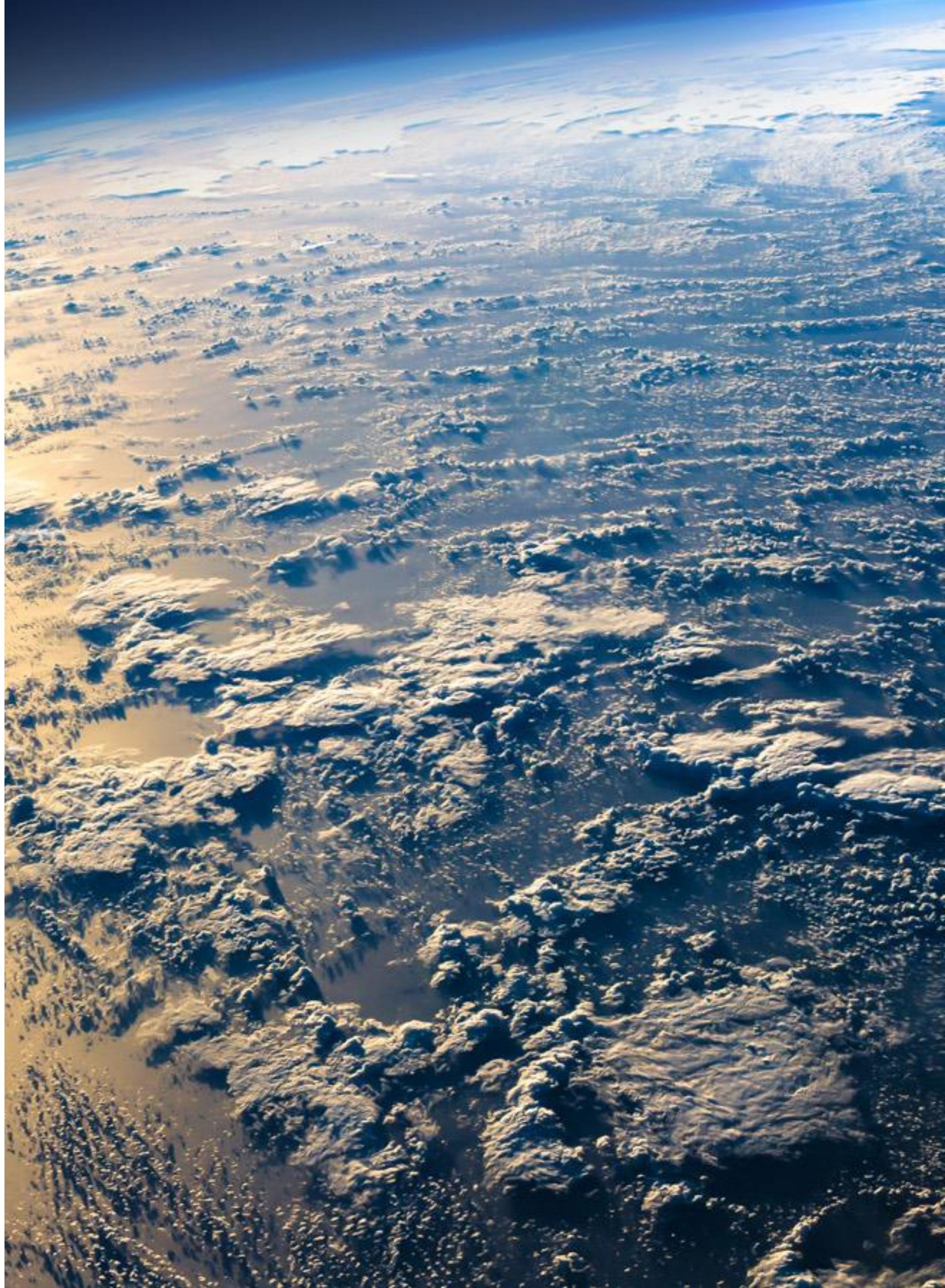
DREAM²
Drizzle Representation via Enclosed Atmosphere Measurements & Modeling @PNNL

Improving ESM Process Representations Using AI/ML

August 8, 2024

Andrew Gettelman, Po-Lun Ma, Susannah Burrows
Pacific Northwest National Laboratory



Summary and Outline

- AI/ML approaches are percolating into parameterizations
- Here, some E3SM Examples
- Theme: Off-Line Training, Emulation of Complex Codes

