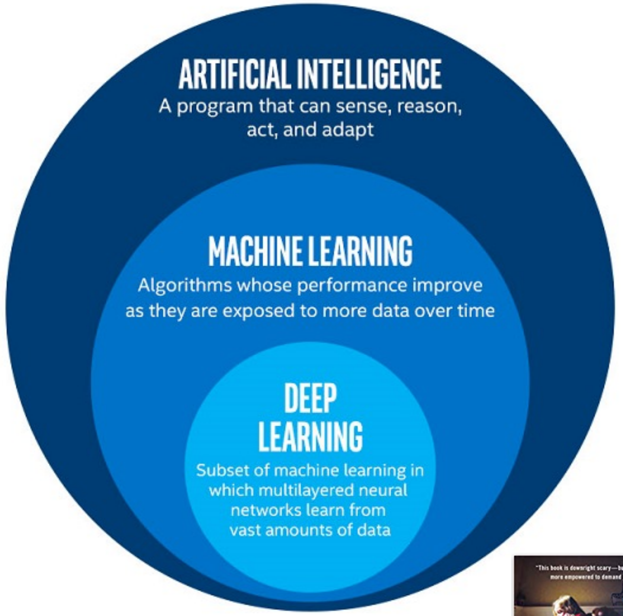
Machine Learning with CESM

Katie Dagon
Climate and Global Dynamics Lab
National Center for Atmospheric Research

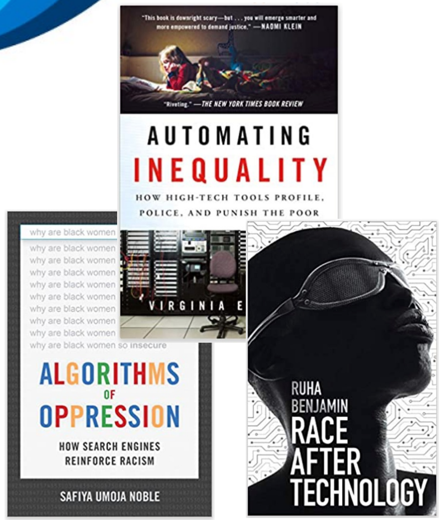
CESM Tutorial – August 2021

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

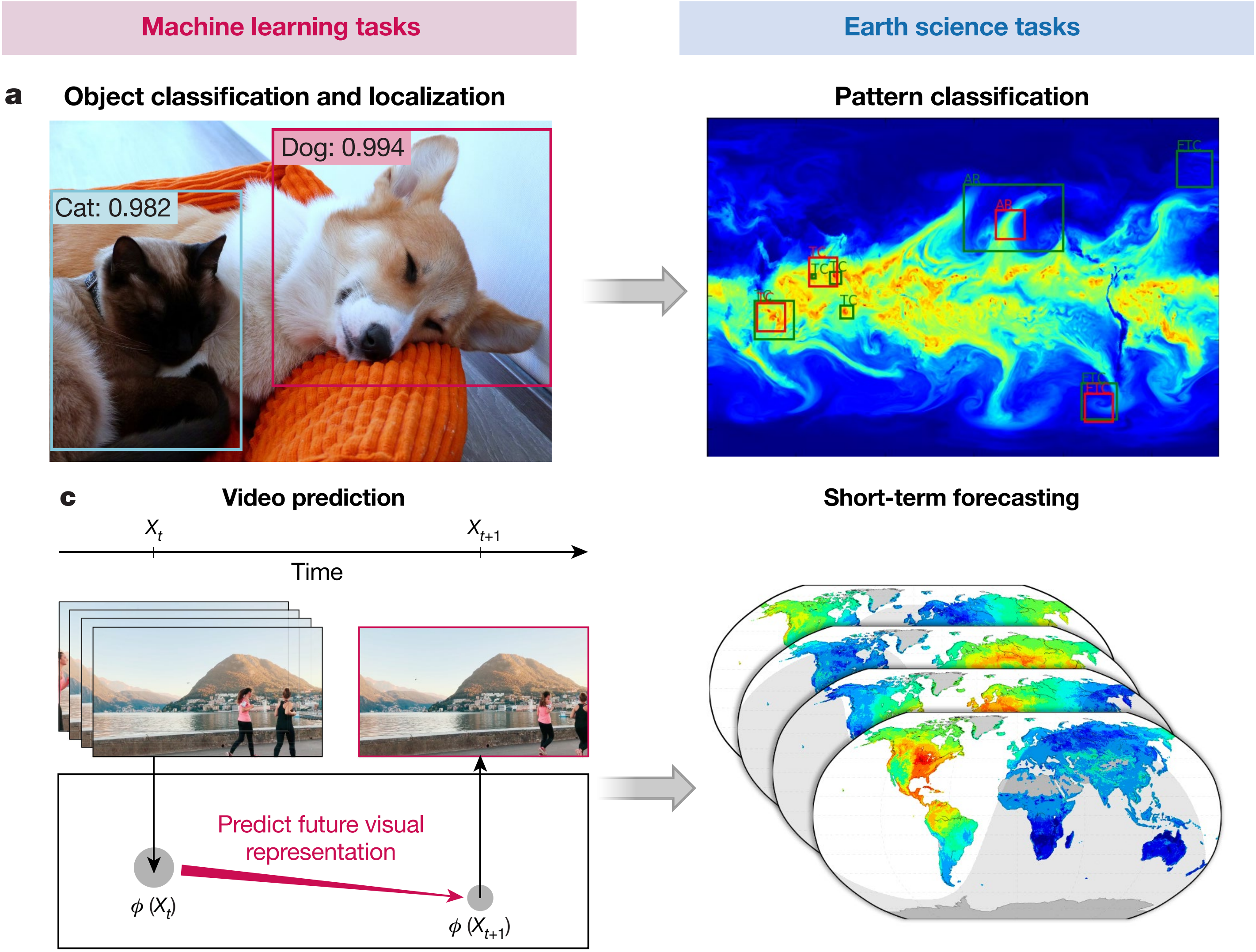
Machine Learning for Climate Modeling



- Artificial intelligence/machine learning is not "neutral" just because it is based on math
- Data and algorithms can reinforce society's prejudices
- Human programmers must ensure the AI/ML we develop is not biased



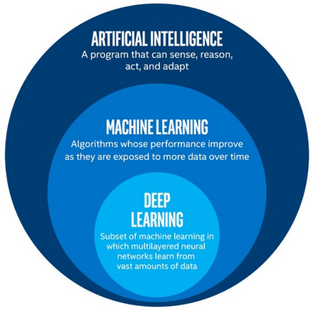
References from Amy McGovern



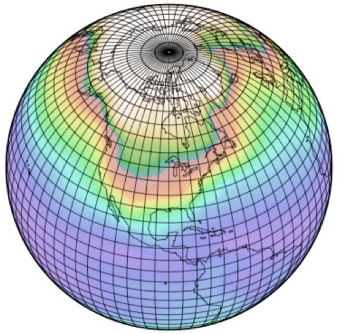
Reichstein et al. (2019)

Machine Learning for Climate Modeling: Applications

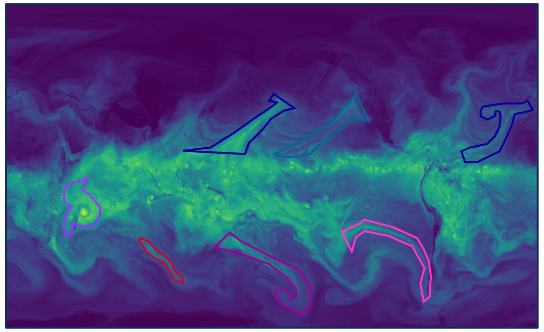
Can machine learning contribute to climate modeling?



