Validating and Enforcing Physical Consistencies in AI Weather Prediction Models

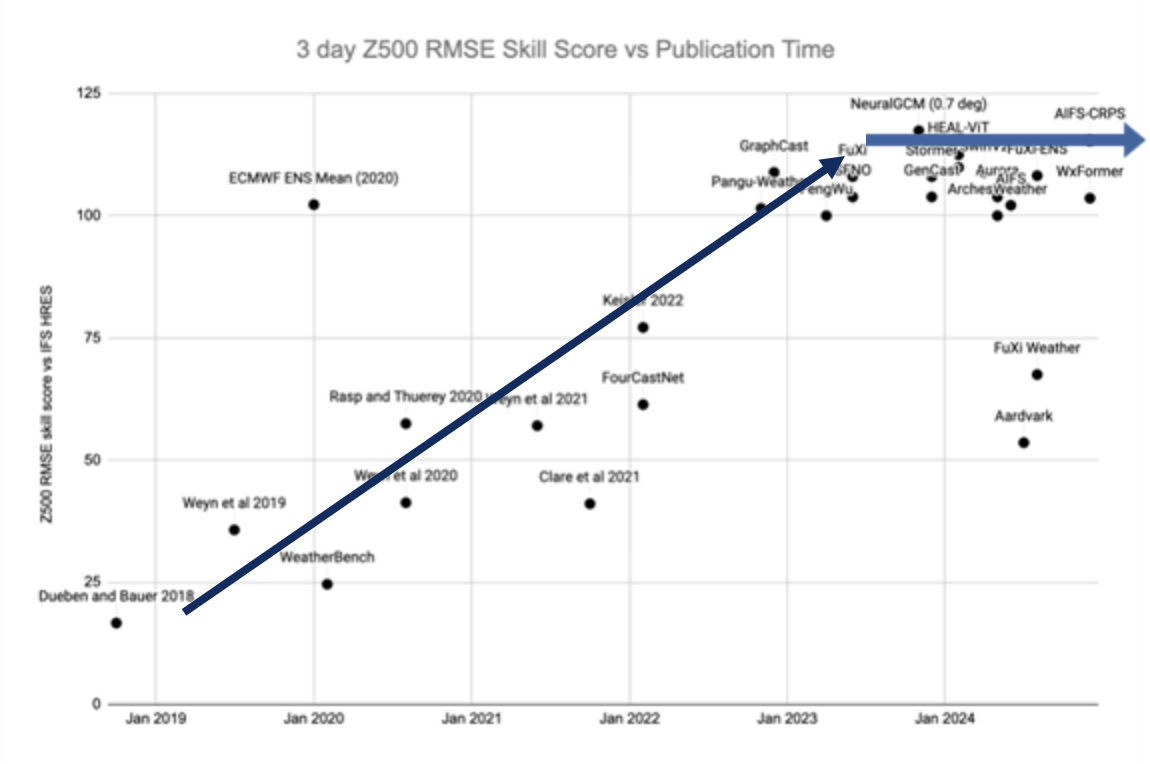
David John Gagne
Machine Learning Scientist V, NSF NCAR

***CREDIT Team**: John Schreck, Will Chapman, **Yingkai Kyle Sha**, David John Gagne, Dhamma Kimpara, Arnold Kazadi, Seth McGinnis, Negin Sobhani, Ben Kirk, Judith Berner, Charlie Becker, Gabrielle Gantos, Kirsten Mayer, Laure Zanna, Stanley Akor, Orhan Eroglu, Nihanth Cherukuru, Katelyn FitzGerald, John Clyne, Chris Snyder*

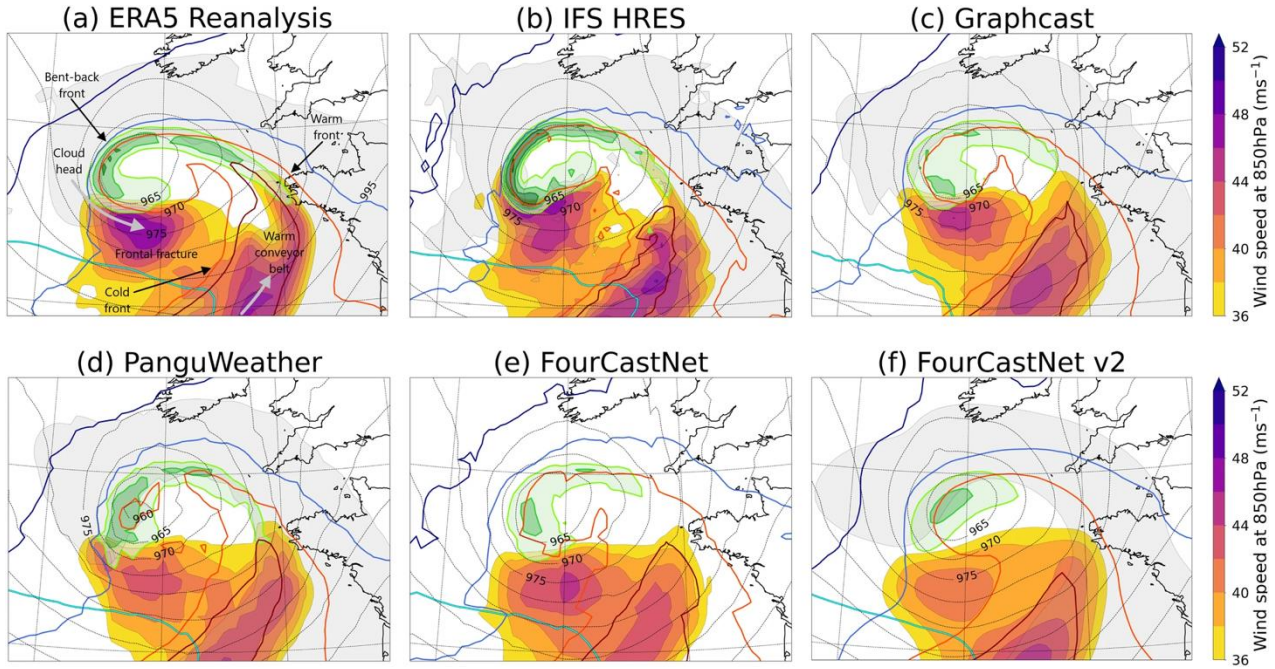
November 6, 2025
Workshop on Correctness and Reproducibility

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

Motivation: AI Weather Prediction Overall Skill vs Physical Structure



From Stephan Rasp's Weather Bench 2 Data



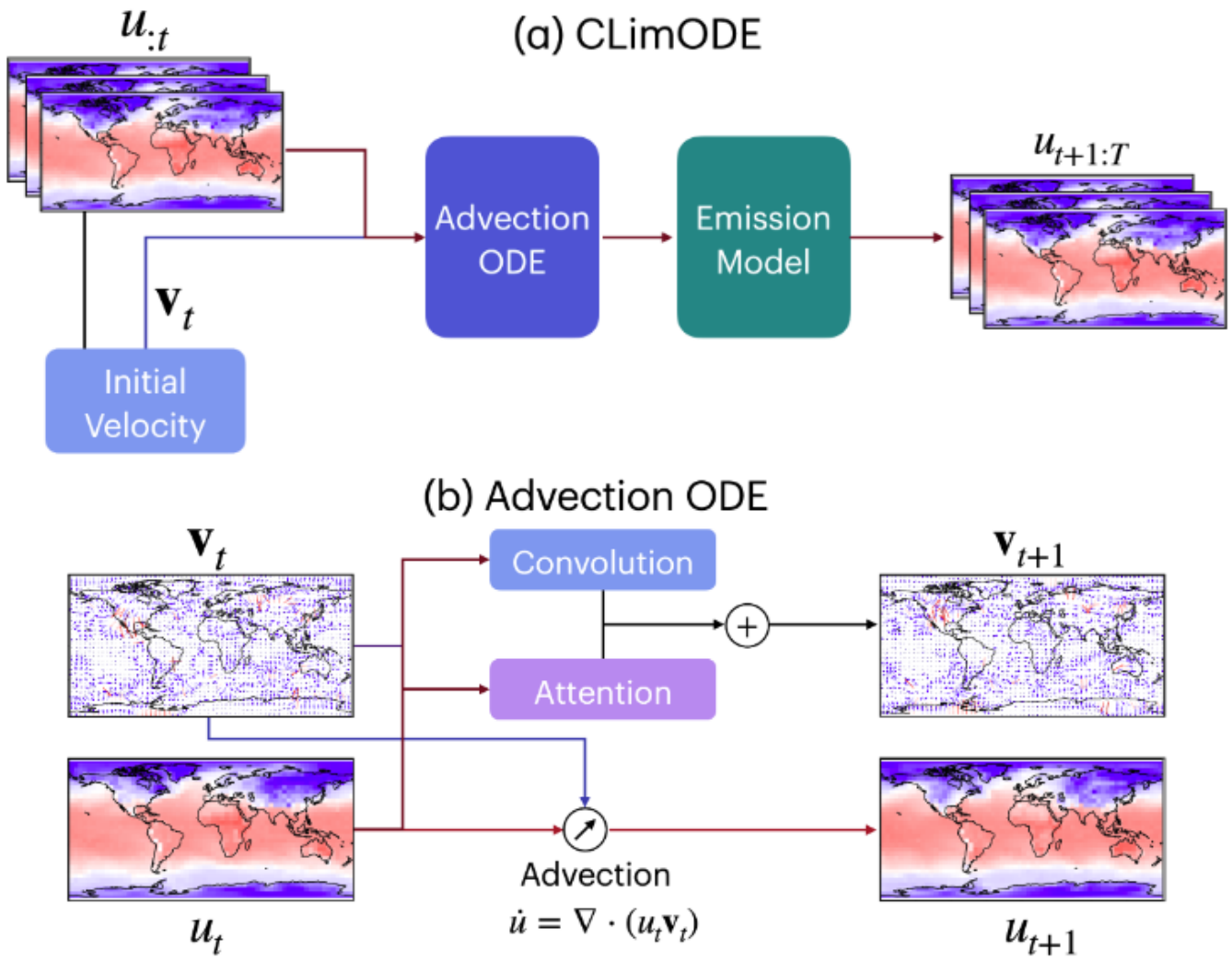
From Charlton-Perez, A.J., Dacre, H.F., Driscoll, S. et al. Do AI models produce better weather forecasts than physics-based models? A quantitative evaluation case study of Storm Ciarán. npj Clim Atmos Sci 7, 93 (2024).

After 4 years of rapid advancement in accuracy, the AI weather prediction ecosystem has matured with a variety of approaches.

Some limits to further improvements in accuracy may be due to issues with correctly resolving and overlaying finer scale storm structures and other physical features.

