National Aeronautics and Space Administration

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

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

Program: FY22 R&TD Strategic Initiative

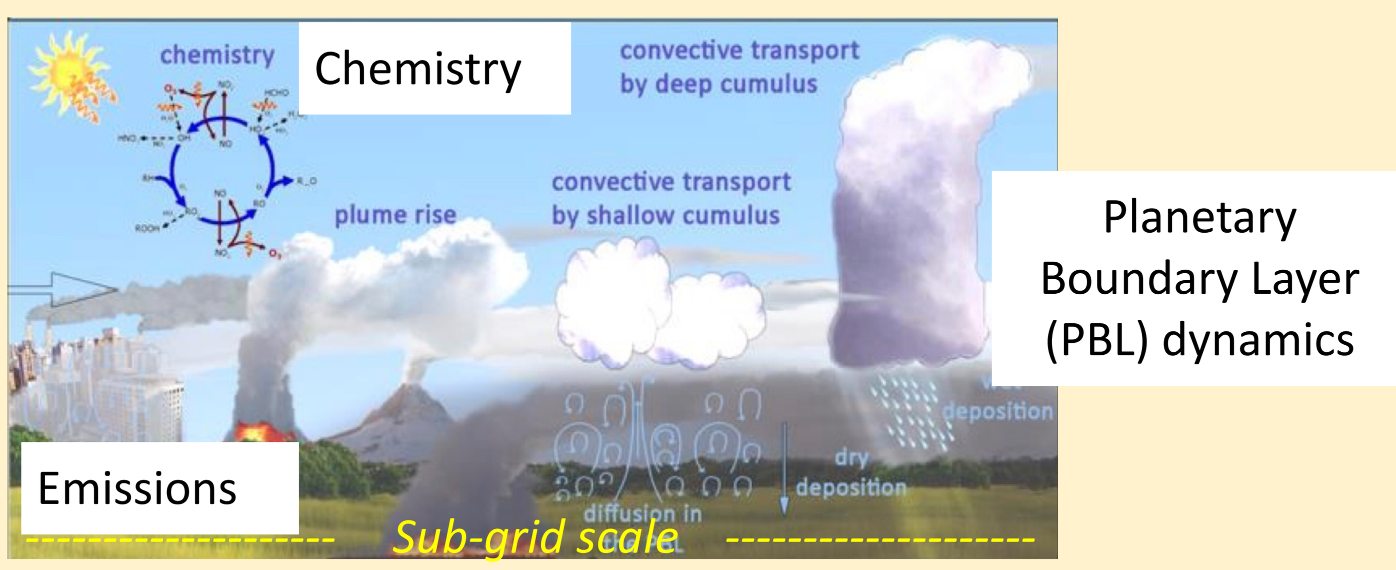
Strategic Focus Area: An Integrated Community of Practice for Scientific Understanding from Data Science (SUDS) - Strategic Initiative Leader: Susan E Owen

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.

