M²LInES: Multiscale Machine Learning In Coupled Earth System Modeling

Laure Zanna, Courant Institute, NYU

CESM Annual Workshop 2022

@M2LInES

https://m2lines.github.io

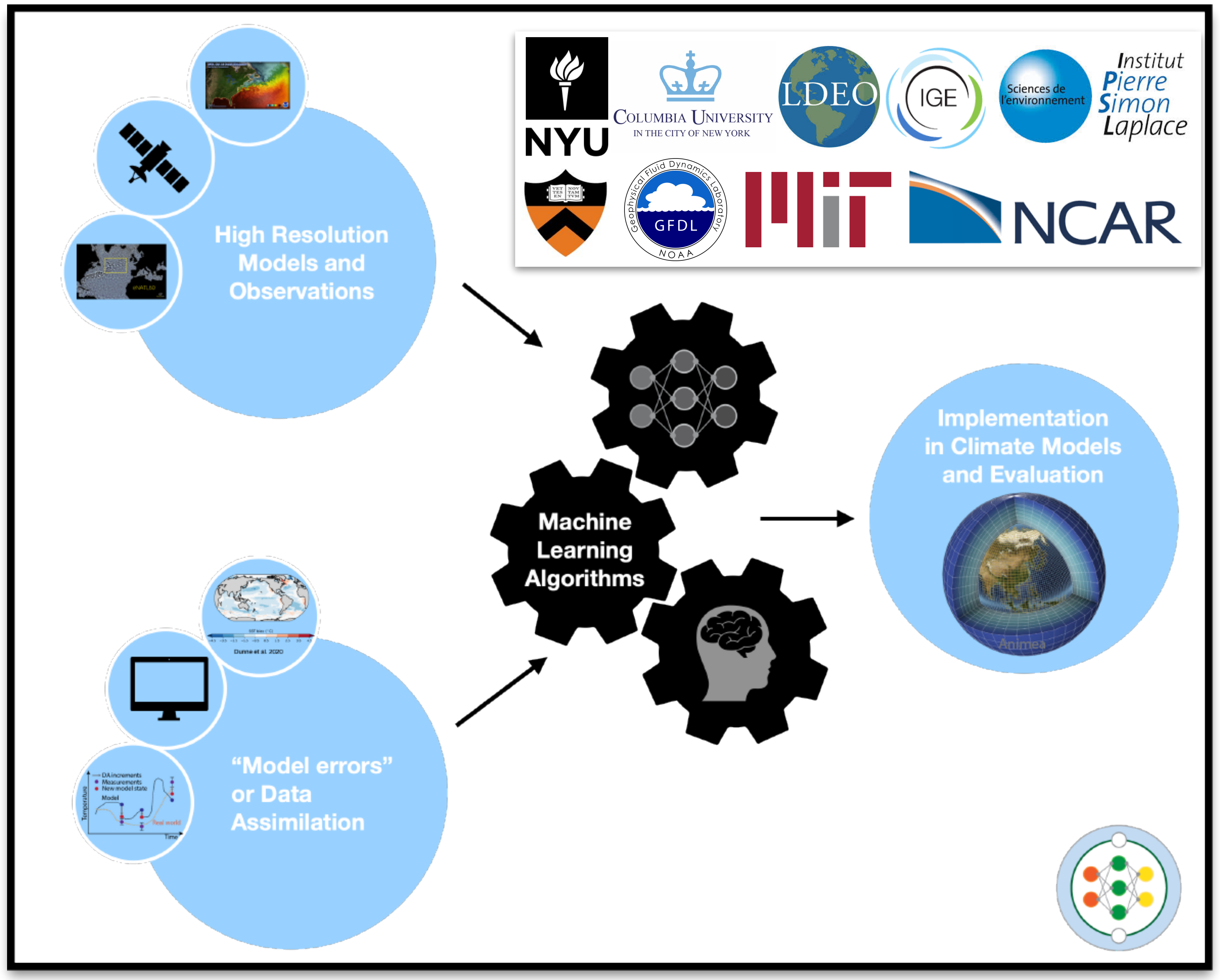
Supported by
SCHMIDT **FUTURES**

NYU
Columbia University in the City of New York
LDEO
IGE
Sciences de l'environnement
Institut Pierre Simon Laplace
Geophysical Fluid Dynamics Laboratory GFDL NOAA
NCAR

M²LInES: Multiscale Machine Learning In Coupled Earth System Modeling

Improving climate model skill at the air-sea interface *using data + ML to* reduce model biases in coupled climate simulations

~ 30 domain scientists, model developers, machine learning experts across 9 institutions in the US + France

