National Aeronautics and Space Administration

FY24 Strategic Initiatives Research and Technology Development (SRTD)

Subgrid scale drivers of pollution inferred from model-based inference and machine learning

Principal Investigator: Kazuyuki Miyazaki (329); **Co-Investigators:** Kevin Bowman (329), Yuliya Marchetti (398), James Montgomery (398), You Lu (398)

Strategic Focus Area: Integrated Community of Practice for Scientific Understanding from Data Science (SUDS) | **Strategic Initiative Leader:** Erika Podest

Objectives:

The central objective of this effort is to provide new scientific insights into (1) the factors that control bias in air quality assessment, and (2) the drivers of global ozone trends and their impact on global air quality at scales relevant for assessing human health impacts. This research will demonstrate to the JPL Scientific Understanding from Data Science (SUDS) Community a generalized approach for using explainable machine learning (ML) to identify, correct, and gain insight from primary drivers of physical model biases while considering uncertainty.

Background:

- Our current knowledge of air pollution suffers from large systematic errors in physical model predictions and insufficient information from the current observing systems, leading to a limited understanding of air quality and its health impact.
- Only a very small change in air pollutant concentrations (by 1 μgm^{-3} for PM2.5 and by 1 ppb for ozone) would change a human health impact estimate by 14,000 deaths per year, whereas the current models have up to 20 ppb bias for ozone.

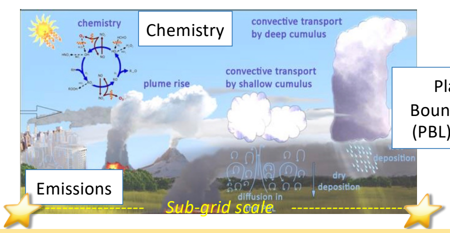


Fig. 1: Schematic picture of processes that control surface ozone. Surface ozone is hard to predict accurately due to errors in advection, chemistry, and sub-grid processes.

Approach:

SUDS AQ offers a unique synthesis of model-based inference and explainable ML techniques to identify mechanisms driving near-surface pollution and correct for their impact on air quality predictions.

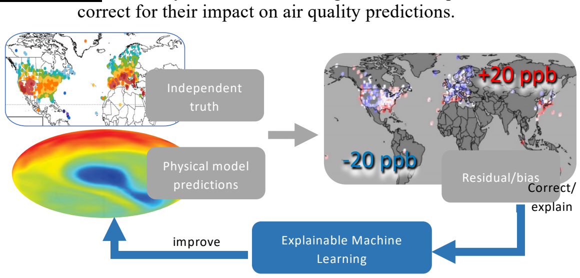


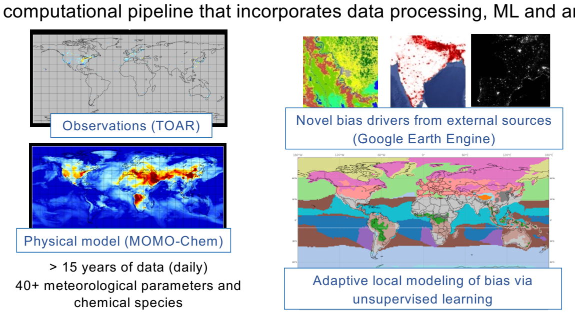
Fig. 2: Schematic picture of the developed ML framework that predicts physical model error and translate them into scientific insights

Significance/Benefit to JPL and NASA:

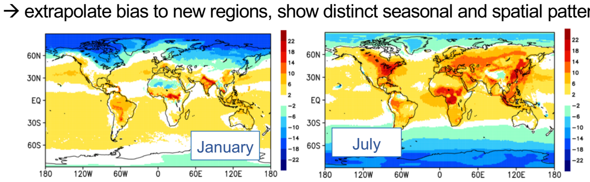
- JPL's parallel architecture studies and pre-formulation studies for Explorer Class air quality and greenhouse gas missions
- OSSE studies can better account for resolving sub-grid processes
- Harnessing current JPL assets, e.g., TROPESS

Data and ML/analysis pipeline:

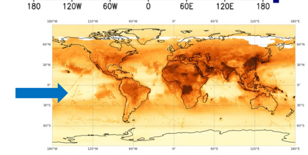
Long-term global datasets, large complex data requires a mature computational pipeline that incorporates data processing, ML and analysis.



Predicted values trained on sparse observations
→ extrapolate bias to new regions, show distinct seasonal and spatial patterns.

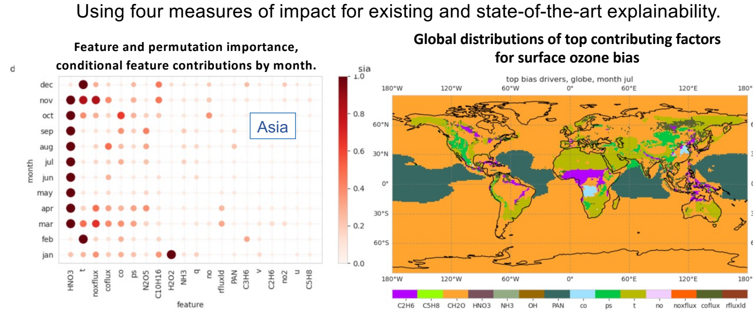


Uncertainty maps for the predicted ozone bias derived from QRF prediction intervals.



Explainable machine learning:

Which input data makes the most impact on ozone prediction?
Using four measures of impact for existing and state-of-the-art explainability.



New insights into the physical model bias of surface ozone.

Publications:

A physical science paper and a data science paper (to be submitted)

PI/Task Mgr. Contact Information:

kazuyuki.miyazaki@jpl.nasa.gov, +1-818-354-3266

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

www.nasa.gov

RPD-000 Clearance Number: CL#00-0000