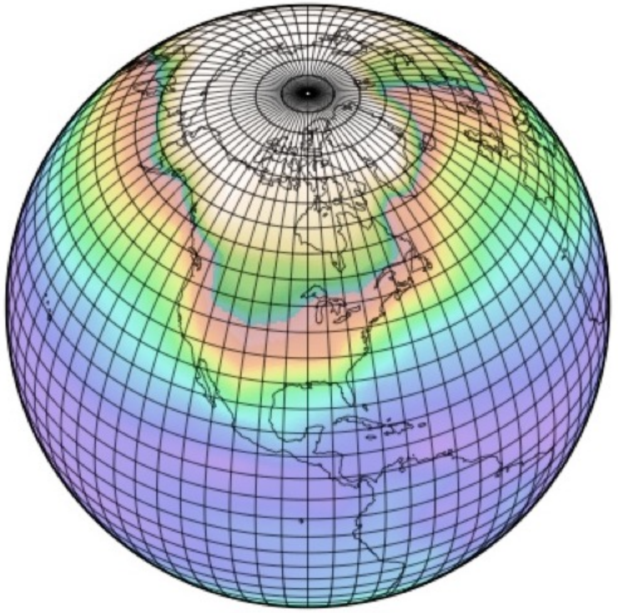
1) Climate Model Uncertainty



2) Detection of Extreme Events



Using climate models to study future climate changes and uncertainties



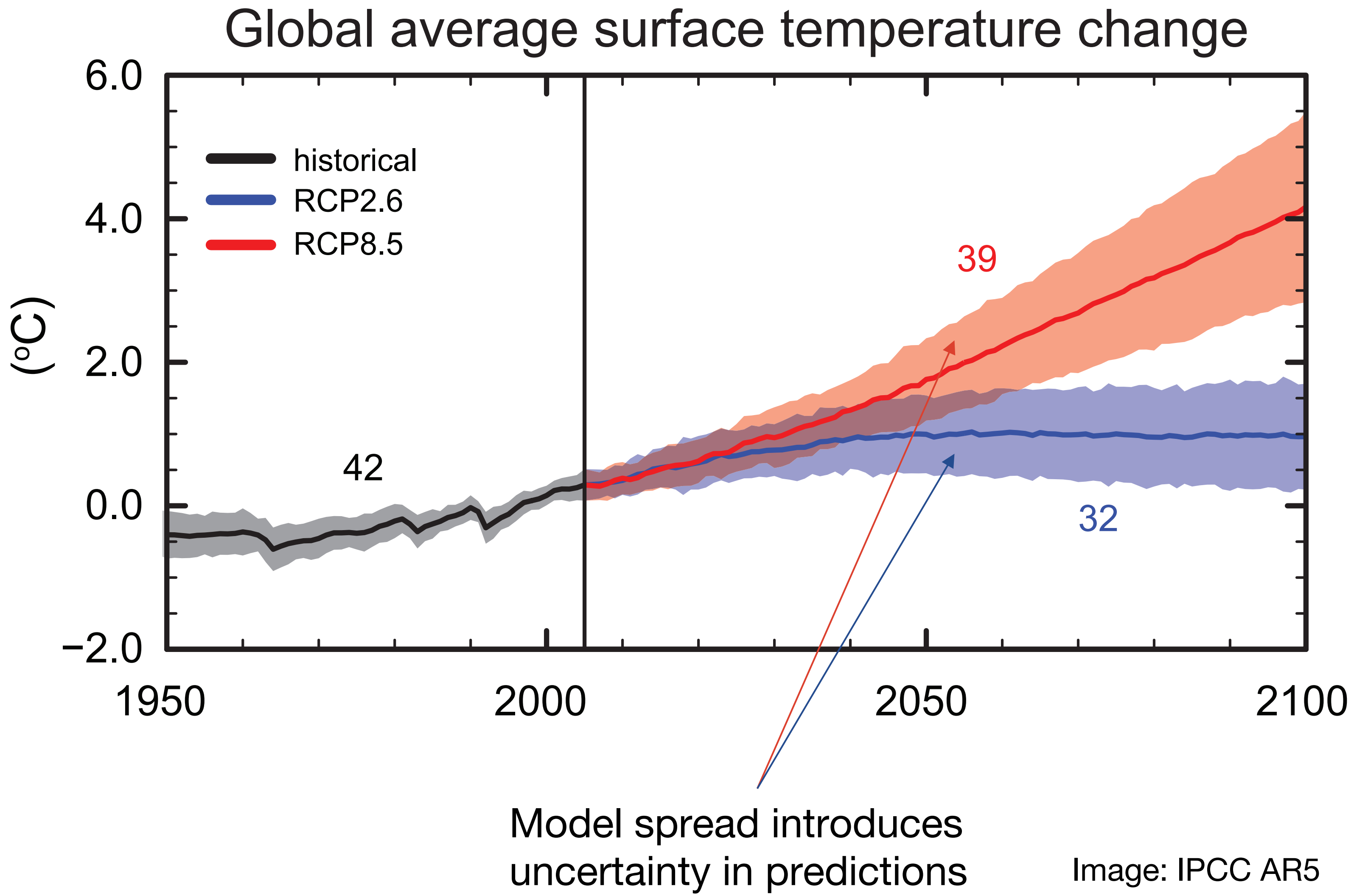
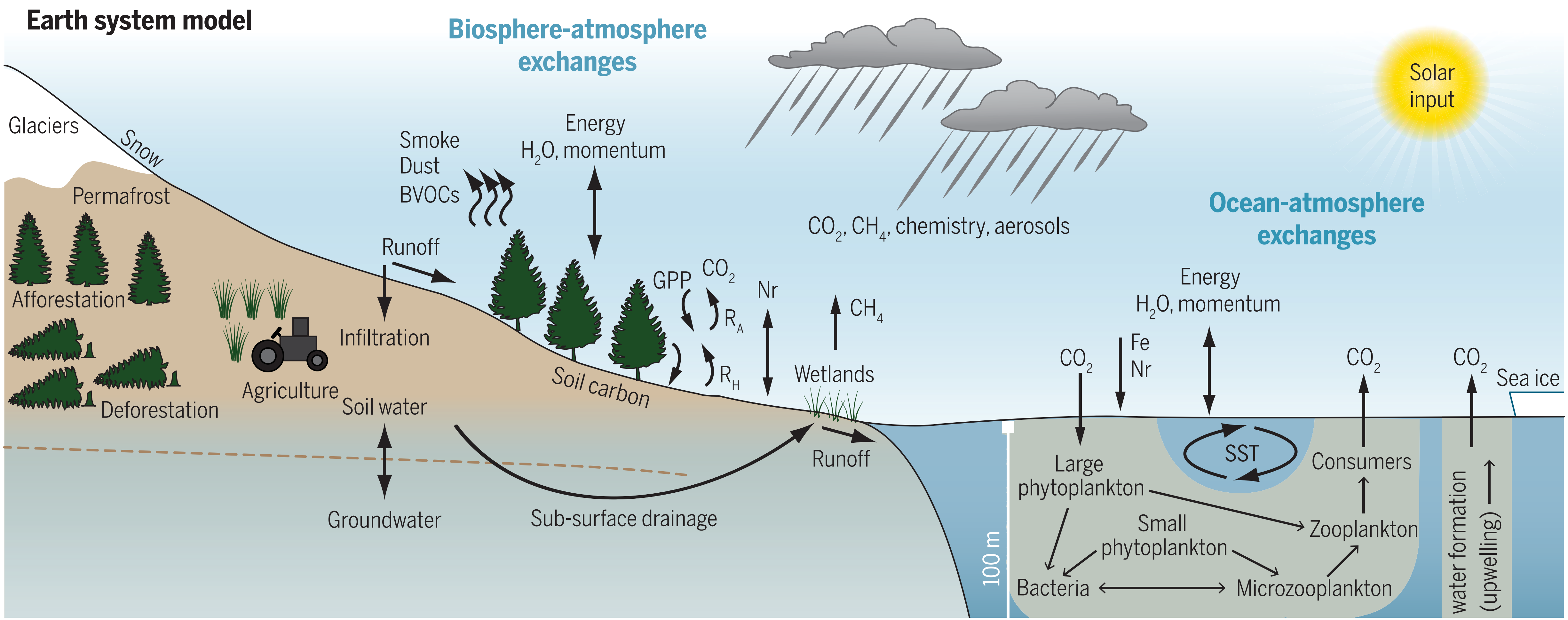