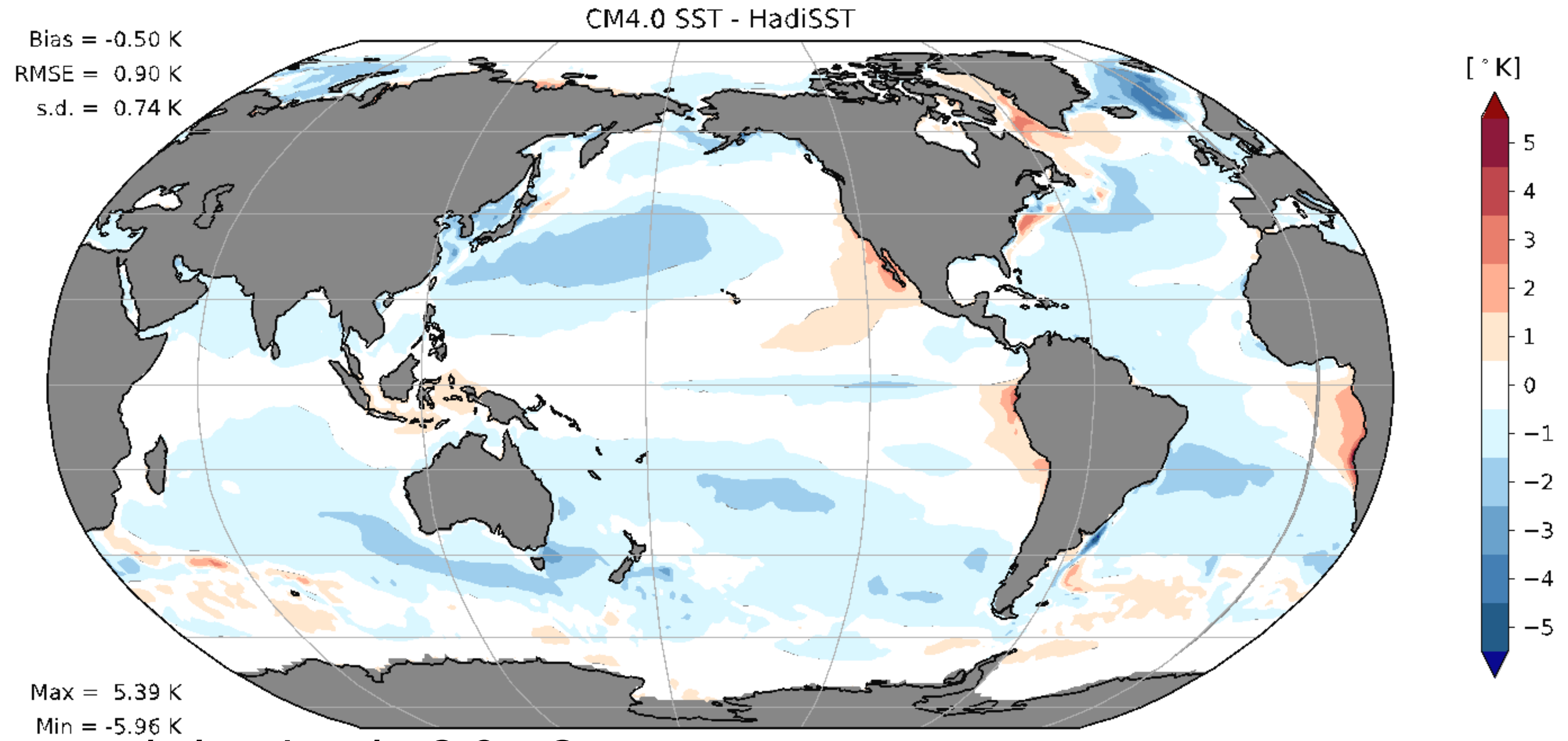


+ another ~15 affiliates

Climate projections uncertainty is due to model error

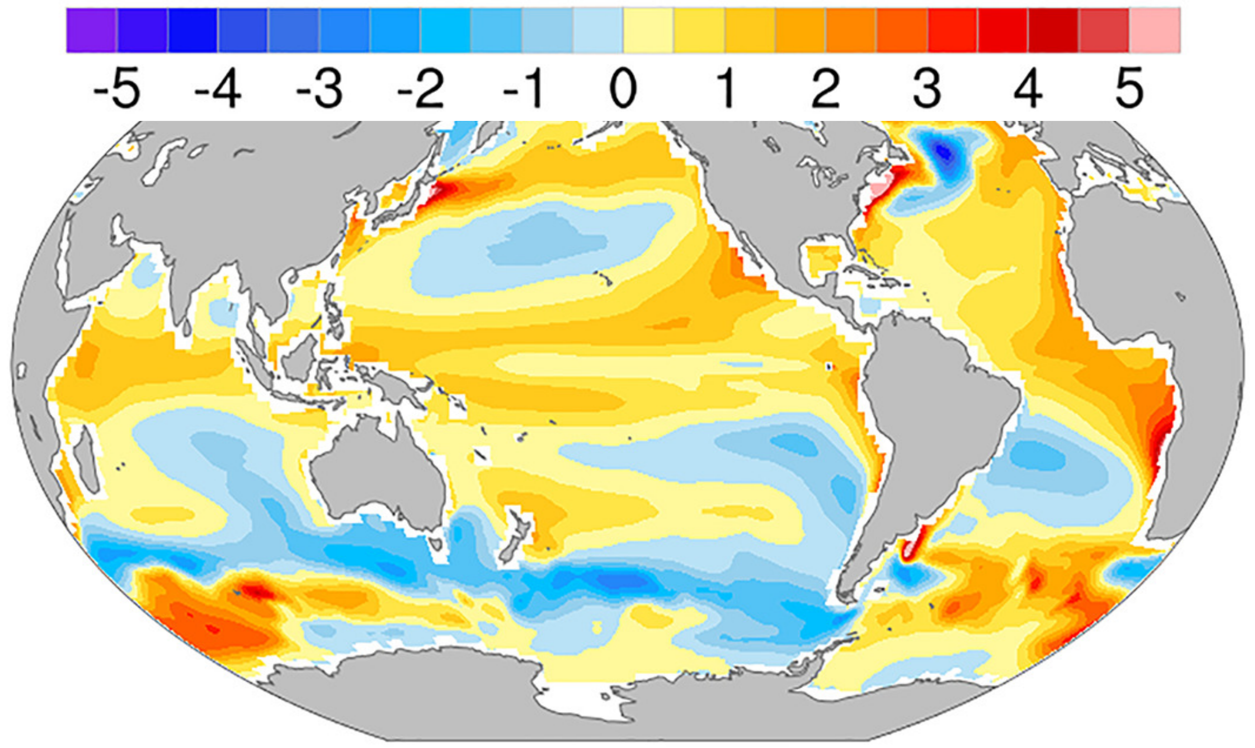
Sea Surface Temperatures Bias

GFDL CM4



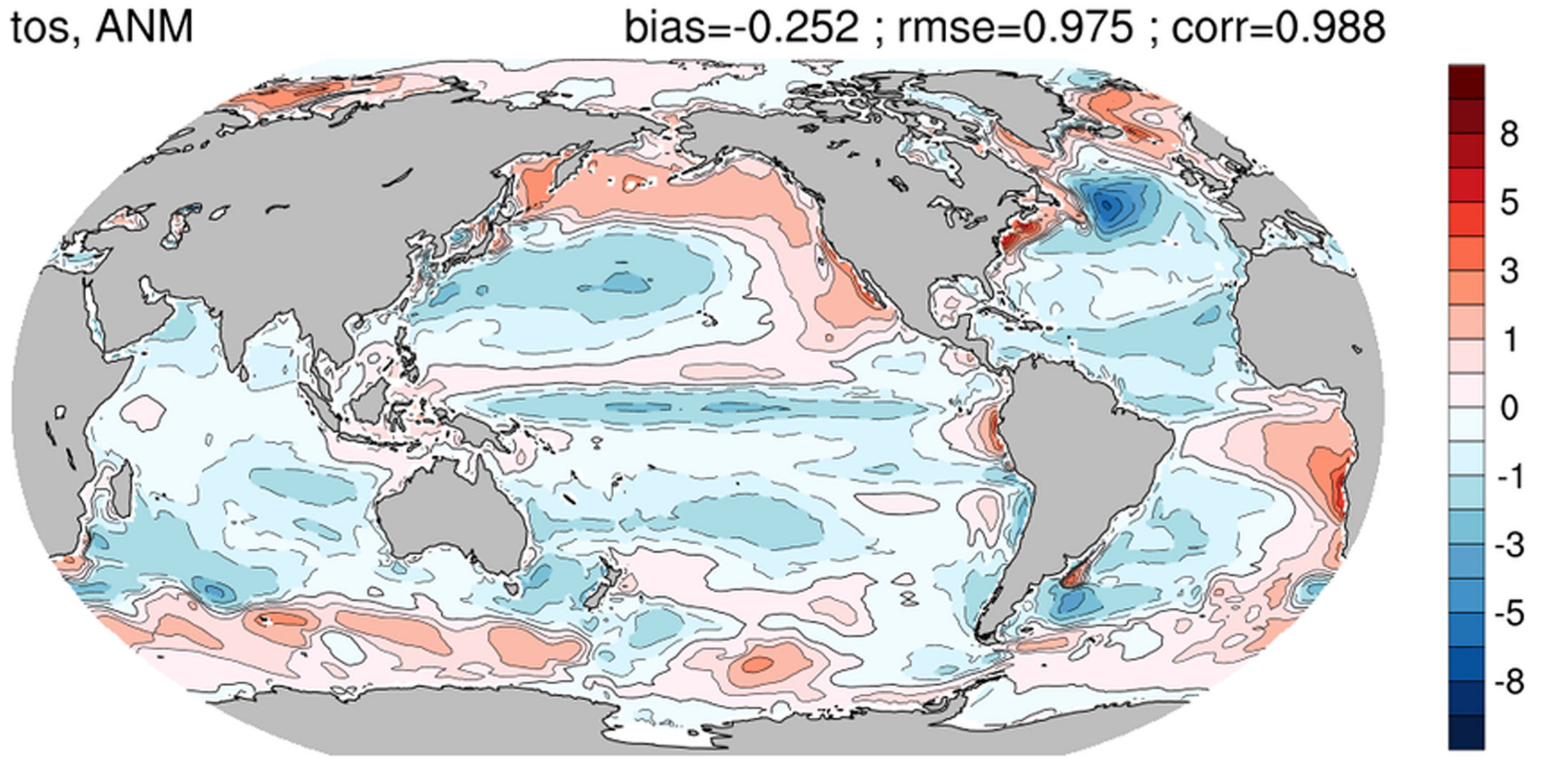
Held et al, 2019

CESM2



Danabasoglu et al., 2020

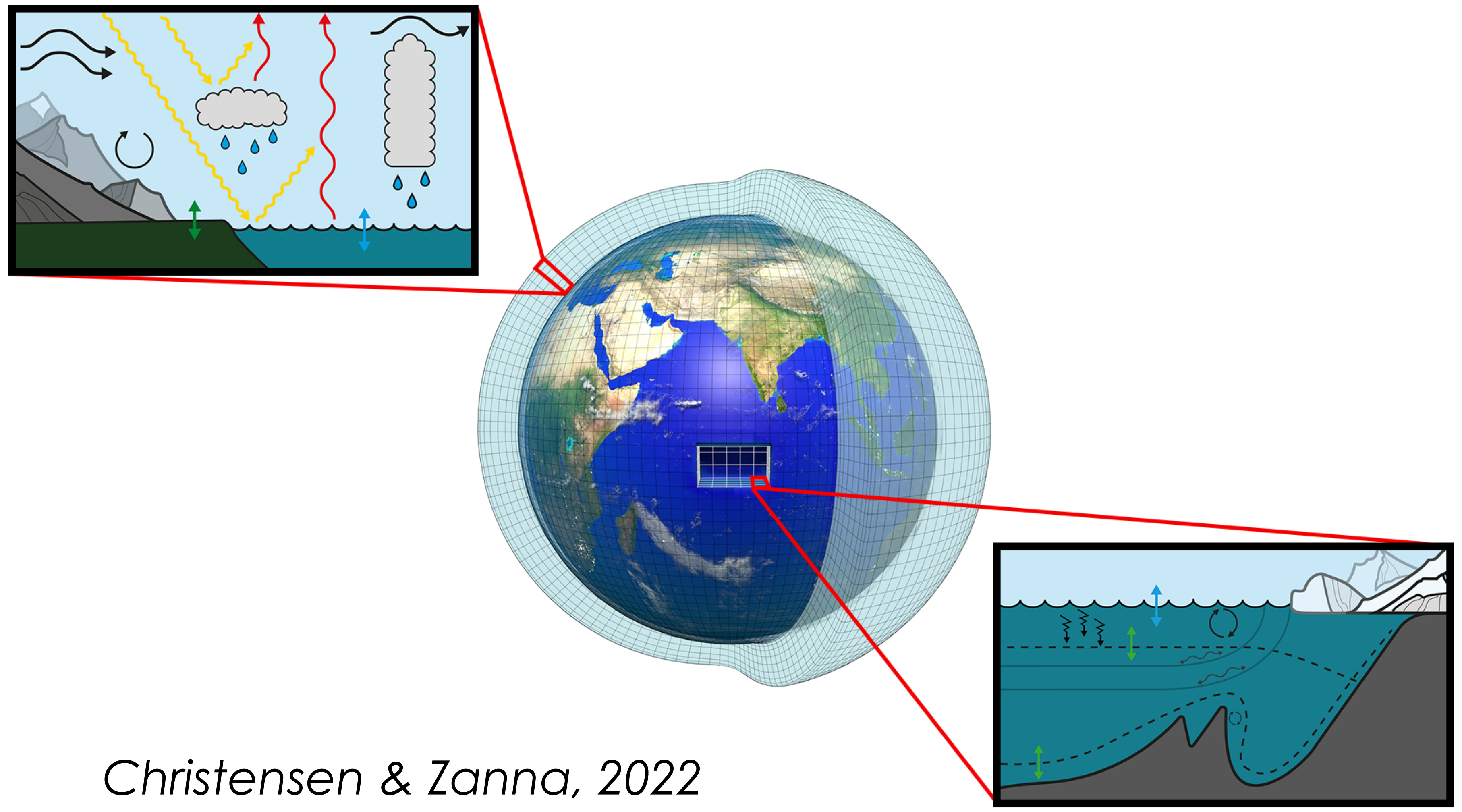
IPSL-CM6-LR



Boucher et al., 2020

Model error

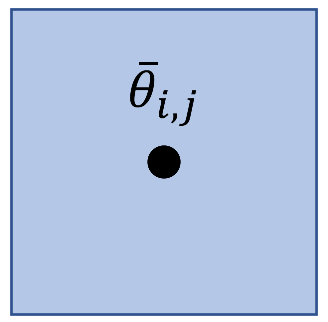
➡ Poor or lacking representation of key processes



Christensen & Zanna, 2022

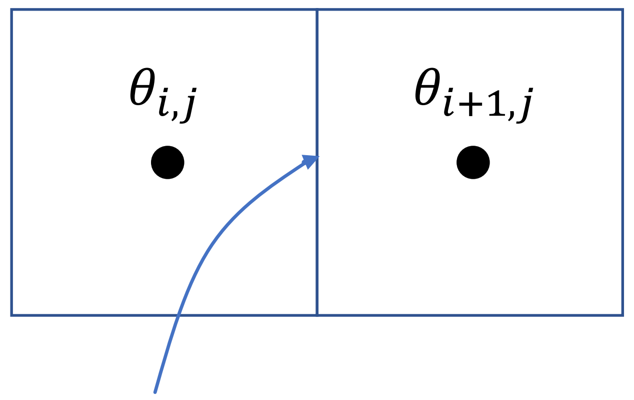
➡ Numerics

a) Finite volume



$$V_{i,j}\frac{\partial \theta_{i,j}}{\partial t} = \oint \kappa \nabla\theta \cdot d\boldsymbol{A}$$

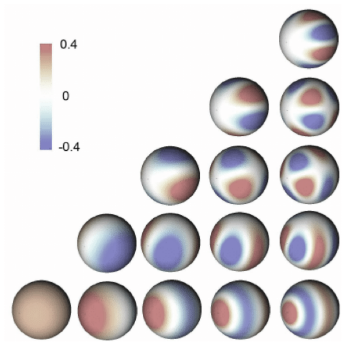
b) Finite difference

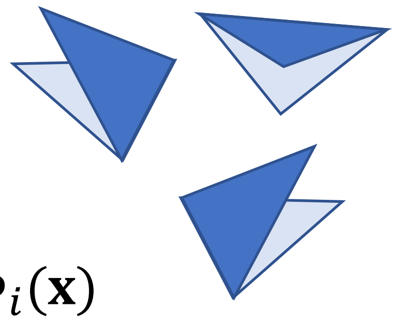


$$\frac{\partial \theta}{\partial x} \approx \frac{\theta_{i+1,j} - \theta_{i,j}}{\Delta x}$$

truncate the functional operators

c) Spectral or finite element





$\Phi_i(\mathbf{x})$

$$\theta(\mathbf{x}) = \sum_{i=1}^{n} A_i \Phi_i$$

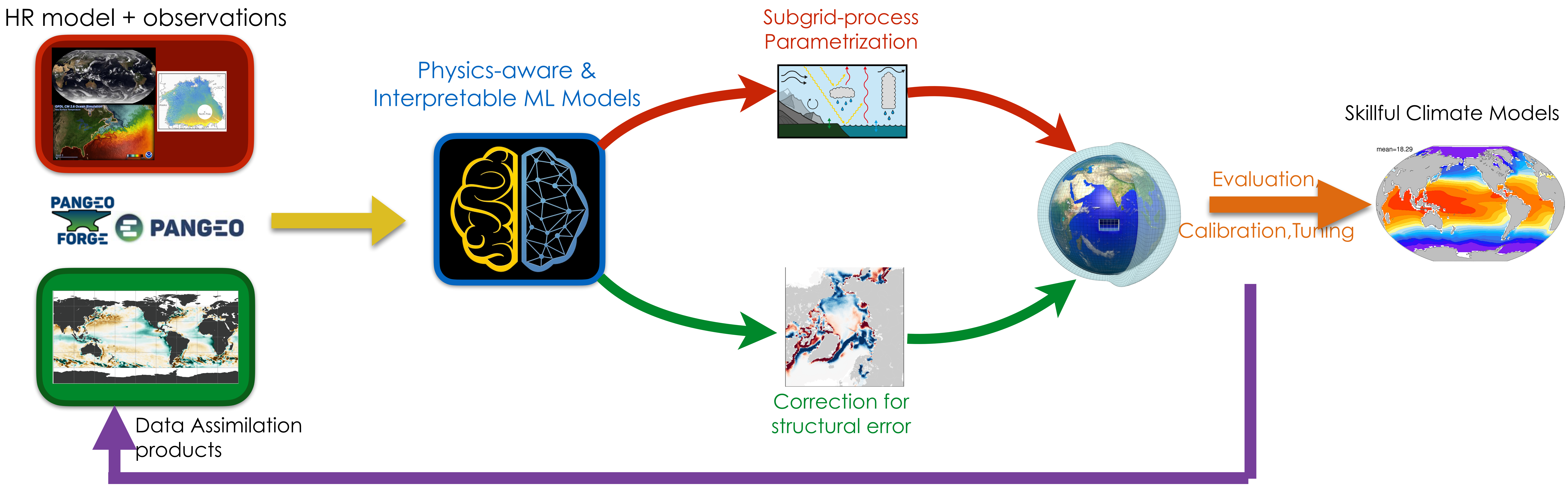
$$\frac{\partial}{\partial t}\int W\theta \, dV = \int \kappa \nabla\theta \cdot \nabla W dV$$

