Machine-learned Gas Optics with a Focus on Geostationary Extended Observations (GeoXO) for Improving Water Vapor Observations in the Lower Atmosphere

In the grand scheme of the earth-atmosphere system, there are few constituents more vital and mysterious than water vapor. Vital because of its interwoven thermodynamic, radiative, and dynamic influence on the weather and climate of the planet, and mysterious because of our limited capacity in observing its time evolution in horizontal and vertical space. The advancements in the spectral and radiometric accuracy of next-generation hyperspectral infrared sounders are expected to bring unprecedented value to our observational capability with improved profiling of lower tropospheric water vapor where it is most abundant.

Essential to performing satellite observations, and their assimilation to dynamical models is the accurate and efficient radiative transfer calculations. In this process, calculating the atmospheric absorption by various gases is the most important step. The ‘line-by-line’ approach of summing the influence of every absorption and emission line is operationally impractical where the radiative transfer solver is evoked multiple times for every observation that is made or assimilated. The existing forward models, therefore, parameterizes the gaseous absorption using methods like pre-computed lookup tables or regression methods.

Here, we present a new method of performing gas absorption calculations using machine learning. We train neural networks to emulate the line-by-line layer optical depths on a consistent grid of 100 atmospheric layers defined by 101 pressure levels spanning from 1100 hPa to 0.005 hPa. We sample a diverse set of 8640 profiles around the globe for the year 2014 from the ERA5 reanalysis dataset and use 80% of these profiles as training data and 20% of the profiles as validation data. We test the performance using a set of 83 profiles from ECMWF for the year 2006-2007, known as ECMWF83.
profiles. These profiles are commonly used for training the existing regression-based parameterization by the remote sensing community around the world. The true optical depths are calculated using the line-by-line model MonoRTM.

Just like any other machine-learning-based application, the performance of our predictions is limited by the representativeness of the training dataset. We found that compared to the testing data, our training data is not well represented for the last 9 layers having atmospheric pressures less than 1 hPa. As a result, the prediction produces a large mean percent error of 12%. However, analyzing the error characteristic we found that the error is mainly concentrated in these last 9 layers. For the rest of the layers with atmospheric pressures greater than 1 hPa, the prediction improves significantly and the mean percent error reduces to 0.8%. The mean percent error further improves to 0.15% when predicted line-by-line values are converted to channel layer-to-space transmittance profiles that are used in the operational forward radiative transfer models. We also demonstrate that the method can be very useful for assessing the usefulness of future hypothetical sensors by showing the ability of predicted values in accurately calculating the channel weighting function. With the incorporation of relevant absorbing constituents, the method presented here can be applied to any spectral region.

**Microsoft Teams meeting**

Join on your computer, mobile app or room device

[Click here to join the meeting](#)

Meeting ID: 260 283 985 71
Passcode: GFtSkN
[Download Teams](#) | [Join on the web](#)

Or call in (audio only)

+1 970-628-0547,,672766274# United States, Grand Junction
Phone Conference ID: 672 766 274#
[Find a local number](#) | [Reset PIN](#)

[Learn More](#) | [Meeting options](#)