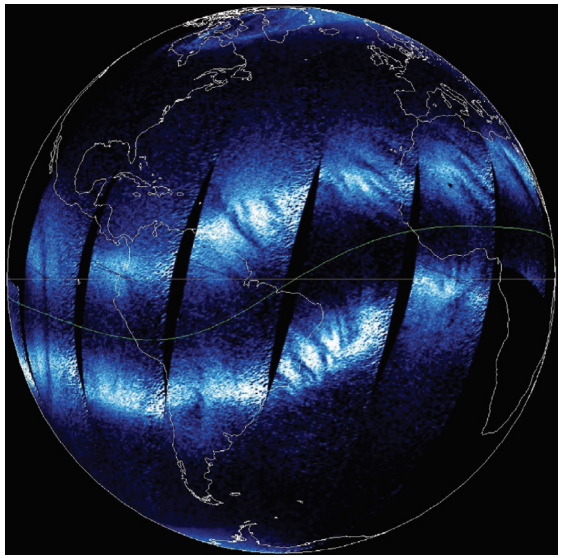
# NCAR's Progress in & Perspectives on Data Science and Machine Learning

**Dr. Rich Loft**
**Director, Technology Development**
**Computational and Information Systems Laboratory**
**National Center for Atmospheric Research**

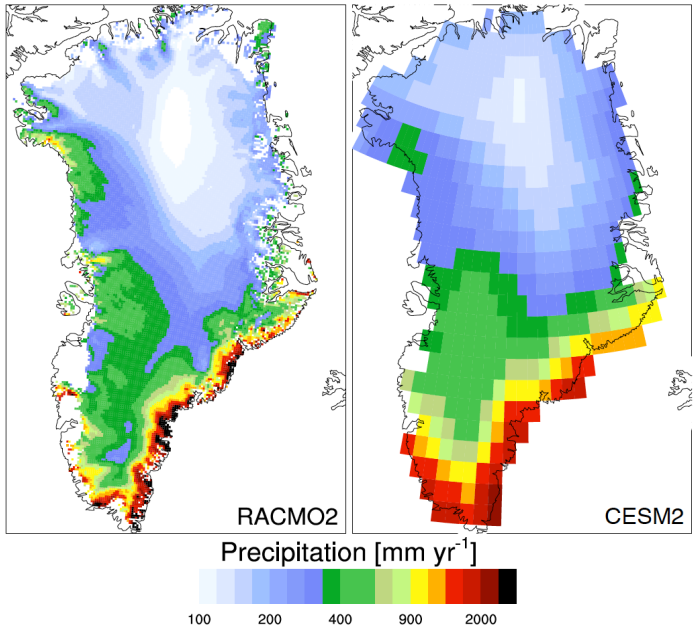
**Emerging Data Science and Machine Learning Opportunities in the Weather and Climate Sciences**
**AGU Washington, DC**
**December 13, 2018**


UCAR

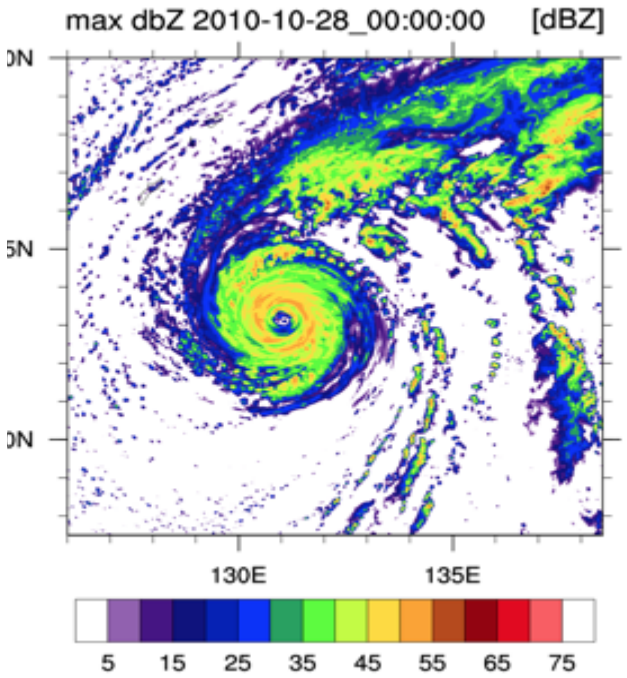
# NCAR future research needs outstripping HPC systems



