Clustering Analysis of Atlantic ATom Observations

Ethan Manninen, Professor Steven Wofsy

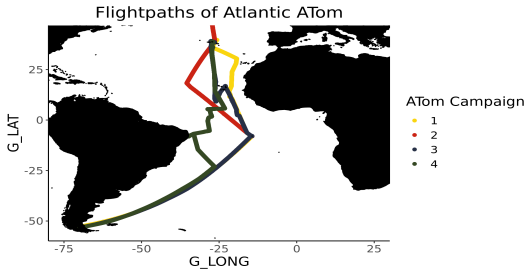
Introduction

Combustion emissions affect the climate and human health. Carbon monoxide is a useful marker for combustion because of its relatively short lifetime[1]. The Atmospheric Tomography Missions (ATom) measured numerous chemical features of the remote atmosphere in four seasons over four years[2]. ATom flights over the Atlantic reveal a plume of high CO air originating from Africa.

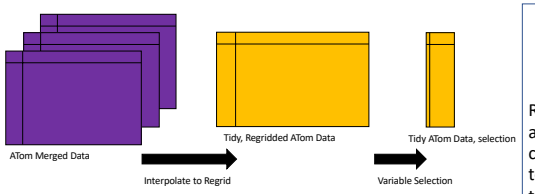
Known sources of CO in Western and Central Africa include wildfires, urban pollution, and emissions from hydrocarbon extraction and refining.

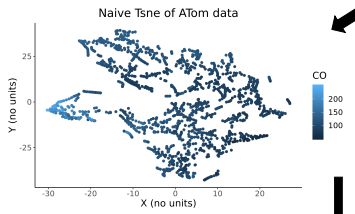
We investigate whether high CO air observed by ATom can be clustered based on chemical data using statistical methods.

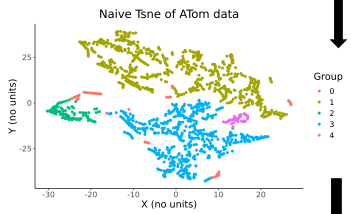
Using an approach built around t-SNE and DBSCAN algorithms, we find varying numbers of clusters within high CO plumes observed by ATom 1-3.

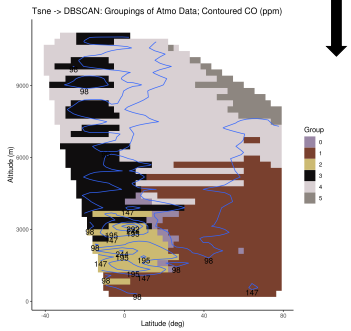


Methods

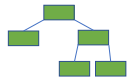






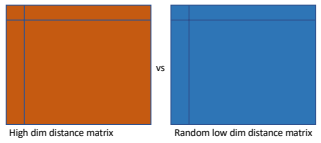


Random Forest[3]



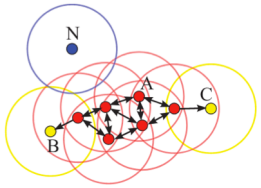
Random subsets of data, and random selections of causal variables are used to build each decision tree.
Used to select most important variables when predicting whether a grid tile is "high CO" or not

T-SNE[4]



Parameters: perplexity, iterations.
Preserves high dimensional distance in two dimensional space by iteratively improving a random projection

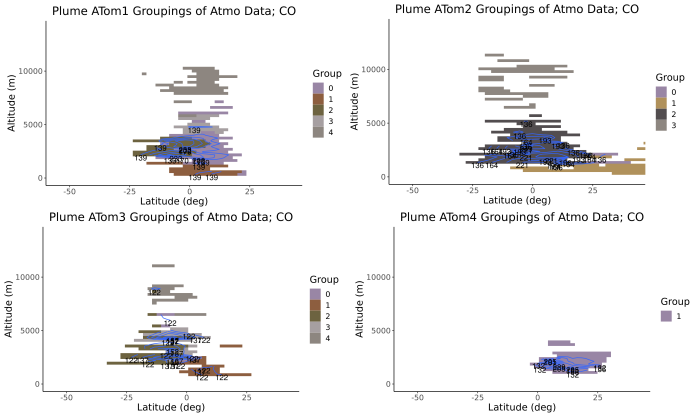
DBSCAN[5]



Parameters: minPts, epsilon.
Density based clusters. Builds neighborhoods based on distance (eps) to number of neighbors (minPts)

Results

We find four clusters within the CO plume seen by ATom 1, three clusters by Atom 2, four clusters by ATom 3, and are unable to find clusters within ATom 4.



Future Directions

Next, we will use boundary layer footprints to understand origins of emissions associated with each cluster of ATom observations. This will advance CO budgets for the region. Ultimately, we hope to develop understanding of the signature trace gases associated with each source, and the chemistry those gases undergo.

Citations

1) Crutzen, P. J.; Andreae, M. O. (1990). Biomass Burning in the Tropics: Impact on Atmospheric Chemistry and Biogeochemical Cycles. *Science; Washington 250* (4988), 1669.
2) https://espo.nasa.gov/atom/content/ATom
3) Tin Kam Ho. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition, 1*, 278-282 vol.1.
4) Van Der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal Of Machine Learning Research, 9*, 2579-2605.
5) Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. (1996). A density-based algorithm for discoveringclusters in large spatial databases with noise. Proceedings of the 2nd ACM International Conference on KnowledgeDiscovery and Data Mining (KDD). 226–231.