truncate the functional space

M2LInES: Multiscale Machine Learning In Coupled Earth System Modeling

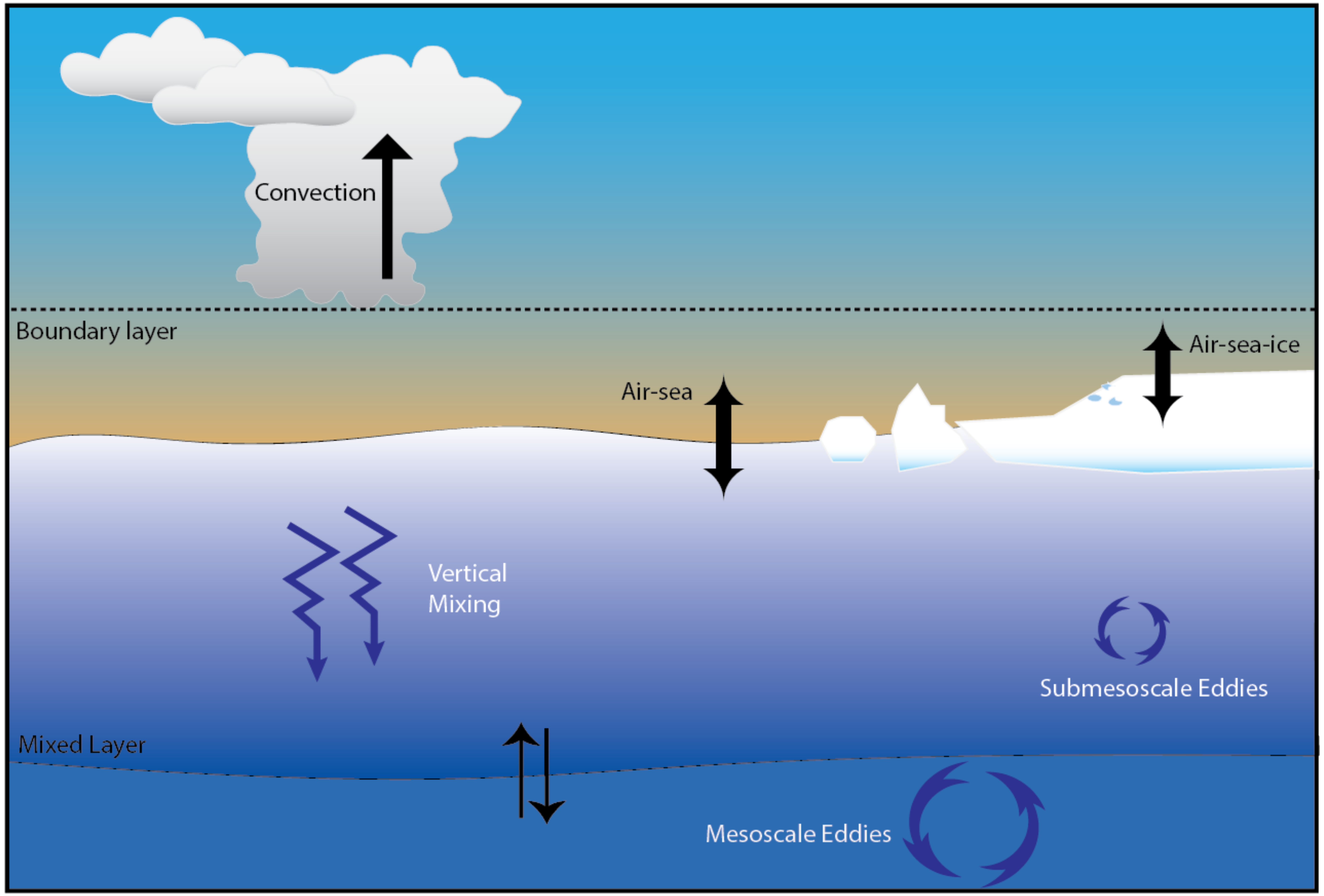
Goal: Improve the skill in surface ocean, ice, atmospheric fields on timescales of hours to centuries in global climate models, *and provide more reliable climate predictions*

- ***Development of new ML, data-driven, physics-aware parameterizations*** of subgrid ocean, sea-ice & atmosphere processes
- ***Reduction of structural model biases*** (numerics, missing physics & poor subgrid parameterizations) in **existing climate models at GFDL, NCAR & IPSL**



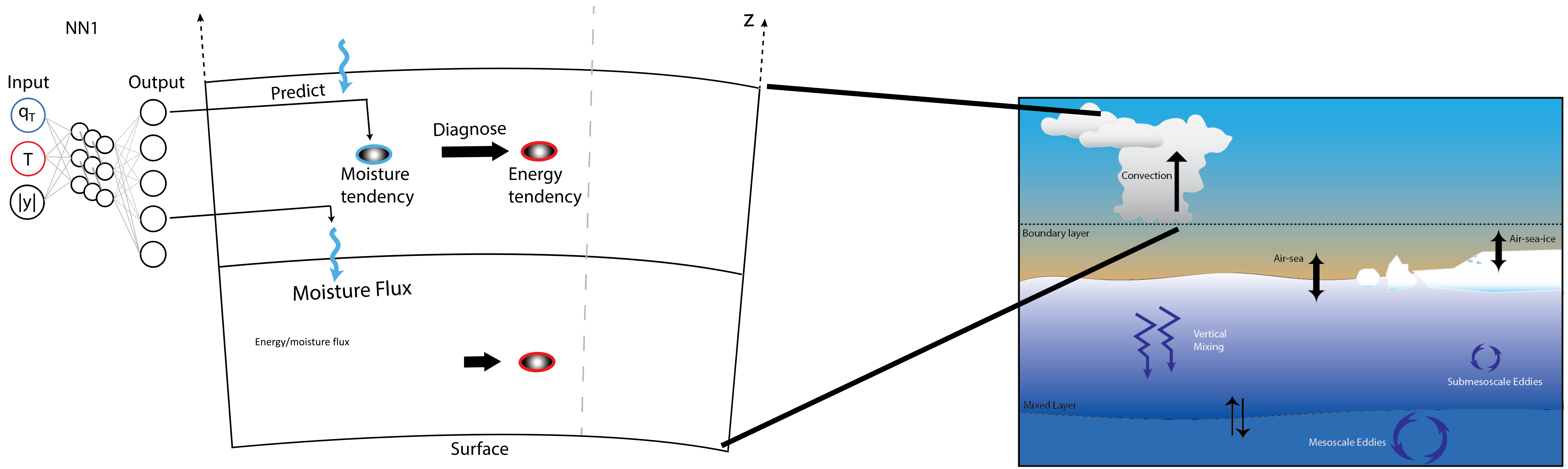
M2LInES: Ocean, Atmosphere, Sea-Ice Key Processes

- **Focus on learning new physics** of processes key for coupled climate dynamics



Convection & microphysics parameterization for improved ECS & precipitation

➡️ Substitutes deep convection & microphysics schemes using data from SAM

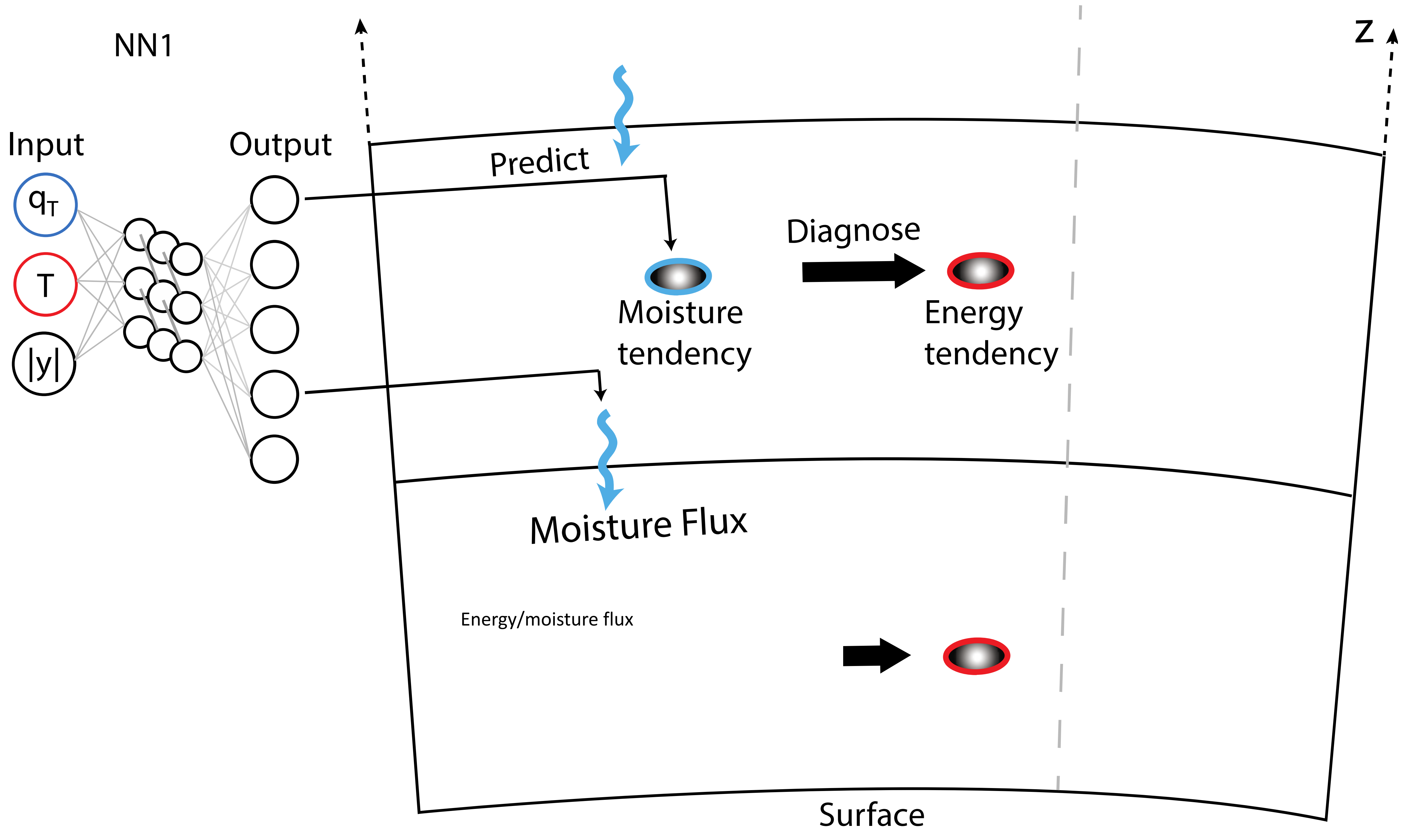


Yuval & O'Gorman (2020), Yuval O'Gorman & Hill (2021)

Yuval, O'Gorman (MIT)