Motivation: How to Incorporate More Physics in AI Weather Prediction

- Multiple groups are experimenting with ways to blend explicit physics into neural network architectures
  - Neural GCM: traditional dycore + NN physics
  - Clim ODE: decomposes prediction into neural ODEs for advection, emissions, and local and global state updates
  - Aurora: physics-informed state embedding
  - HEALPix grid: equal area with consistent directionality for all grid cells
- What if there were ways to add physical constraints that were more architecture agnostic?
- Goal: Analyze data and post-model constraints for how they can improve overall performance and physical structures of weather predictions.



Yogesh, V., Markus, H., & Vikas, G. (2024, April 15). *ClimODE: Climate and weather forecasting with physics-informed neural ODEs. arXiv [cs.AI]*. Retrieved from http://arxiv.org/abs/2404.10024

CREDIT: Community Research Earth Digital Intelligence Twin

What is CREDIT?

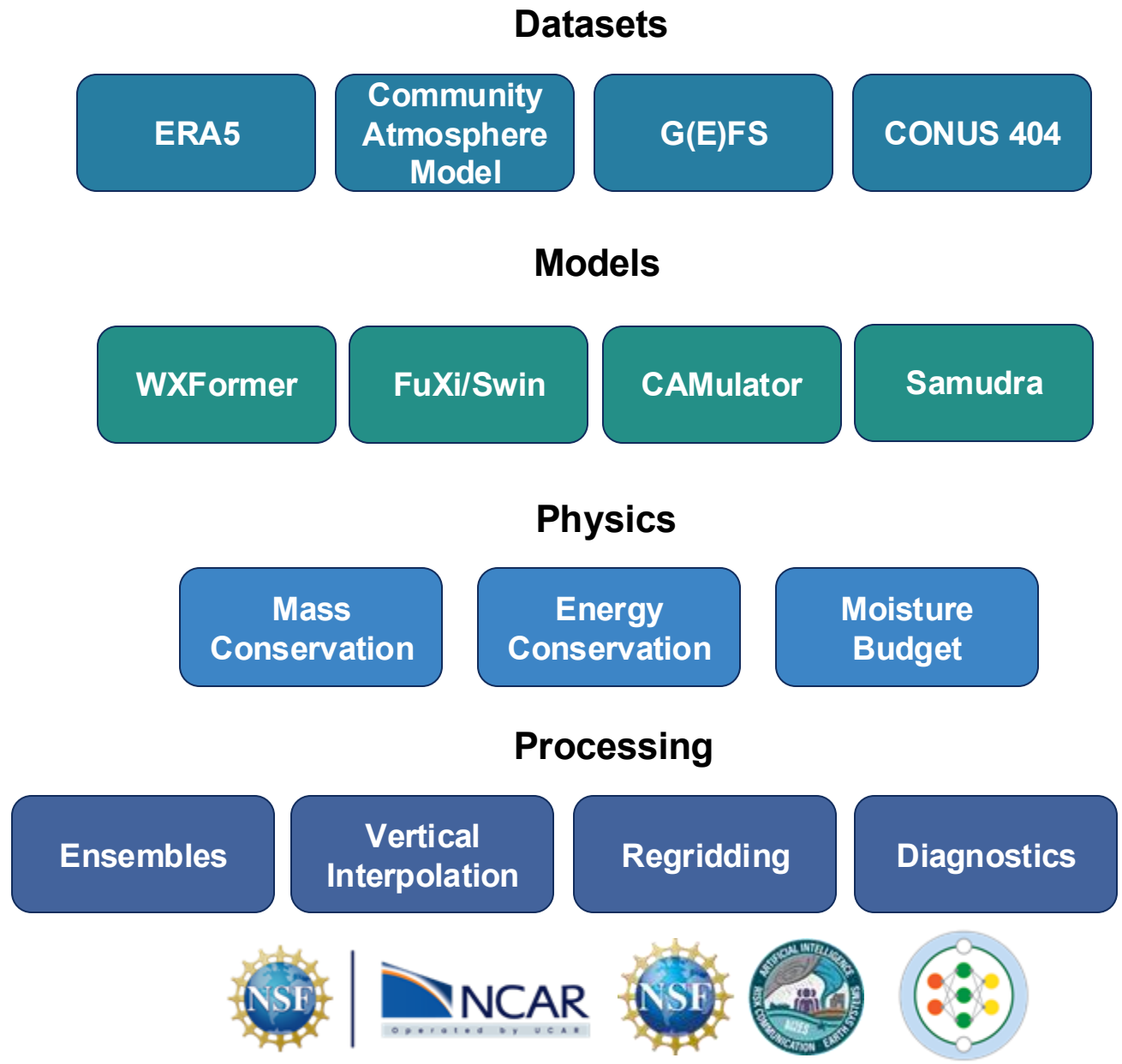
An **open, foundational research platform** for developing AI Earth System Models.

CREDIT enables users to

- Load large datasets.
- Train customizable AI models at scale.
- Deploy models for forecasting, hindcasting, or projections.

CREDIT's full pipeline is open source and is being openly developed.

Code: github.com/NCAR/miles-credit

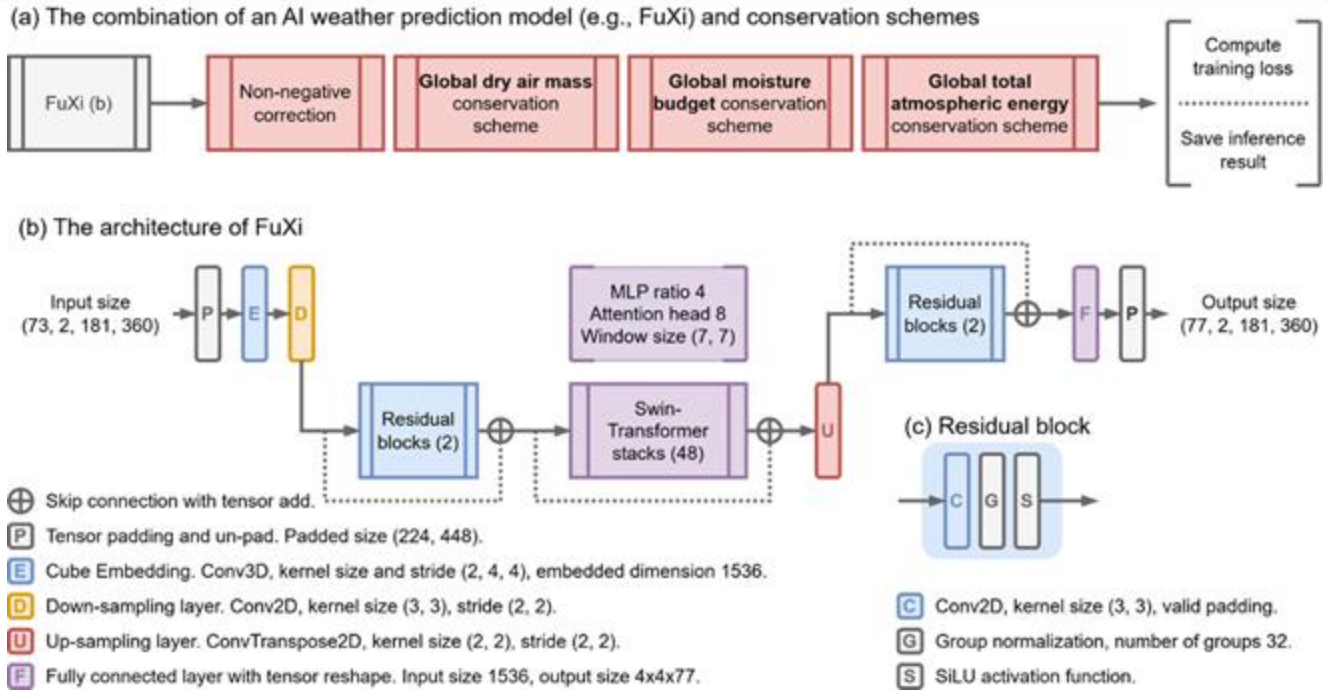


CREDIT Physics Conservation

Table 1. The variables of interest in this study.

Type	Variable Name	Units	Role
Pressure level	Zonal Wind	$m \cdot s^{-1}$	Prognostic, Instantaneous
	Meridional Wind	$m \cdot s^{-1}$	
	Air Temperature	K	
	Specific Total Water[a]	$kg \cdot kg^{-1}$	
	Geopotential height	m	
Single level	Mean Sea Level Pressure	Pa	Prognostic, Instantaneous
	2-Meter Temperature	K	
	10-Meter Zonal Wind	$m \cdot s^{-1}$	
	10-Meter Meridional Wind	$m \cdot s^{-1}$	
Flux form[b]	Total Precipitation	m	Diagnostic, Cumulative
	Evaporation	m	
	Top-of-atmosphere Net Solar Radiation	$J \cdot m^{-2}$	
	Outgoing Longwave Radiation	$J \cdot m^{-2}$	
	Surface Net Solar Radiation	$J \cdot m^{-2}$	
	Surface Net Longwave Radiation	$J \cdot m^{-2}$	
	Surface Net Sensible Heat Flux	$J \cdot m^{-2}$	
	Surface Net Latent Heat Flux	$J \cdot m^{-2}$	
	Top-of-atmosphere Incident Solar Radiation	$J \cdot m^{-2}$	Input-only, Cumulative
Others	Sea-ice Cover	n/a	Input-only, Instantaneous
	Geopotential at the Surface	$m^2 \cdot s^{-2}$	Input-only, Static
	Land-sea Mask	n/a	Input-only, Static
	Soil Type	n/a	Input-only, Static

[a] Specific total water is the combination of specific humidity, cloud liquid water content, and rainwater content.
[b] Flux form variables are accumulated every 6 hours. Downward flux is positive.



