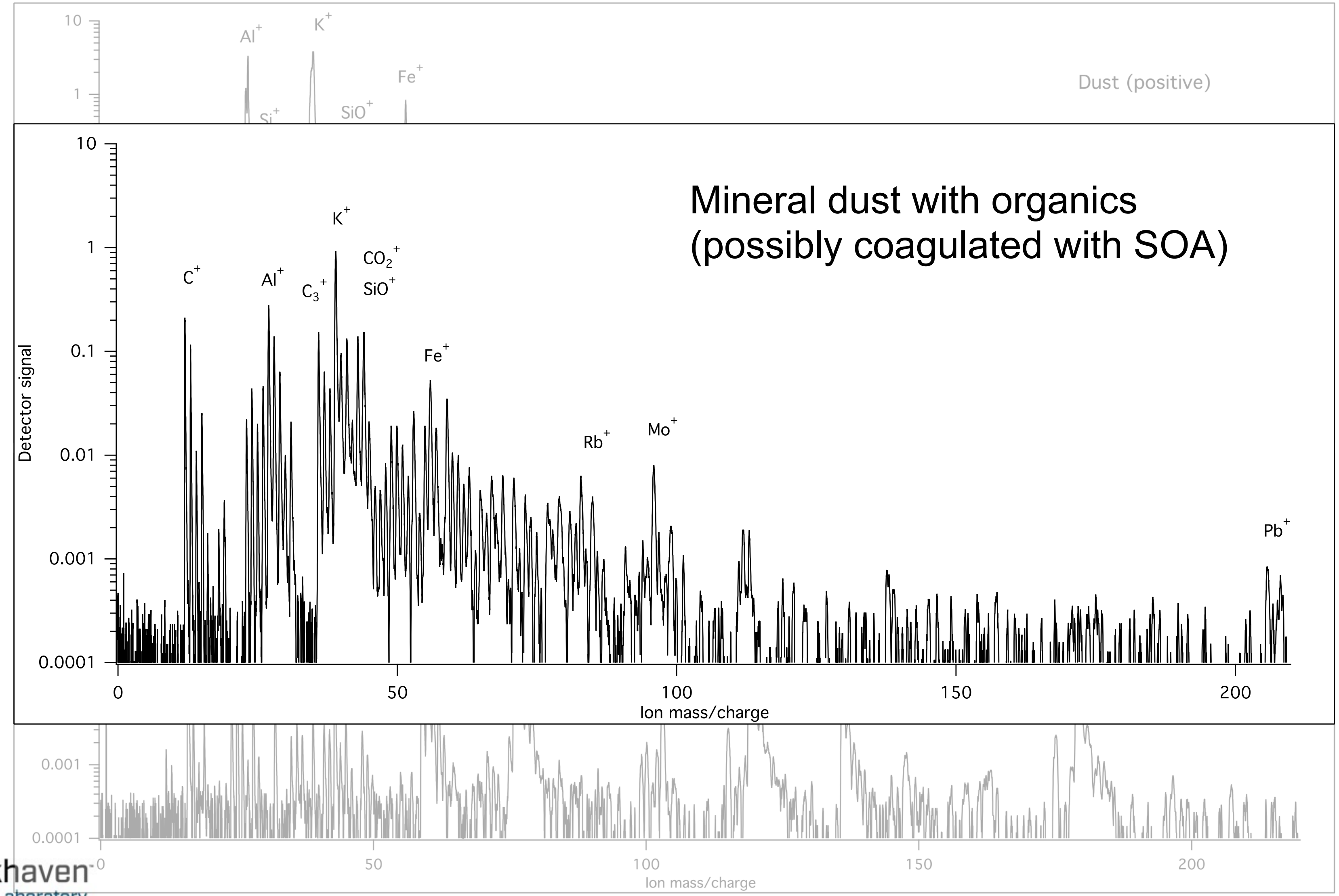
Single-particle mass spectra are challenging to interpret