Air quality prediction

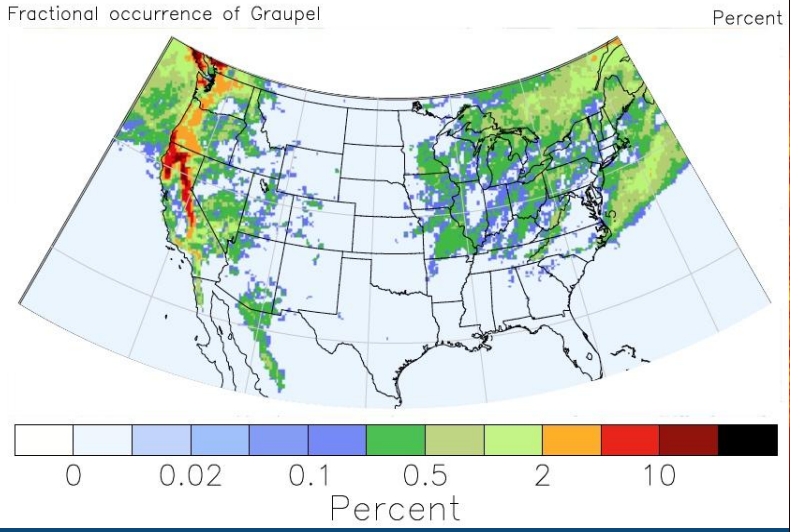



Coupled prediction of the Arctic

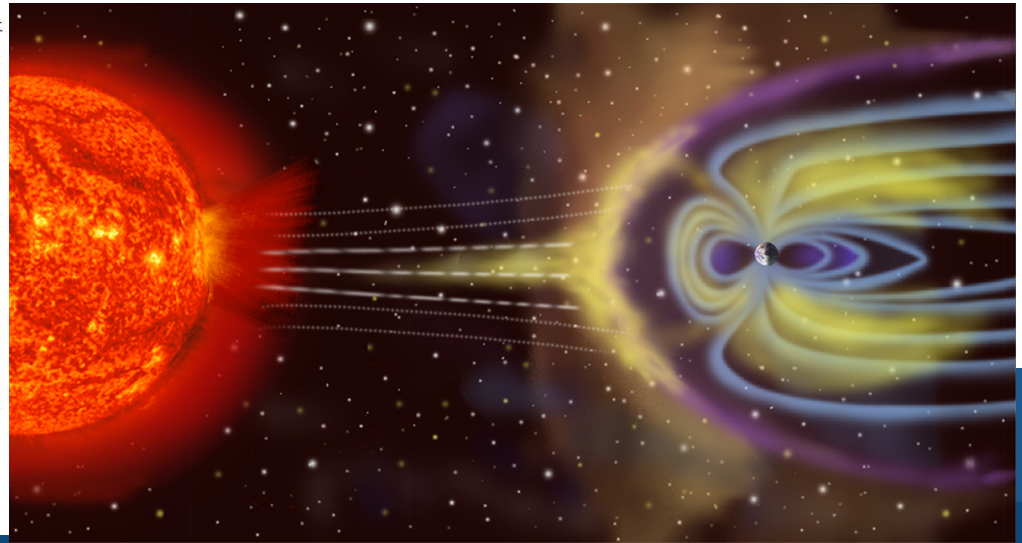



Tropical cyclone predictability

Extreme events in climate



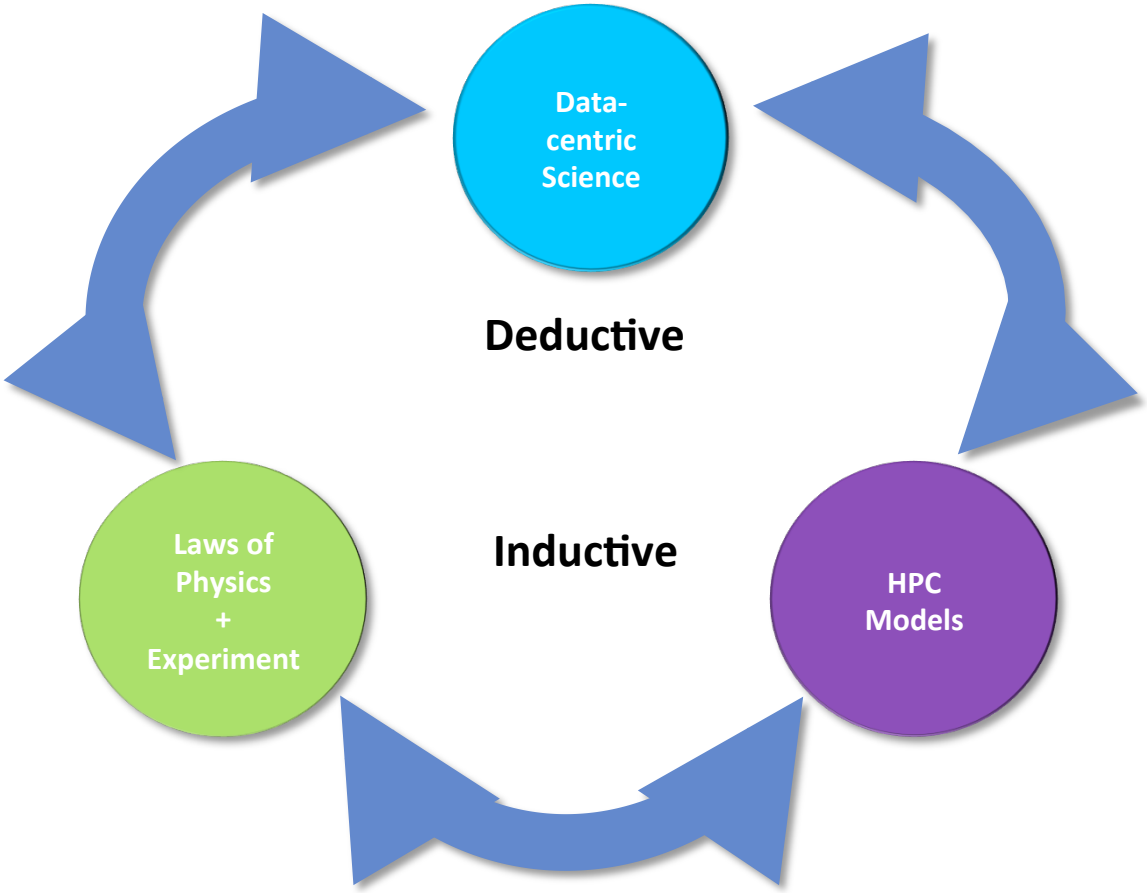

Space Weather Predictability