Data: ERA5 conservatively regridded to 1 degree

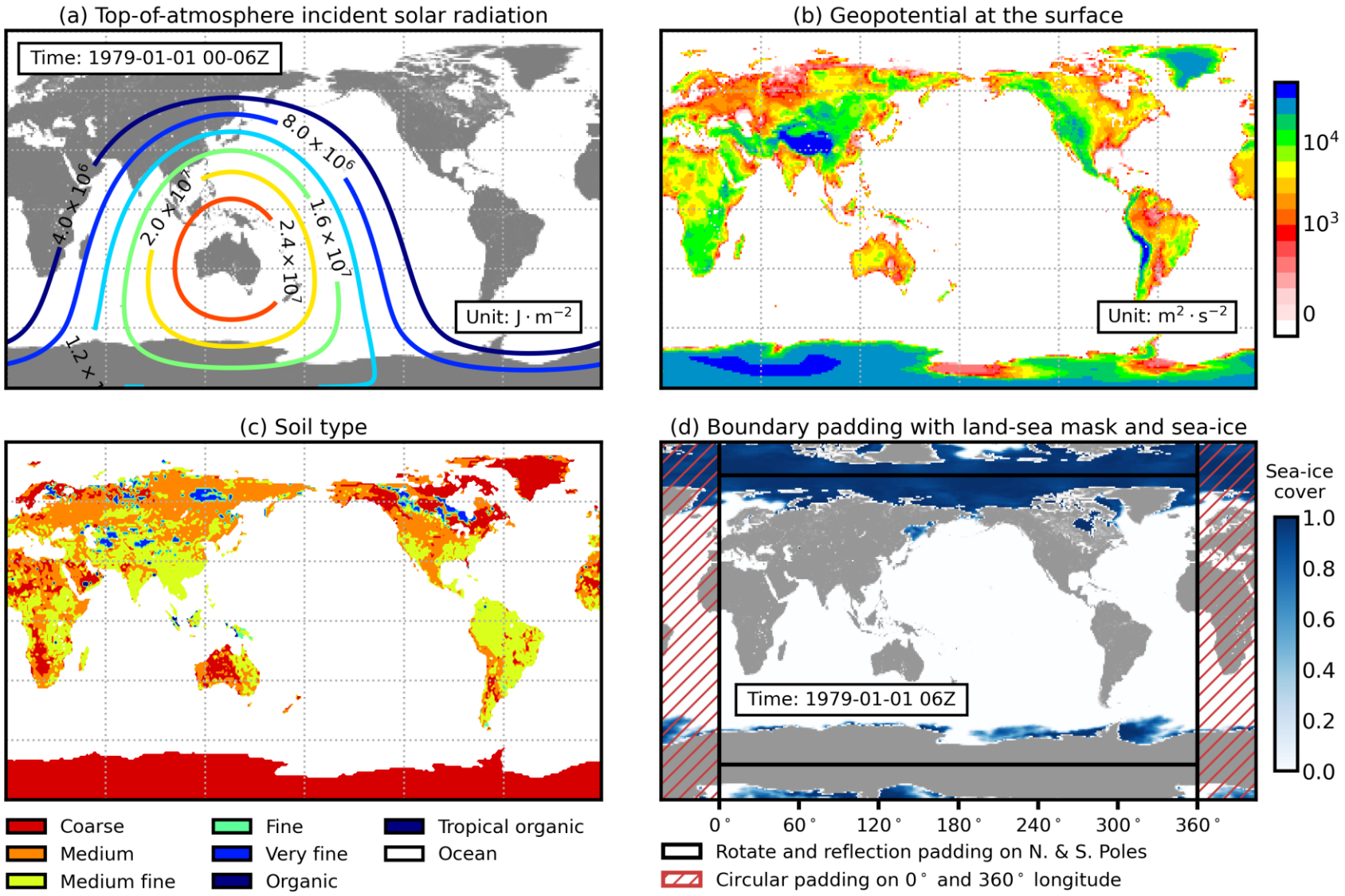
Loss: Latitude-weighted MSE

Ingredients for Physics Constraints in AI Weather Prediction

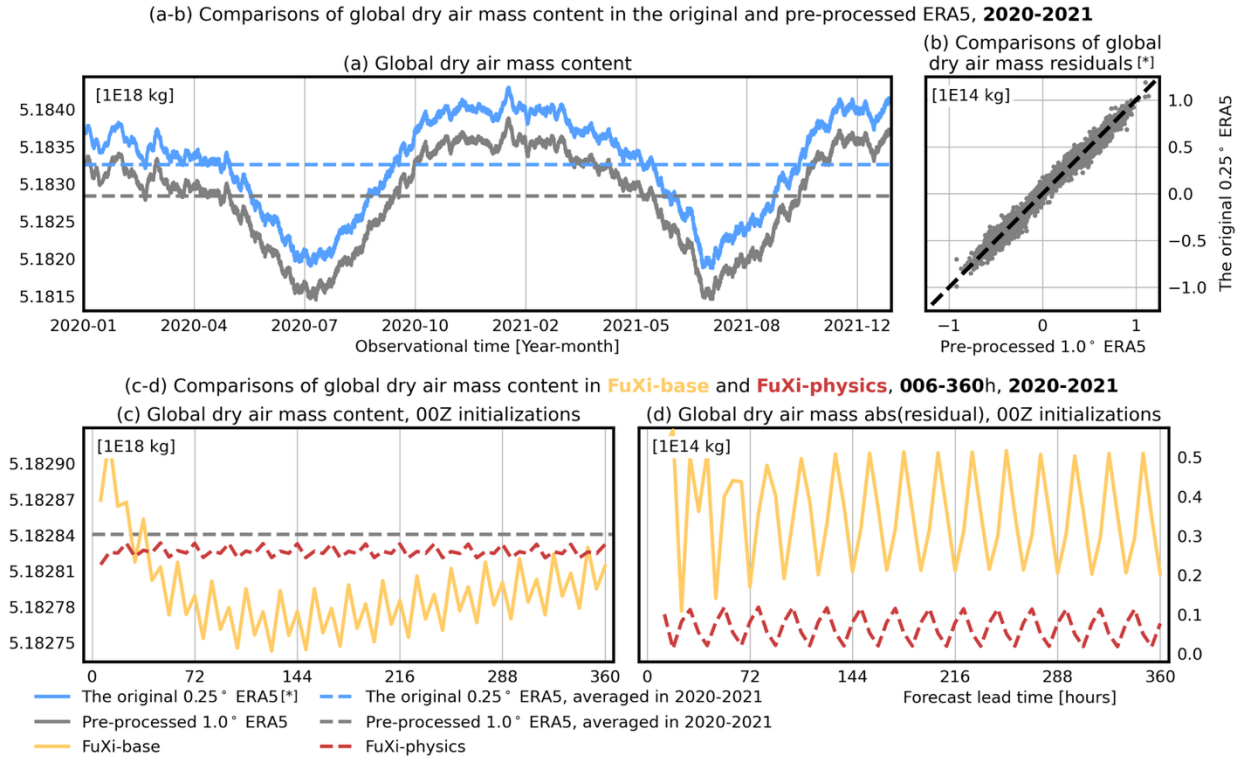
1. Sufficient variables to calculate mass, moisture, and energy budgets
2. Conservation layers that adjust data to conserve mass, moisture, and total energy across the globe to match initial values with multiplicative scaling

Sha, Y., J. S. Schreck, W. Chapman, and D. J. Gagne II, 2025: Improving AI weather prediction models using global mass and energy conservation schemes. *JAMES,*