Image: IPCC AR5

Earth system models are complex and process-rich



Bonan & Doney (2018)

Community Land Model (CLM) component of CESM

CESM Components

Focus on the land model component

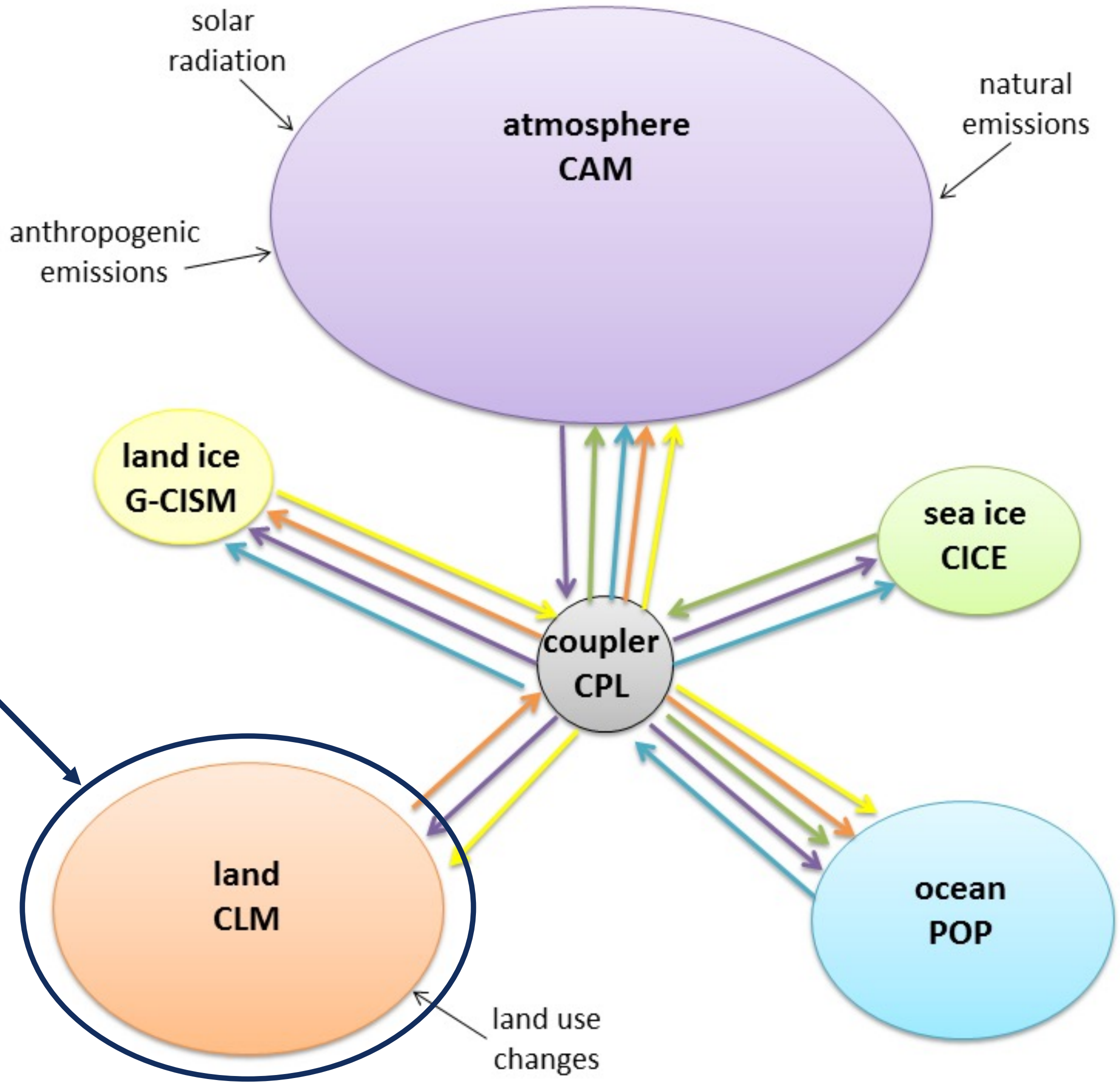
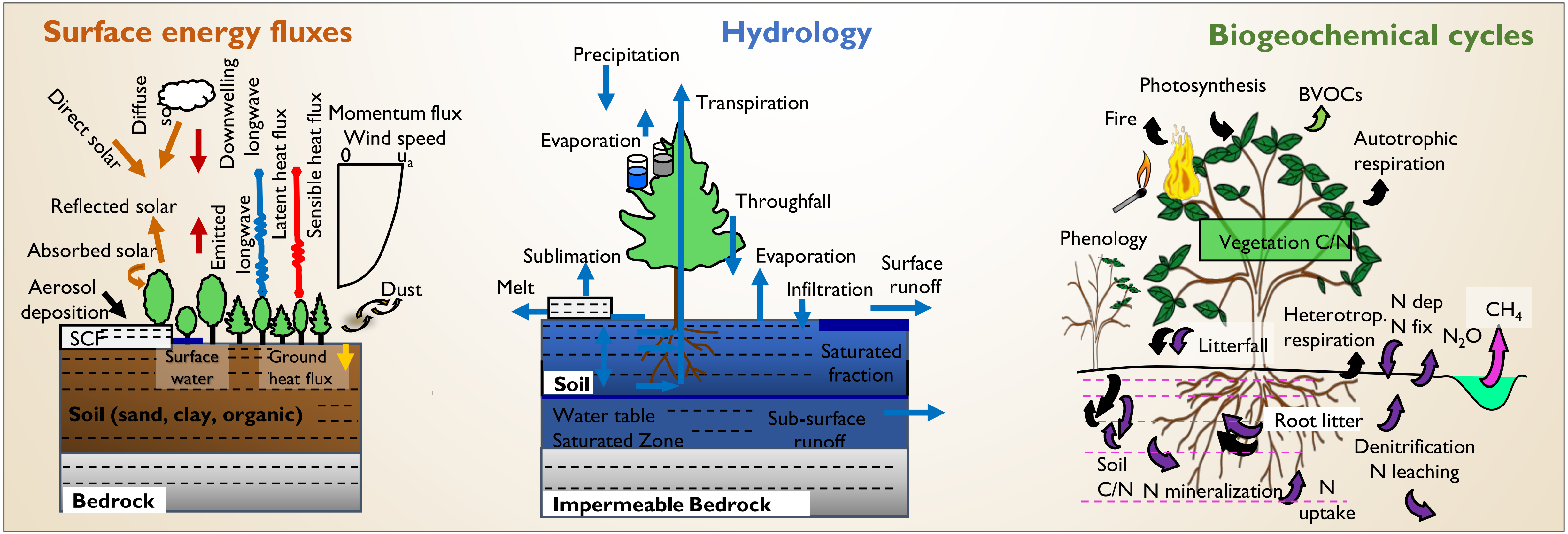


Figure from Alexander and Easterbrook (2011)

