Parameterization of Vertical Cloud Distribution from C3M and MERRA Data Using ML Method

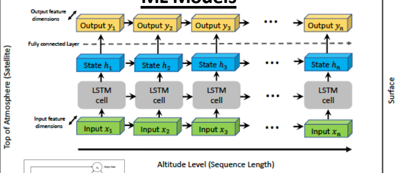
Shan Zeng[1,2], Kuan-Man Xu[2], Yongxiang Hu[2],

INSTITUTIONS : 1. Coherent Applications, Inc. , Hampton, VA, United States
2. NASA Langley Research Center, Hampton, VA, United States.

ABSTRACT

Clouds play a key role in regulating the hydrological cycle and Earth's radiative budget. However, global climate models (GCMs) with horizontal gridspacings of ~100 km cannot represent the subgrid-scale cloud dynamics on the order of kilometers, which can introduce uncertainties in cloud radiative feedback on a global scale. In our study, we will use a Deep Machine Learning (DML) method for physical parameterizations to duplicate the subgrid-scale cloud physical processes. We investigate the volumetric cloud fraction (VCF), which is the frequency of occurrence on a grid volume accumulated in the horizontal and vertical directions obtained from the NASA CALIPSO-CloudSat-CERES-MODIS (CCCM) satellite data and 3-D MERRA-2 meteorological data (i.e., Wind, Relative Humidity, Temperature), and we are able to recover their complicated relationships with the Sequence-to-Sequence DML method on a day-to-day-based analysis. Our preliminary results show that the DML model can learn the cloud physical processes and represent well their relationships based on the statistical, horizontal and vertical distribution results.

ML Models



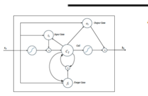


Figure 1. Schematic of LSTM Sequence-to-Sequence Model.

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)

Inputs and Outputs

Inputs (MERRA)	Outputs (CCCM)
U Profiles	CCCM Volume Cloud Profiles
V Profiles	
ω Profiles	
T Profiles	
RH Profiles	
P Profiles	
Turbulent Latent Heat Flux	
Turbulent Sensitive Heat Flux	
Surface T, RH, P	

Table 1. List of ML training inputs and outputs.

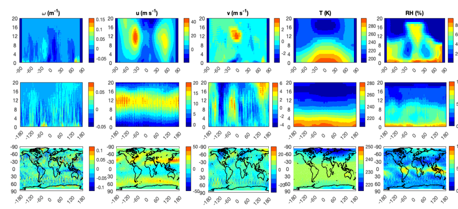


Figure 2. Distribution of Inputs: 3-D winds, temperature and relative humidity.

Training data	(2008-01) 90%
Validation data	(2008-01) 10%
Test data	(2008-02)

Table 2. List of data sample for training, validation and testing.

Discussion of Results

Vertical Distribution

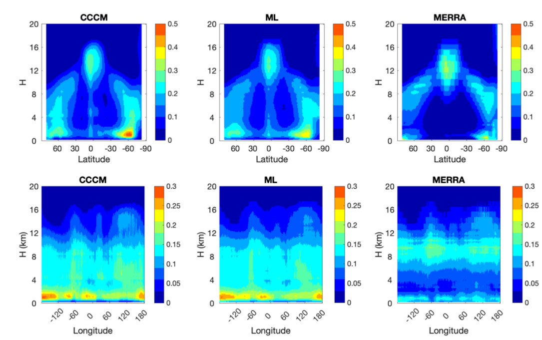


Figure 3. Vertical distributions of cloud cover against latitude (left raw) and longitude (right raw) from CCCM, ML Predictions and MERRA simulations.

Geographical Distribution

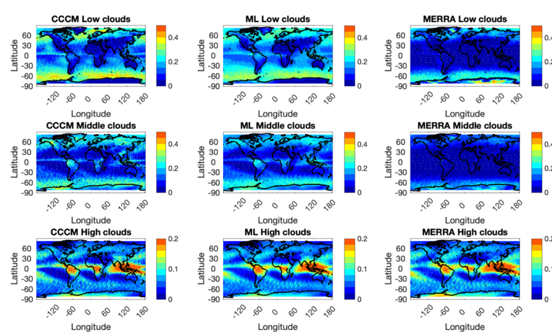


Figure 4. Geographical distributions of cloud cover of low (top row), middle (middle row) and high (bottom row) clouds calculated from CCCM, ML Predictions and MERRA-2 simulations.

Orbit case

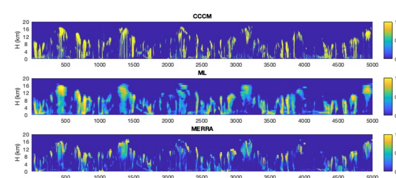


Figure 5. Example of Vertical Cloud Mask from CCCM observation (top), ML prediction (middle) and MERRA simulations (bottom) on May 3rd, 2008.

Uncertainties

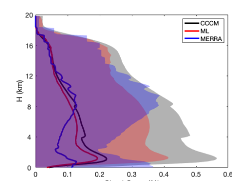


Figure 6. One-month of statistical cloud cover profiles from CCCM (black), ML Predictions (red) and MERRA simulations (blue).

Seasonal Cycle

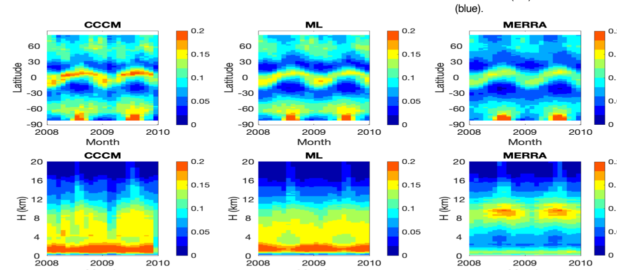


Figure 7. Seasonal variations of cloud cover against latitude (top row) and altitude (bottom row) from CCCM, ML Predictions and MERRA simulations.

Enso Cycle

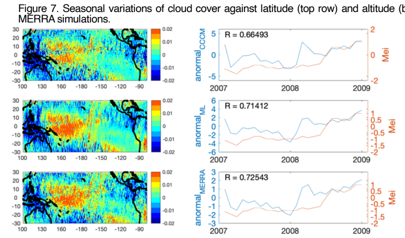


Figure 8. Principal component analysis of cloud cover for CCCM, ML Predictions and MERRA simulations.

Future work

- Merge into climate Foundation Model or GCM to estimate the improvement of radiative forcing due to the improvement of CF
- Merge into satellite Vision Foundation model or Radiative Transfer simulations to reconstruct lidar signal
- Continue to parameterize the vertical liquid and ice cloud fraction, drizzle fraction as well as cloud microphysics

Conclusion:

Sequence-to-Sequence ML model can learn the cloud dynamics and correctly represent cloud cover at different altitude, longitude, and latitude in different large-scale dynamic conditions. We also tried feedforward neural network, from which model, the training could not succeed. Selecting a correct ML model that can learn and memorize the beneath physical processes is the first key for ML studies. From this study, we can see relative humidity (RH) is the most important meteorological parameter. Our future work includes adding small or larger neighboring meteorological profiles to see how advection impacts the parameterization. We will also use the trained relationships to parameterize clouds in global climate model and see how it can improve the global climate model simulation.