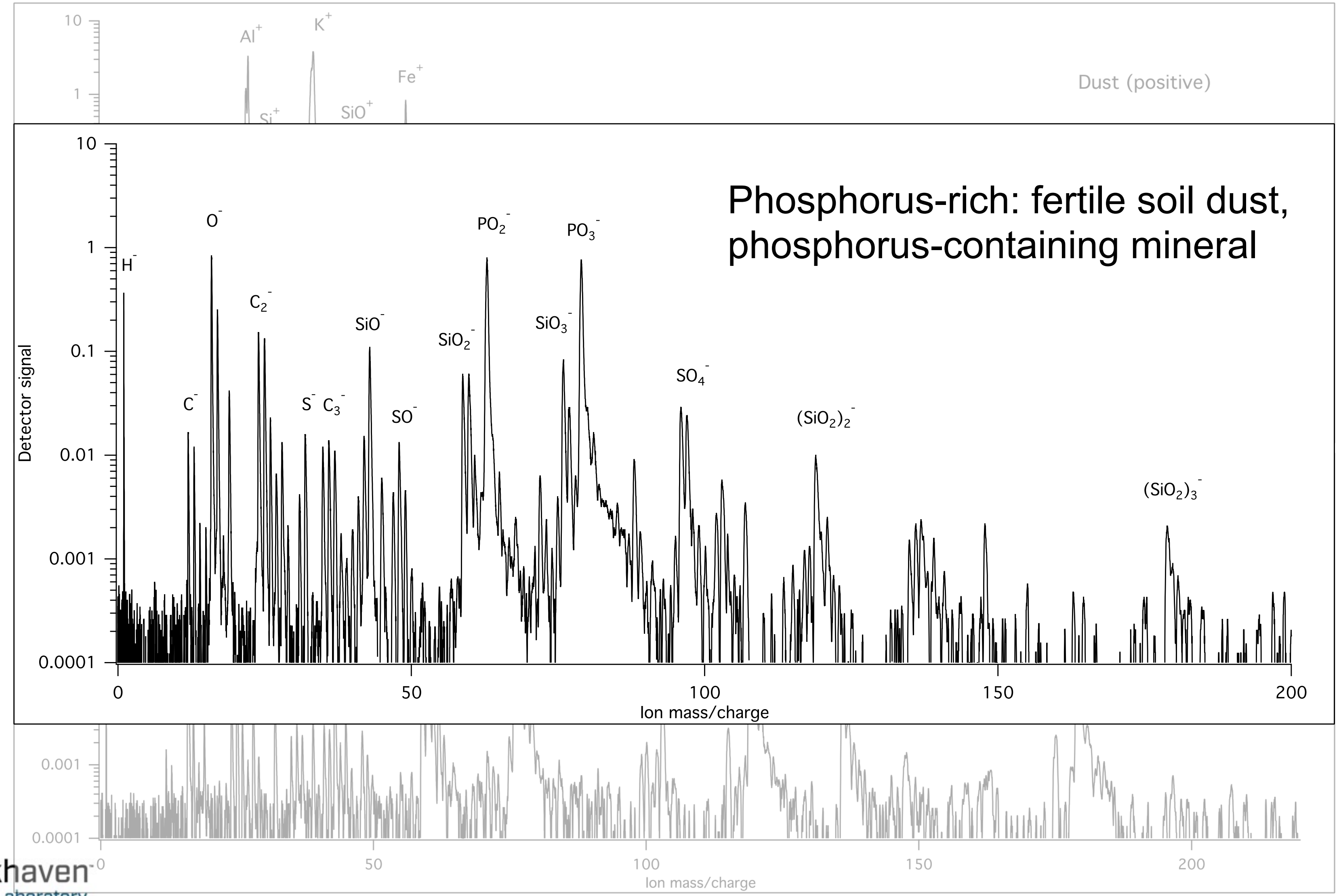
Brookhaven National Laboratory

Single-particle mass spectra are challenging to interpret



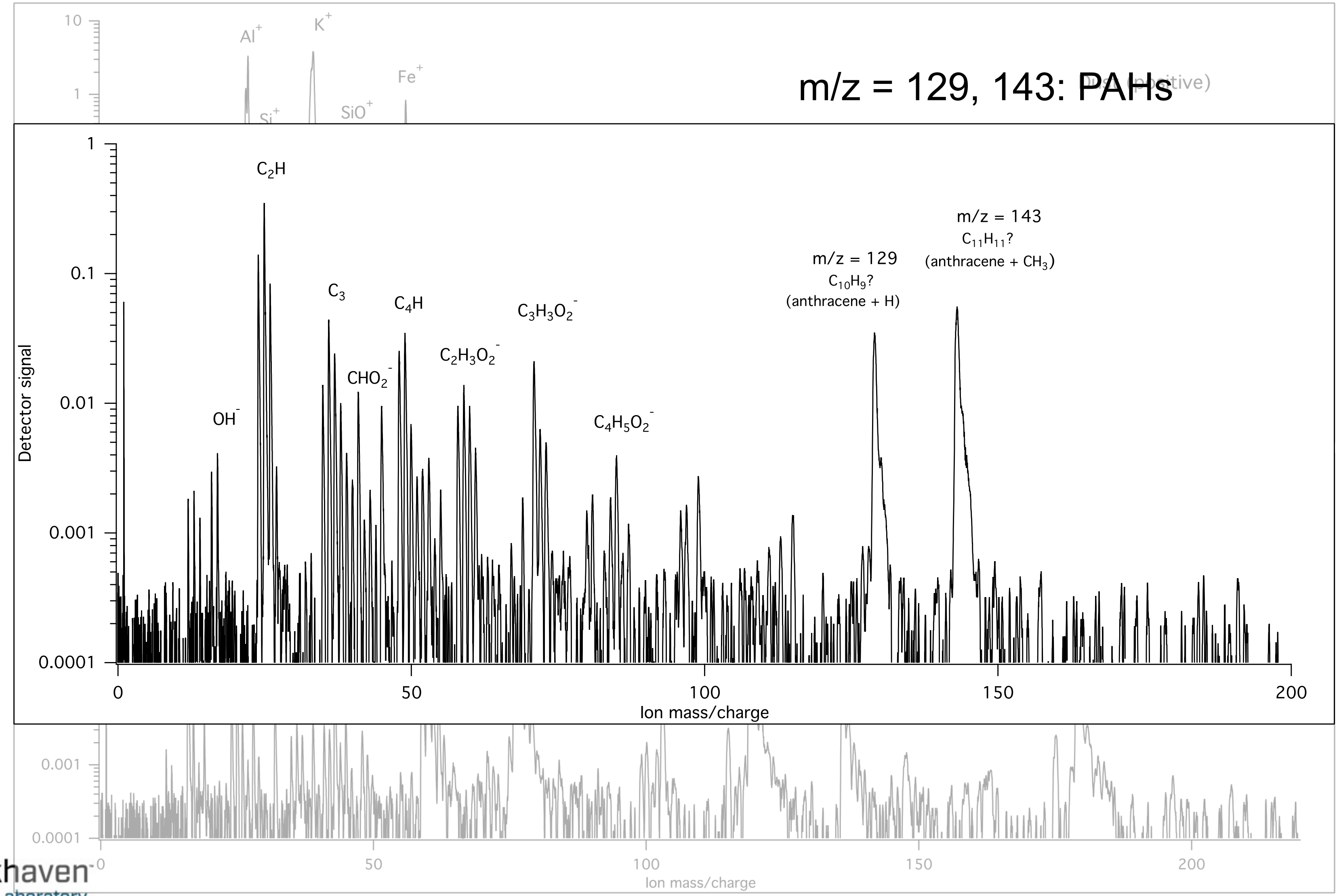
Brookhaven National Laboratory