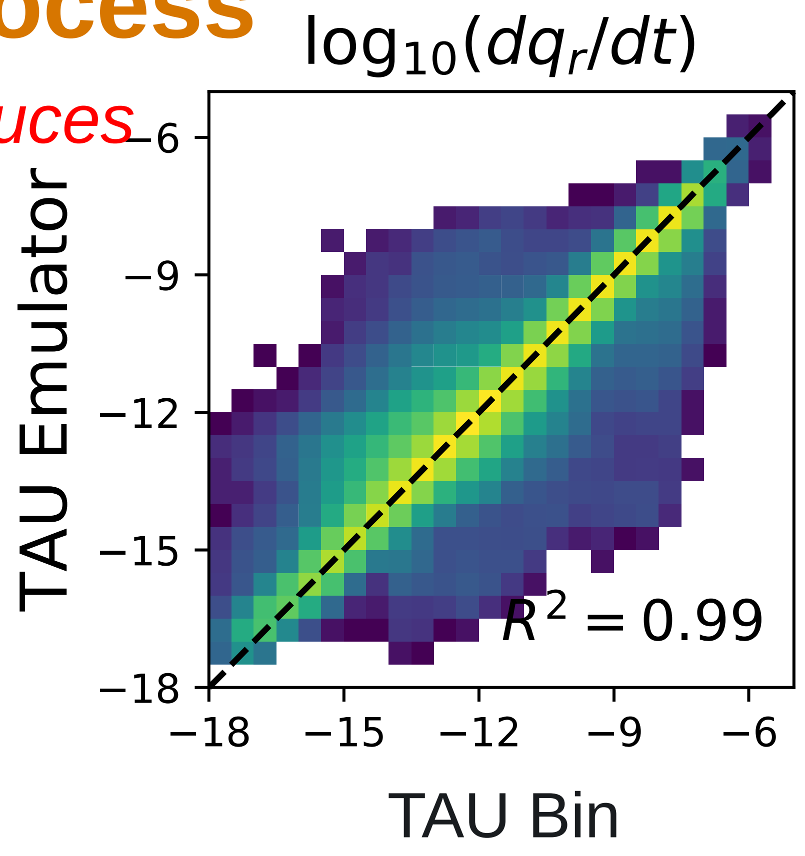
Outline:

1. A better treatment of rain formation
2. Learning Aerosol Optics and Aerosol Activation
3. A broader view of AI/ML challenges and opportunities

Machine Learning the Warm Rain Process

Initial CESM results

Emulator reproduces Bin code

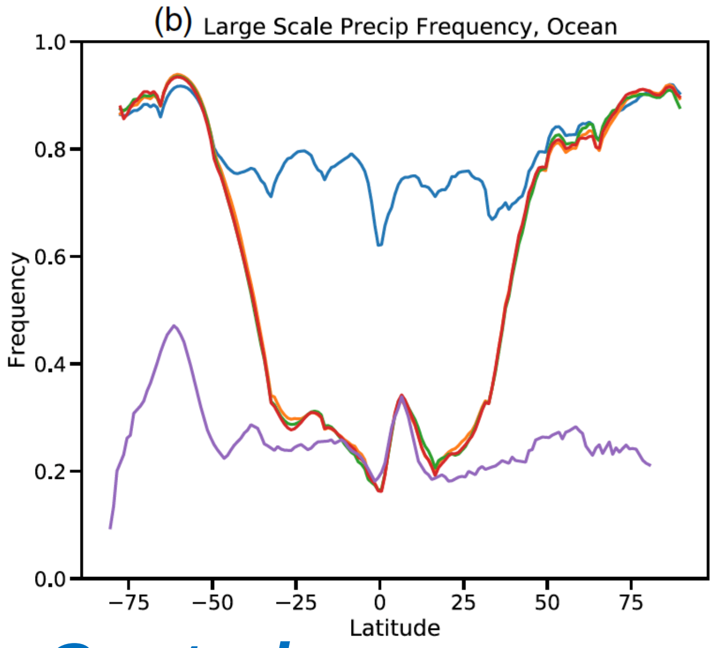


Can we do the warm rain process better with ML?

Replace traditional GCM bulk rain formation with a bin model formulation for stochastic collection. This is too expensive for climate use. Emulate it with a neural network.