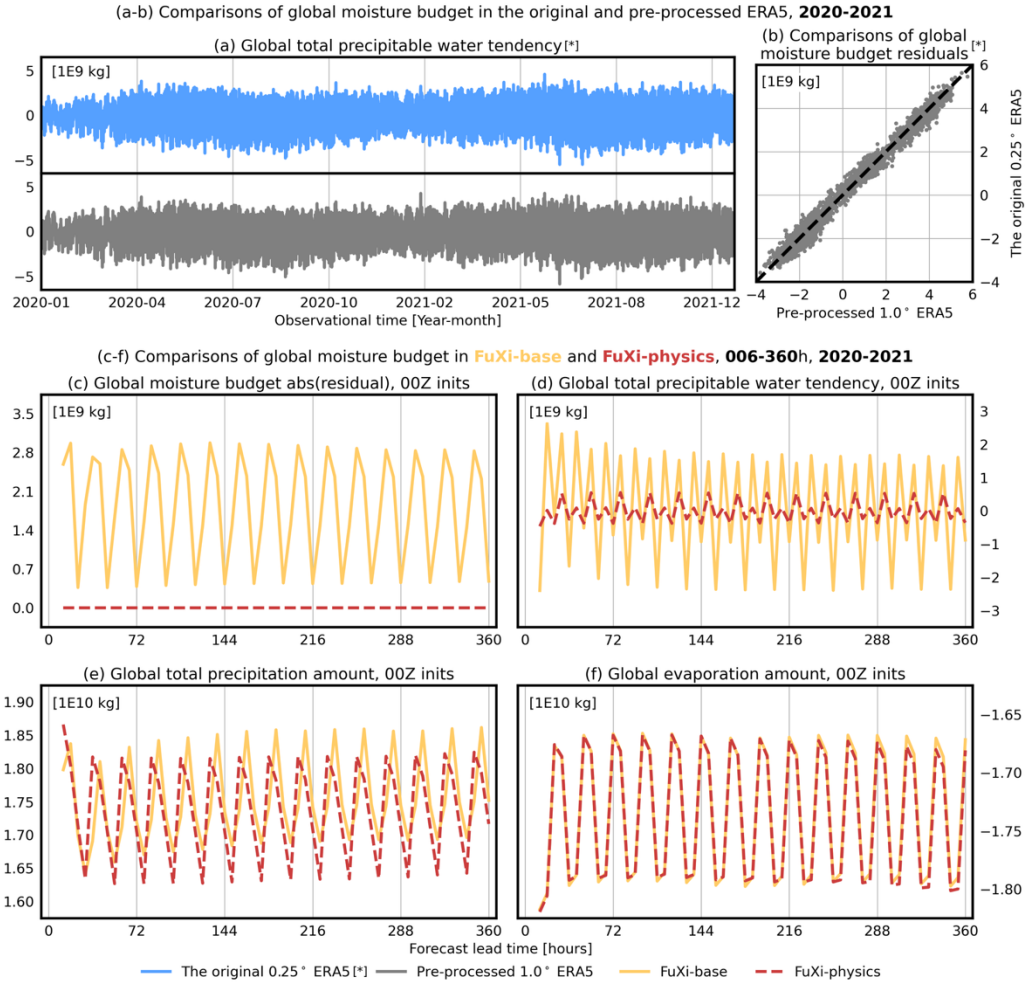
Static Fields



Dry Air Mass and Moisture Budget Comparisons



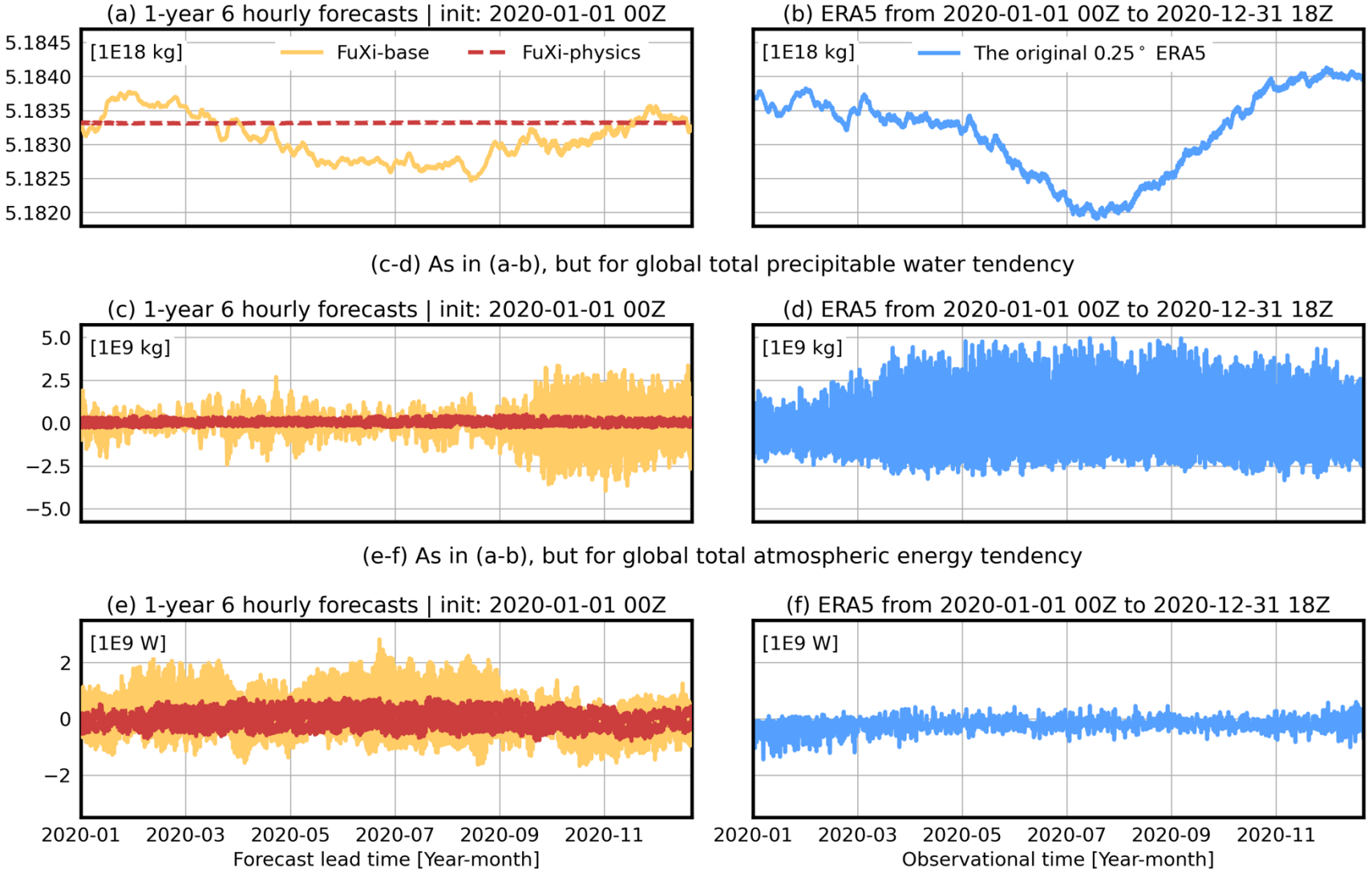
[*] The global dry air mass content calculated from the oringal 0.25° ERA5 considers 37 pressure levels and all moisture compontents.
For the pre-processed 1.0° ERA5, 13 pressure levels and specific total water are used.



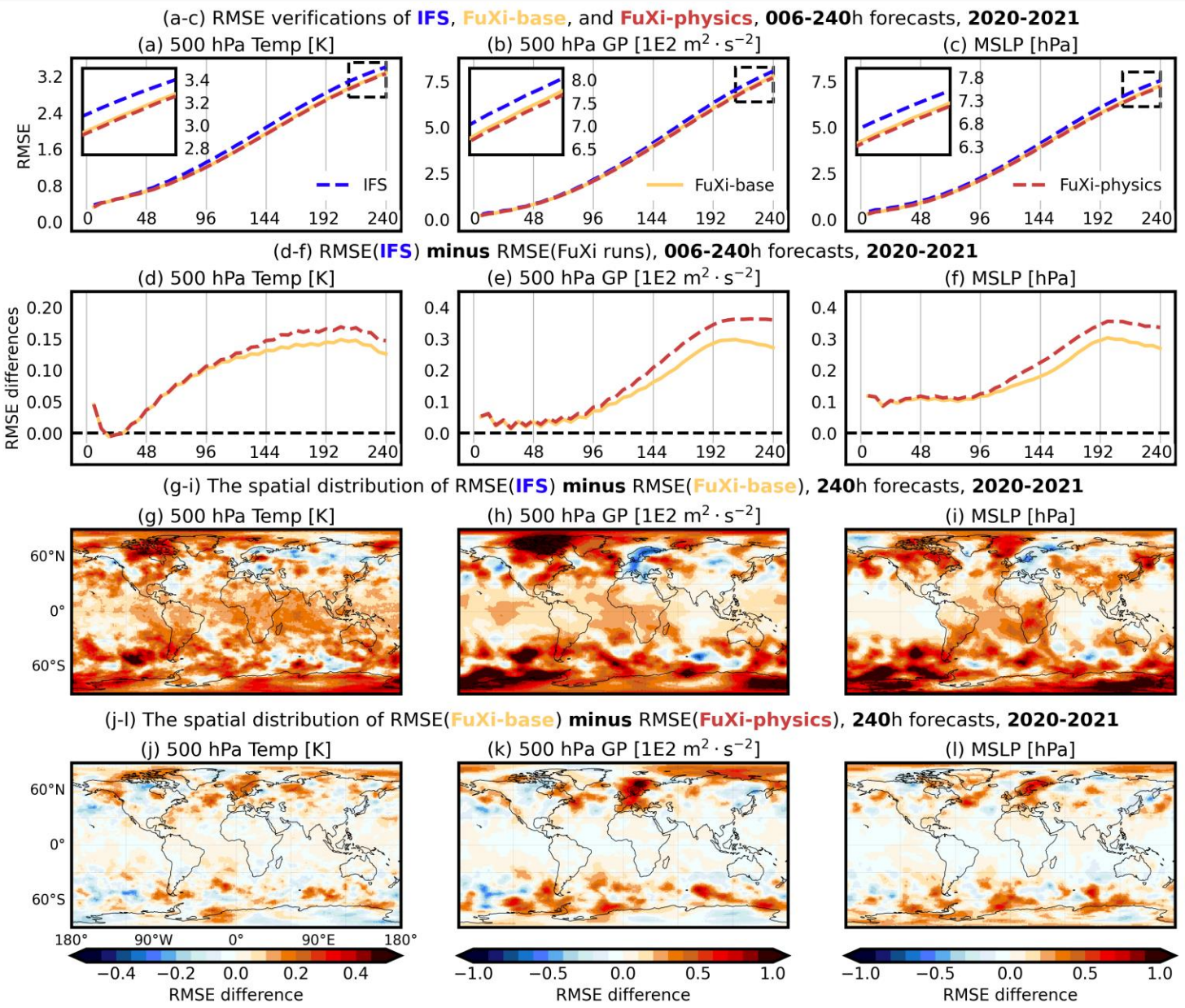
[*] The global total precipitable water calculated from the oringal 0.25° ERA5 considers 37 pressure levels and all moisture compontents.
For the pre-processed 1.0° ERA5, 13 pressure levels and specific total water are used.

Physics Conservation over a Yearlong Run

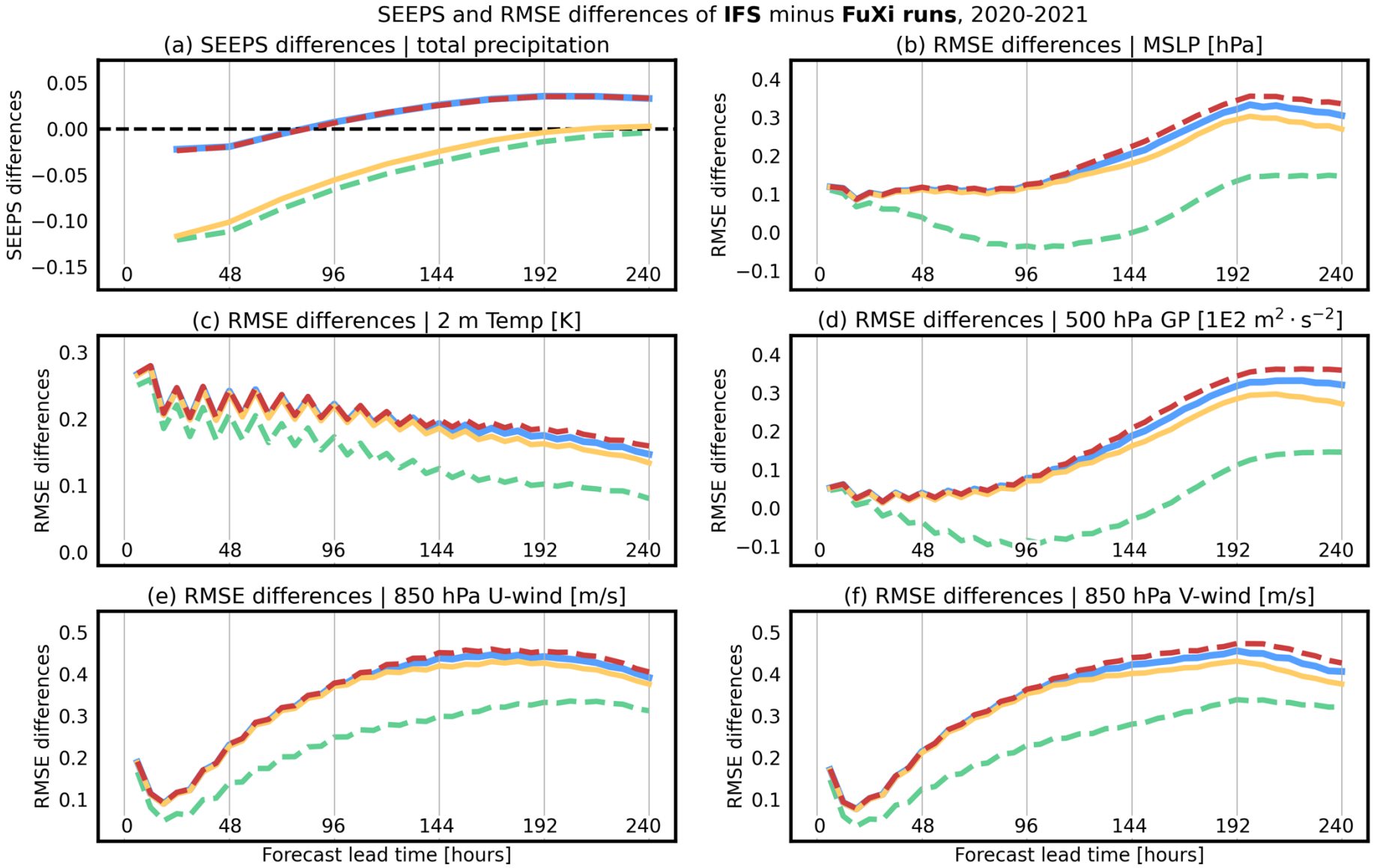
(a-b) Global dry air mass content comparisons between 1-year forecasts and the original 0.25° ERA5



