

Fast Machine Learning Lidar Surrogate Simulator: Pristine Clear Sky

Shan Zeng^{1,2}, Bing Lin², Ali Omar², Yongxiang Hu²
 1. Coherent Applications, Inc., Hampton, VA, United States
 2. NASA Langley Research Center, Hampton, VA, United States.

ABSTRACT

The simulations of lidar signals and retrievals rely on a range of opto-physical models, such as radiative transfer models, particle scattering and absorption models, along with the output data from atmospheric physical models. Integrating these different models to represent signals of a lidar system is computationally expensive, and performing backward retrievals can be complex and ambiguous. However, with the advantages of Machine Learning, there is a new potential for building effective lidar signal database for a CALIPSO-like lidar (355 nm, 532 nm, and 1064 nm), and a CO₂ differential absorption lidar (DIAL) near 1571 nm. Specifically, we utilize a long short-term memory (LSTM) model to map the relationships between atmospheric profiles (pressure, temperature, air density and CO₂ mixing ratio) and lidar signals. This approach allows us to build machine learning based simulators that can reconstruct lidar signals at specific bands from MERRA reanalysis data, and perform retrievals of atmospheric profiles using lidar signals at various wavelengths. As a first step, the results show the potential of this method to establish a foundational model for sensor signals. This model offers the promise of enabling both accurate predictions and rapid retrievals, providing a more efficient approach to signal processing and analysis.

Workflow & Methodology

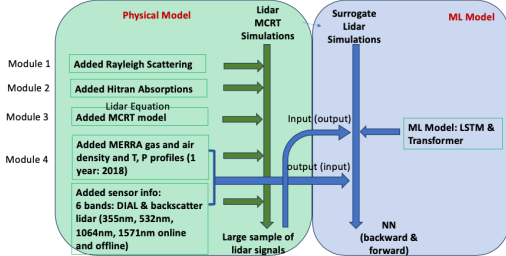


Fig. 1: Physics-based lidar simulator vs. ML-based surrogate lidar simulator

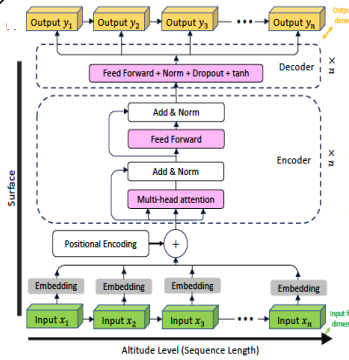


Fig. 2 Schematic of the Transformer Model

ML Model and its Configurations

Model Structure:

Hidden_layer_size = 200
 embedding_size = 512/2048
 Num_layer = 3
 Attention head = 4

6-D Loss Functions:

1. Min mean (Pred minus Ground Truth)
2. Min standard deviation (Pred minus Ground Truth)
3. Max correlation between (pred and Ground Truth)
4. Min of (Pred max - Ground Truth max)
5. Min of (Pred min - Ground Truth min)
6. Weighted mean (Pred minus Ground Truth)

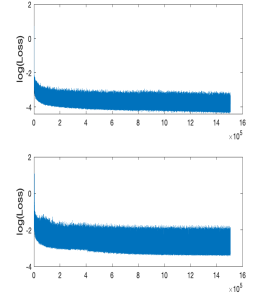


Fig. 3 The loss function in the training processes

Module 2=>

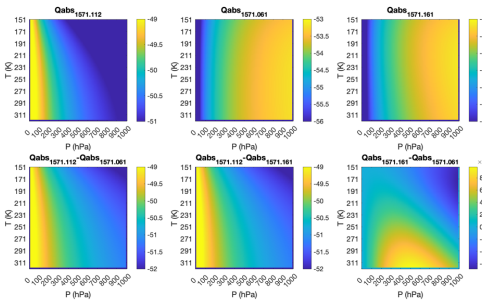


Fig. 4 The dependence of the CO₂ absorption coefficients (derived from HIRAN) on temperature and pressure at the CO₂ absorption lines near 1571 nm

Module 4=>

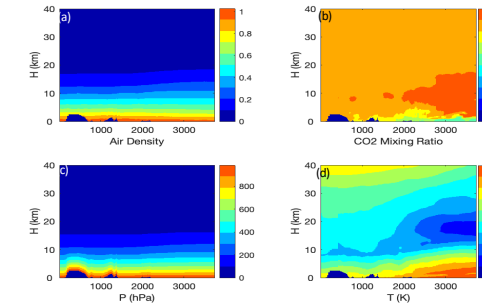


Fig. 5 Collocated MERRA outputs along one CALIOP track: Air Density (a), CO₂ Mixing Ratio (b), Pressure (c) and Temperature (d)