Results:

- Bin code can improve rain formation results
- Recover speed and results with neural network emulator
- Embedded NN in the microphysics: maintains conservation with series of checks
- Implemented first in CESM, now ported to E3SM (not in master)



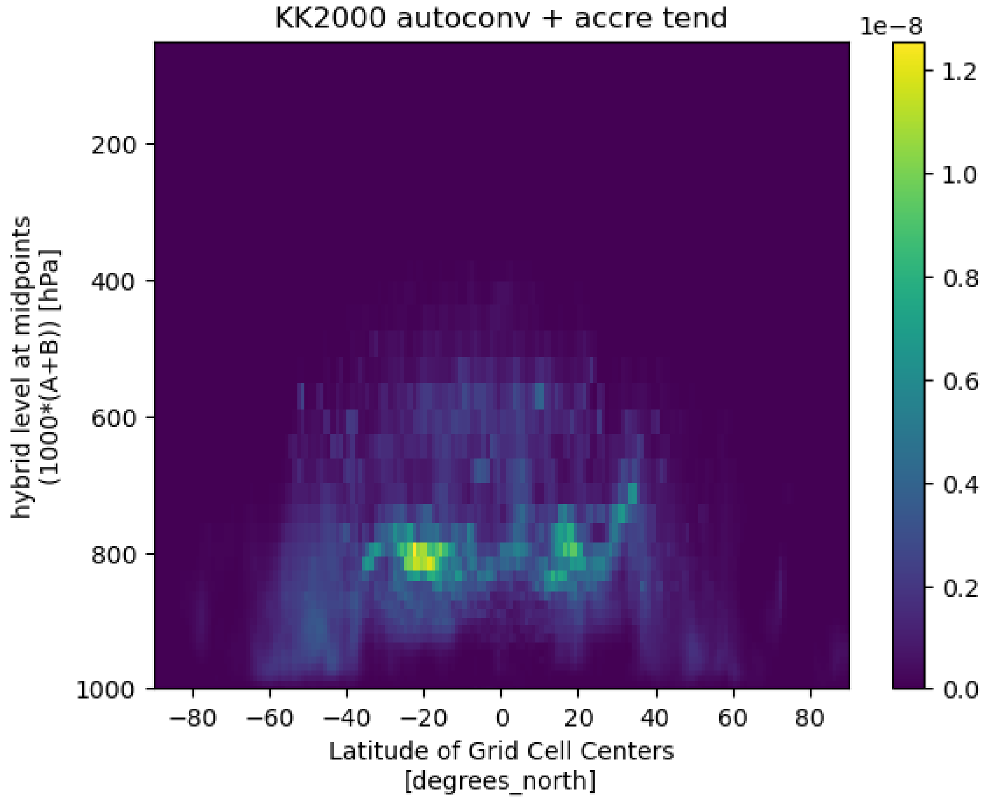
BIN** code closer to **Observations** than **Control