Single-particle mass spectra are challenging to interpret

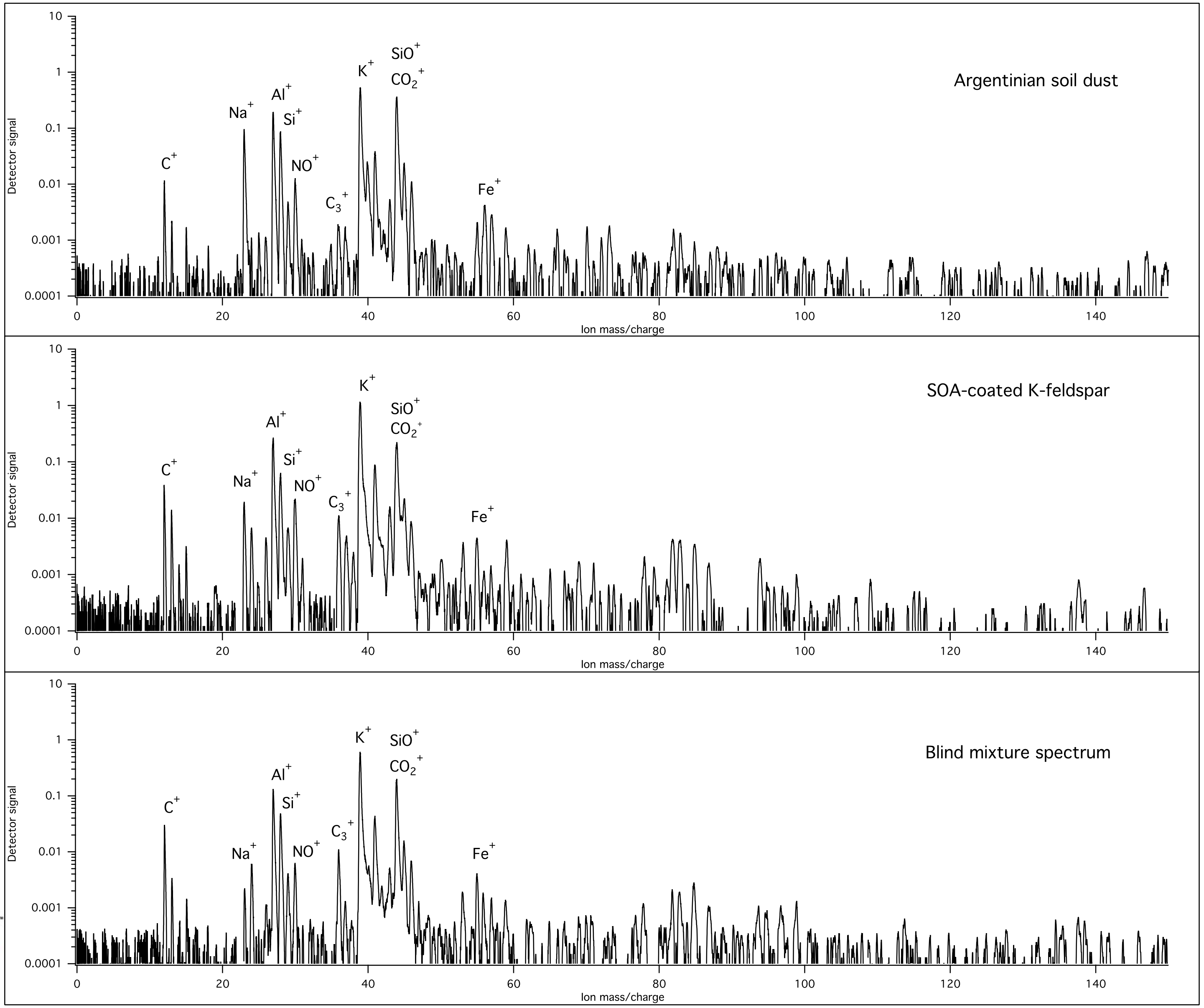


Brookhaven National Laboratory

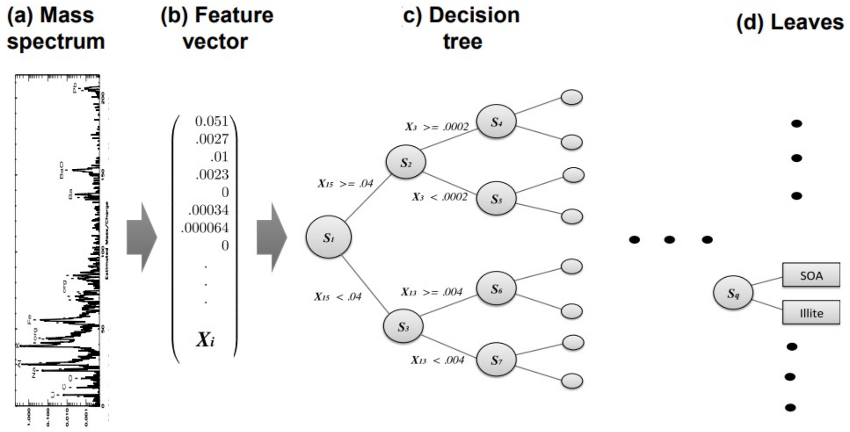
Single-particle mass spectra are challenging to interpret