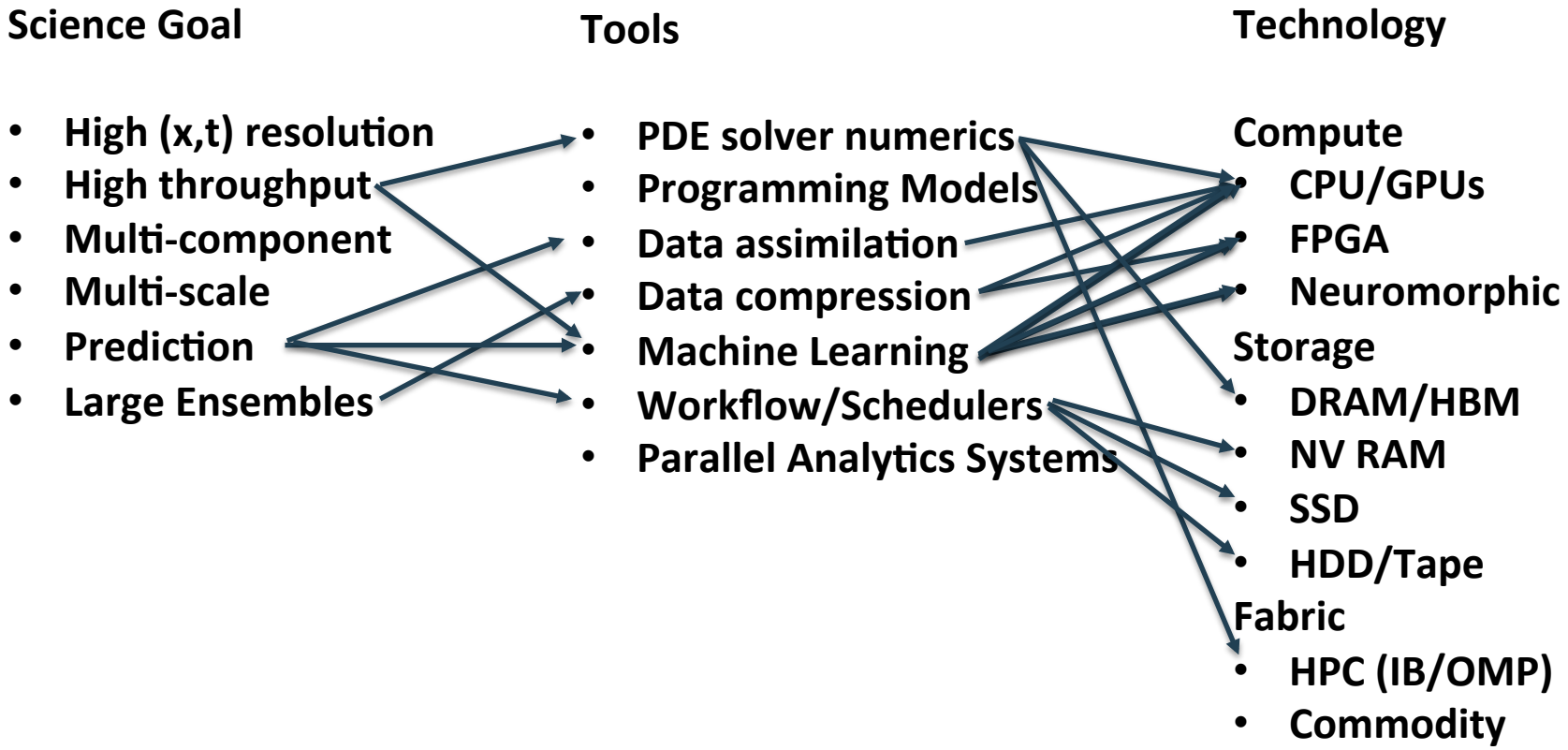
# Earth System Modeling Catch 22

- Due to insufficient computing power ESMs can't resolve key phenomena.
- Scientists try to describe the unresolved scales using human crafted physics parameterizations.
- ESM's *software complexity* grows, driven by the increasing complexity of these parameterizations.
- Growing *architectural complexity* hinders the ability to port and optimize ESM codes on new architectures.
- Due to insufficient computing power ESMs can't resolve key phenomena.

