In-situ and Remote-Sensing Data Fusion Using Machine Learning Techniques to Infer Urban and Fire Related Pollution Plumes.

Russell, P.B.[1], Segal-Rosenheimer, M.[1,2], Schmid, B.[3], Redemann, J[1]., Livingston, J.M.[4], Flynn, C.J[3]., Johnson, R.R.[1], Dunagan, S.E.[1], Shinozuka, Y.[2,5], Kacenelenbogen, M.[1,2], Chatfield, R.B.[1]

Key-words: urban pollution, biomass-burning, neural-network, clustering techniques, SEAC[4]RS

Abstract

Airmass type characterization is key in understanding the relative contribution of various emission sources to atmospheric composition and air quality and can be useful in bottom-up model validation and emission inventories. However, classification of pollution plumes from space is often not trivial. Sub-orbital campaigns, such as SEAC^{4}RS (Studies of Emissions, Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys) give us a unique opportunity to study atmospheric composition in detail, by using a vast suite of in-situ instruments for the detection of trace gases and aerosols. These measurements allow identification of spatial and temporal atmospheric composition changes due to various pollution plumes resulting from urban, biogenic and smoke emissions. Nevertheless, to transfer the knowledge gathered from such campaigns into a global spatial and temporal context, there is a need to develop workflow that can be applicable to measurements from space. In this work we rely on sub-orbital in-situ and total column remote sensing measurements of various pollution plumes taken aboard the NASA DC-8 during 2013 SEAC^{4}RS campaign, linking them through a neural-network (NN) algorithm to allow inference of pollution plume types by input of columnar aerosol and trace-gas measurements. In particular, we use the 4STAR (Spectrometer for Sky-Scanning, Sun-Tracking Atmospheric Research) airborne measurements of wavelength dependent aerosol optical depth (AOD), particle size proxies, O_3, NO_2 and water vapor to classify different pollution plumes. Our method relies on assigning a-priori "ground-truth" labeling to the various plumes, which include urban pollution, different fire types (i.e. forest and agriculture) and fire stage (i.e. fresh and aged) using cluster analysis of aerosol and trace-gases *in-situ* and auxiliary (e.g. trajectory) data and the training of a NN scheme to fit the best prediction parameters using 4STAR measurements as input. We explore our misclassification rates as related to our "ground-truth" labels, and with multi-layered pollution plume cases. The next step in our analysis is to optimize parameter selection for a scheme that can be applied to space-borne aerosol and trace-gas observation platforms such as OMI, and future geostationary satellites such as TEMPO and GEO-CAPE.