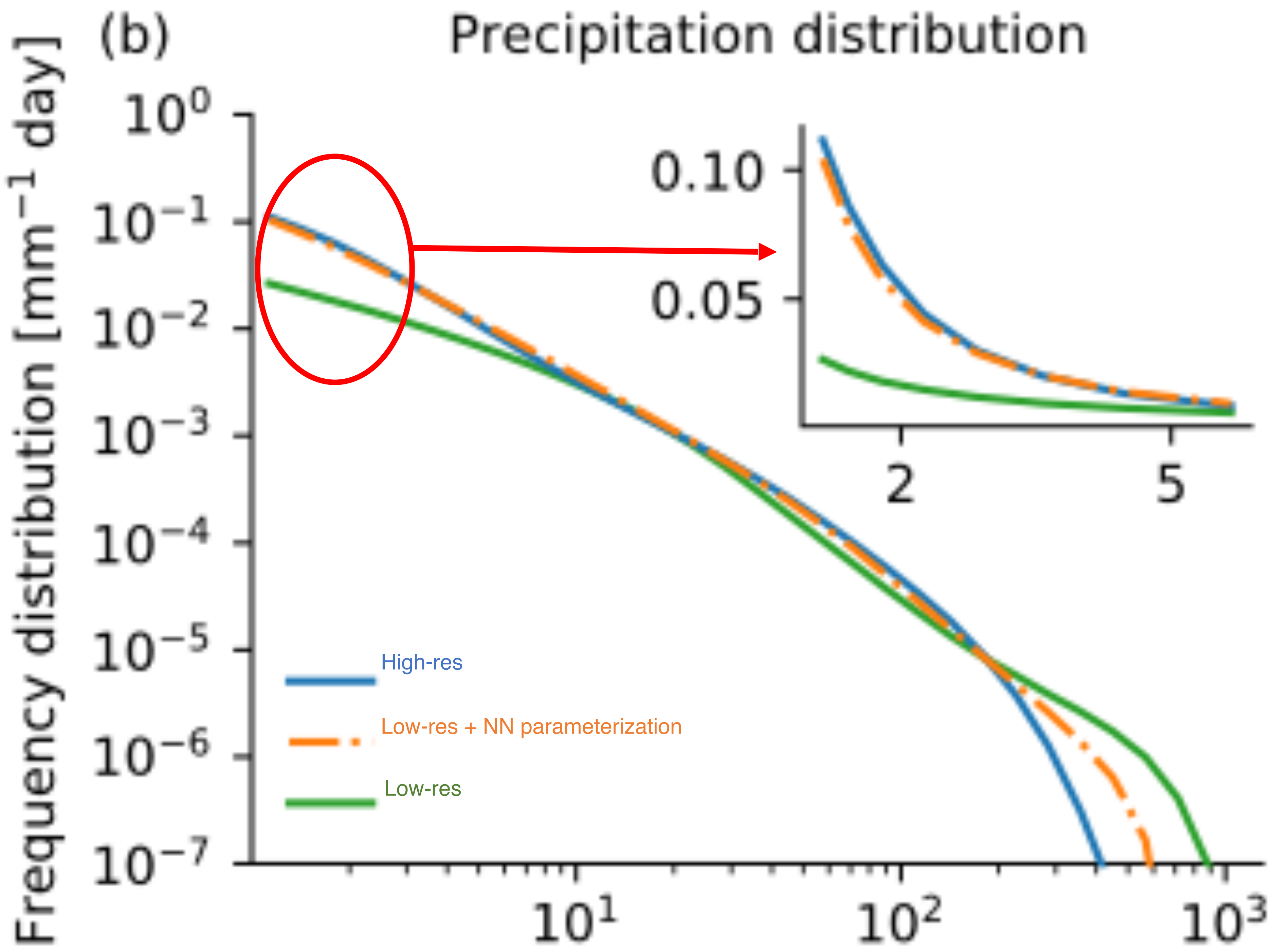
Convection & microphysics parameterization for improved ECS & precipitation

➡️ Substitutes deep convection & microphysics schemes



Yuval & O'Gorman (2020), Yuval O'Gorman & Hill (2021)

➡️ Potential to improve the representation of precipitation

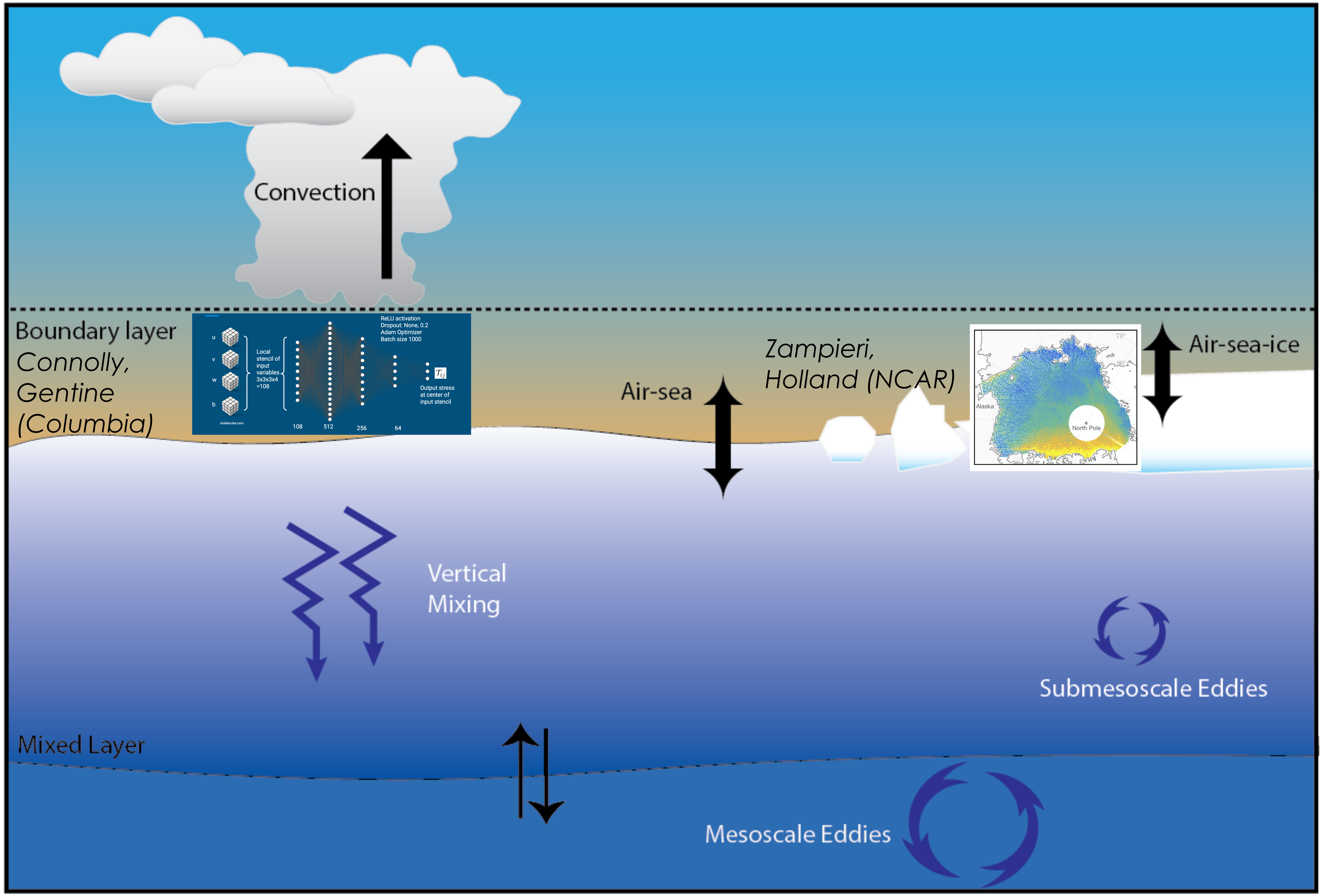


Precipitation rate [mm day^{-1}]

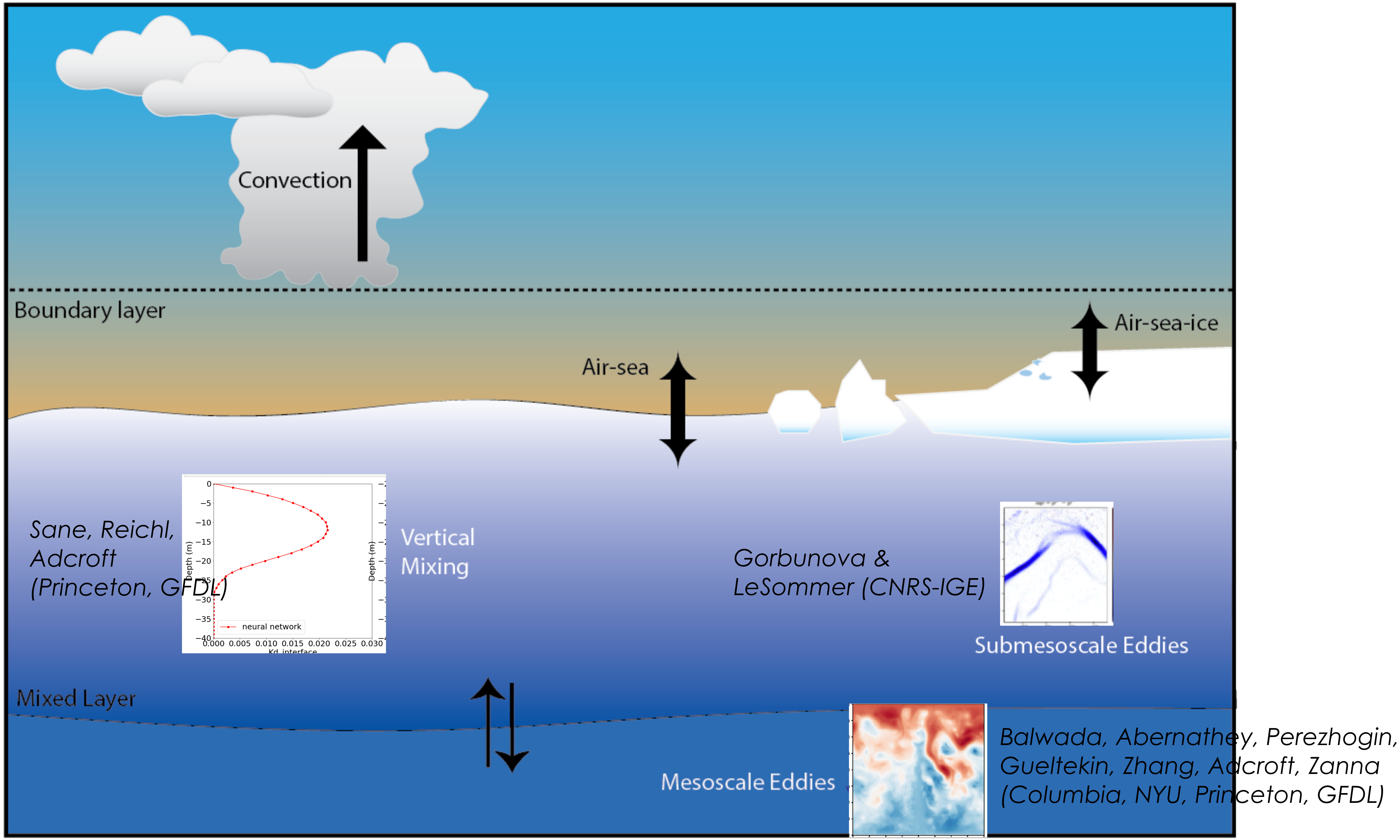
Parameterization conserves energy and water & leads to stable simulations

➡️ **In progress:** Implementation of ML convection & microphysics into CAM *(led by Paul O'Gorman & Janni Yuval)*

APBL & Sea-ice parameterizations to improve radiative feedbacks



Ocean parameterizations to reduce uncertainty in TCRE, OHU, tracer transport

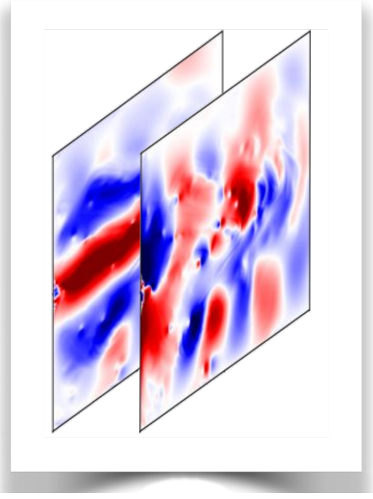


Developing & Implementing Ocean Backscatter Parameterizations

- Novel ML backscatter parameterizations:

Equation-discovery (BZ20)

Zanna & Bolton 2020



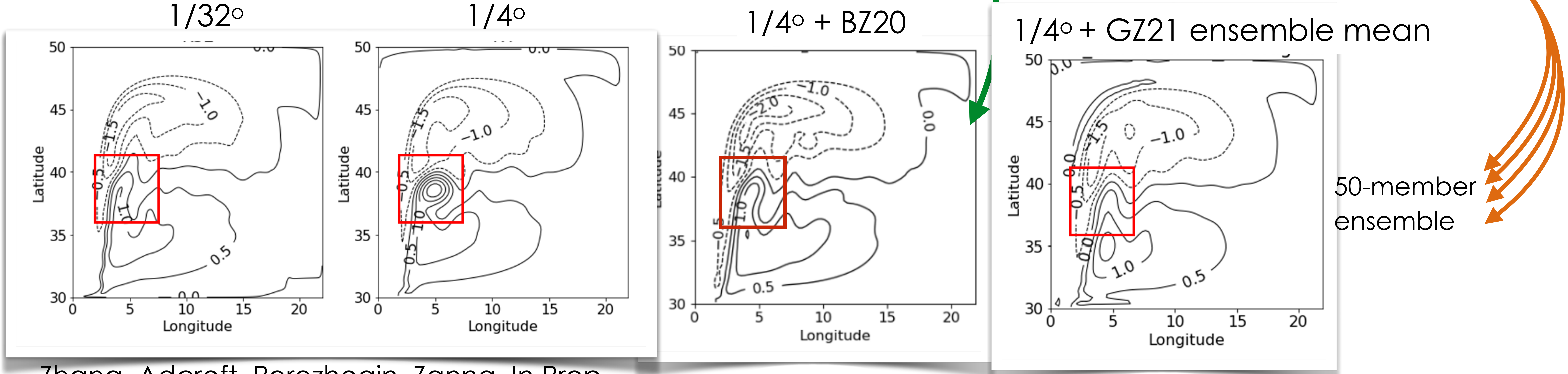
$$\hat{\mathbf{S}} = \kappa \overline{\nabla} \cdot \begin{pmatrix} \zeta^2 - \zeta D & \zeta \tilde{D} \\ \zeta \tilde{D} & \zeta^2 + \zeta D \end{pmatrix}$$

(scalar)

Stochastic parametrization (GZ21)

Guillaumin & Zanna 2021

- MOM6 implementation in idealized config (testing of numerics, stability, python-fortran, ...)