# The new HPC: Blending Deductive and Inductive Science



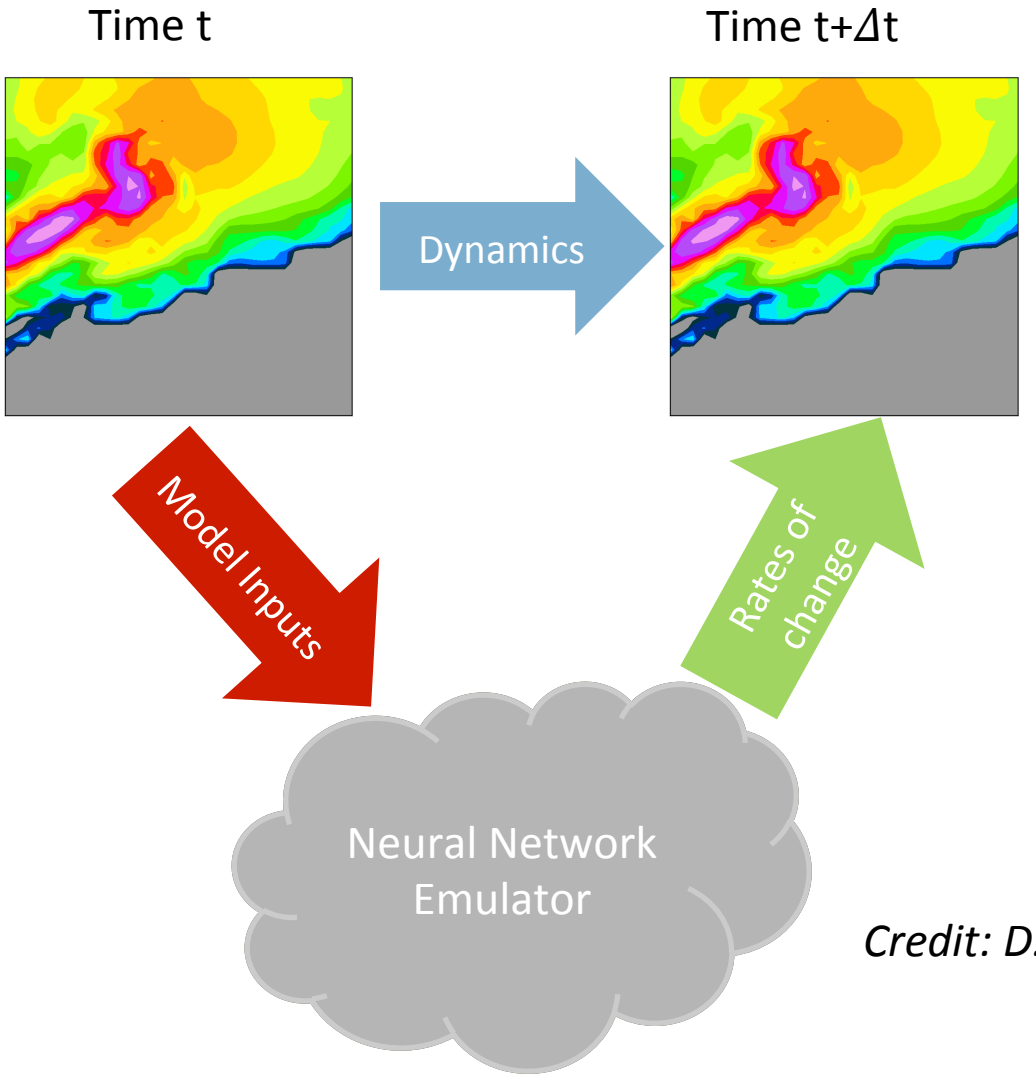
# Mapping Science to Tools and Technology



# NCAR Goals in Machine Learning

- Machine learning is clearly taking off in many sectors: how will it affect scientific computing?
- **Emulation**
  - Replacement of models or model components with learned components.
- **Augmentation**
  - Enhancing model results -> fewer runs, less resolution.
- **Data Analysis**
  - Feature detection, causality, etc…
- **Define a role appropriate to a national center!**