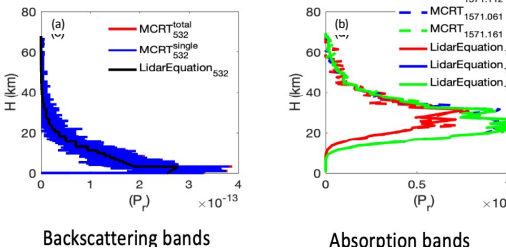


Fig. 6: Lidar signals obtained from MCRT simulations and lidar equation for backscatter lidar (a) and CO₂ DIAL (b)

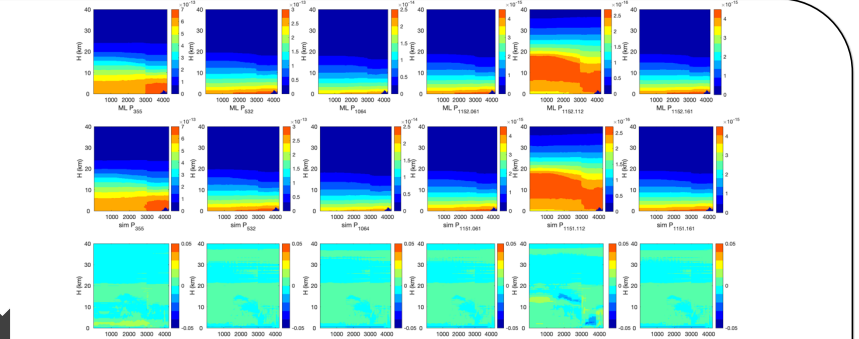


Fig. 7 Forward lidar simulations at 355nm (first column), 532nm (second column), 1064nm (third column), 1151.061nm (fourth column), 1151.112nm (fifth column) and 1151.161nm (sixth column) based on ML predictions (first row) and physical model simulations (second row), and the differences between the two (third row).

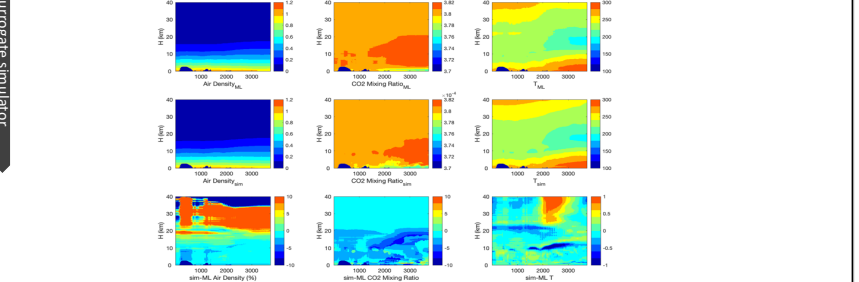


Fig. 8 Backward lidar retrievals of air density (first column), CO₂ mixing ratio (second column) and T (third column) based on the ML predictions (first row) and the ground truth (second row) and the differences between the two (third row).

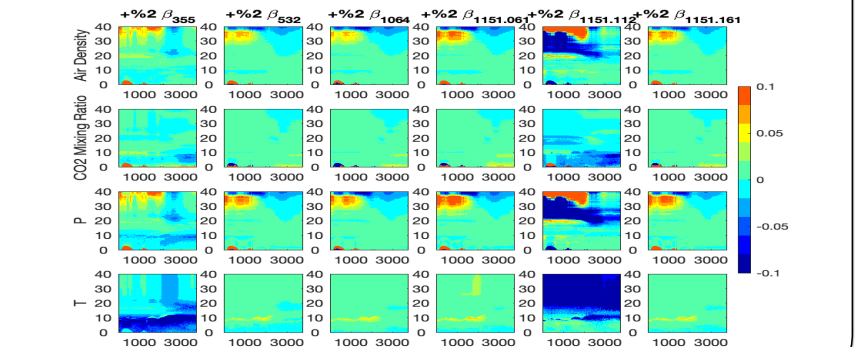


Fig. 9 Propagation of uncertainties in the retrievals of air density (first row), CO₂ mixing ratio (second row), pressure (third row) and temperature (third row). The uncertainties are within 2% for return signals at 355 nm (first column), 532 nm (second column), 1064 (third column) and CO₂ absorption band (online fifth column; offline, forth and sixth columns)

Conclusion The ML Transformer model proves to be an efficient surrogate for lidar radiative transfer simulations, and can handle both forward lidar signal simulations and backward retrievals of the atmospheric states. The uncertainty in forward simulations is under 2%, while the retrieval uncertainties are less than 10%, which may be acceptable for retrievals. Looking ahead, our future work will involve clouds, aerosols and other atmospheric compositions to complete the physical model and train it into a comprehensive lidar foundation model. The study also identifies certain key parameters for retrievals. CO₂ mixing ratio profiles are primarily derived from the CO₂ absorption band, air density profiles are influenced by signals at 355nm, and temperature profiles retrievals depend on signals at both the 355nm and CO₂ absorption bands. These insights provide a theoretical foundation for future work in profiling CO₂ mixing ratio, temperature and air density, offering a promising direction for advancing sensor signal modeling and retrieval techniques.