Pacific Northwest
NATIONAL LABORATORY

PNNL-31981

Applying novel analytical tools for analyzing multidimensional secondary organic aerosol measurements

September 2021

Manish Shrivastava
Paritosh Pande
John Shilling
Alla Zelenyuk
Quazi Rasool

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <https://www.ntis.gov/about>
Online ordering: http://www.ntis.gov

Applying novel analytical tools for analyzing multidimensional secondary organic aerosol measurements

September 2021

Manish Shrivastava
Paritosh Pande
John Shilling
Alla Zelenyuk
Quazi Rasool

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Introduction

In the atmosphere, secondary organic aerosols (SOA) are often the major components of fine particulate matter and interact with clouds and radiation. SOA comprises a mixture of thousands of organic compounds. There is tremendous complexity and uncertainty in understanding SOA formation, since it is formed by oxidation and gas to particle conversion of a variety of sources: natural biogenic, anthropogenic (vehicles, cooking coal combustion) and biomass burning. The Aerosol Mass Spectrometer (AMS) produces multidimensional chemical information about SOA but analyzing this data to understand SOA sources relies on time consuming analyses (~months to years) such as the positive matrix factorization (PMF) (1). PMF also becomes difficult for aircraft data where signal to noise ratio is weaker. There is a critical need to develop fast machine learning techniques that can analytically provide information about SOA sources using AMS data on the same timescales as the data is being collected (~minutes). We apply a machine learning supervised classification approach: the multinomial logistic regression (2) to rapidly classify AMS data obtained from aircraft measurements.

Approach

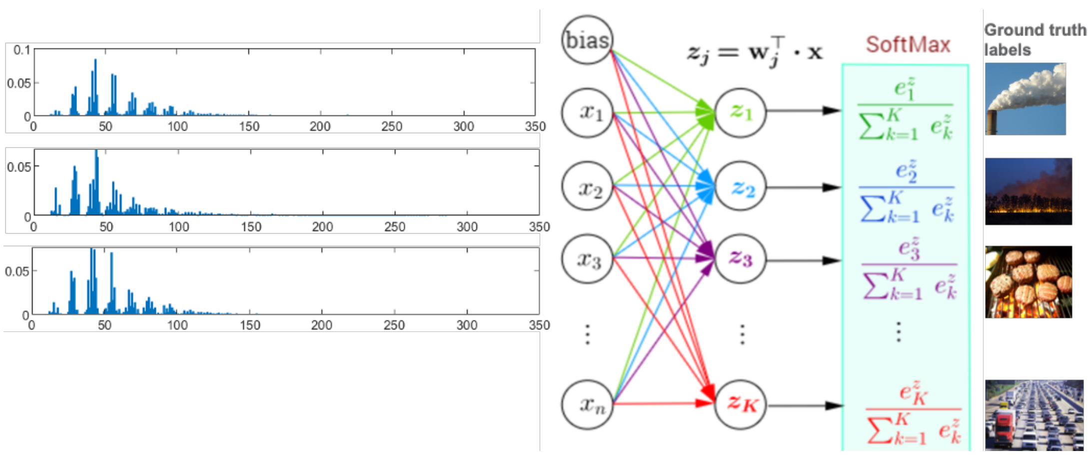

Figure 1: Schematic illustrating development and application of a new machine learning approach: the multinomial logistic regression (middle) to classify AMS data. Ground truth AMS mass spectra of pure single SOA components from laboratory (leftmost) are used to train the classifier to identify the various SOA sources (rightmost).

As illustrated above the multinomial logistic classifier is a supervised classification technique that leverages ground truth laboratory mass spectra for single components and PMF deconvolved mass spectra (in case of more complex sources like biomass burning) to rapidly classify SOA mass spectra. The classifier is then applied to classify aircraft-based AMS measured SOA mass spectra. This method has been applied for the first time to AMS mass spectra and shows a lot of promise.

<u>Key results</u>
The classifier was applied on aircraft-based aerosol mass spectra measurements during the HI-SCALE 2016 (3) field campaign.

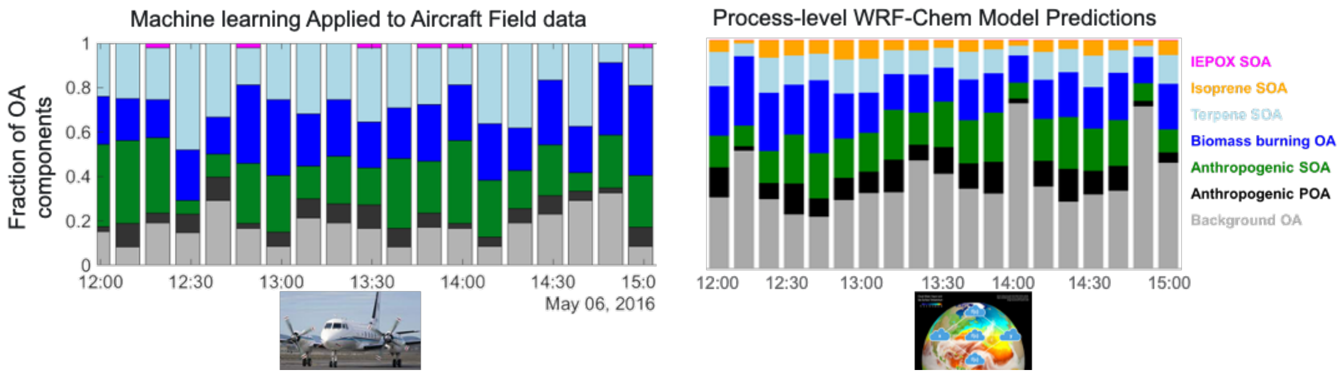

Figure 2: (left) Machine learning (logistic regression) classified fractional distribution of SOA sources along aircraft flight transects on 6 May 2016 during HI-SCALE; (right) WRF-Chem process based simulated distributions of SOA sources. May 6 was a dry day with significant influence of biomass burning (blue). Both machine learning and WRF-Chem results identify the significant influence of wildfires on aircraft sampled organic aerosol mass spectra.

We compared machine learning (ML) results with a completely independent process-based modeling approach using the three-dimensional regional Weather Research and Forecasting Model coupled to chemistry (WRF-Chem). The ML approach identified significant contribution of wildfires (blue) to organic aerosol mass spectra, consistent with WRF-Chem results.

<u>Impacts</u>
Importantly, the classifier can rapidly classify SOA sources in the field and can be deployed online to classify aircraft data. Thus, we developed a new machine learning capability to classify AMS data, which could be of great interest to DOE BER, ARM and aircraft data analyses capabilities.

<u>References</u>

1. I. M. Ulbrich, M. R. Canagaratna, Q. Zhang, D. R. Worsnop, J. L. Jimenez, Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data. *Atmos. Chem. Phys.* **9**, 2891-2918 (2009).
2. J. Friedman, Trevor Hastie, and Robert Tibshirani, *The elements of statistical learni*, New York: Springer series in statistics (New York, 2001), vol. 1.
3. J. D. Fast *et al.*, Overview of the HI-SCALE Field Campaign: A New Perspective on Shallow Convective Clouds. *Bulletin of the American Meteorological Society* **100**, 821-840 (2019).

Pacific Northwest
National Laboratory
902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)
www.pnnl.gov