# Replacing Models with Emulation




*Credit: D.J. Gagne, NCAR*

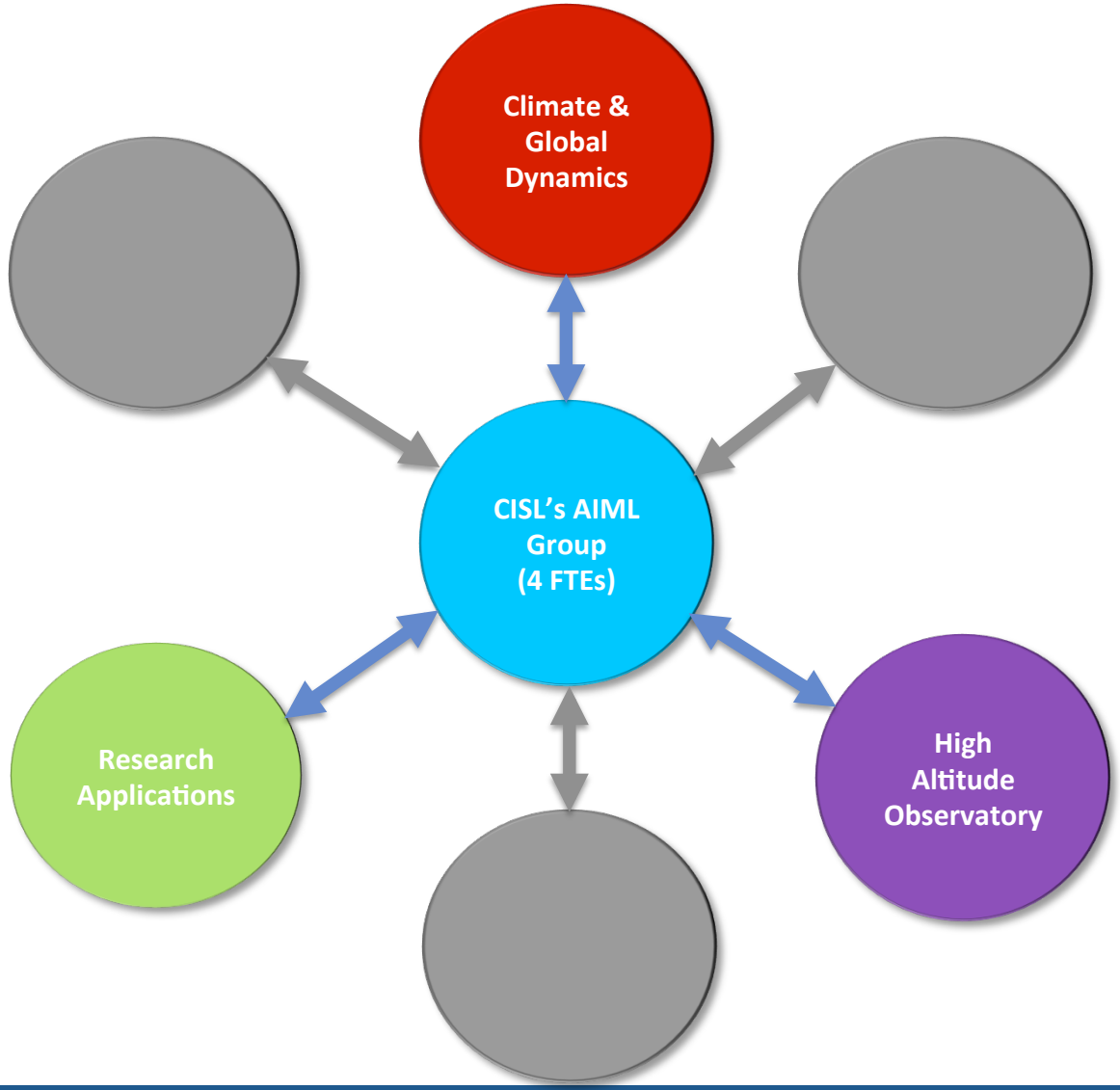
# AIML: New Machine Learning Group at NCAR

**AIML Founding Research Focus: model emulation**

**Why machine-learned emulation?** The *per-core performance* of conventional computer architectures has stagnated, and models are getting *increasingly complex*. Replacing human-crafted parameterizations with machine learning algorithms may simplify, accelerate and improve models.