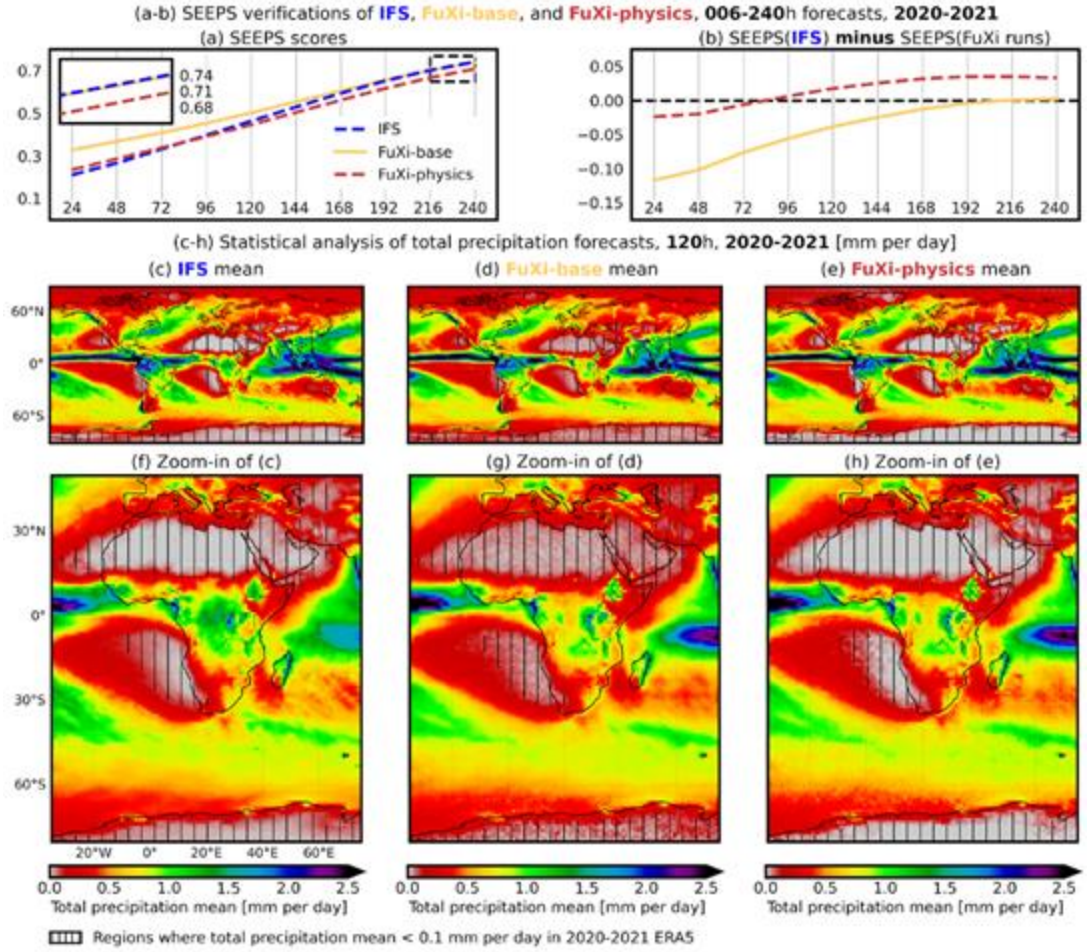
RMSE of Forecasts

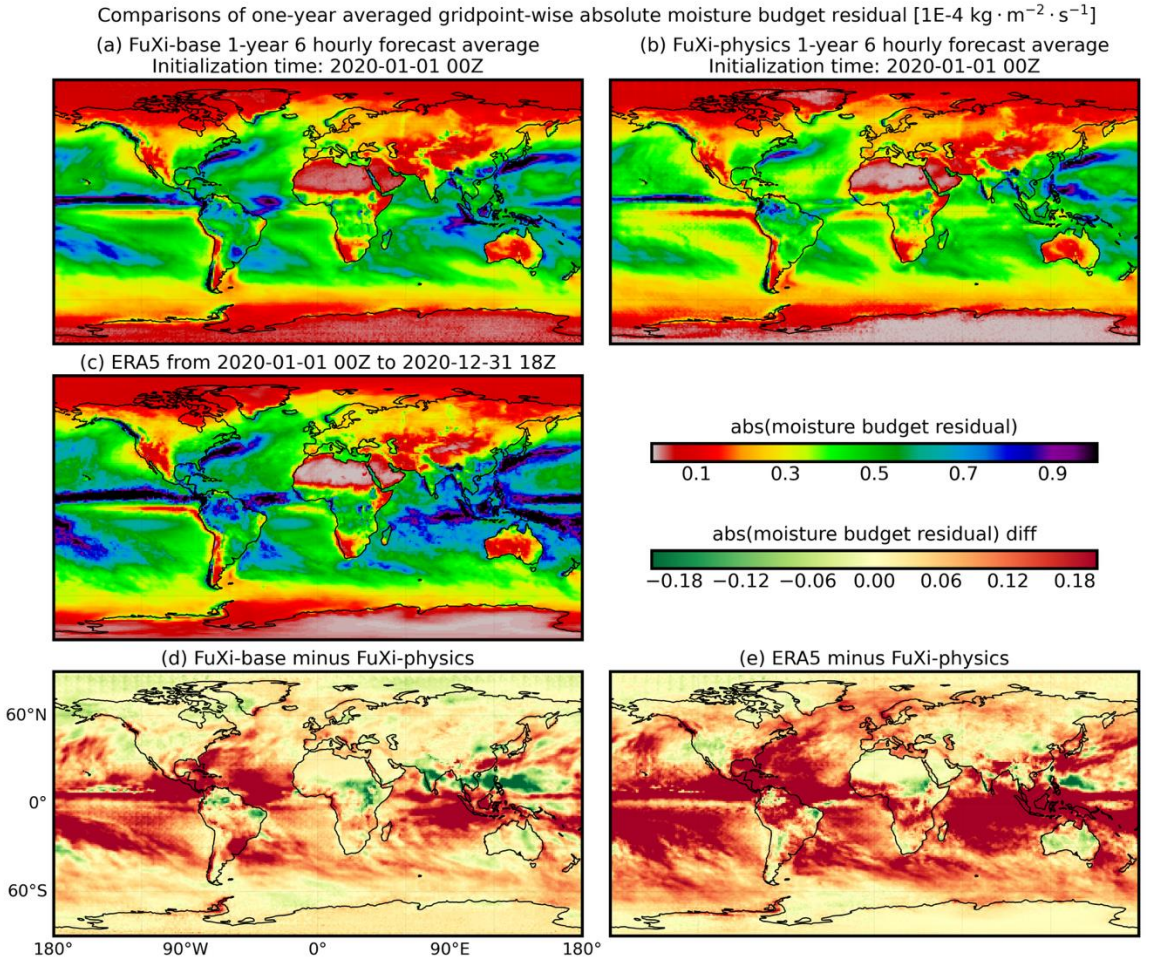


Why not only apply physics constraints during inference?

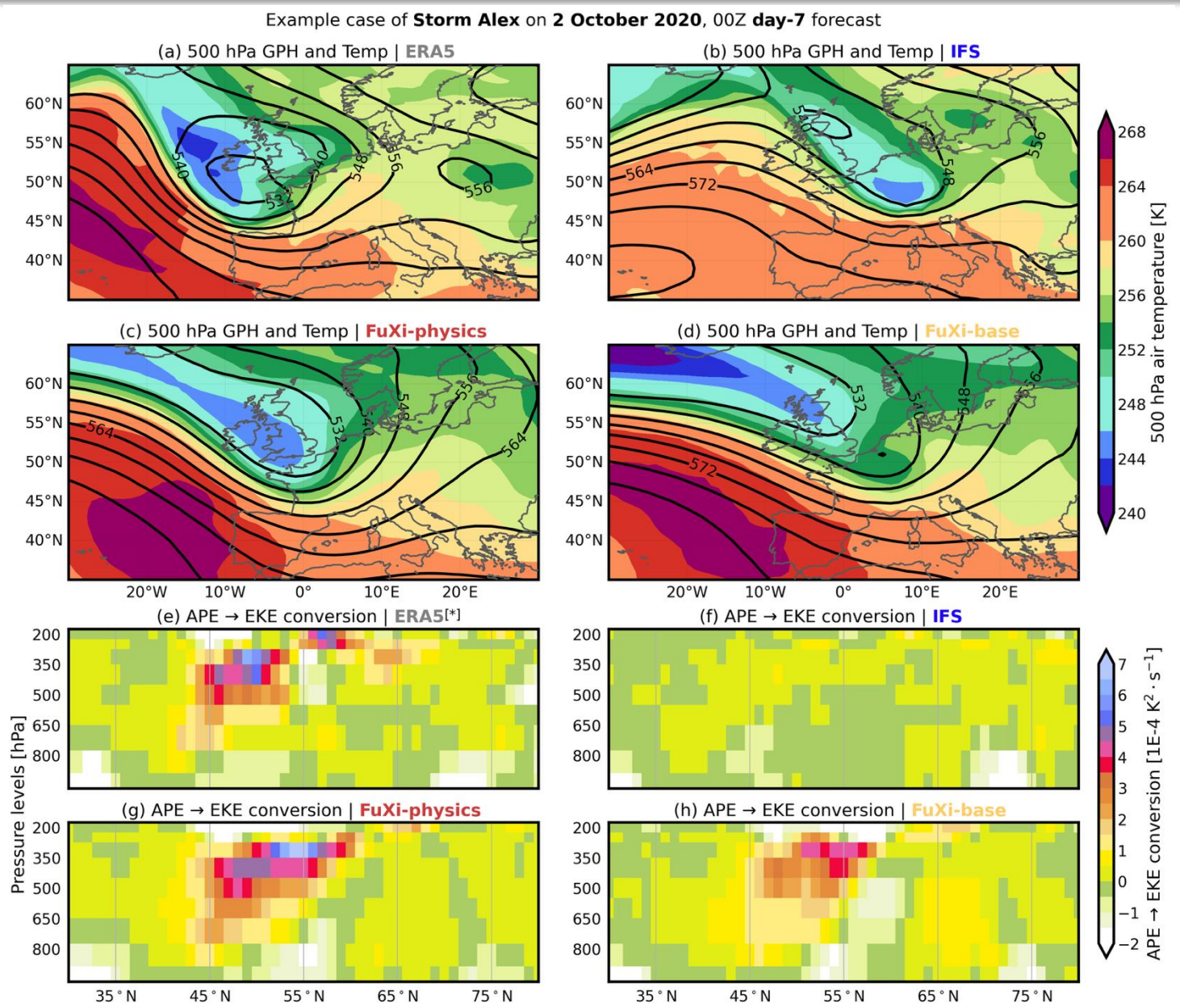


1-Year Moisture Budget Differences





Physics Constraint Case Study



[*] APE → EKE conversion is computed as meridional eddy heat flux multiplies temperature gradient $-(\overline{v'T'})(\partial T/\partial y)$.

