Representing Aerosol-Cloud Interactions Using Machine Learning Techniques in Energy Exascale Earth System Model

Po-Lun Ma

October 1, 2020
Aerosol effect is a major source of uncertainty

IPCC AR5 (2014)

Kiehl (2007)

Smith et al (2020)

• CMIP3-5: Stronger aerosol effects lead to weaker total forcing and higher climate sensitivity
• CMIP6: New (no?) relationship
Increase resolution improves agreement between model and observational estimates of ACI

Ma et al. (2015)

- Increasing resolution improves small-scale meteorological features critical for aerosol-cloud interaction, increasing the agreement of cloud and precipitation susceptibility with observations even though the same physical parameterizations are used.

Terai et al. (2020)

- Weaker ERFaci is caused by a combination of weaker increase in LWP in non-raining clouds and a smaller fraction of raining clouds in UPCAM.
EAGLES

3 model development themes (Aerosol, Clouds, Computation)
1 cross-cutting activity (Testbeds)
Deliver improved representation to run on DOE’s GPU-based computers

Vision: To increase confidence in and understanding of the role of aerosols and aerosol-cloud interactions in the evolution of the Earth system using new modeling techniques that are scientifically robust and computationally efficient for global convection-permitting simulations.
Activation emulation
Silva et al.

Ridge
MSE = 0.048
$R^2 = 0.69$

XGBoost
MSE = 0.0014
$R^2 = 0.99$

DNN
MSE = 0.0008
$R^2 = 0.99$

ARG
MSE = 0.007
$R^2 = 0.96$
Use physics to regulate the emulator

Silva et al.
Physics regularization: reducing emulator complexity
Silva et al.

No Regularization
ARG Regularization

XGBoost Model Mean Squared Error

Number of Iterations
New DNN-based activation in E3SM runs smoothly
Silva et al.

**Cloud fraction**
- **DNN**
  - v1pg2_f2000 (yrs 2000)
  - Mean: 65.32, Percent: 65.40
  - ANN: Min = 7.88, Max = 96.67

- **ARG**
  - v1pg2_act_f2000 - v1pg2_f2000
  - Mean = -0.08, RMSE = 0.96
  - ANN: Min = 4.75, Max = 4.15

**Precipitation rate**
- **DNN**
  - v1pg2_f2000 (yrs 2000)
  - Mean: 3.06, Mean/day: 3.05
  - ANN: Min = 0.00, Max = 20.06

- **ARG**
  - v1pg2_act_f2000 - v1pg2_f2000
  - Mean = 0.01, RMSE = 0.19
  - ANN: Min = -3.27, Max = 1.57
Some new problems indicates (hidden) issues in the emulator
Silva et al.

- Column integrated droplet number concentration ($10^9$)
  - DNN: avg = 33.31
  - ARG: avg = 14.10
  - DNN-ARG: avg = 19.21

- ERFaci
  - DNN: avg = -2.45 W·m$^{-2}$
  - ARG: avg = -1.70 W·m$^{-2}$
  - DNN-ARG: avg = -0.75 W·m$^{-2}$
Software considerations when implementing emulators in E3SM
Singh et al.

• We use Fortran keras bridge library
  • Easy: Provide a text file, link the library, write an interface routine
  • Flexible: do not need to recompile the whole E3SM after replacing the emulator

• Numerical considerations
  • Clipping (input)
    • Assumptions made in other parts of the model: updraft velocity
    • Inconsistent bounds: minimum hygroscopicity in the emulator \(10^{-4}\) vs in E3SM \(10^{-10}\)
  • Clipping (output)
    • Fraction between 0 and 1
  • Sampling of training data
    • Comprehensive (Latin Hypercube, MCMC) vs. realistic multivariate PDF

• Computational cost
  • DNN on CPU machines is slightly cheaper than the default E3SM with ARG
Autoconversion emulation
Pressel et al.

Objective
• To develop robust training data sets to be used for machine learning emulation of autoconversion rates for a diverse set of aerosol conditions and boundary layer cloud regimes

Approach
• A new computationally efficient LES model, Predicting INteractions of Aerosol and Clouds in Large Eddy Simulation (PINACLES)
• Coupled with spectral bin microphysics (SBM) to explicitly predict autoconversion rates
• Perform a very large LES ensemble for a wide range of aerosol conditions and cloud types

Results
• A proof-of-concept exercise builds a reasonable emulator
• In addition to traditional variables, meteorological variables play a significant role in autoconversion rates and need to be included in the training
Summary

• Successful development of physically regularized DNN that outperforms all existing parameterizations

• Successful implementation of DNN-based parameterization in E3SM with ample flexibility. Computational cost is slightly reduced

• Interface routine has some numerical considerations needed for both the DNN emulator and for the E3SM

• Climate simulations are stable and reasonable, though we see new biases in droplet numbers, pointing to DNN’s relatively large bias in low activation scenarios.
Lessons learned

- Tools are available for inserting a DNN-based parameterization (python-based) in a climate model (FORTRAN-based), but interface routine needs to be designed carefully.

- Emulator development requires close collaboration between domain scientists and data scientists. E3SM with the DNN-based parameterization produces reasonable climate simulations, but new biases were produced. Some iterations between emulator-building (parameterization development) and climate modeling are needed.

- What are we learning? Statistical characterization of model states and process rates or real-world physics?
Thank you.