Zhang, Adcroft, Perezhogin, Zanna, In Prep

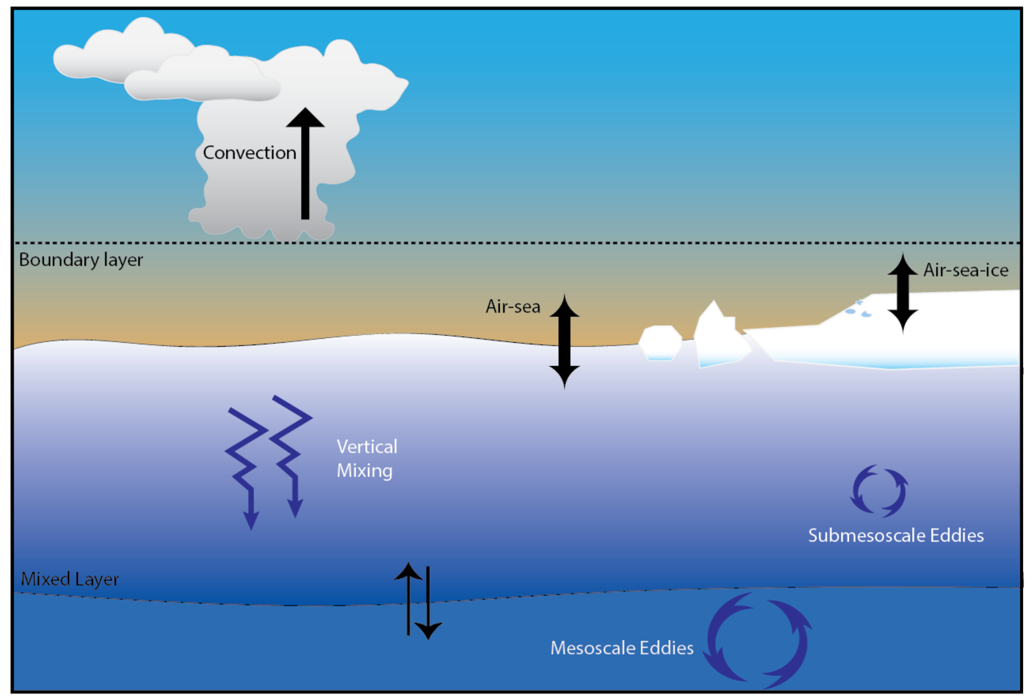
Dual Strategy: HR & DA to reduce structural model error

Pros

- Targeting physics: *improved transport*
- Agnostic to the coarse-res GCM
- Transferable (for the most part, **some tuning necessary!**)

Challenges

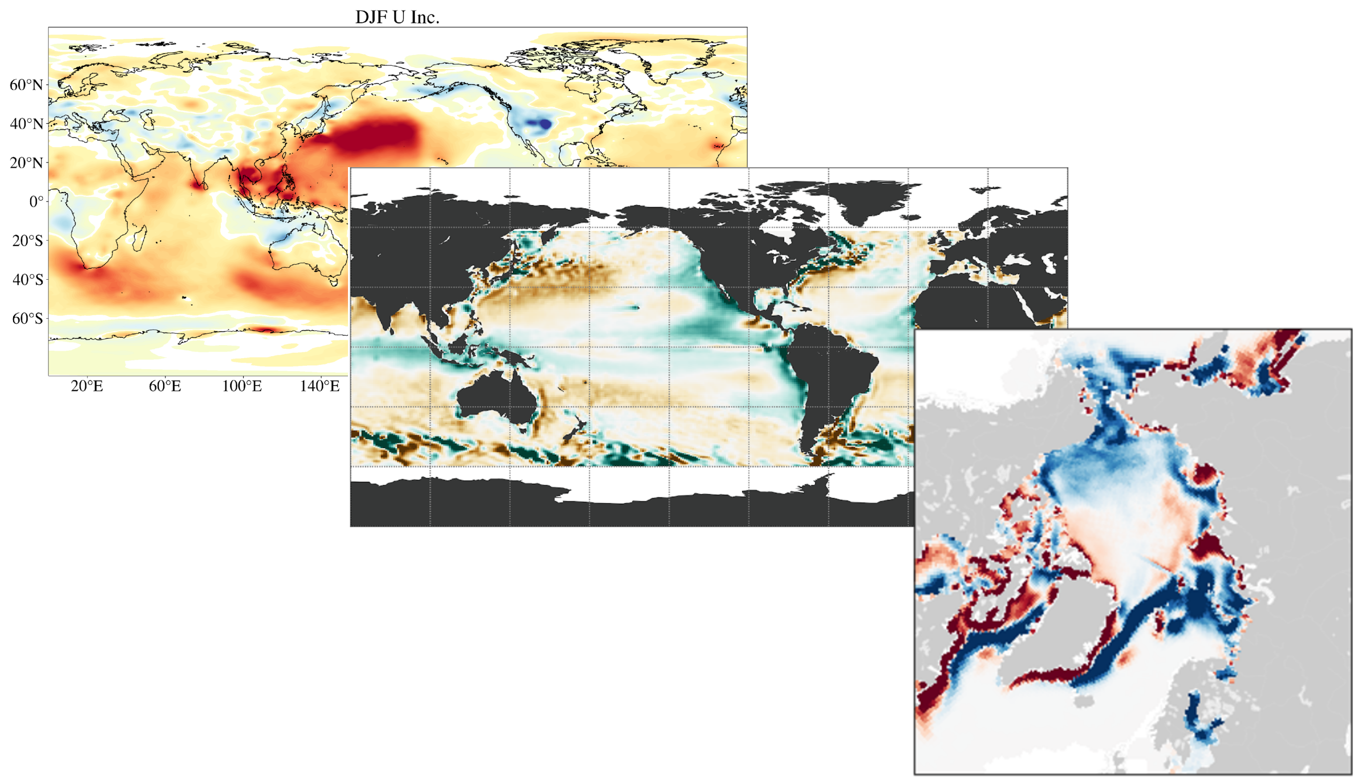
- Compensating errors
- Relies on high-res model (not obs)
- Limited domain, regime, etc



Our Strategy

- Simultaneously targeting many processes
- Adapting our metrics & diagnostics for learning

- **Machine learning model error (model - observation) from DA increments**
  - Integrated errors which projects onto the long term bias
  - Contains different common & model-specific biases
  - Focus on existing or missing parameterizations

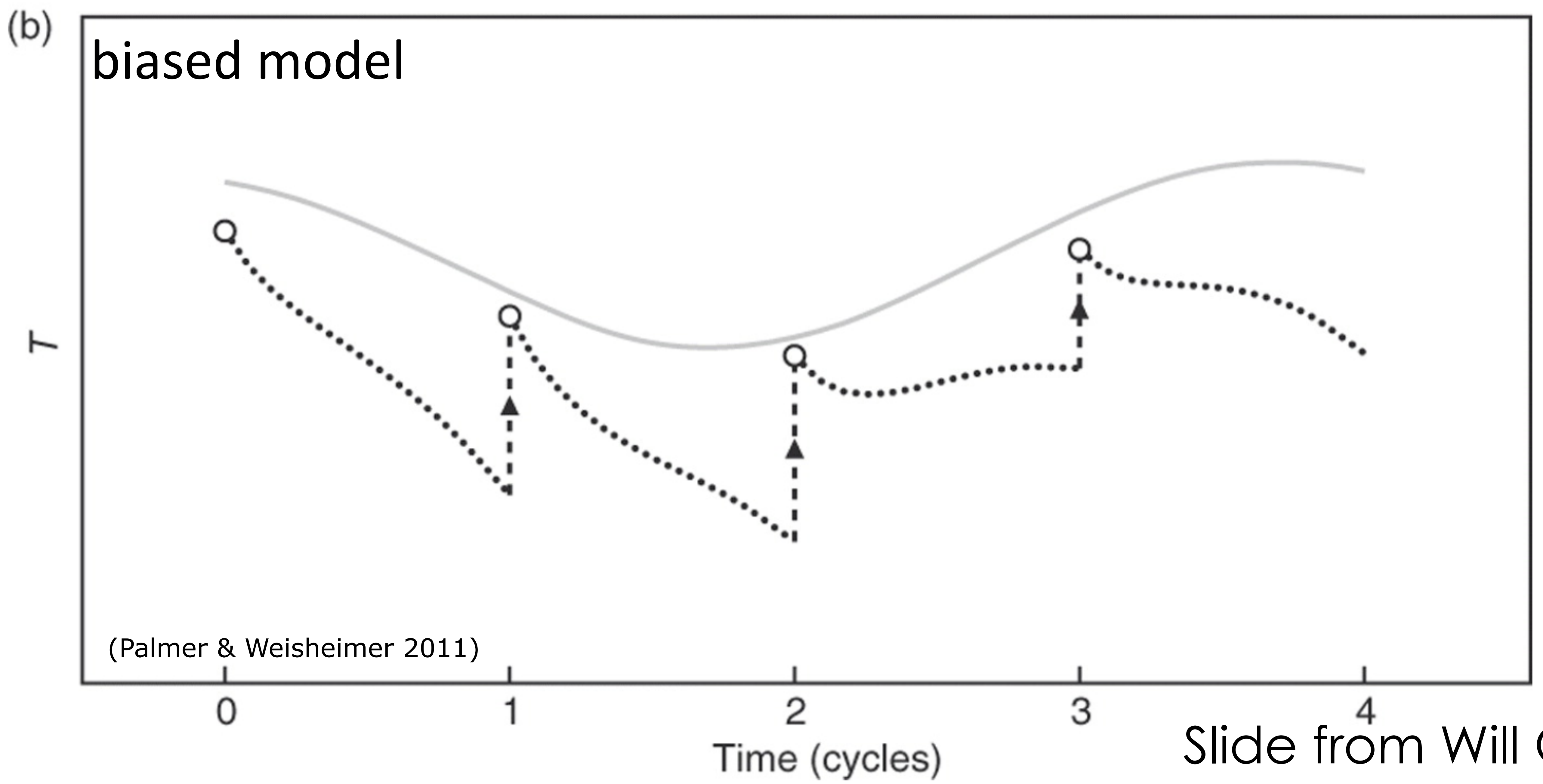


M²LInES: Learning Model Error from Data Assimilation Increments

- **Focus on learning structural model error** from data assimilation (model - observation)

*Although the amplification of the effect of **FORCING** on **BIASED PROCESS** will occur on **timescales of decades**, the intrinsic timescale associated with **BIASED PROCESS** itself is typically on the order of hours. Hence it should in principle be possible to assess whether the anomalously small values of **PARAMETRIZED PROCESS** are realistic or not, by studying the performance of such models in short-range weather prediction mode.*

