



Simulations of atmospheric trace gases using machine learning

Imran A. Girach^(1, 2), Narendra Ojha⁽³⁾, Sankar Jyoti Nath⁽⁴⁾, S. Hareethasree^(3, 5), S. Suresh Babu⁽⁶⁾, Radhika Ramachandran⁽⁷⁾

- (1) Space Applications Cntre, ISRO, Ahmedabad-380015, India, e-mail: imran.girach@gmail.com
- (2) Previously at Space Physics Laboratory, Vikram Sarabhai Space Centre, ISRO, Thiruvananthapuram-695022, India; e-mail: imran.girach@gmail.com
- (3) Physical Research Laboratory, Ahmedabad-380009, India; e-mail: ojha@prl.res.in; harithasree97@gmail.com
- (4) Centre for Atmospheric Studies, Dibrugarh University, Dibrugarh-786004, India; e-mail: sankar6apr@gmail.com
- (5) Indian Institute of Technology, Gandhinagar, India; e-mail: harithasree97@gmail.com
- (6) Space Physics Laboratory, Vikram Sarabhai Space Centre, ISRO, Thiruvananthapuram-695022, India; e-mail: s_sureshbabu@vssc.gov.in
- (7) Former scientist at Space Physics Laboratory, Vikram Sarabhai Space Centre, ISRO, Thiruvananthapuram, India; e-mail: 601sree@gmail.com

Trace gases in the troposphere are imperative in context of global warming, air pollution and geosphere-biosphere interactions. Modeling is essential to unravel the dynamical and chemical processes governing the variabilities in trace gases. The conventional models employing physics and chemistry are computationally intensive and can typically have stronger biases owing to uncertain emissions or parameterizations. In this context, the Artificial intelligence (AI) and Machine learning (ML) have emerged as a powerful and computationally less-expensive modeling technique in various research fields including the earth system science. Here, we explore the capability of ML to simulate the variabilities in carbon dioxide (CO_2) and ozone (O_3) at surface level. The model was tuned by adjusting the hyperparameters and avoiding the overfitting. The model, trained with various predictors (i.e., input parameters) reproduced 72% of the variabilities in CO_2 with small root mean square errors (RMSE) of 0.45 ppm over a background site, Mauna Loa. Variabilities in surface O_3 over two distinct urban cities, Ahmedabad in western India and Dehradun in northern India are successfully simulated with squared correlation $r^2 \geq 0.83$ and RMSE ≤ 6 ppbv. Interestingly, though the ML model does not include chemistry, the analysis of feature importance (i.e., importance of input parameters) could reveal the crucial role of precursor gases. Sensitivity simulations showed that model performance is depended strongly on the length of training data and the number of relevant input parameters. Our study demonstrates the potential of ML for simulations of key trace gases which can be extended for chemistry-climate studies to complement the conventional earth system models. We recommend ensemble of various ML models for more robust modeling of trace gases by integrating ground- and satellite-based observations.