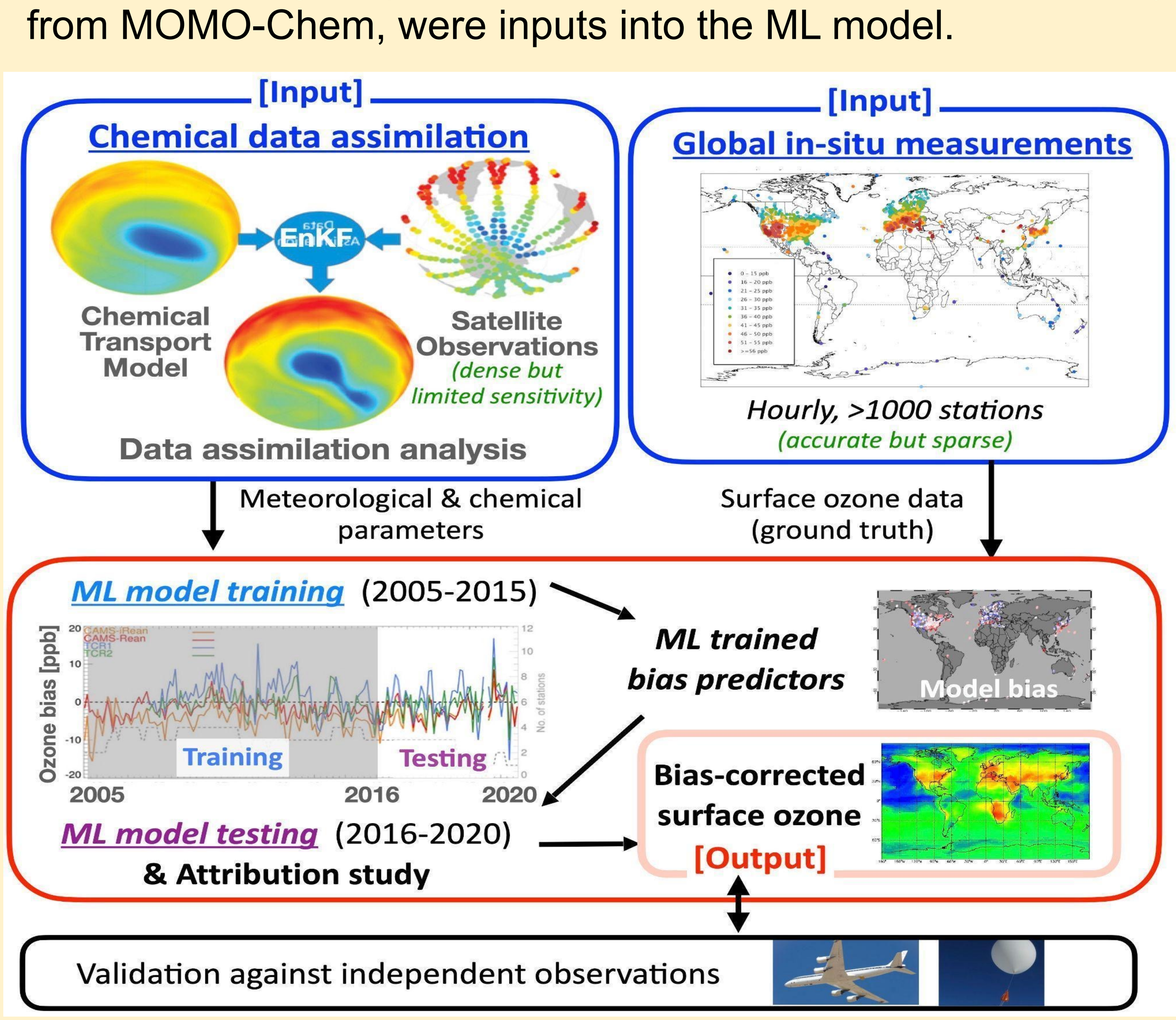
Background:

- Our current knowledge of air pollution suffers from large systematic errors, i.e. biases, 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 typically have up to 20 ppb bias for ozone.



Approach:

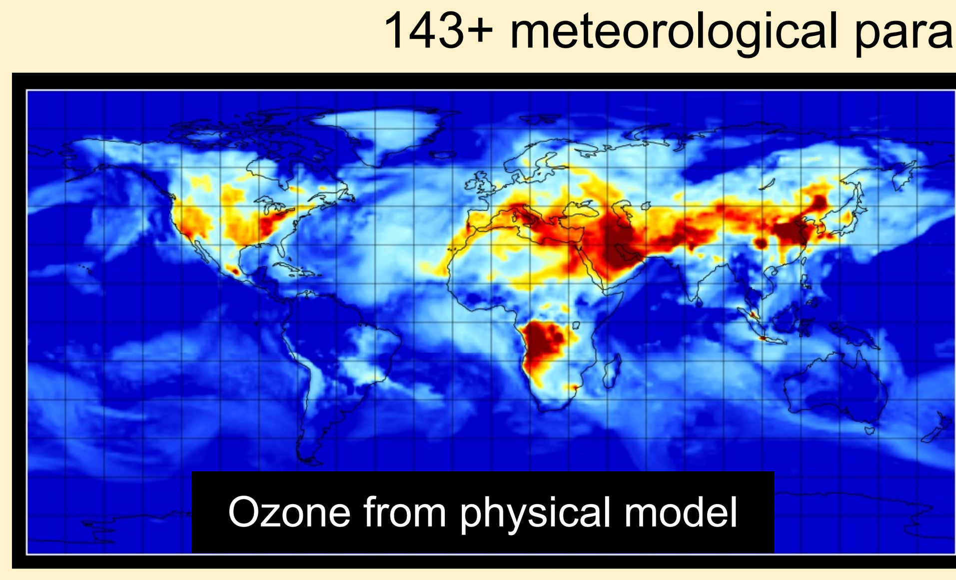
SUDS AQ offers a unique synthesis of model-based inference and explainable ML techniques to identify mechanisms driving near-surface pollution and correct for their impact on air quality predictions. We used ML to model the patterns of the residual bias of the output of JPL's chemical DA system, MOMO-Chem. We obtained the differences between DA analysis and the independent observations from the surface TOAR-2 network. These differences, i.e. biases, were the outputs of interest of the baseline ML model, while various chemical concentrations and meteorological parameters, such as pressure and humidity, from MOMO-Chem, were inputs into the ML model.