(see Klinker & Sardeshmukh 1992; Rodwell & Palmer 2007; Palmer & Weisheimer 2011)

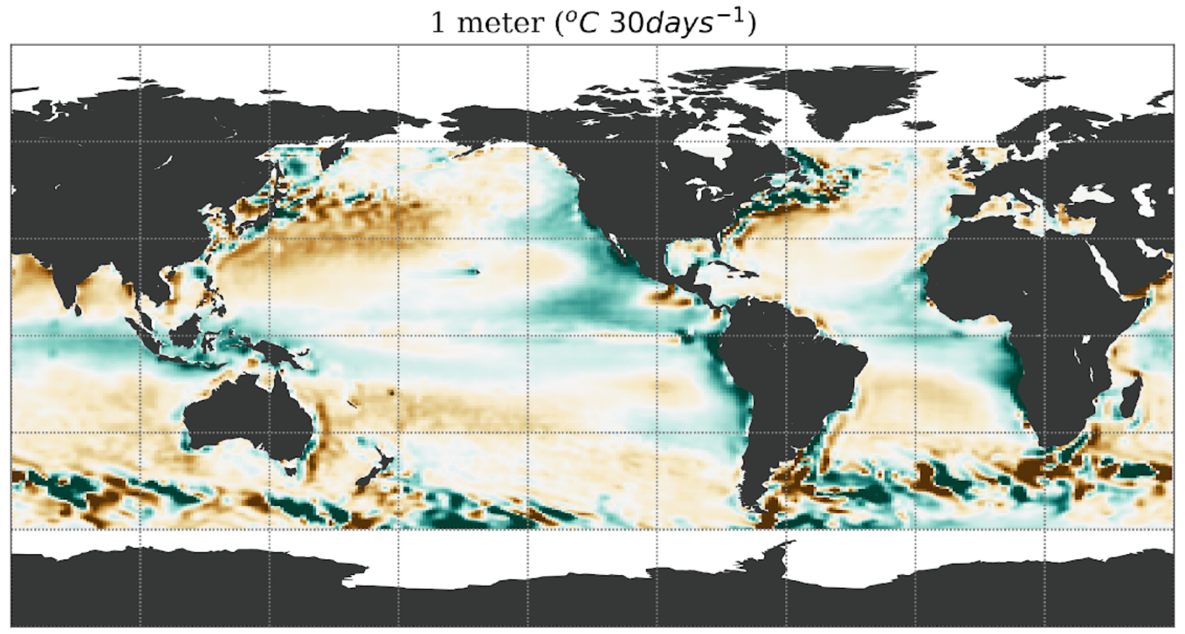


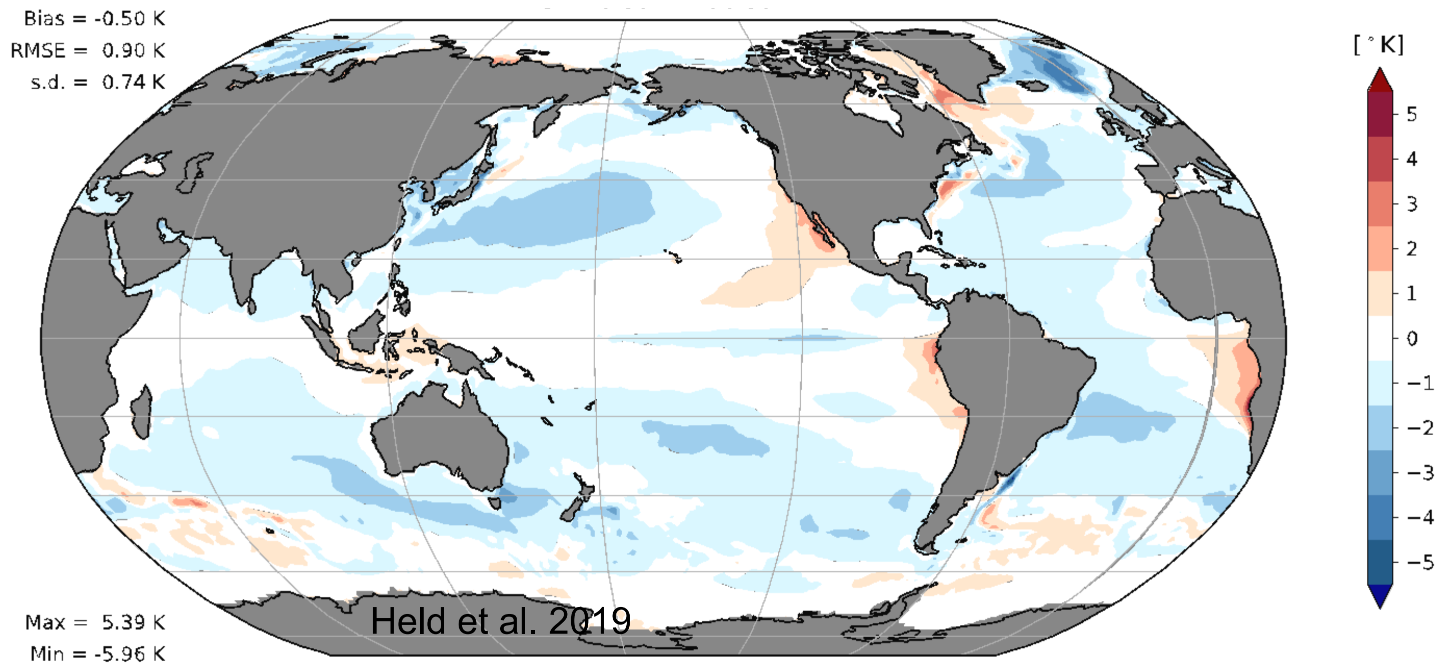
Can we machine-learn model error arising from Gent-McWilliams, convective rate, KPP, sea-ice conductivity, from DA increments to improve models ?

Slide from Will Chapman (NCAR), Will Gregory & Tarun Verma (GFDL)

DA Increments project onto Mean Model Bias?

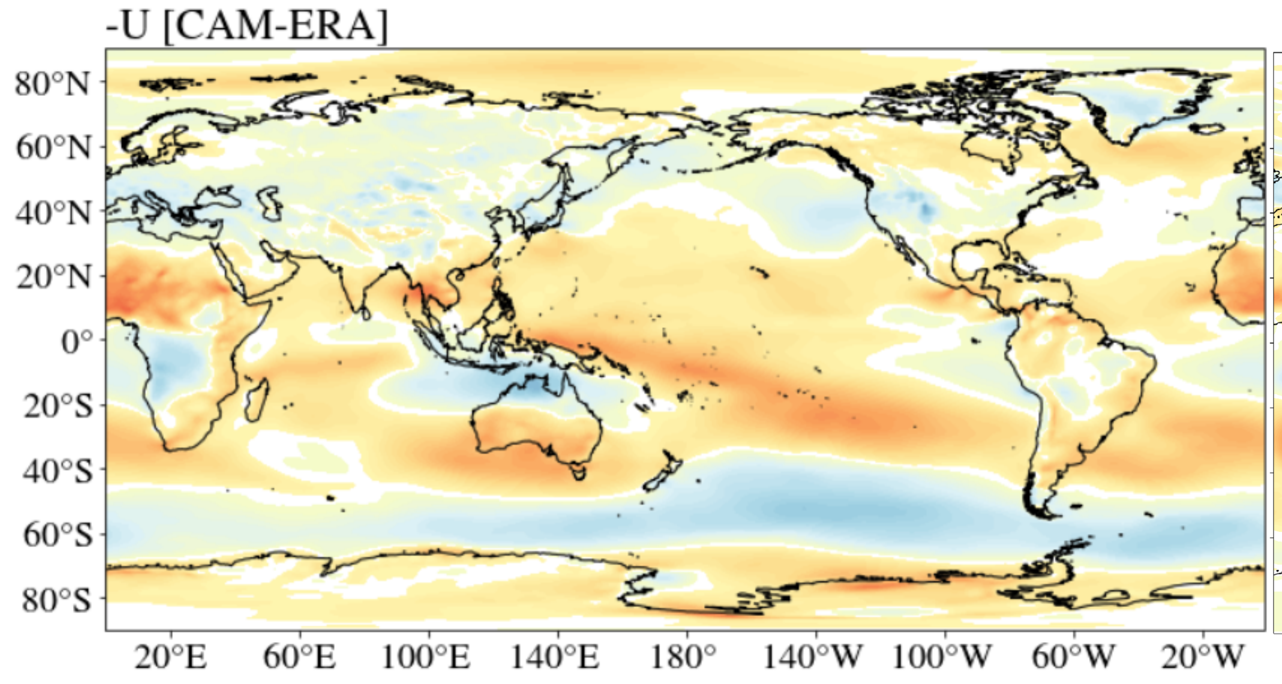
Mean Temperature Increments from SPEAR (GFDL)



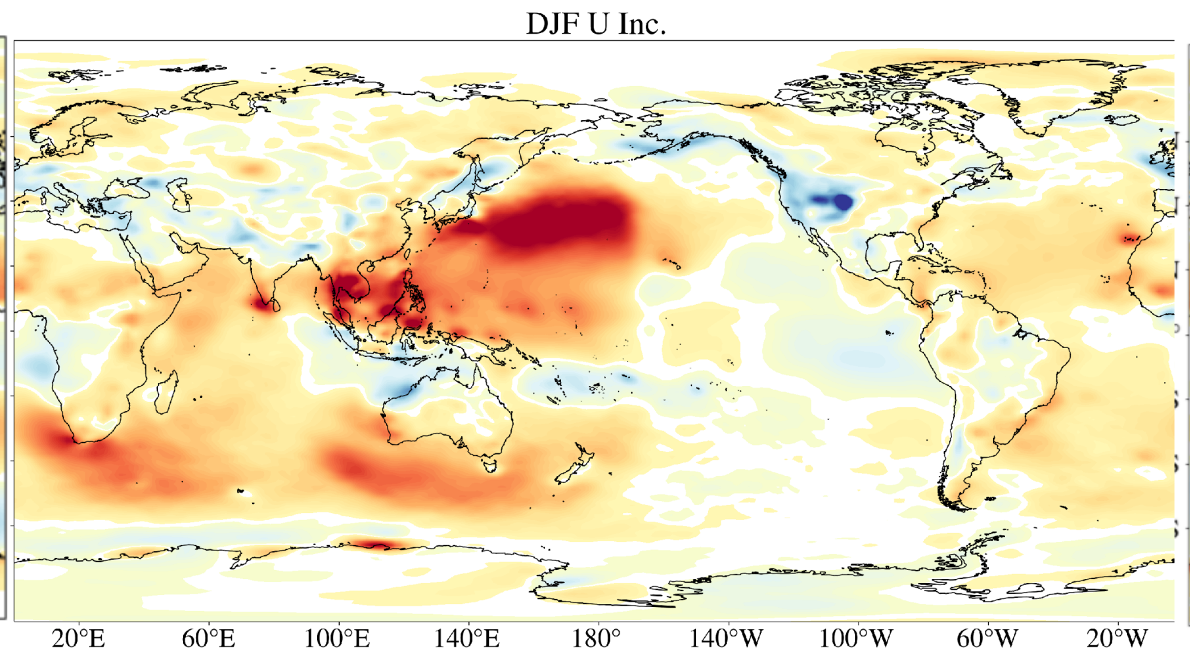


Verma, Liu, Adcroft (GFDL)