Forecast Improvements from Physics Conservation + Vertical Coordinate

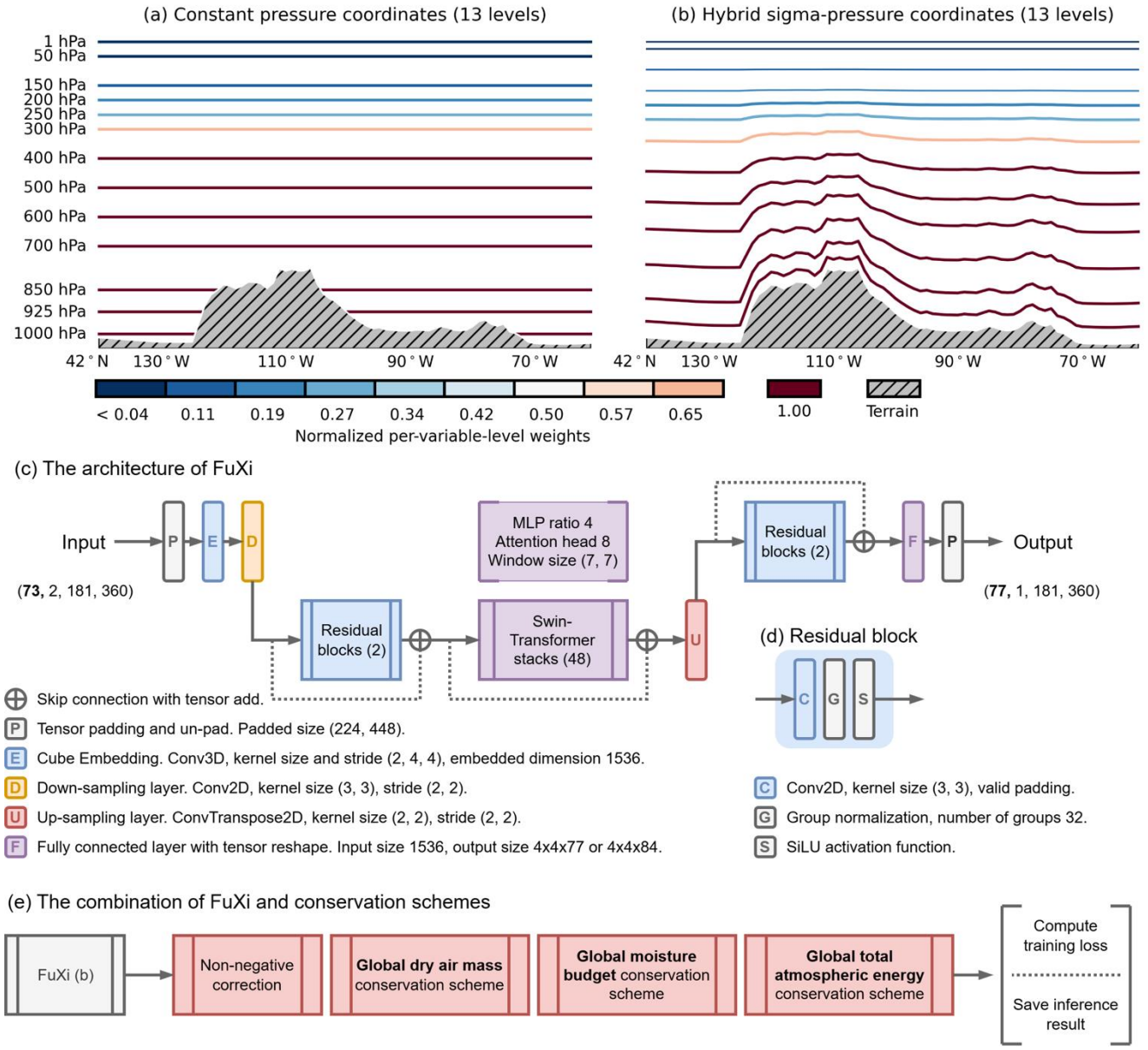
Most AI Weather Prediction models use a pressure-level vertical coordinate system.

- Pros:
  - Real-time analyses from major met. services released on pressure levels
  - Smooth spatial variations at mid and upper levels
- Cons:
  - Intersections with terrain
  - Inconsistent procedures from extrapolated below surface
  - Less vertical resolution near surface

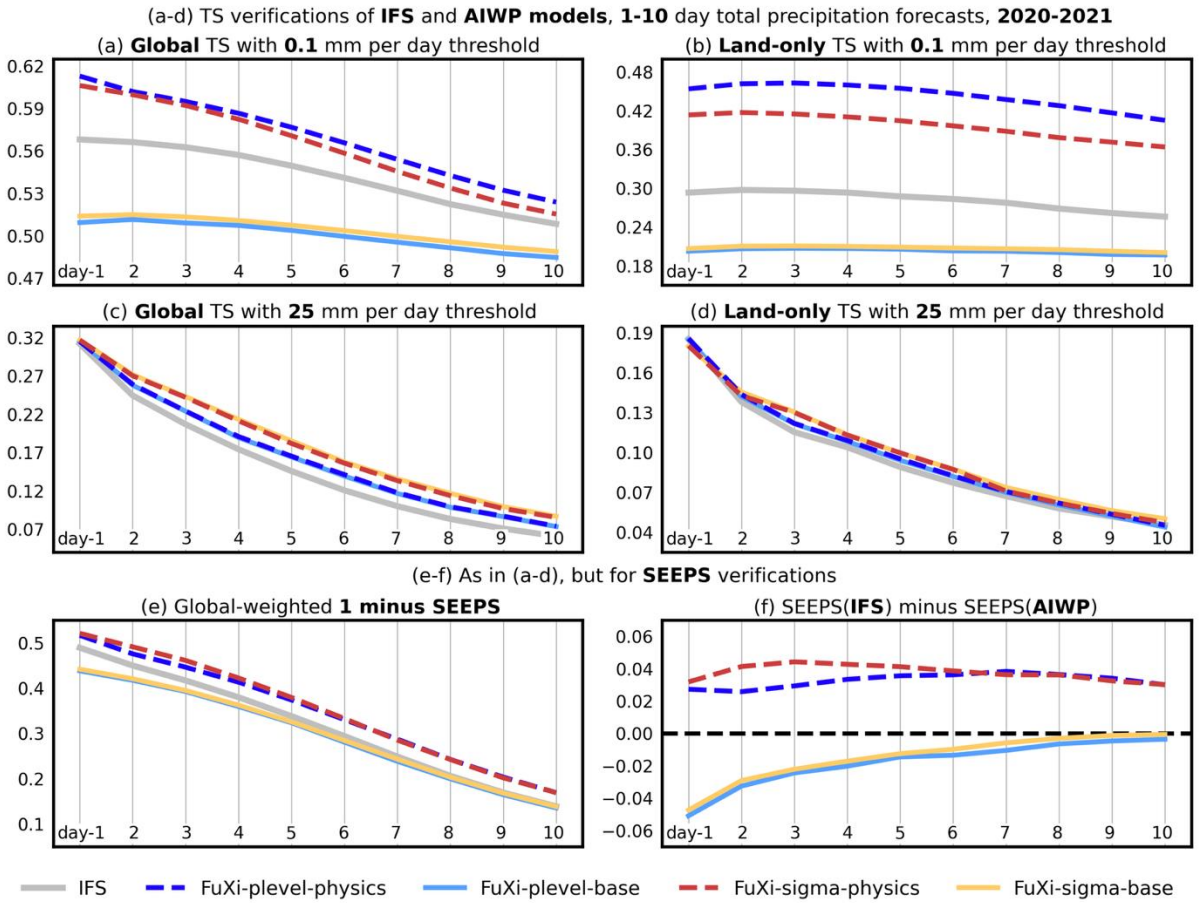
Hybrid sigma-pressure terrain following coordinates address some of these issues

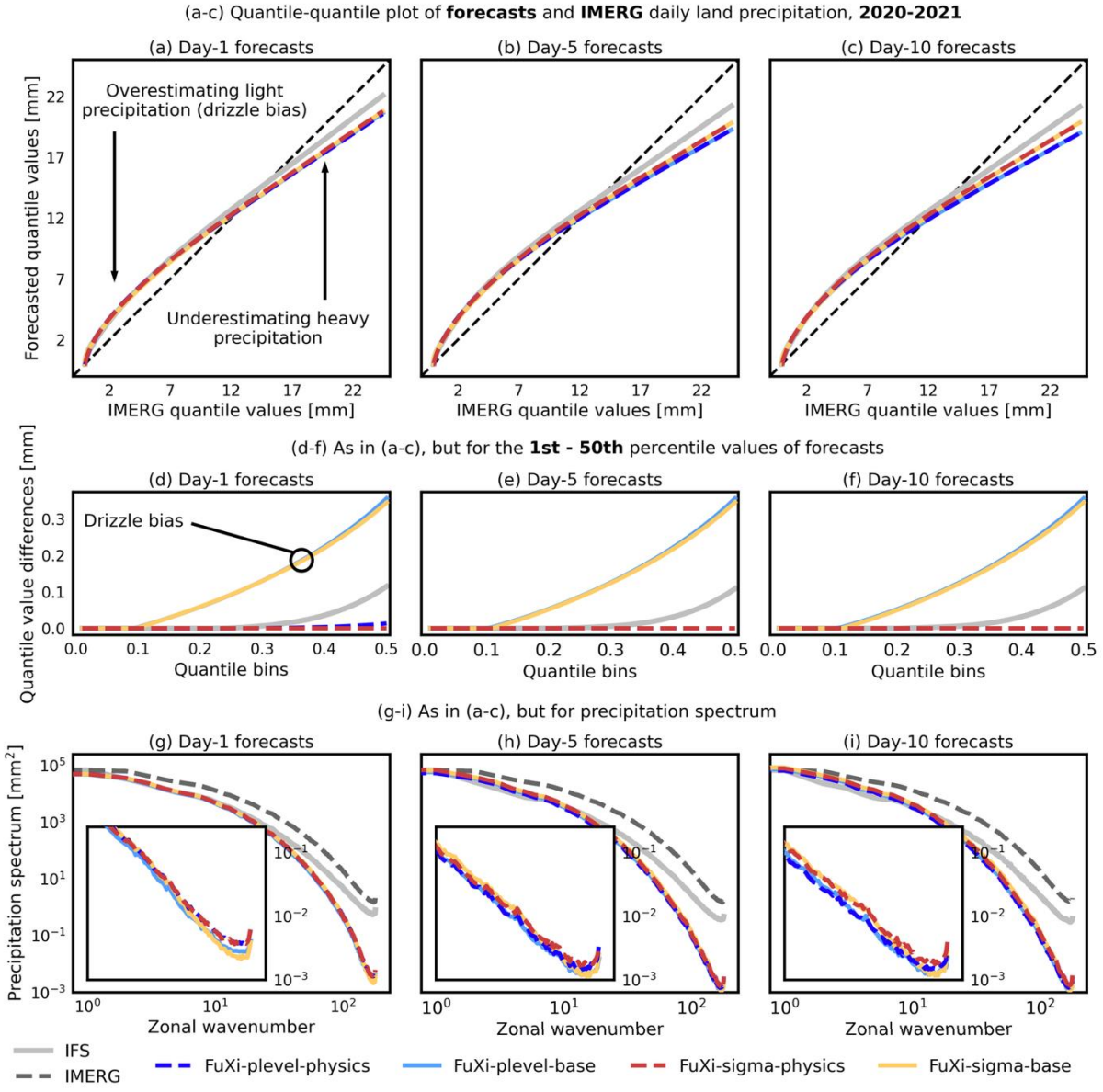
- Pros:
  - Always above ground
  - Native coordinate of most modern NWP models
- Cons:
  - Terrain variations impact spatial smoothness
  - Terrain variations present up to 200 hPa in ERA5 to accommodate Himalayas

