Predicting OH Measurements from In-Situ Aircraft Campaigns Using Machine Learning

Adiel Felsen

Code 614, Julie Nicely, Bryan Duncan

NASA Goddard Space Flight Center

Introduction

In 2020 and 2021, global methane concentrations have been increasing at an accelerated rate. While CO2 is still the largest contributor to climate change, methane (CH4) can be 25x as effective at trapping heat in our atmosphere [1].

The hydroxyl radical (OH) is a major sink for methane, but its volatility in the atmosphere makes it difficult to track and measure [2]. Understanding the presence of OH in the atmosphere is an important step towards explaining methane concentrations and combatting climate change.

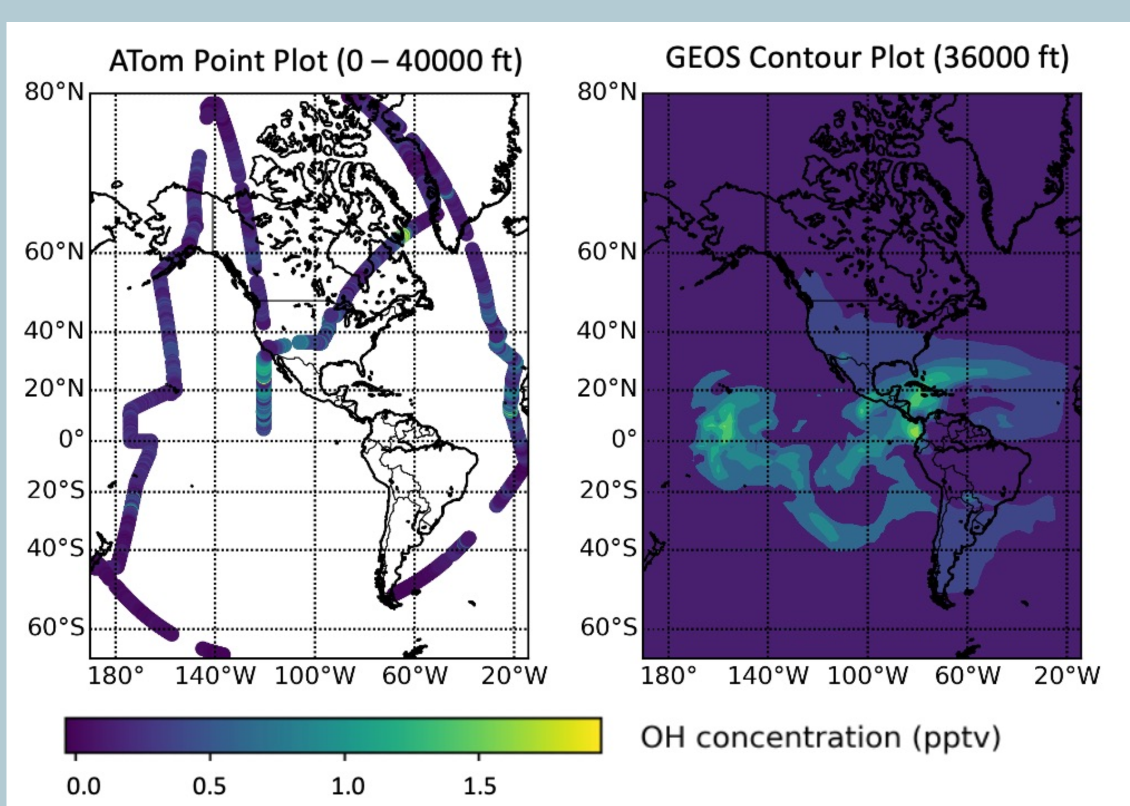

Fig 1. Graph of OH concentrations over North and South America . The NASA ATom campaigns (left) gathered in-situ measurements of several chemicals in the atmosphere. The GEOS model (right) calculates chemical concentrations based on known chemistry.

Over the past decade, machine learning has become a popular technique for extracting information from large and complex datasets. Using deep neural networks, this research takes advantage of large datasets produced by a global atmospheric model and in-situ aircraft campaigns in order to better understand atmospheric OH concentrations.

Results

Using GEOS model data, a neural network was trained to predict OH concentrations based on the concentrations of other atmospheric variables – CO, NO, O3, H2O, J_O3_O1D and J_NO2. The model not only performs well on a separate GEOS testing dataset, but also produces fair predictions on Aircraft measurements.