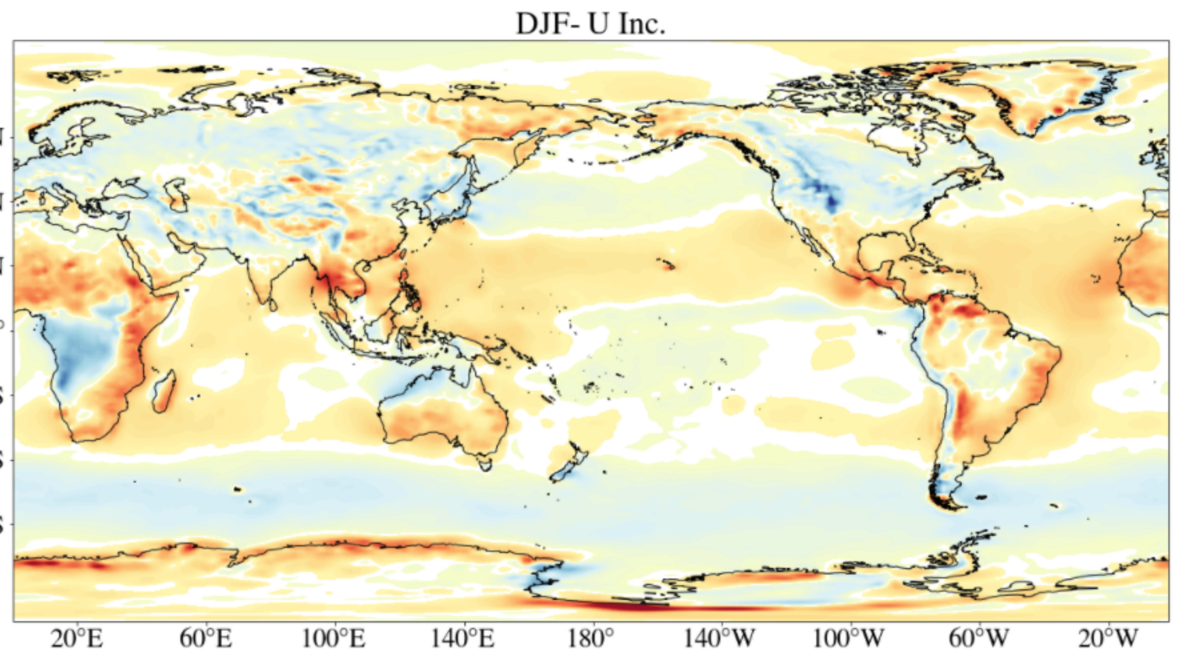
Free Bias



DA Increment



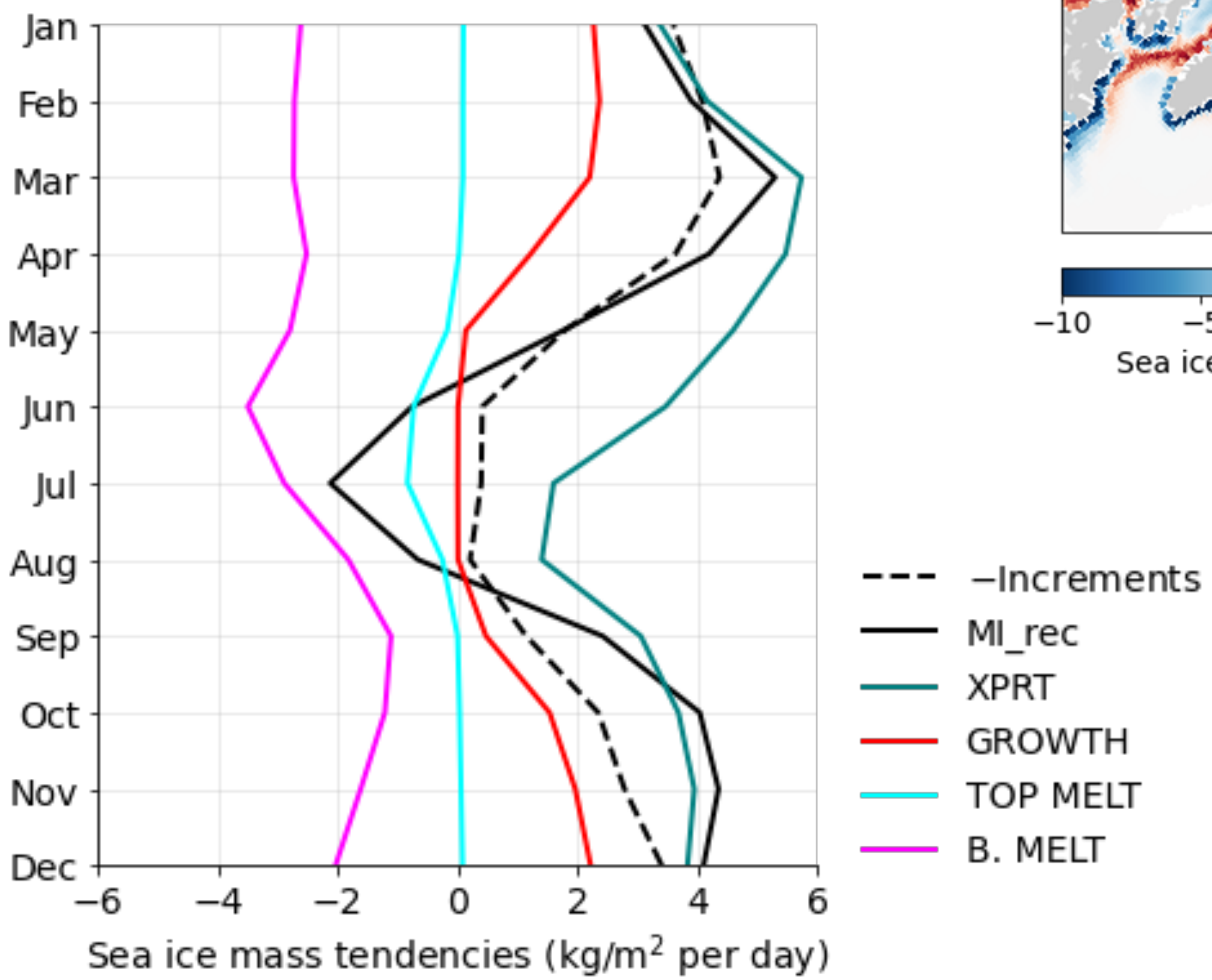
Nudge Increment

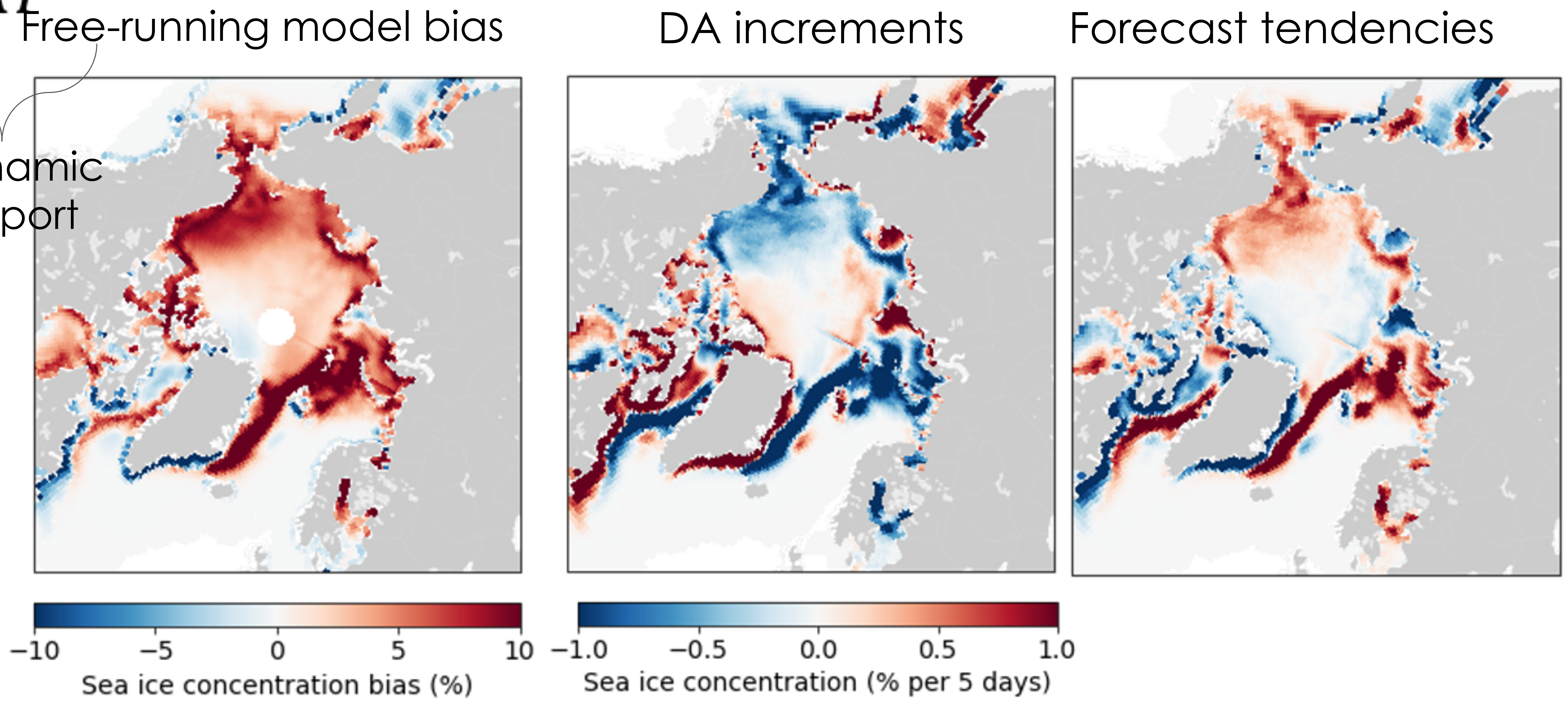


Chapman, Berner (NCAR)

Learning state-dependent parameterizations from DA

$$\underbrace{\frac{\partial}{\partial t}MI}_{\text{Change in mass of snow \& ice}} = \underbrace{TSNK + BSNK}_{\text{Top \& bottom melt}} + \underbrace{LSRC}_{\text{Thermo-dynamic ice growth}} + \underbrace{XPRT}_{\text{Dynamic export}}$$





Gregory, Bushuk (GFDL)

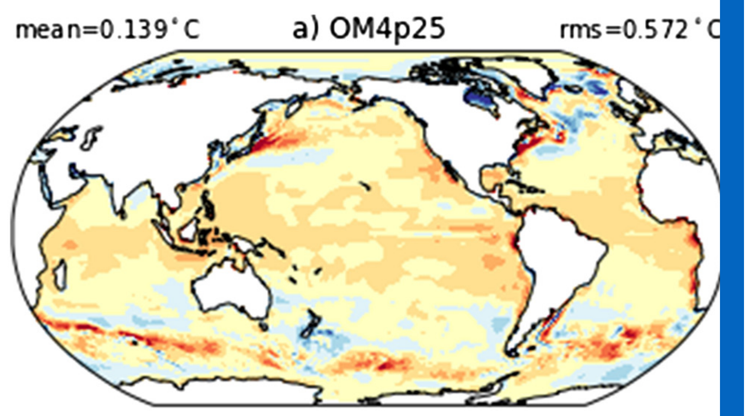
- **Next step:** machine learned state dependent corrections for coarse-resolution models
- **M²LInES Strategy:** learning in // from HR & DA; interpretability to identify common biases & iteratively target remaining/ model-specific biases

A few CESM-relevant plans for next year @ M²LInES

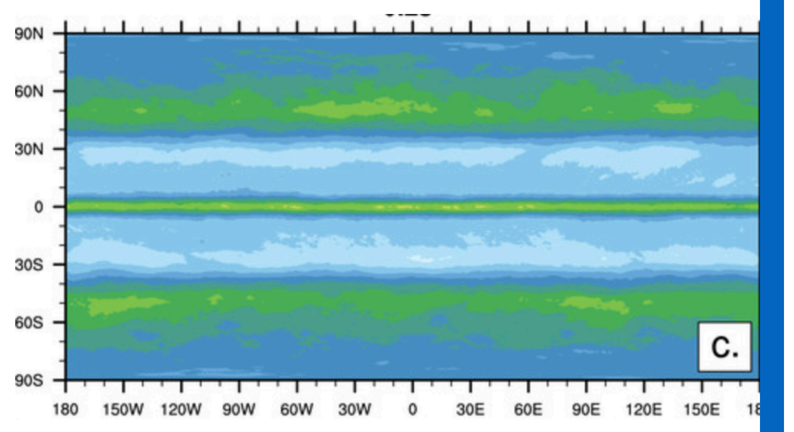
- ML parameterizations from HR + DA: some of the new problems are the same as the old ones: calibration, tuning, evaluation, stability … but a lot of exciting work to be done