Gettelman et al 2021, JAMES

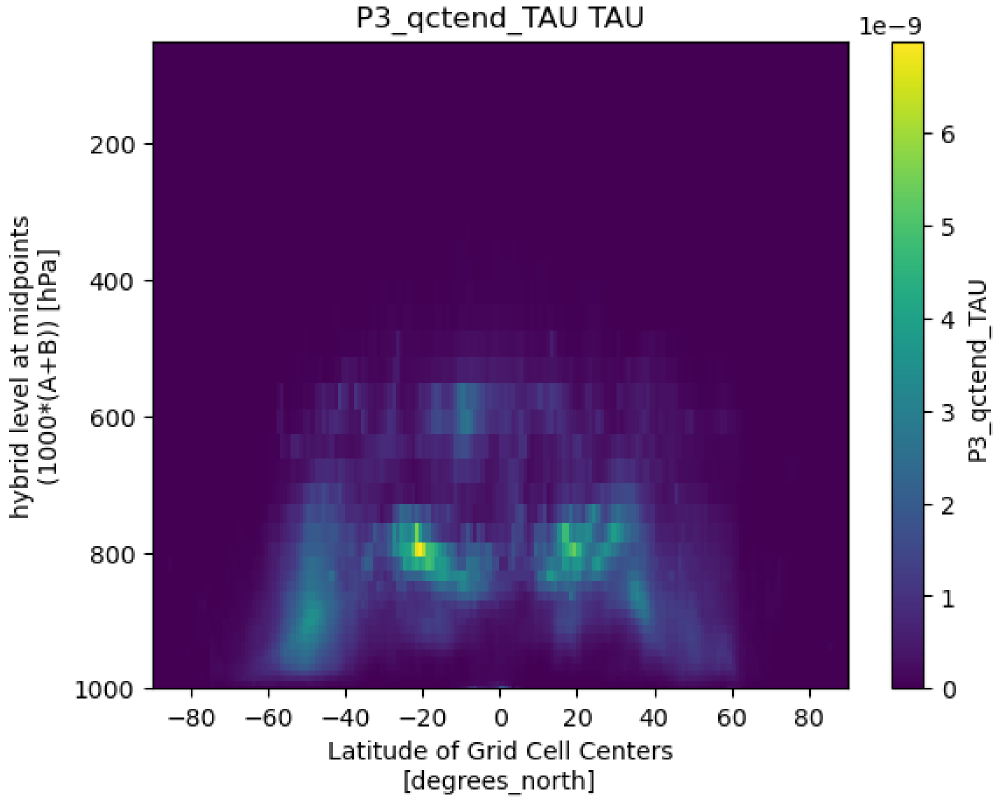
Pacific Northwest NATIONAL LABORATORY

Emulator Running Interactively in E3SM

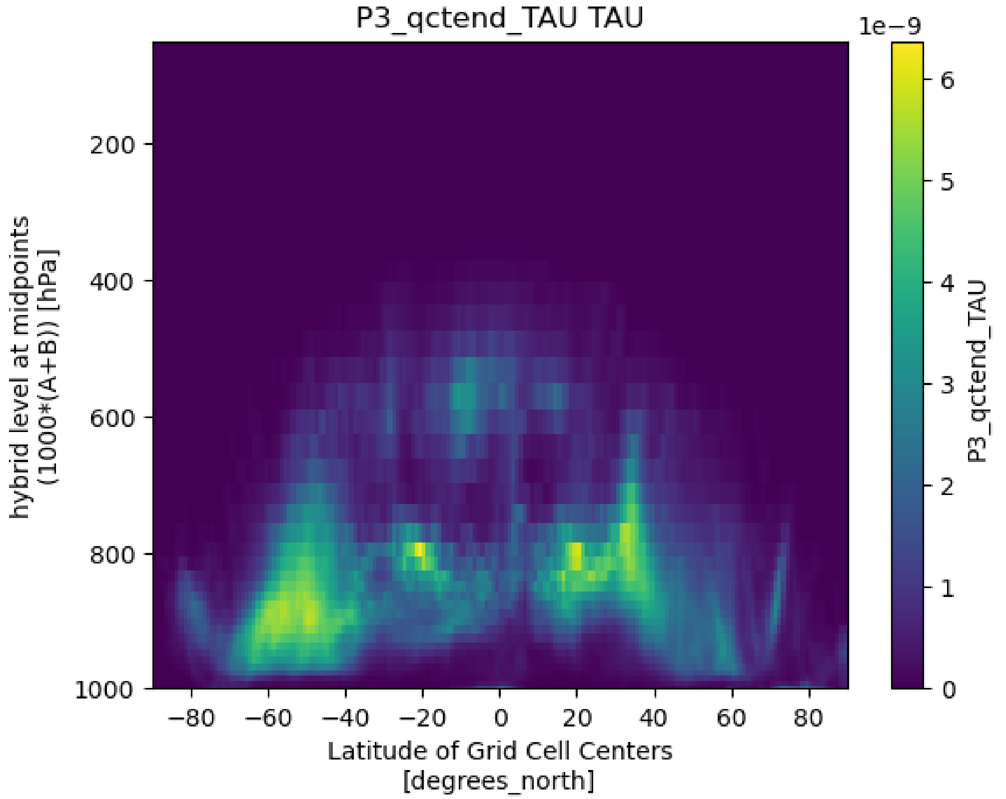
Bulk Rain Formation



Bin Rain Formation



