Using Supervised Machine Learning and HYSPLIT Backwards Trajectories to Predict Airborne VOC Concentrations

Victor Geiser, Dr. Donald Blake

Extended Abstract:

It is known that Volatile Organic Compounds (VOCs) in the atmosphere have harmful and adverse effects on human health. Recent literature suggests some VOCs may also have an effect on global radiative forcing, despite making up a very small fraction of the total atmospheric composition. Due to this, making accurate predictions of these concentrations within the context of the dynamic atmosphere is both a challenging yet imperative problem. While the fields of atmospheric chemistry and meteorology are often addressed in adjacent but related contexts, this study looks to examine the importance of meteorological variables on the concentrations of these VOCs at parts per trillion level precision, based on fully open-source meteorological data alone. Through this study it was found that although considering meteorology data alone is not enough to predict these concentrations, meteorological variables such as potential temperature, solar radiation, and mixing depth appear to have elevated importance across a standard array of meteorological variables. Moreover, this study proves that additional considerations are necessary to address the differences in data types of whole air sampling data and meteorology data when incorporated into machine learning workflows. The results of this study show that it is possible to predict airborne VOC concentrations using meteorological variables, but the accuracy of these results is currently limited by discrepancies between the spatial resolution of meteorology data and the discrete nature of whole air samples. However, given that these structural challenges exist, applying modern techniques in imbalanced regression may help address these challenges.