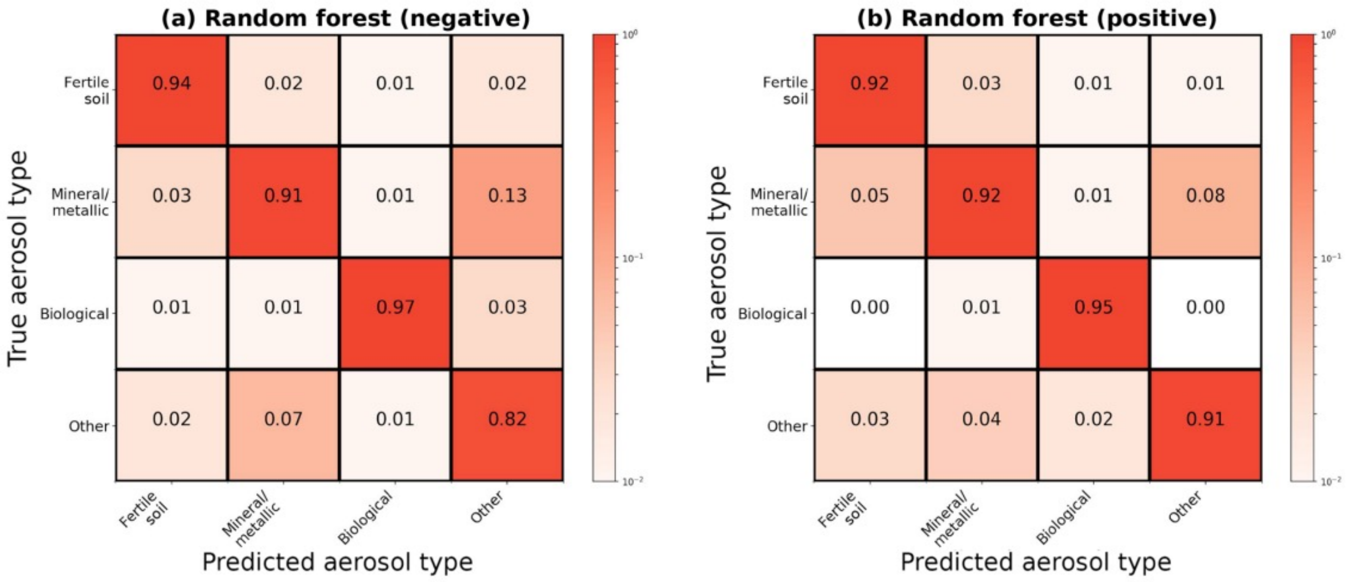


Single-particle mass spectra are challenging to interpret



Supervised classification approaches



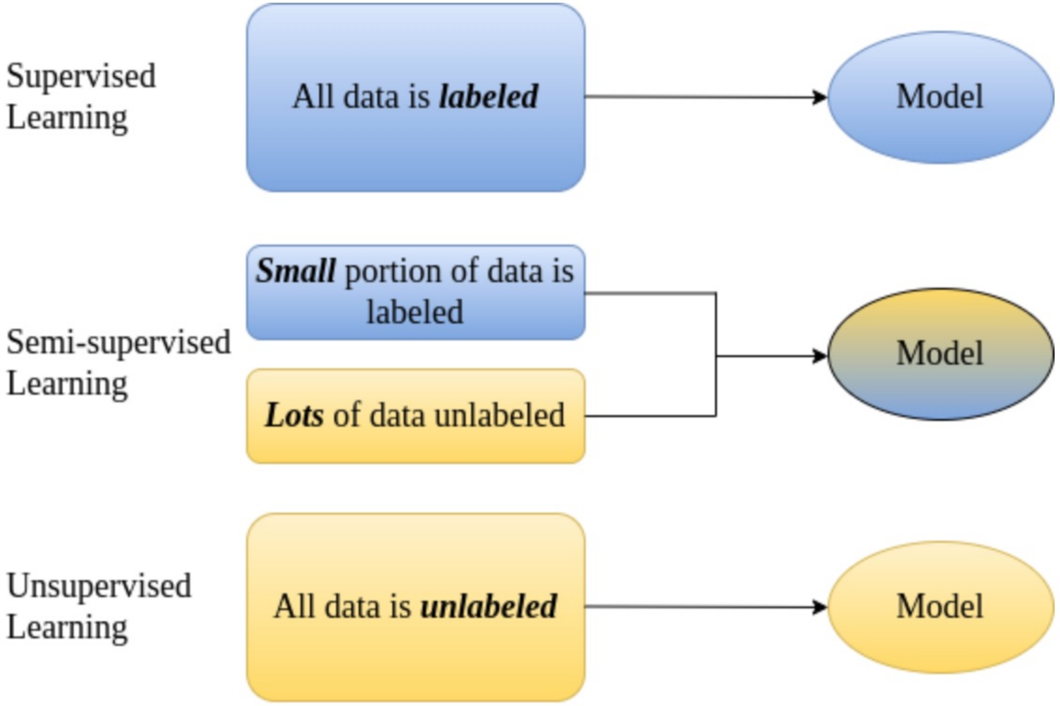


Random forest supervised classifier produced encouraging results when tested on pure laboratory mass spectra but struggled when deployed on a ”blind” mixture of aerosols.

Christopoulos et al. (2018)

Brookhaven National Laboratory

Semi-supervised classification approaches



Strategy called “self-training” is used to train a stacked autoencoder neural network:

(1) labeled dataset is used to train the classifier.

(2) the trained classifier is used on the unlabeled dataset to classify it.

(3) spectra with the highest classification certainty are selected and reviewed for any errors.

(4) correctly classified spectra are then incorporated into the labeled training dataset.

(5) this procedure is repeated five times. Five iterations are generally sufficient to minimize the error frequency.

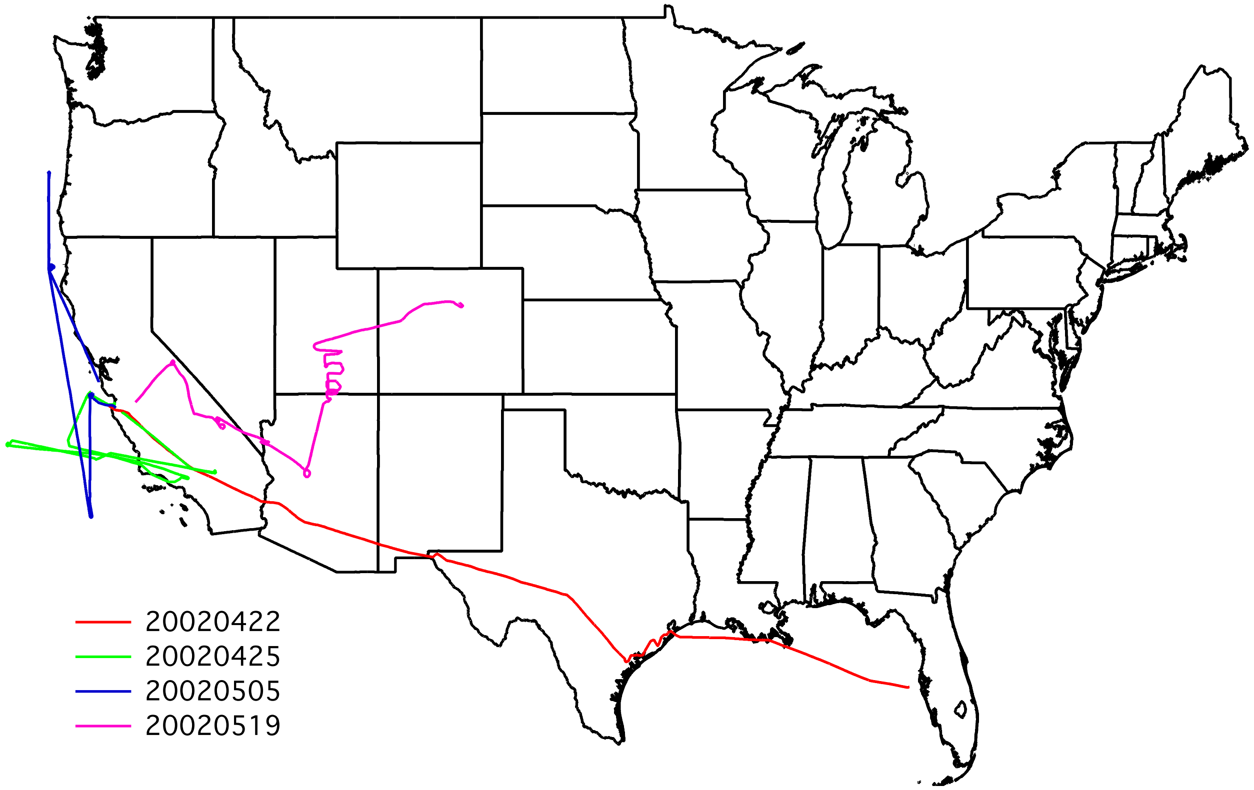
Semi-supervised classifiers produced outstanding results in controlled lab experiments