- **Sub-grid-scale turbulence -Drs. Kosovic & Haupt (RAL), McCandless (AIML)**
  - **improved representation of the surface layer in meteorological models**
- **Cloud microphysics - Drs. Gettelman (CGD), Gagne & Sobhani (AIML)**
  - **improved weather and climate modeling**
- **Interplanetary coronal mass ejection (CME) - Drs. Gibson (HAO), Flyer (AIML)**
  - **space weather prediction**
- **Seasonal weather patterns - Drs. Sobhani (AIML) & DelVento (CISL)**
  - **Seasonal prediction of dangerous hot weather in the Eastern U.S.**

# AIML: An ML+HPC Research Hub for NCAR



# AIML Group: Interdisciplinary Group Dynamics

Mathematician

Dr. Natasha Flyer

Meteorologists

Dr. David John Gagne*

Dr. Tyler McCandless*

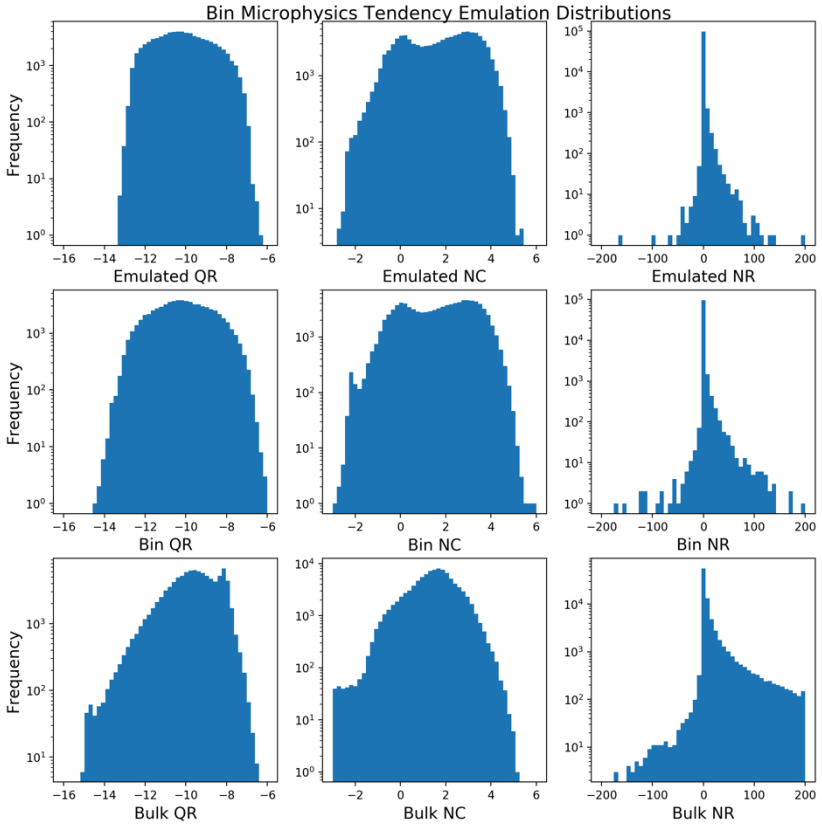
Statistician

Dr. Dorit Hammerling

Plus there are pockets of interest/activity in ML all over NCAR, CISL and beyond…

*Joint appointment between the Research Applications Lab and the Computational Lab

# Microphysics Emulator Results




**Emulated** ✓

Neural network microphysics emulates distribution and exact values of bin microphysics more closely than bulk microphysics

**Bin - too expensive for climate**

**Bulk - affordable for climate**

Credit: Gagne & Gettelman, NCAR

# Outstanding emulator challenges

- Ensuring interpretability & reproducibility of ML emulator results.
- Conditioning/scaling inputs are critical to the successful formulation of a successful emulator.
- Tuning emulator hyper-parameters for optimal performance.
- Representing extreme/unusual events in the emulator's training data.
- Getting ML emulators to respect constraints.
- Ensuring ML model robustness under iterative maps (time integration).

# Thanks!