Land Model Parameters

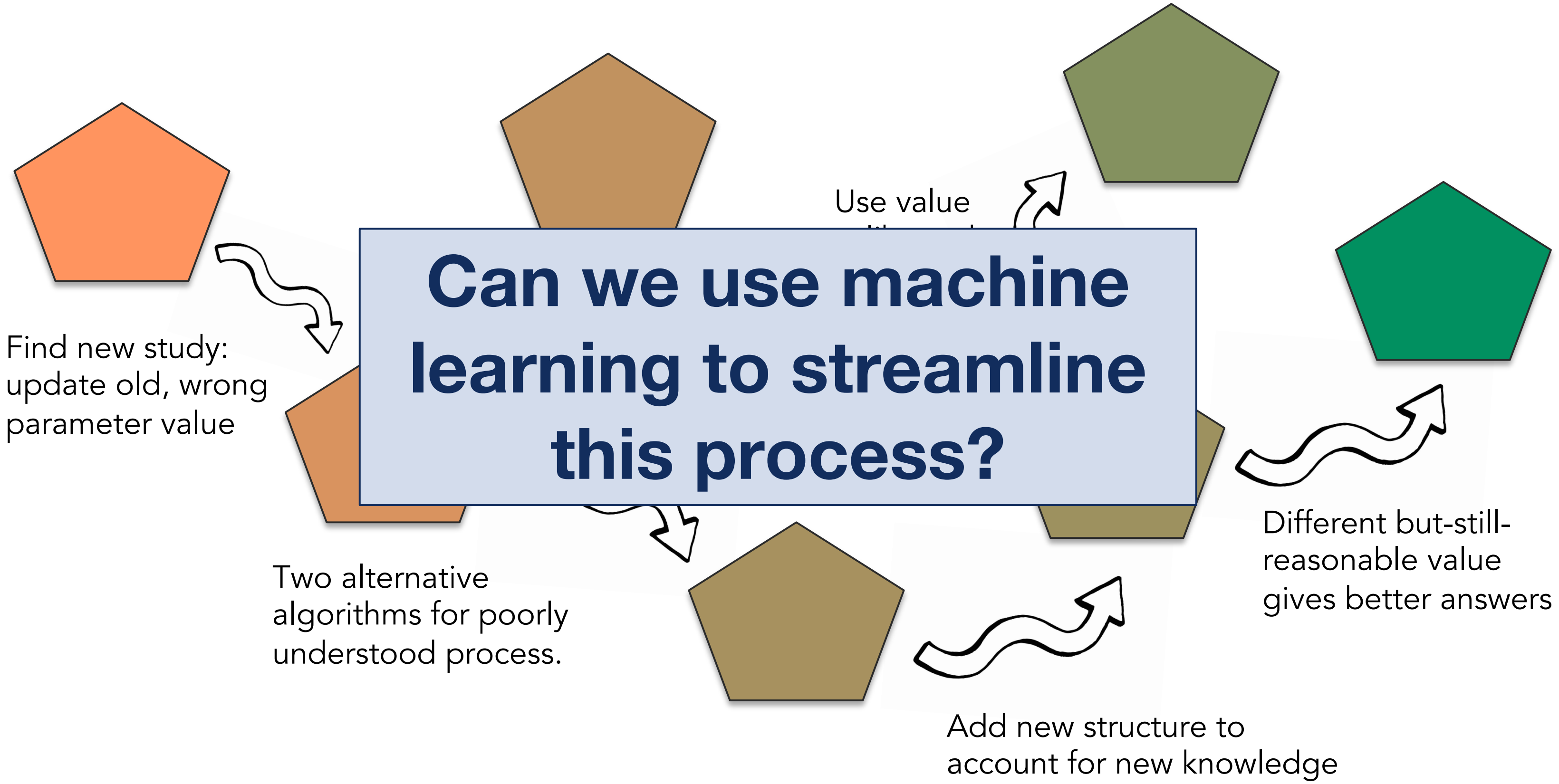


Schematic of NCAR's Community Land Model (CLM), version 5

Lawrence et al. (2019)

Can we use machine learning to investigate model uncertainties?

Hand-tuning parameter values takes a long time (many model runs, trial and error).



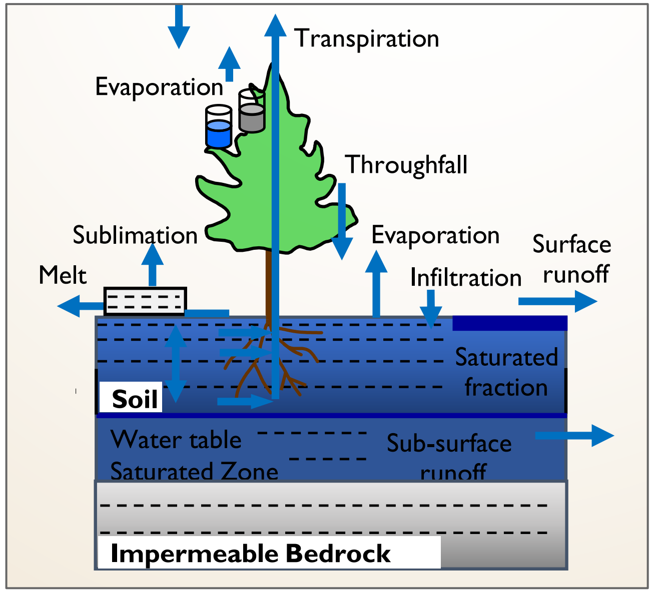
Can we use machine learning to streamline this process?

Figure from Rosie Fisher

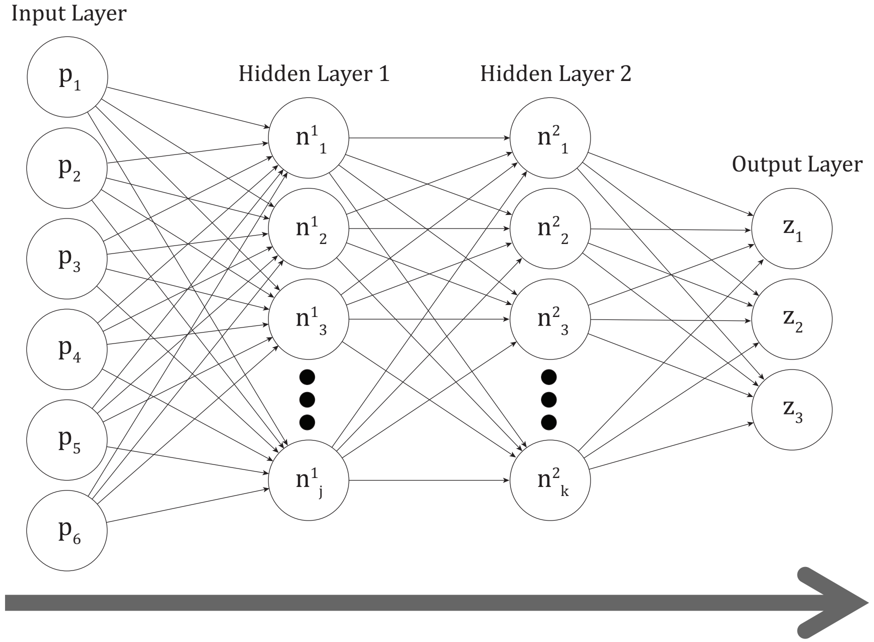
Land Model Emulation for Parameter Calibration

Objective: Train neural networks to emulate CLM5 output, allowing for many fast computations with different parameter values.