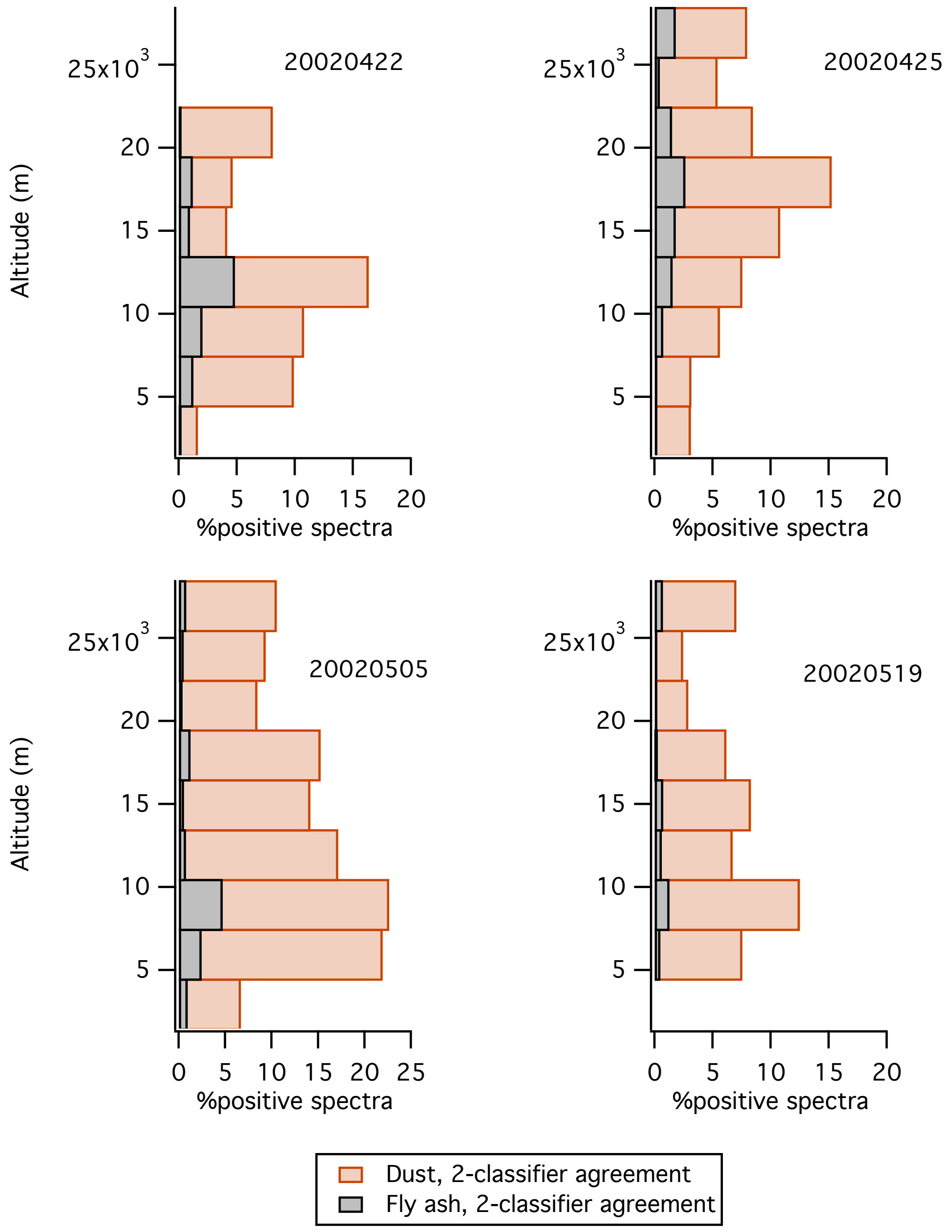
Accuracy: 96.94%

Output Class \ Target Class	Fly ash	Apatite	Ca-rich	Illite	Feldspars	Monazite	Na-Mont	BioBurn	Organic	Sea salt	Soot	Metallic
Fly ash	97.9% 2329	0.4% 2	0.1% 1	0.0% 0	0.8% 21	0.0% 0	0.8% 3	0.3% 2	0.0% 0	0.8% 5	0.0% 0	0.2% 1
Apatite	0.0% 1	99.6% 543	0.0% 0	0.1% 1	0.0% 1	2.0% 3	0.3% 1	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0
Ca-rich	0.0% 1	0.0% 0	99.7% 706	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.2% 1	0.0% 0	0.0% 0
Illite	0.1% 2	0.0% 0	0.1% 1	95.2% 1406	3.3% 91	1.3% 2	1.0% 4	0.4% 3	0.0% 0	0.2% 1	0.0% 0	0.2% 1
Feldspars	1.0% 24	0.0% 0	0.0% 0	4.3% 64	95.4% 2643	0.0% 0	0.5% 2	0.0% 0	0.0% 0	0.0% 0	0.2% 1	0.0% 0
Monazite	0.1% 2	0.0% 0	0.0% 0	0.1% 2	0.0% 0	95.3% 142	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.5% 2