Emulated Rain Formation

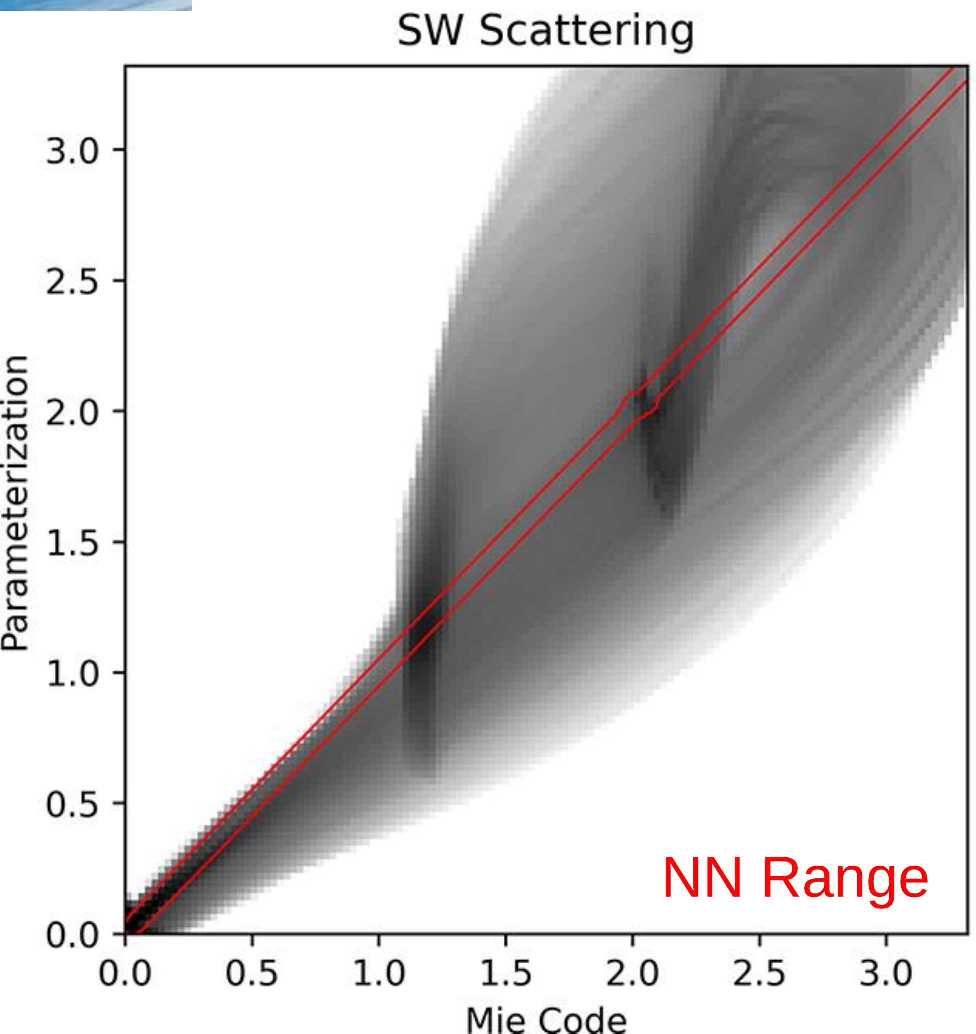


TAU and updated emulator (single NN) are running interactively in E3SM.
Emulator (from CESM) looks different than bin code in E3SM it is trying to emulate