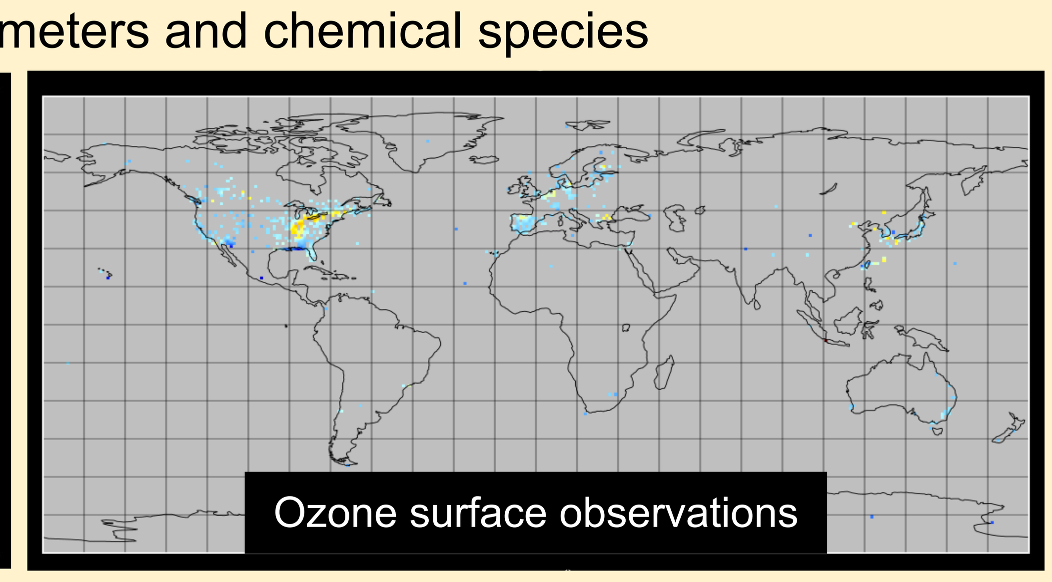


Large scale data:

>2TB of physical model data at daily resolution for 10 years

143+ meteorological parameters and chemical species





Explainable machine learning:

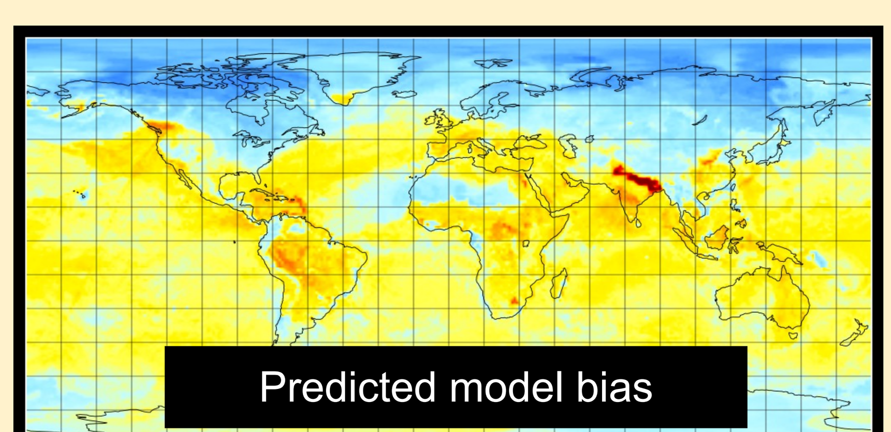
We implemented a baseline ML model based on randomized ensemble tree methodology and trained it seasonally to account for strong temporal variability as well as to characterize distinct seasonal trends of the bias and to derive driver explainability, which in turn facilitated the driver impact analysis for the physical model. We implemented four ML explanability methods.

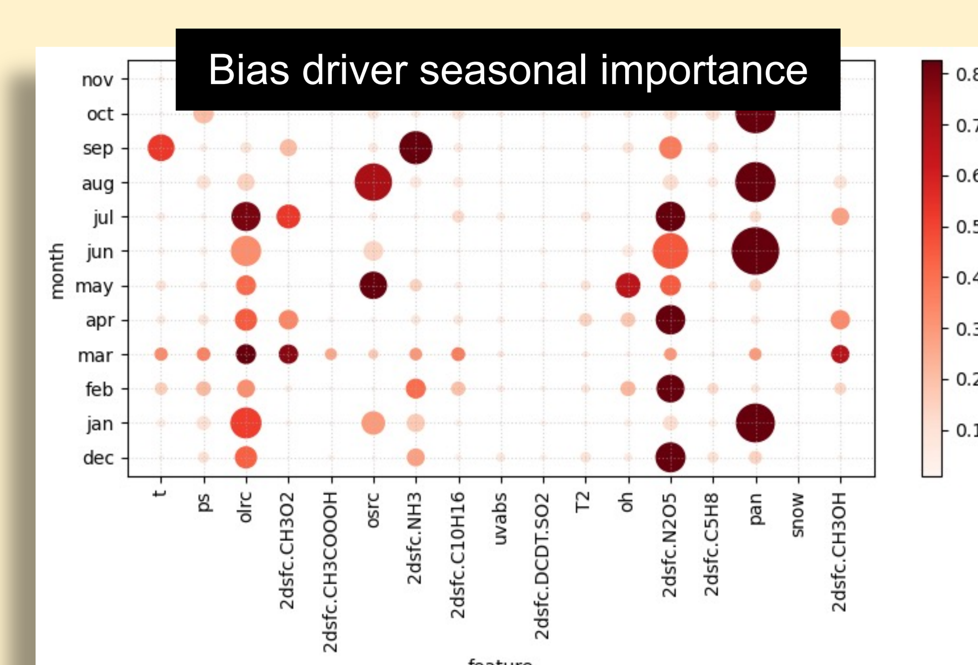
Seasonal baseline ML ET model:
- Conditional feature importances
- Conditional feature contributions
- Permutation importances
- SHAP explanations