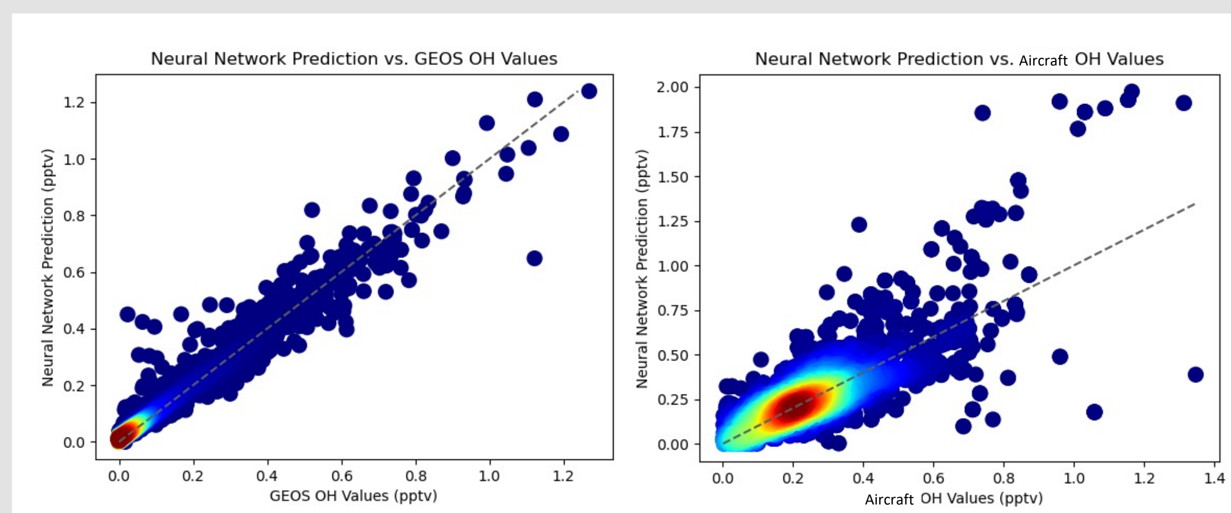

Fig 2. Accuracy comparison of a neural network trained on the GEOS training dataset. the neural network's predictions are plotted against GEOS OH values (left) and Aircraft OH measurements (right). Color represents the density of datapoints (red is high density).

Neural networks are a great regression tool, but their predictions are generally viewed as black boxes. Shap Values and Layerwise Relevance Propagation (LRP) are two metrics that can help explain a network's prediction. Figure 3 indicates that H2O and J_O3_O1D are the most influential variables for our model to predict OH in the atmosphere.

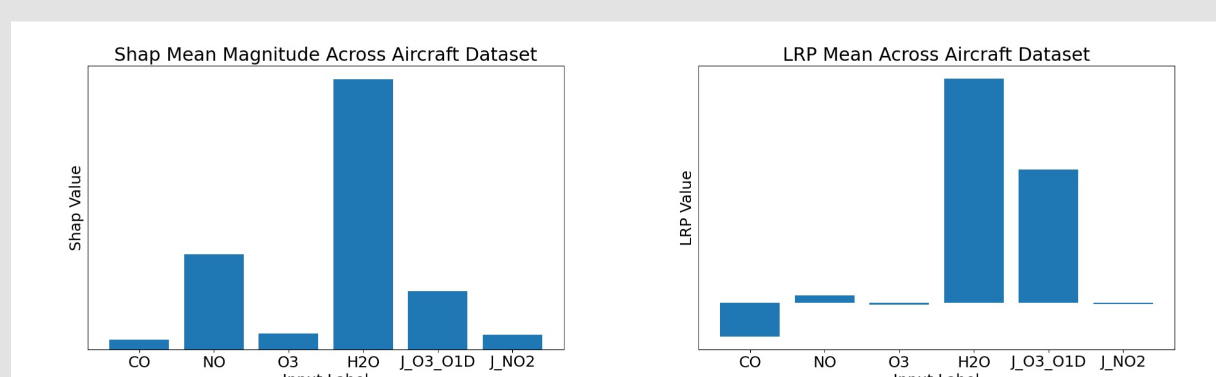

Fig 3. Mean shap values (left) and LRP values (right) attempt to explain the neural network's prediction on the Aircraft data. These values provide some insight, but they can be inconsistent and are not fully reliable.

Methodology

This research considered the use of boosted trees and neural networks for detecting OH concentrations. Both machine learning models are generally capable of producing accurate results for this project. Ultimately, a neural network was used for the final model due to its increased flexibility with data processing.