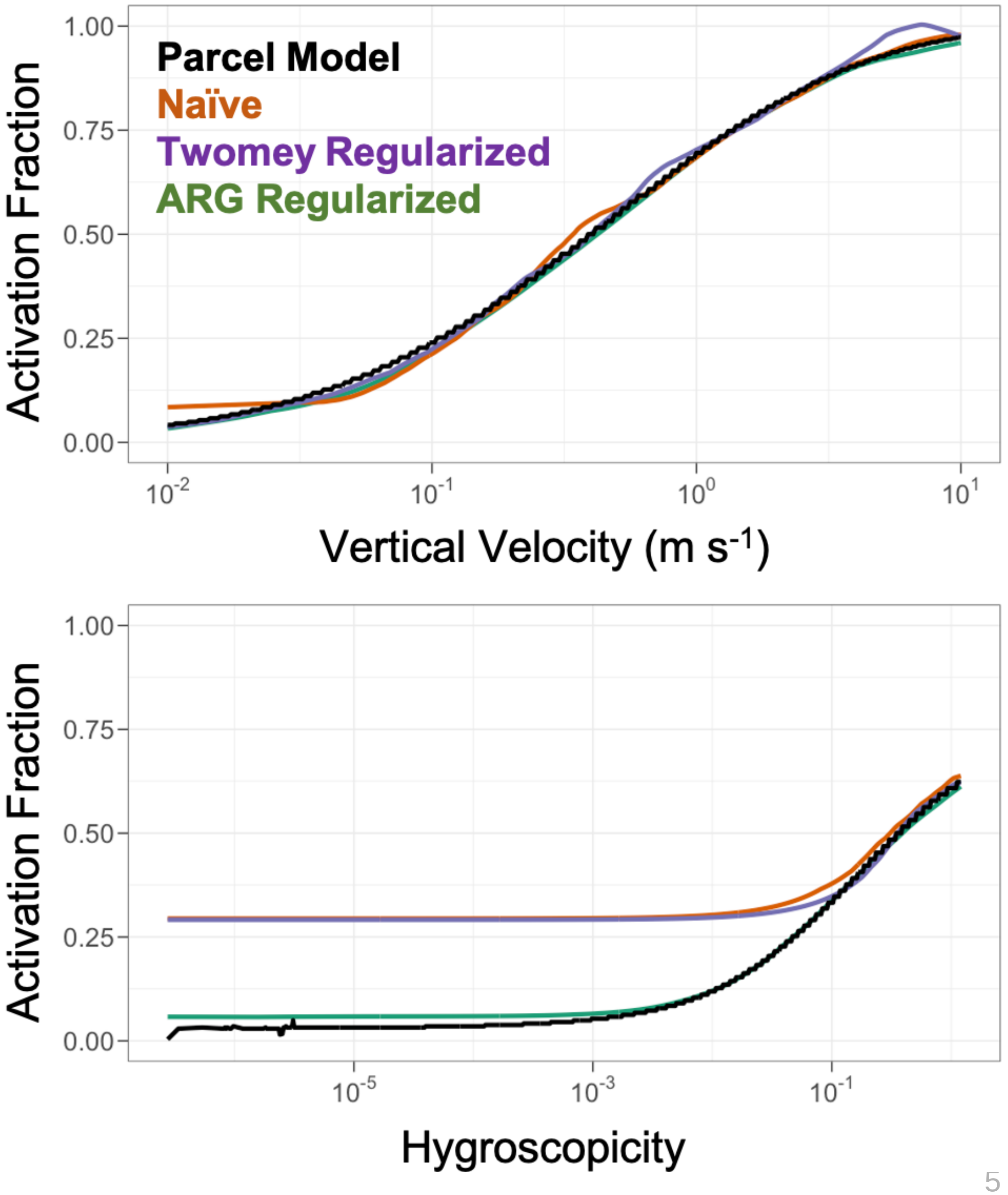
DREAM² Drizzle Representation via Enclosed Atmosphere Measurements & Modeling @PNNL

Machine Learning Aerosol Parameterizations

RIGHT: Machine Learning aerosol activation reproduces existing scheme when regularized (Silva, Ma et al 2021, GMD)



Left: Machine Learning aerosol optics better reproduces detailed Mie calculations than current parameterizations (Geiss, Ma et al 2023, GMD)



Strategery: Paths Forward with AI/ML
Opportunities, Challenges, Gaps

- Opportunity: AI Can simplify the model hierarchy for domain science
  - Train parameterizations off detailed models to use across scales
  - Use LLMs (ChatGPT, etc.) to do code translation between languages
    - Python <-> C++ <-> Fortran (JAX and Julia also options)
- Opportunity: Exploring ‘Foundation Modeling’ (mixed modeling) with AI/ML
  - Need to partner with Industry (Google, Microsoft, NVIDIA): develop coherent relationships
    - NeuralGCM type differential approaches (learning physics or components)
    - Foundation emulators: GraphCast, FourCastNet, etc: learning the whole model solutions
    - DOE offers expertise, computation, training data, interoperability. Industry: AI expertise, computation
  - Foundation models may replace ‘core’ parts of models (dycore, thermodynamic state)
  - Develop further models (aerosols, climate extremes, energy grid) from foundation models
  - *Will ML ‘eat the model’?* (now learning the solution still relies on the IFS/Reanalysis)
- Gap: AI and Observations: Simulators and Assimilation approaches
  - Very relevant for ARM data: to better compare to models, and to help with data processing
- Challenge/Gap: Robustness, Uncertainty, Interpretability (DOE can lead)