- Focus on component models before moving to the coupled system

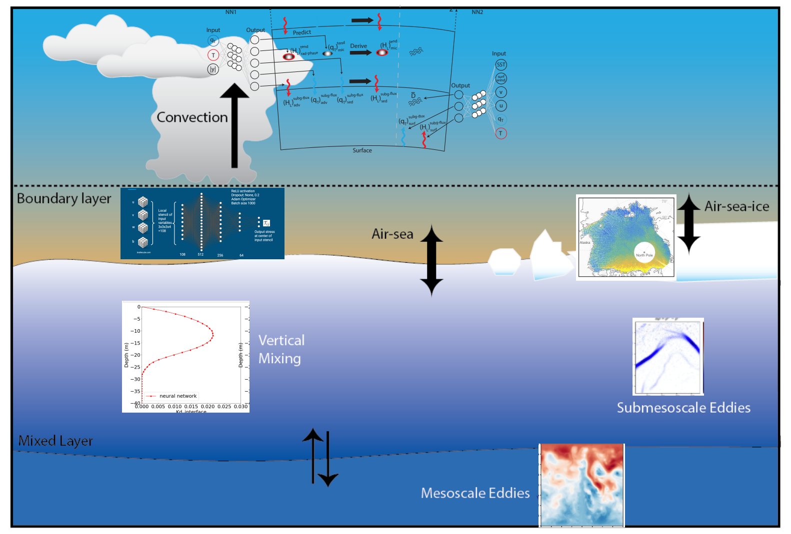
- **Global MOM6 OMIP runs with ML parameterizations:**
  - ePBL + ML
  - mesoscale closures (CNN & equation-discovery)



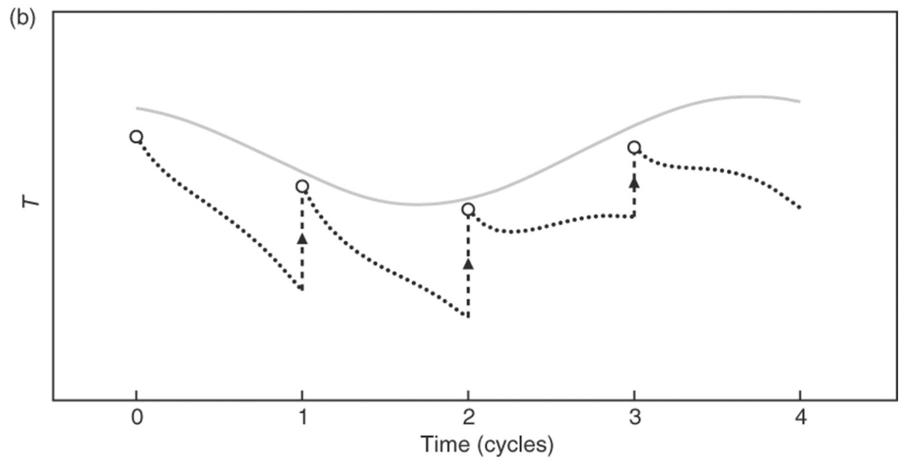
- **Aquaplanet CAM with ML convection** & microphysics



- ***Refining other ML closures + implementation in CAM, NEMO, MOM6***: APBL, sea-ice. submeso, mesoscales, nonlinear equation of state, MLE, …



- **Further developing ML DA** increments for state-dependent parameterizations for sea-ice, ocean, atmosphere models



M²LInES: Advances & Tools for the Community

m2lines

https://github.com/m2lines

Overview | Repositories | Projects | Packages | Teams | People

Find a repository... | Type | Language

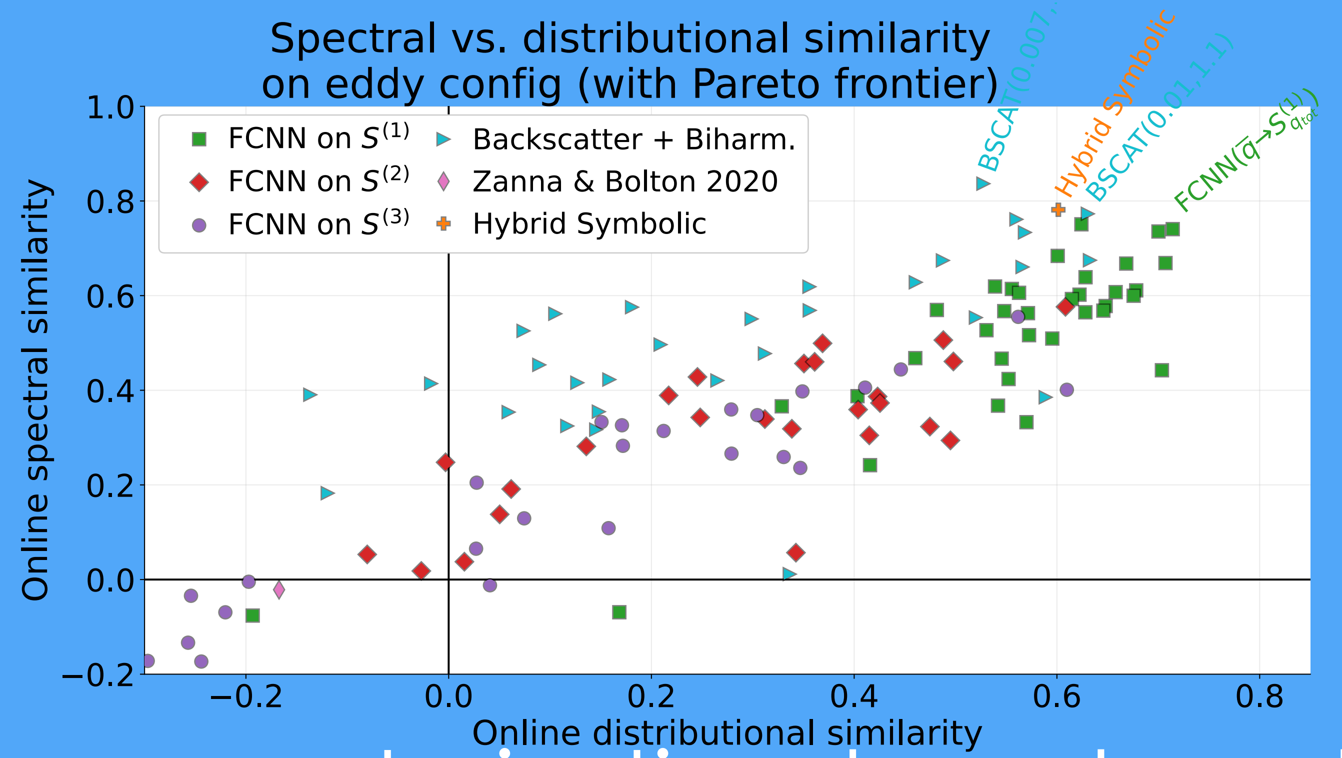
Jupyter Book: Learning Machine Learning with Lorenz-96

Learning Machine

Introduction

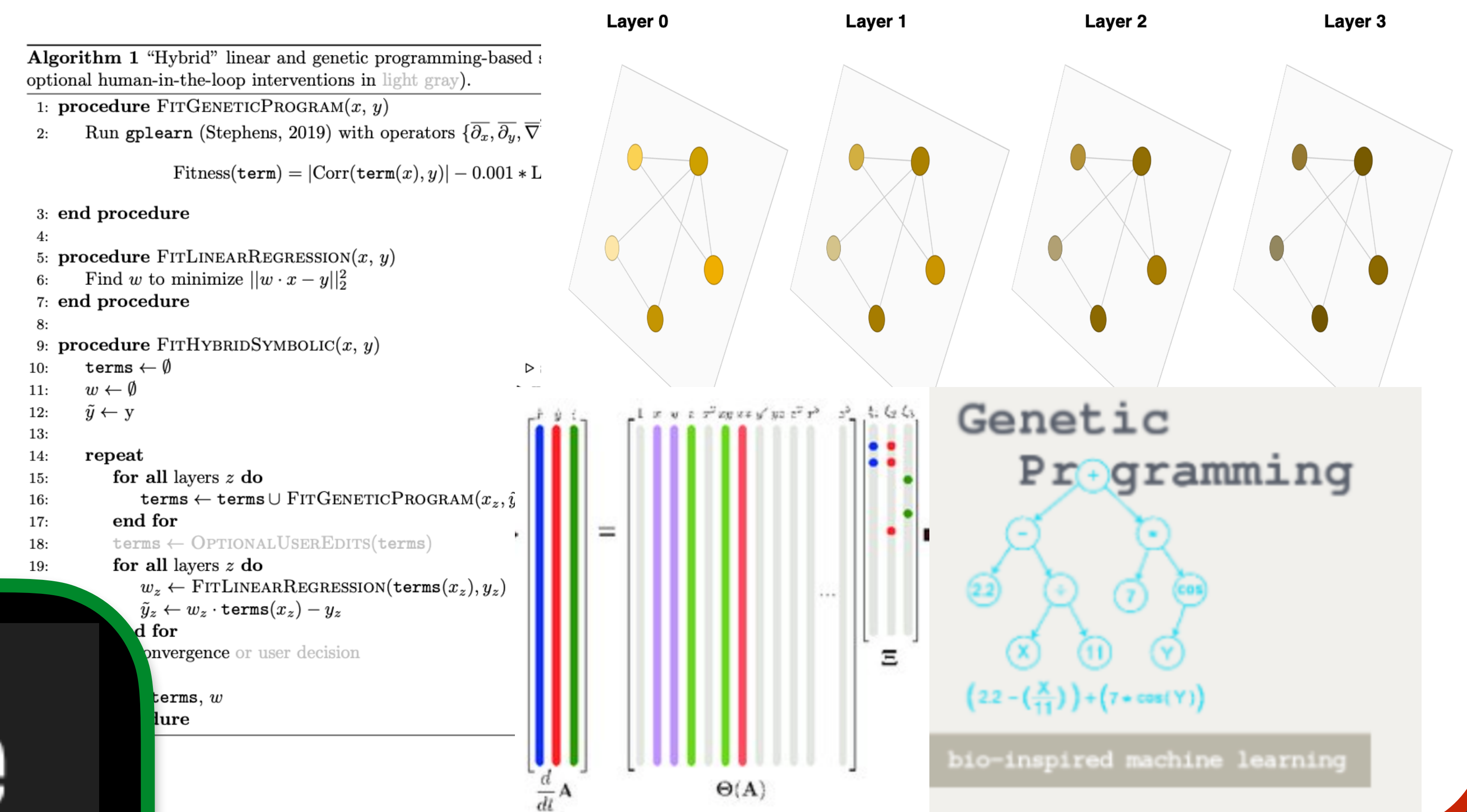
The climate system is composed of many interacting components and described by complex nonlinear equations. These equations are solved numerically under a number of simplifications, therefore leading to errors. The errors are the result of numerics used to solve the equations and the lack of an approximate ...resentation of processes occurring below the resolution of the climate model grid.

✅ Developed a testbed with quantitative metrics for ML parametrization benchmarks



https://github.com/m2lines/pyqg_parameterization_benchmarks

ML Algorithms & trained models



YouTube

HOME | VIDEOS | PLAYLISTS | CHANNELS | ABOUT

Created playlists

Intro to M²LInES — VIEW FULL PLAYLIST (2)

Machine Learning — VIEW FULL PLAYLIST (9)

Big Data and Cloud — VIEW FULL PLAYLIST (4)

Climate Modeling — VIEW FULL PLAYLIST (6)

ML for Physics Discovery — VIEW FULL PLAYLIST (18)

@M2LInES

https://m2lines.github.io

NYU · Columbia University in the City of New York · LDEO · IGE · Sciences de l'environnement · Institut Pierre Simon Laplace · Princeton · GFDL NOAA · MIT · NCAR

Machine Learning for Climate Modeling: M^{2}LInES and LEAP

Galen A. McKinley (Columbia/LDEO)
and
Laure Zanna (NYU)

27th CESM Workshop, 13 June 2022

https://m2lines.github.io https://leap.columbia.edu/

Machine Learning for Climate Modeling: M^{2}LInES and LEAP

Galen A. McKinley (Columbia/LDEO)
and
Laure Zanna (NYU)

27th CESM Workshop, 13 June 2022

https://m2lines.github.io

https://leap.columbia.edu/