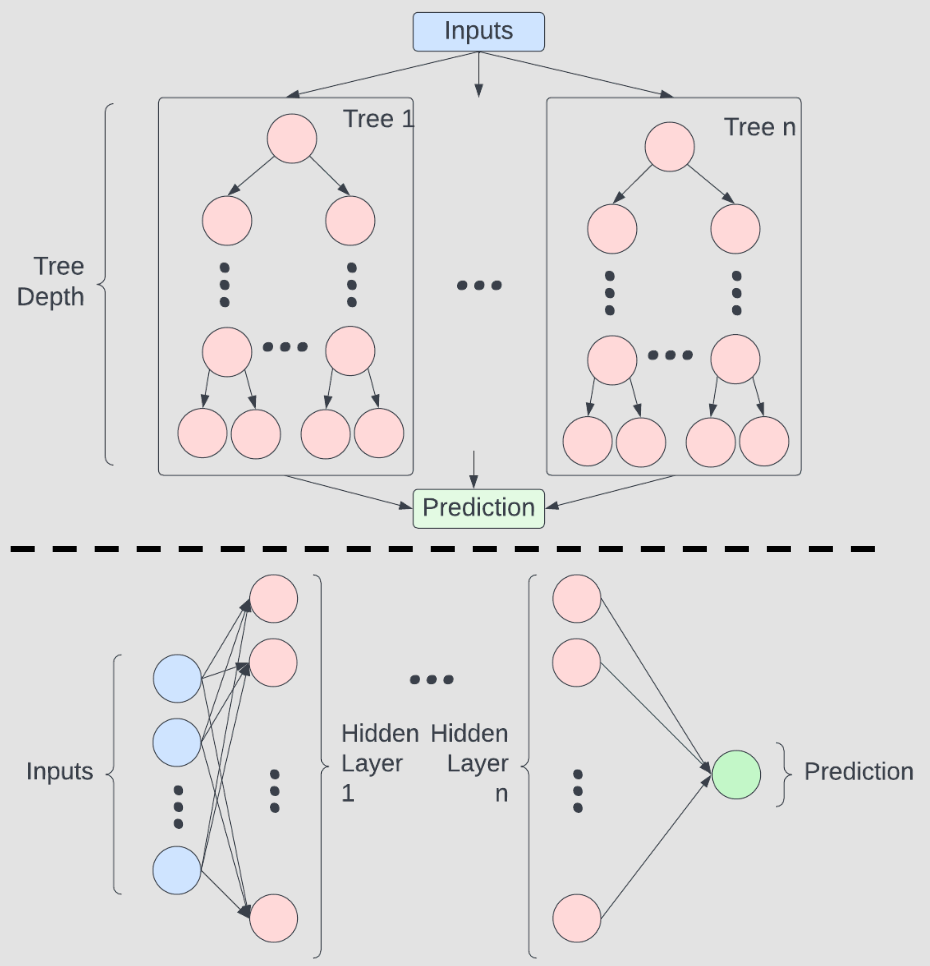

Fig 4. Basic architecture of a boosted tree (top) and a neural network (bottom)

This work utilizes two datasets: (1) a 24-hour run of the GEOS model and (2) a compilation of several aircraft campaigns.

The GEOS model output was adjusted to closely match the format of the aircraft campaign data. Latitude, longitude, pressure, and temperature were excluded from the datasets to encourage the model to focus on chemical interactions. Methane was also excluded due to unreliable methane values produced by the GEOS model.

The model was trained exclusively on a GEOS training dataset. While training our model, normalization, weighted sampling, and random data jitter were used to improve the performance. The model was tested on a separate GEOS testing dataset and on the aircraft measurements.