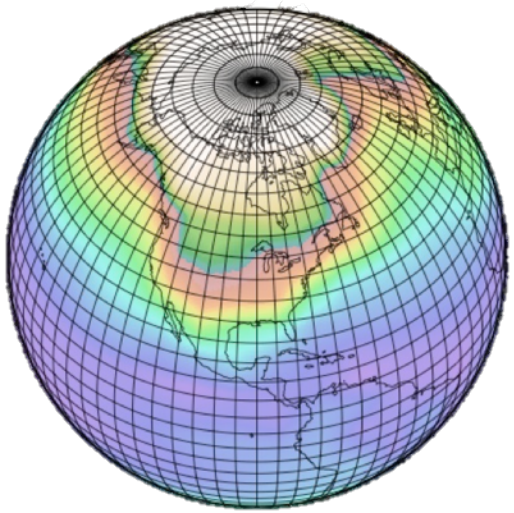
Input: CLM5 sensitive parameter values



Neural network emulator



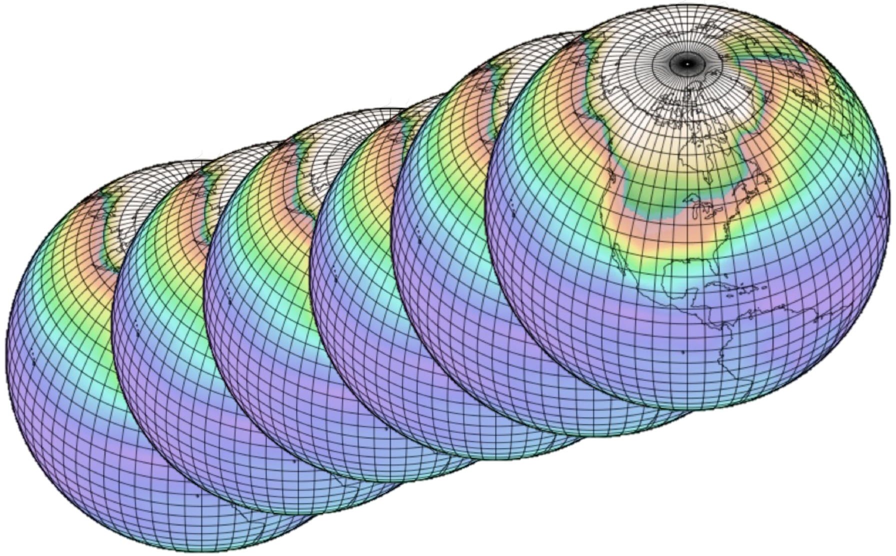
Output: CLM5 predictions



Dagon et al. 2020

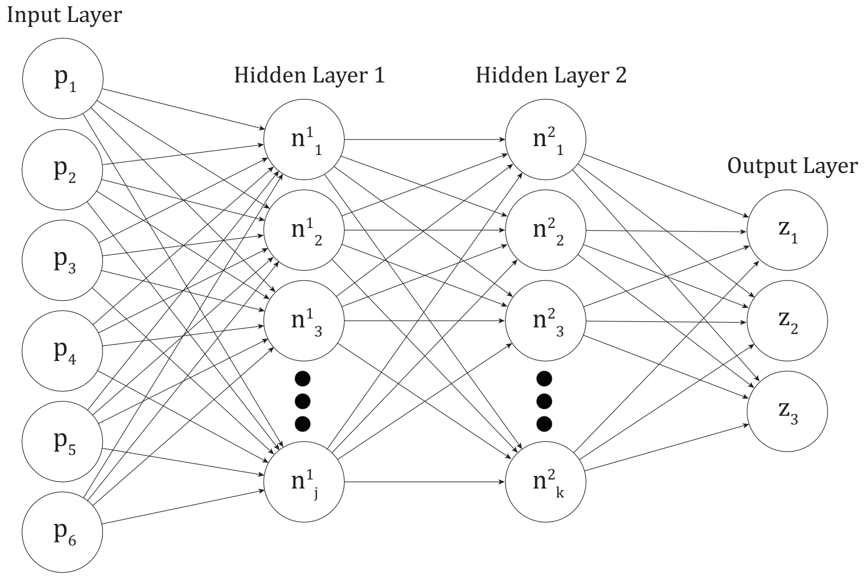
Increase in Computational Efficiency

Land model perturbed parameter ensemble



~2 hours per simulation

Machine learning emulator

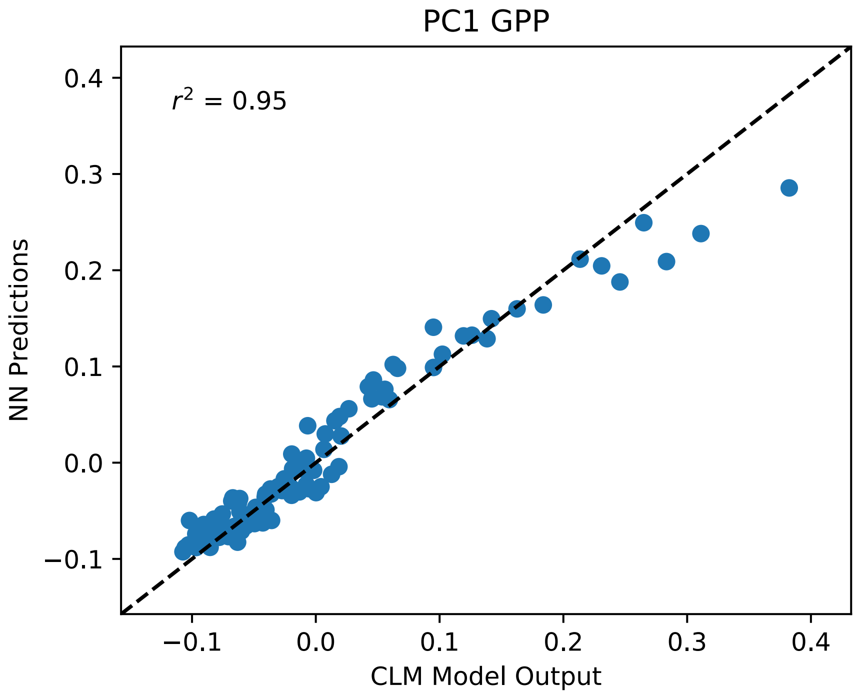


2.6 seconds to generate predictions!

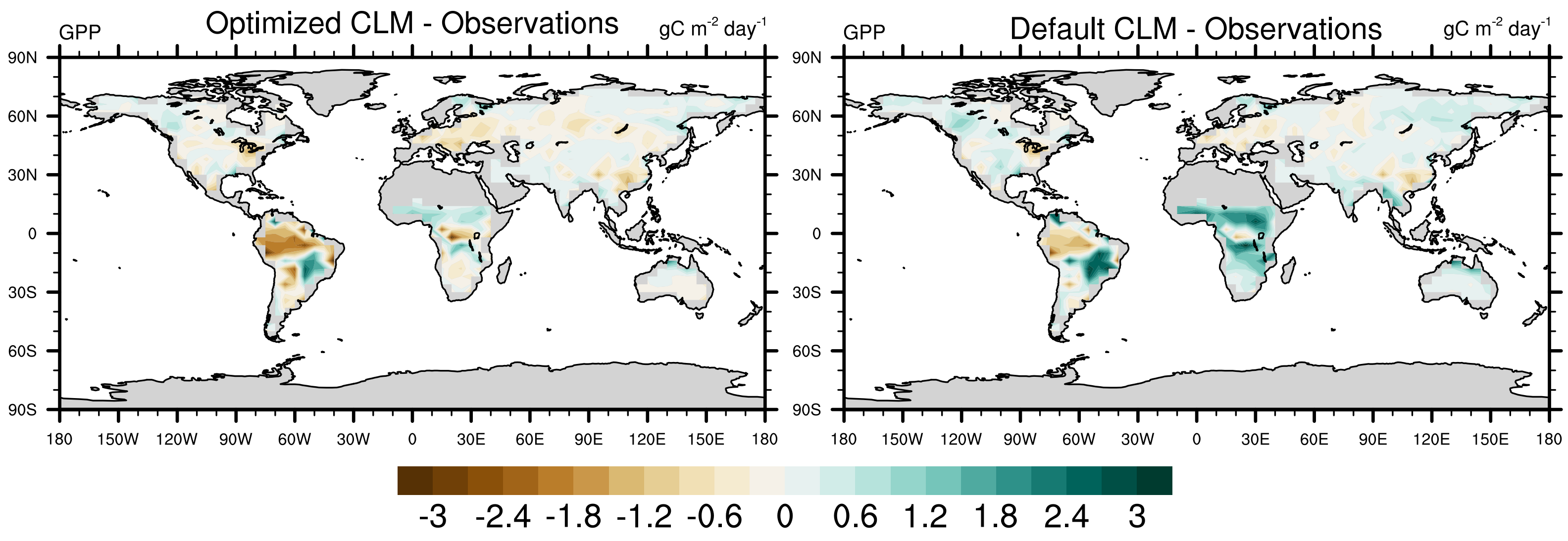
Land Model Emulation for Parameter Calibration

Approach: Emulate, calibrate, test.

