Contribution ID: **5** Type: **Oral presentation**

Driving mechanisms of global surface ozone and its bias in the chemical reanalysis products using machine-learning approach

Monday, 6 March 2023 14:10 (20 minutes)

Providing accurate global estimates of pollution is essential to evaluate the global public health burden of disease associated with air pollution exposure, which in turn will help environmental policy making. Nevertheless, our current knowledge of air pollution suffers from large biases in model predictions and insufficient information from the current observing system that includes surface in situ and satellite measurements. Chemical data assimilation (DA) has made substantial progress in reproducing regional and global ozone and health impacts. Nevertheless, chemical DA is fundamentally limited by model fidelity both in terms of representation resolution and unresolved processes that can exhibit systematic patterns in model and data mismatches. Thus, the analysis and prediction of air quality at regional scales has been stymied by unresolved and highly non-linear physical and chemical processes.

In this study, we develop and utilize a novel, explainable machine learning (ML) model to break down regional bias dependence and provide scientific interpretation of ozone and its bias drivers by analyzing a large set of exogenous and input data sources provided by chemical DA and various observations. By doing so, we aim to provide new scientific insights into the factors that control bias in air quality assessment, and the drivers of global ozone trends and their impact on global air quality.

We obtained the differences between the analysis from JPL's chemical DA system, MOMO-Chem, and the available independent observations from the surface Tropospheric Ozone Assessment report (TOAR) network for ozone for 2011-2015. These differences are used as the outputs of interest in the ML model, while various global chemical concentrations and meteorological (>100) parameters from MOMO-Chem, as well as high-resolution satellite measurements of relevant parameters, are used as inputs into the ML model. A regression-tree randomized ensemble ML approach was applied to model the patterns of the residual bias of the MOMO-Chem output. Importantly, we extracted and combined local and global measures of how inputs affected the predicted bias in MOMO-Chem output, therefore, providing ML model explanations and quantification of the impacts. Subsequently, we are able to identify a set of distinct areas and their corresponding drivers, which in turn will shed light on both resolved and unresolved model physical and chemical processes and their importance to air pollutant predictions.

Our analysis suggests that the developed ML framework is able to predict the overall model ozone bias magnitude and variations over North America, Europe, and East Asia. Our results also highlight that adding high-resolution satellite data, MODIS land cover data in this case, provides additional important constraints to improve the ML prediction, especially for highly polluted events. The framework is then used to identify global patterns of surface ozone and physical model bias drivers, demonstrating substantial contributions from parameters related to, for example, PBL mixing, sea breeze, photolysis rates, precursors, and topography, with significant seasonal and regional variations. This study offers a unique synthesis of model-based inference and explainable ML techniques for chemical transport modeling and data assimilation to identify regionally-dependent and process-level mechanisms driving near-surface pollution and correct for their impact on air quality predictions.

ML method

Random forest

Main air pollutant of interest

Tropospheric ozone and precursors

Primary author: MIYAZAKI, Kazuyuki (NASA JPL)

Co-authors: PATEL, Eshani (Caltech); MONTGOMERY, James (NASA JPL); LI, Joanna (UCLA); DOERKSEN, Kelsey (University of Oxford); BOWMAN, Kevin (NASA JPL); LU, Steven (NASA JPL); MARCHETTI, Yuliya (NASA JPL)

Presenter: MIYAZAKI, Kazuyuki (NASA JPL)

Track Classification: Machine learning applications