Na-Mont	0.4% 10	0.0% 0	0.0% 0	0.1% 2	0.4% 12	1.3% 2	96.3% 369	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0
BioBurn	0.0% 0	0.0% 0	0.0% 0	0.1% 2	0.1% 2	0.0% 0	0.0% 0	97.9% 694	2.4% 17	0.0% 0	0.2% 1	0.0% 0
Organic	0.2% 4	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.7% 5	95.0% 668	0.2% 1	1.2% 6	0.0% 0
Sea salt	0.3% 6	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.3% 1	0.4% 3	0.1% 1	98.2% 606	0.0% 0	0.2% 1
Soot	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.5% 2	0.1% 1	2.4% 17	0.2% 1	98.4% 507	0.2% 1
Metallic	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.3% 1	0.1% 1	0.0% 0	0.3% 2	0.0% 0	98.6% 413

…but tend to fall short in real-life situations

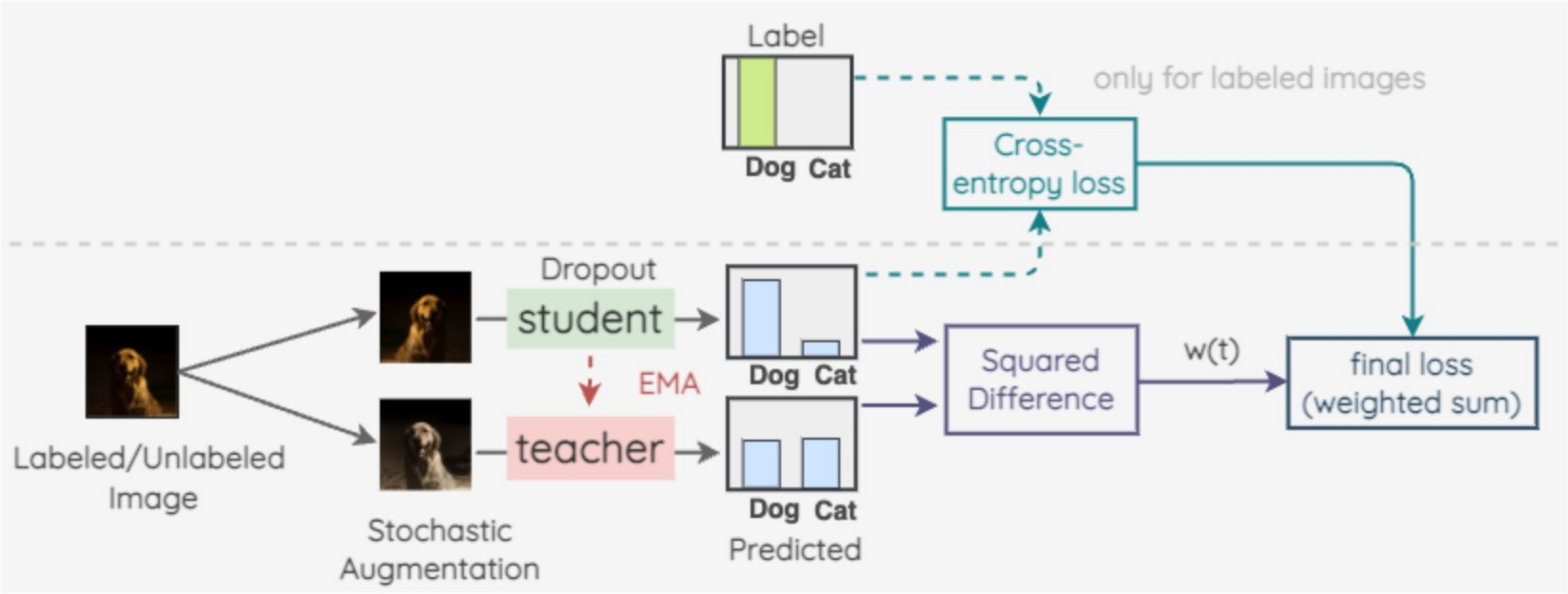
Four example flights from ITCT that had the most fly ash hits. Fly ash very strongly correlates with dust. No fly ash-only plumes.