Sha, Y., Schreck, J. S., Chapman, W., & Gagne, D. J., II (2025). Investigating the use of terrain-following coordinates in AI-driven precipitation forecasts. *Geophysical Research Letters*, 52, e2025GL118478. https://doi.org/10.1029/2025GL118478

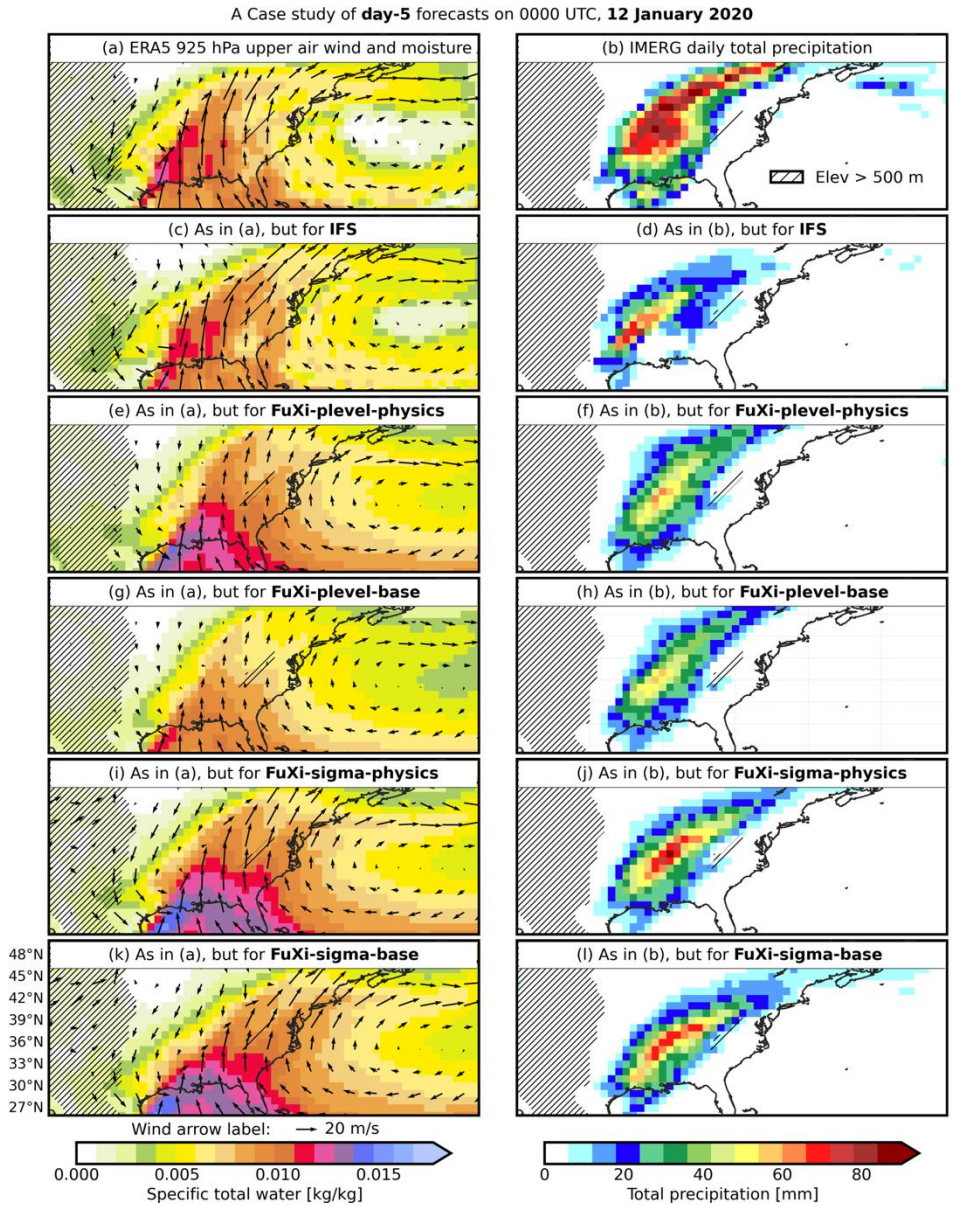


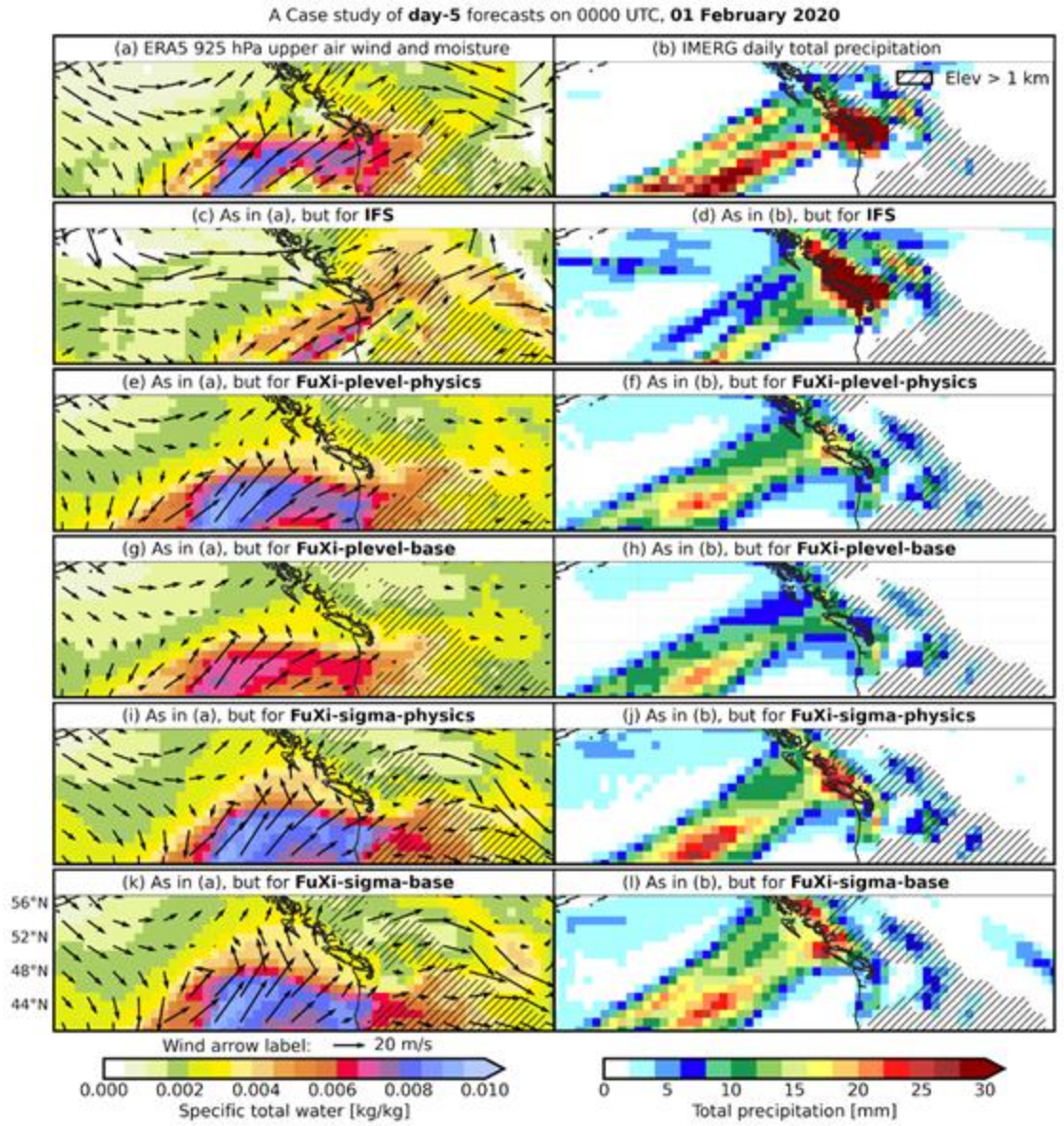
Precipitation Verification Results





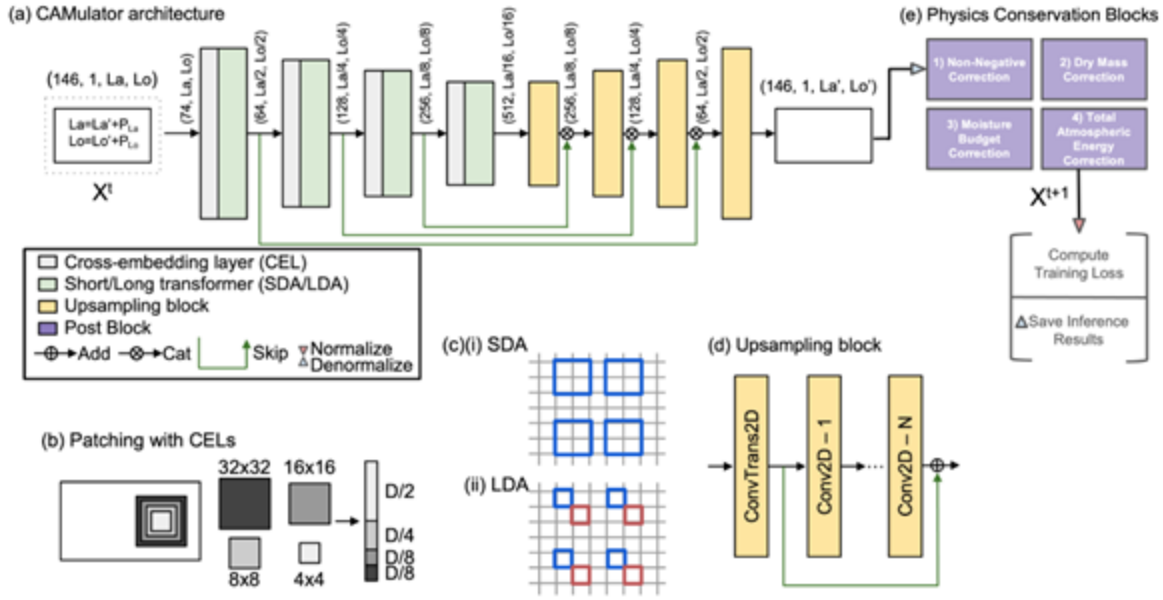
Physics Case Studies





CAMulator Overview

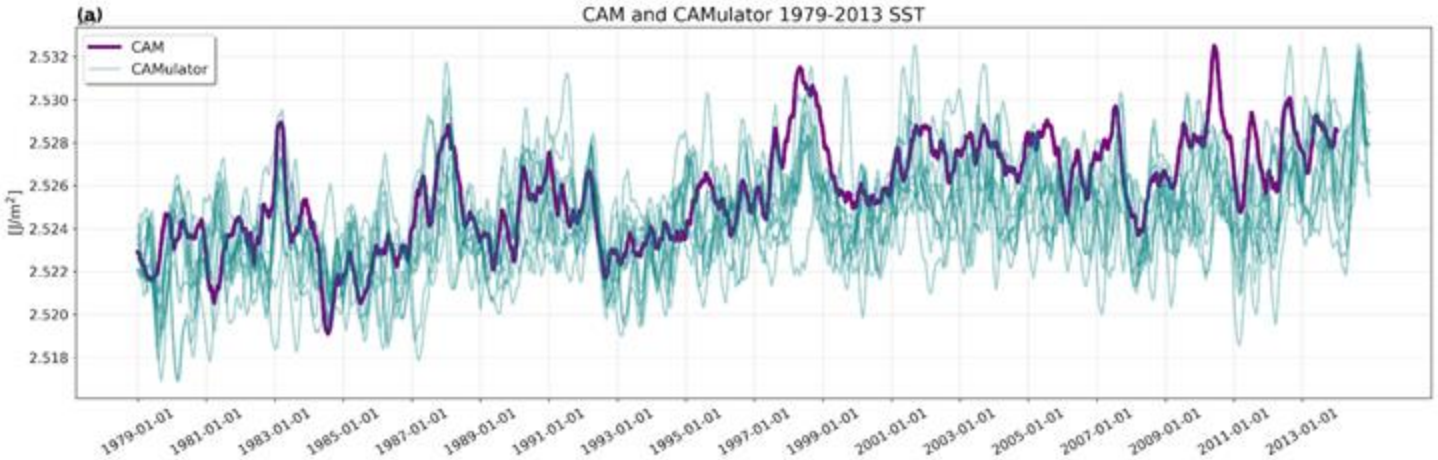
Variable	Description	Units	Levels	Category
Input Variables (Forcing)				
Static Forcing (Time-Invariant Input)				
Surface Geop.	Surface Height	m^2/s^2	Single Level	Input
Land-Sea Mask	Land Mask	Unitless	Single Level	Input
Dynamic Forcing (Time-Varying Input)				
SOLIN	Incoming Solar Radiation	J/m^2	Single Level	Input
SST	Sea Surface Temperature	K	Single Level	Input
Prognostic Variables (Input and Output)				
U	Zonal Wind	m/s	32 Levels	Input/Output
V	Meridional Wind	m/s	32 Levels	Input/Output
T	Temperature	K	32 Levels	Input/Output
*Qtot	Specific Total Water	kg/kg	32 Levels	Input/Output
PS	Surface Pressure	Pa	Single Level	Input/Output
TREFHT	Near-Surface Air Temperature	K	Single Level	Input/Output
Diagnostic Variables (Output Only)				
PRECT	Accumulated Precipitation	m	Single Level	Output
CLDTOT	Total Cloud Cover	Fraction	Single Level	Output
CLDHGH	High Cloud Cover	Fraction	Single Level	Output