Emulator predictions vs. CLM output



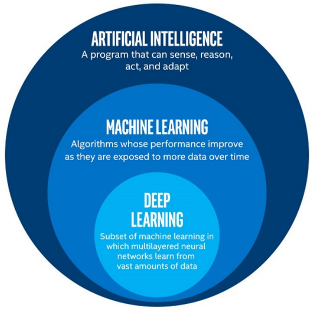
Comparing model bias with calibrated *(left)* and default *(right)* parameters



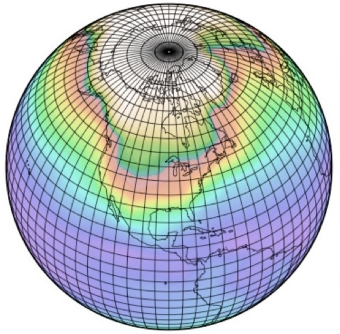
Dagon et al. 2020

Machine Learning for Climate Modeling: Applications

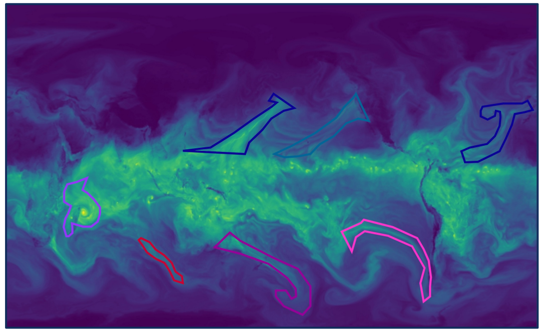
Can machine learning contribute to climate modeling?



1) Climate Model Uncertainty

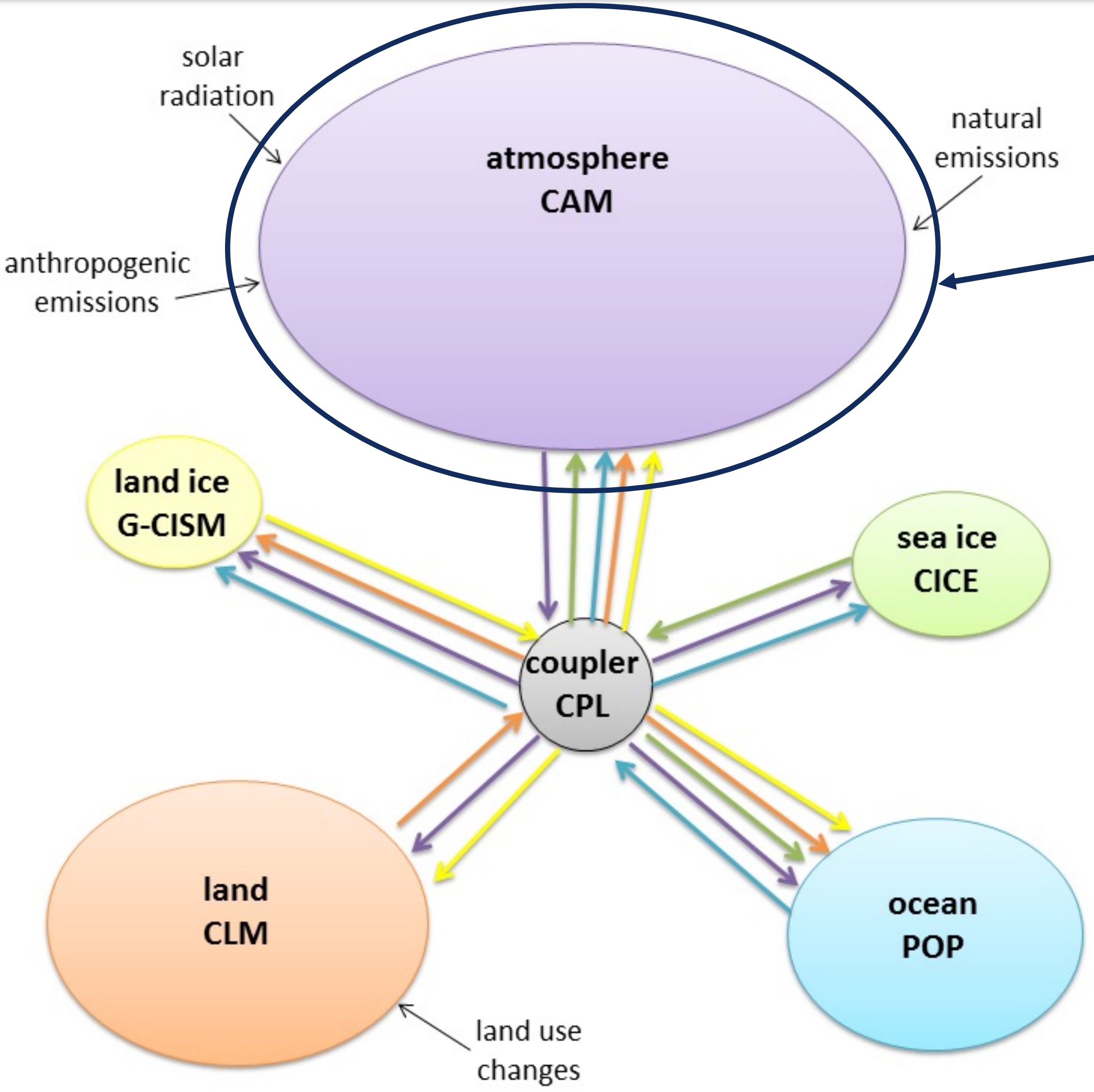


2) Detection of Extreme Events



Community Atmosphere Model (CAM) component of CESM

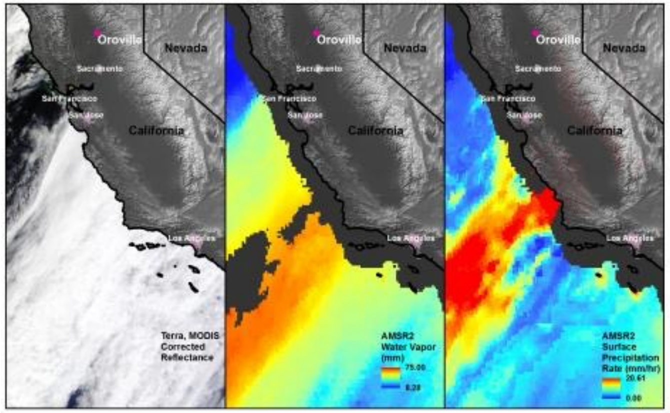
<u>CESM Components</u>



Focus on the atmosphere model component

Extreme Precipitation Has Significant Consequences

Oroville Dam spillway overflowing in February 2017 following an atmospheric river event in California.



Flooding after Hurricane Harvey in August 2017.

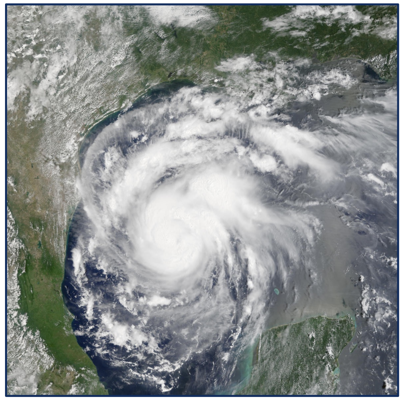
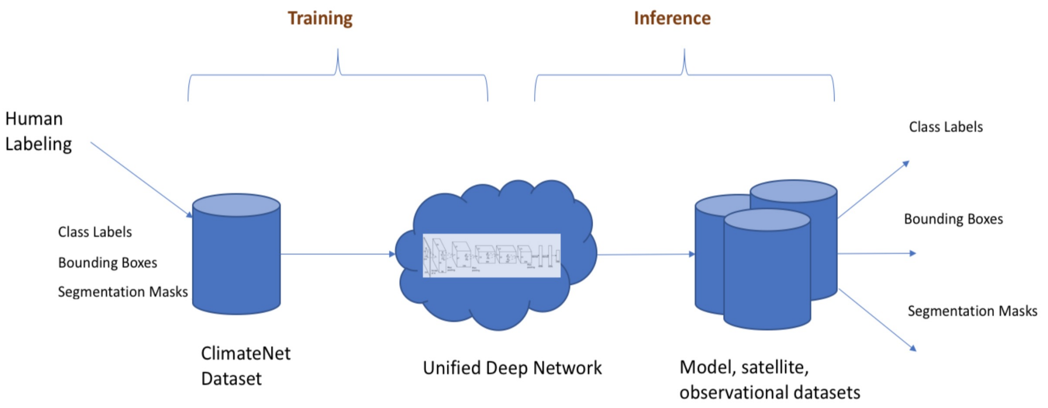