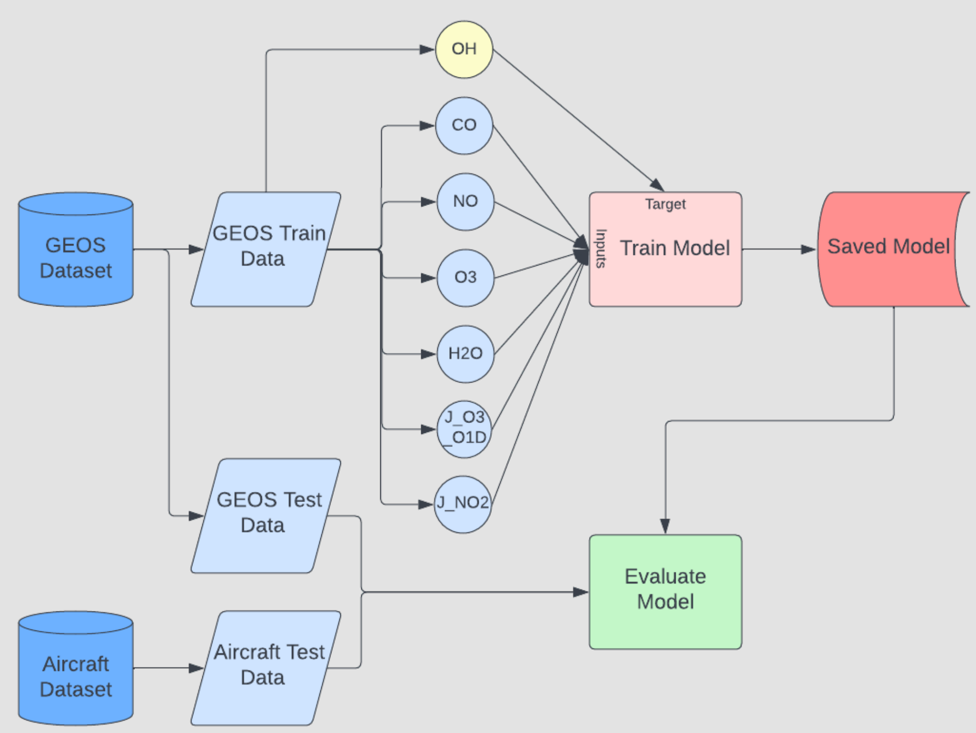

Fig 5. Diagram of the data division and training/testing procedure.

Conclusion

This work demonstrates that artificial model data can be used with machine learning to predict real measurements of OH in the atmosphere.

The GEOS model data is more consistent than the relatively noisy Aircraft data. Table 1 demonstrates that data jitter can improve the model's prediction on the Aircraft data.

Data Jitter	0%	5%	10%	20%
GEOS R^2 Value	**0.9697**	0.9693	0.9629	0.9543
Aircraft R^2 Value	0.1858	0.2285	0.2580	**0.3577**

Table 1. Random data jitter to the GEOS training data impacts model performance. Evaluating on the GEOS testing set, the model performs best with minimal data jittering. Evaluating on the Aircraft data, some data jittering appears to be beneficial.

Even with some jitter, the GEOS model data is not fully compatible with the Aircraft measurements. Figure 6 shows one example of a difference between the GEOS model output and aircraft measurements. In the future, improving the data compatibility could lead to more accurate results.

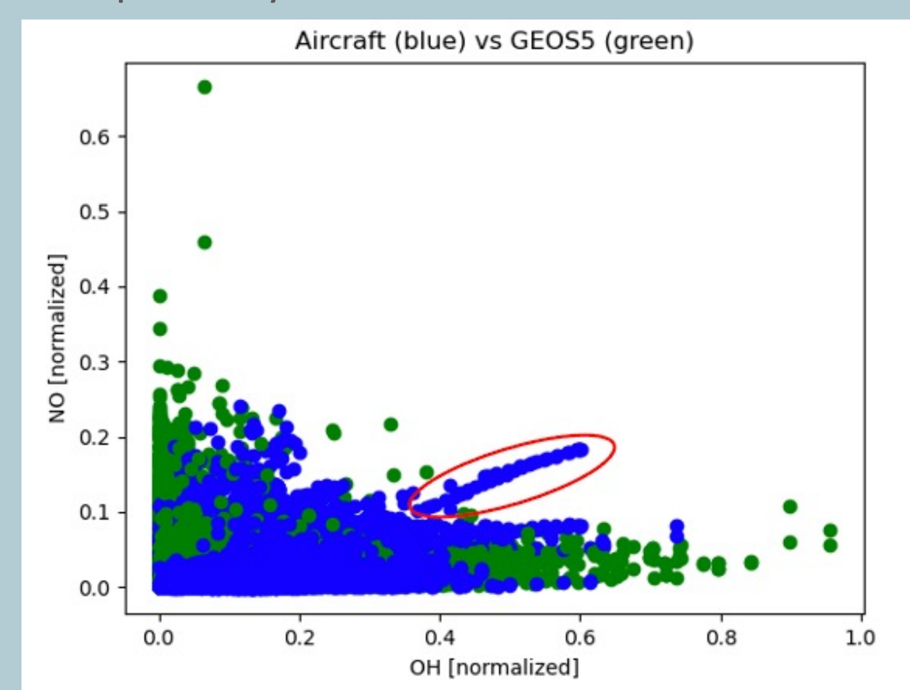

Fig 6. NO vs. OH values for Aircraft data (blue) and GEOS data (green). A section of Aircraft data (circled in red) does not overlap with the GEOS data.

Acknowledgements

[1] Stein, Theo. "Increase in Atmospheric Methane Set Another Record during 2021." National Oceanic and Atmospheric Administration, 7 Apr. 2022, https://www.noaa.gov/news-release/increase-in-atmospheric-methane-set-another-record-during-2021.
[2] Saunois, Marielle, et al. "The global methane budget 2000–2017." Earth system science data 12.3 (2020): 1561-1623.

Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center.

We acknowledge the Science Teams of the following NASA aircraft campaigns, which provided the publicly available observational data used in this study: ATom, INTEX-B, TRACE-P, and PEM-Tropics B.

We acknowledge USRA for administration of the NASA Summer Internship program