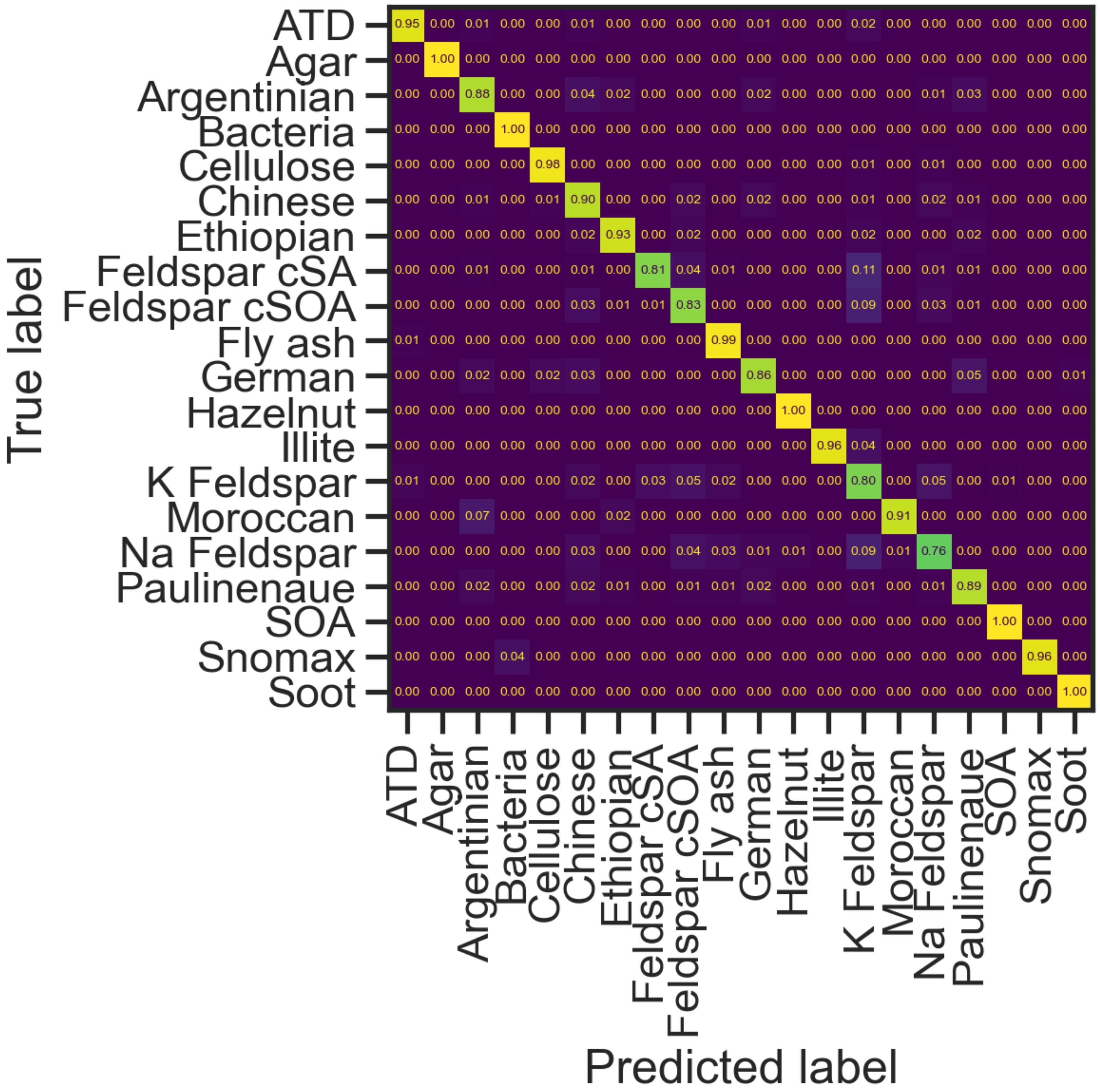
Application to field datasets: ITCT (2002)

Current approach: using mean-teacher training strategy



- Early results of training stacked autoencoders using mean teacher approach show better accuracy than supervised classification (decision trees, SVM).

- Next step is to increase the training dataset size with more unlabeled examples.



Brookhaven National Laboratory