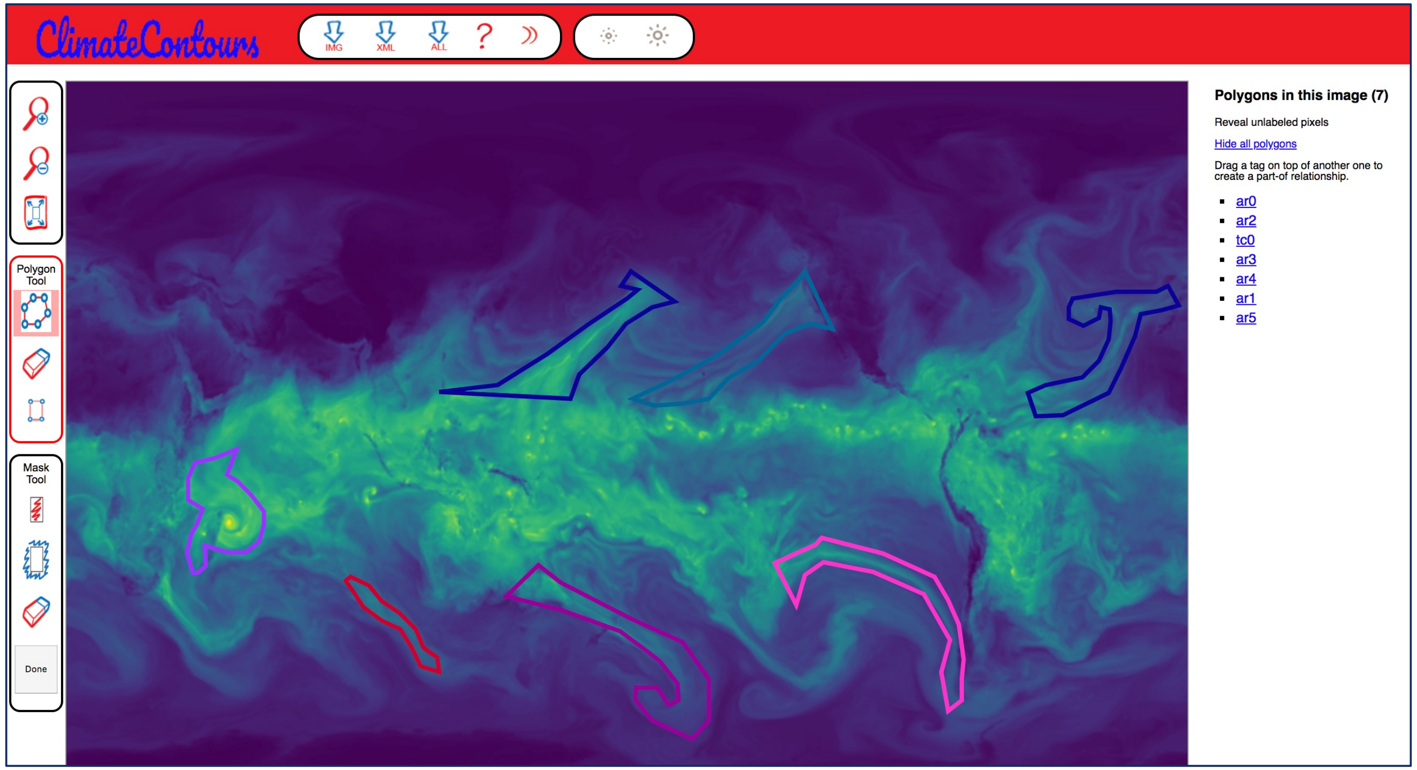
Image Recognition for Detecting Extreme Events

ClimateNet: a community-sourced expert-labeled dataset to improve and accelerate machine learning applications in weather and climate

- *Focus on detecting atmospheric rivers (ARs) and tropical cyclones (TCs).*



Prabhat et al. (2021)



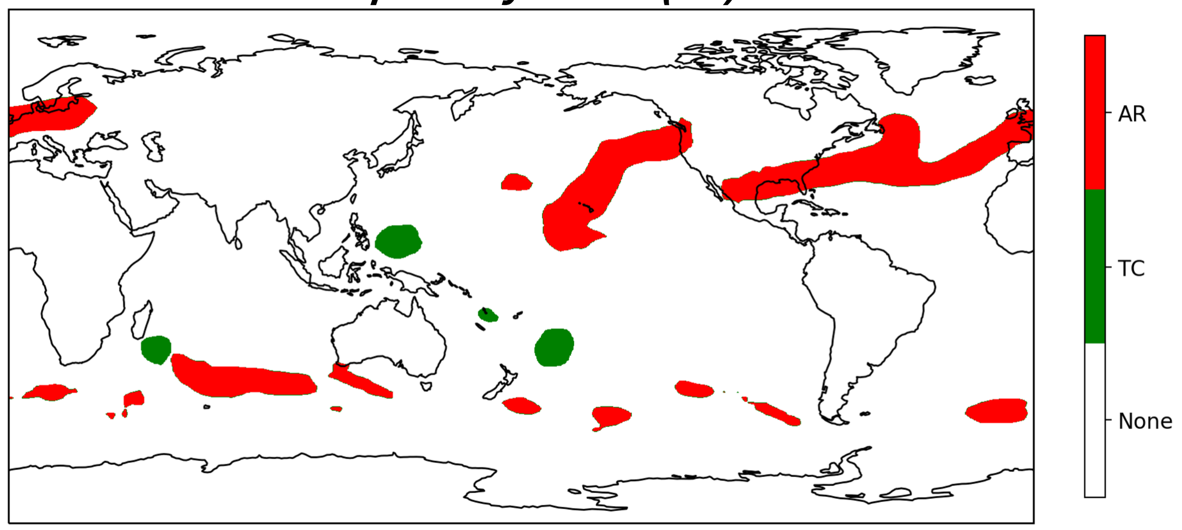
Images courtesy of Karthik Kashinath, NERSC
https://www.nersc.gov/research-and-development/data-analytics/big-data-center/climatenet

U.S. DEPARTMENT OF ENERGY | Office of Science
NERSC
BERKELEY LAB

Extreme Weather Detection

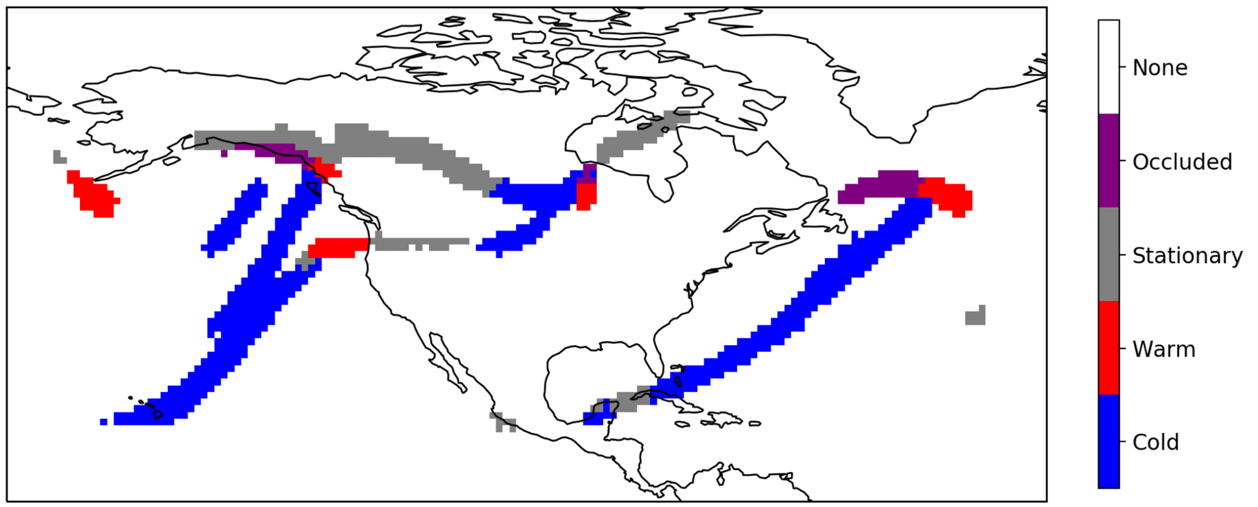
Objective: Apply existing machine learning-based detection algorithms to automate the classification of synoptic-scale weather features.