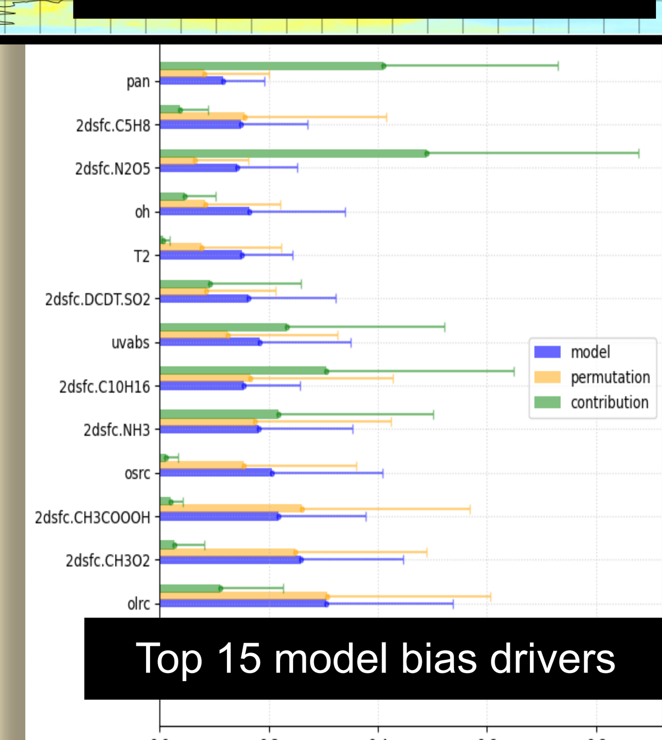
→ Local/regional ML model → Validated ML model with uncertainty

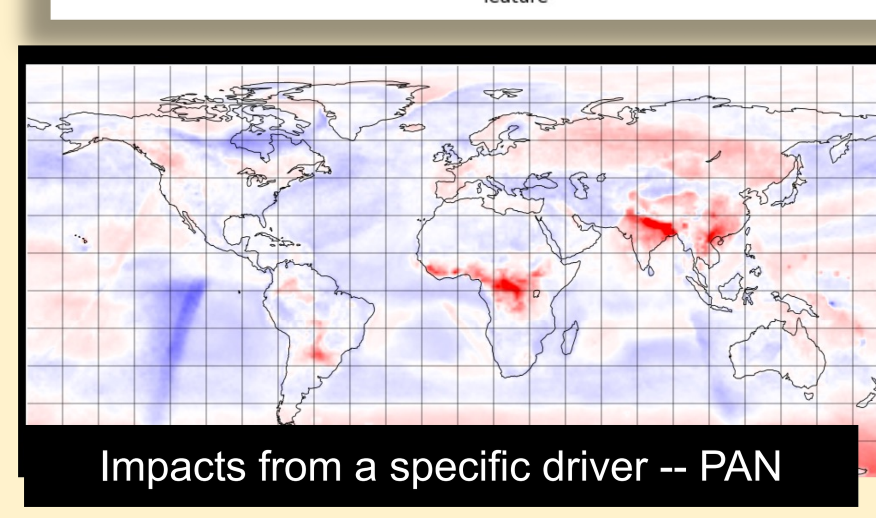
Results:

The key results obtained are 1) Global predictions of daily model ozone bias for 5 years, 2) global/local explanations of bias drivers, 3) physical interpretation of the drivers.









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
- Generalizable ML framework for evaluating any Earth system models

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California
www.nasa.gov

Clearance Number: CL#
Poster Number: RPC#R22004

PI/Task Mgr. Contact Information:

Email: Kazuyuki.Miyazaki@jpl.nasa.gov