Atmospheric Rivers (AR) and Tropical Cyclones (TC)



Applying trained ClimateNet algorithm (Prabhat et al., 2021) to detect ARs and TCs globally in high resolution (0.25°) coupled CESM simulations.

Frontal ID

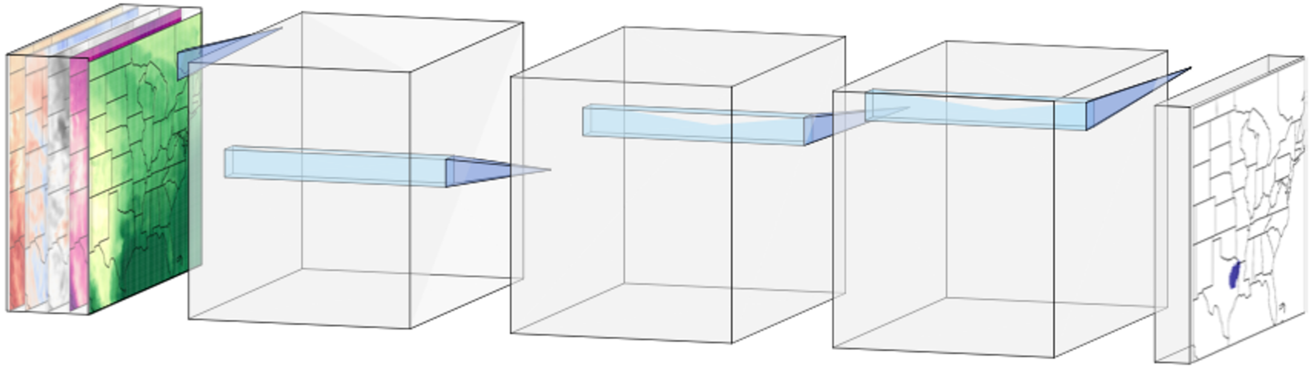


Applying trained DL-Front algorithm (Biard and Kunkel, 2019) to detect front types over North America in coupled CESM simulations.

Extreme Weather Detection

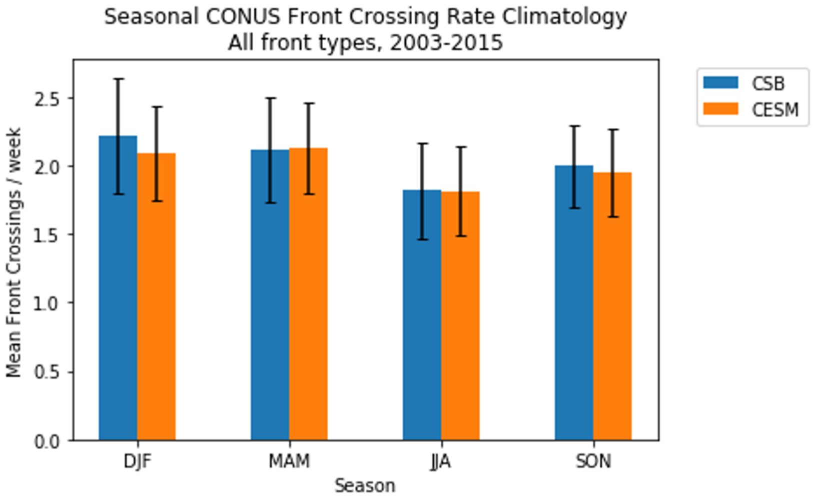
Project Goals:

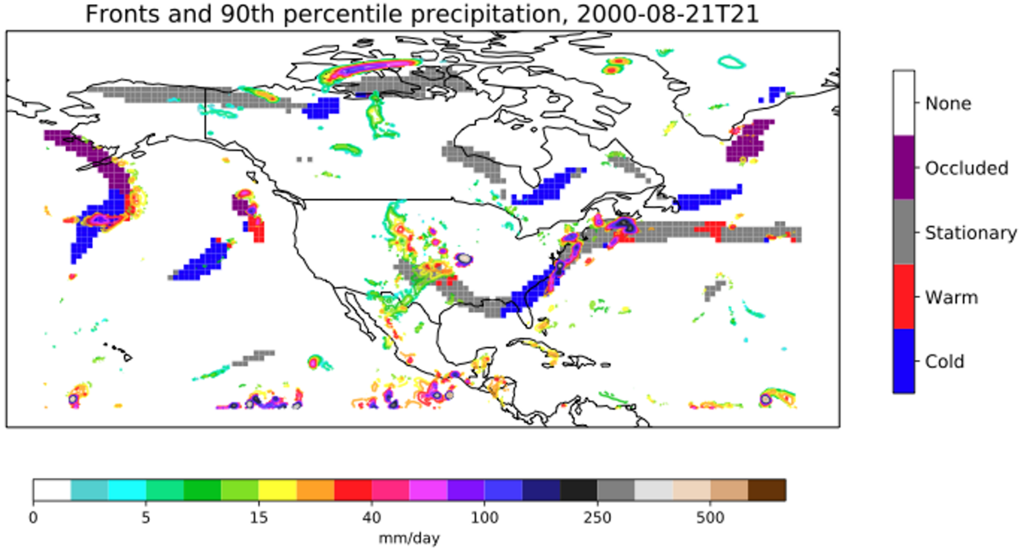
- Explore validation and explainability of ML-based algorithms to build confidence and trust.
- Develop additional detection algorithm for mesoscale convective systems (MCS).
- Connect identified features with extreme precipitation events.



Deep Learning Infrastructure for MCS Detection (Maria Molina)

Validation of DL-Front using NWS Coded Surface Bulletin (CSB)





Machine Learning Challenges and Opportunities

Challenges

- Interpretability (i.e., machine learning as a “black box”)
- Obtaining high quality training data
- Physics-driven and data-driven models
- Working with “Big Data”
- Research across disciplines

Opportunities

- Interdisciplinary projects
- Uncertainty quantification
- Parameterization
- Climate prediction
- Detection and attribution

See also Karpatne et al. (2019), Rolnick et al. (2019)

via XKCD

ML/AI Online Learning and Workshop Opportunities

- September 13-17, 2021 NOAA 3rd Workshop on Leveraging AI in Environmental Sciences: https://2021noaaaiworkshop.sched.com/info
- July 2021 Trustworthy Artificial Intelligence for Environmental Science (TAI4ES) virtual Summer School: https://www2.cisl.ucar.edu/tai4es
- July 2020 Artificial Intelligence for Earth System Science (AI4ESS) Summer School: https://www2.cisl.ucar.edu/events/summer-school/ai4ess/2020/artificial-intelligence-earth-system-science-ai4ess-summer-school
- 2019 and 2020 AGU Tutorials on Machine Learning and Deep Learning for Environmental and Geosciences: https://sites.google.com/lbl.gov/ml4egs/
- 2nd NOAA Workshop on Leveraging AI in Environmental Science (2020-2021): https://www.star.nesdis.noaa.gov/star/meeting_2020AIWorkshop.php
- US CLIVAR Data Science Working Group Webinar Series on Machine Learning (2020-2021): https://usclivar.org/working-groups/data-science-working-group
- ECMWF 2020 Machine Learning Seminar Series: https://www.ecmwf.int/en/learning/workshops/machine-learning-seminar-series

Summary

- Machine learning emulators trained to reproduce land model output with greater computational efficiency; emulator predictions are **optimized to minimize error** between model and observations.
- Machine learning-based detection algorithms are developed and applied to **capture high-impact weather events**; validation and interpretation are key ongoing steps to building confidence in predictions.
- **Ongoing CESM-related machine learning projects:** Earth system predictability, model component parameterizations (e.g., CAM6 and MOM6), process understanding for sea ice.

Thanks!

Questions?

kdagon@ucar.edu

@katiedagon