CLDLOW	Low Cloud Cover	Fraction	Single Level	Output
CLDMED	Medium Cloud Cover	Fraction	Single Level	Output
TAUX	Zonal Wind Stress	N/m^2	Single Level	Output
TAUY	Meridional Wind Stress	N/m^2	Single Level	Output
U10	10m Wind Speed	m/s	Single Level	Output
QFLX	**Surface Moisture Flux**	m	Single Level	Output
FSNS	**Net Solar Flux** at Surface	J/m^2	Single Level	Output
FLNS	**Net Longwave Flux** at Surface	J/m^2	Single Level	Output
FSNT	**Net Solar Flux** at TOA	J/m^2	Single Level	Output
FLNT	**Net Longwave Flux** at TOA	J/m^2	Single Level	Output
SHFLX	**Sensible Heat Flux**	J/m^2	Single Level	Output
LHFLX	**Latent Heat Flux**	J/m^2	Single Level	Output

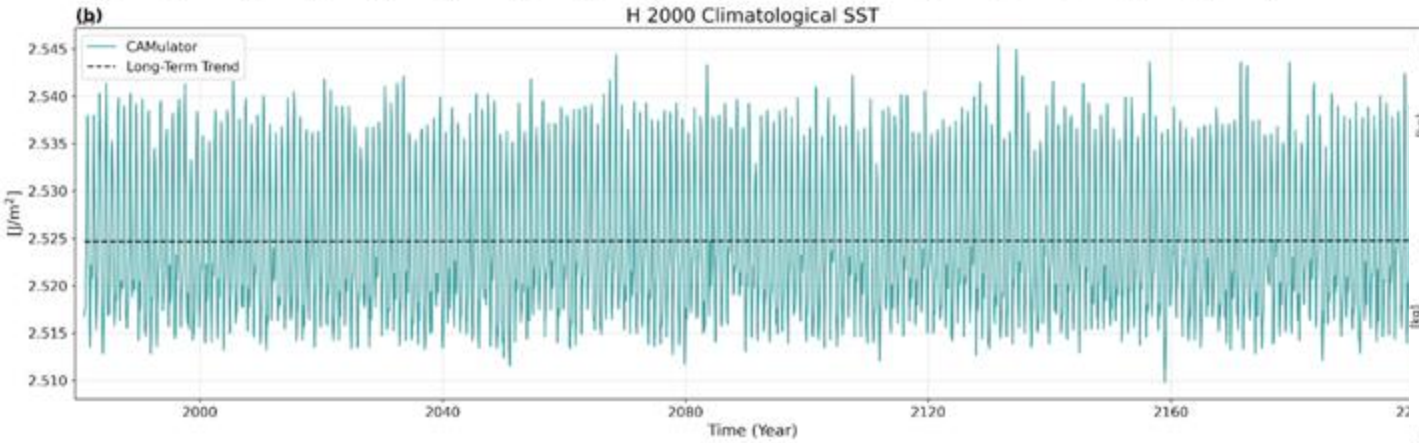


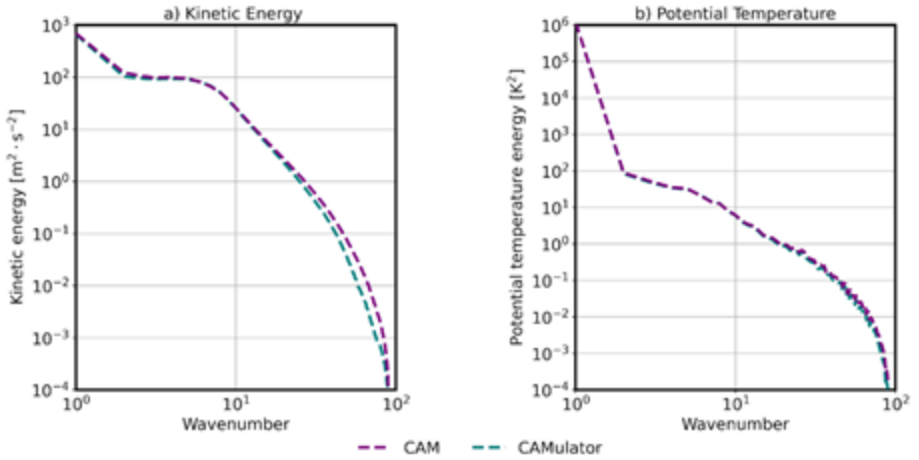
- Trained on CAM over 30 years with SST forcing from observations
- Throughput of 480 Sim. Years/Day

Chapman, W. E., J. S. Schreck, Y. Sha, D. J. Gagne II, D. Kimpara, L. Zanna, K. J. Mayer, and J. Berner, 2025: CAMulator: Fast emulation of the Community Atmosphere Model. *arXiv [physics.ao-ph]*,.

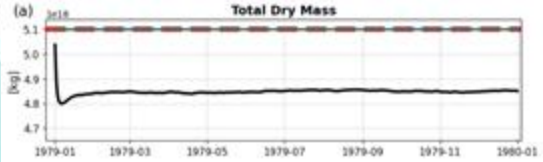
CAMulator Stability



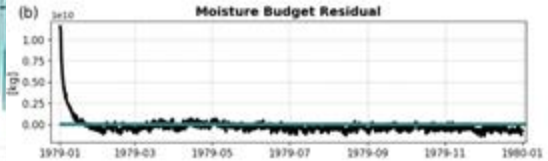


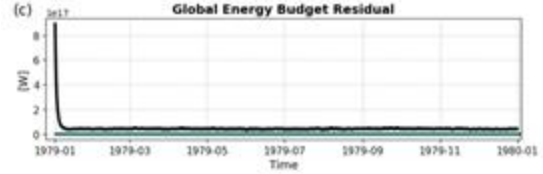


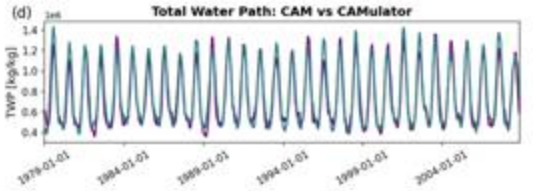
[*] Energy spectrum was computed using spherical harmonic transform on the total rollout and average

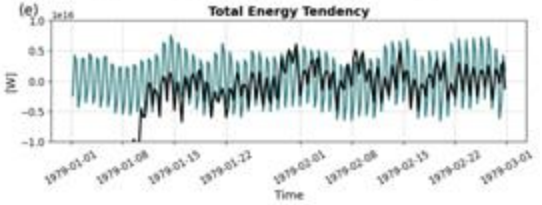




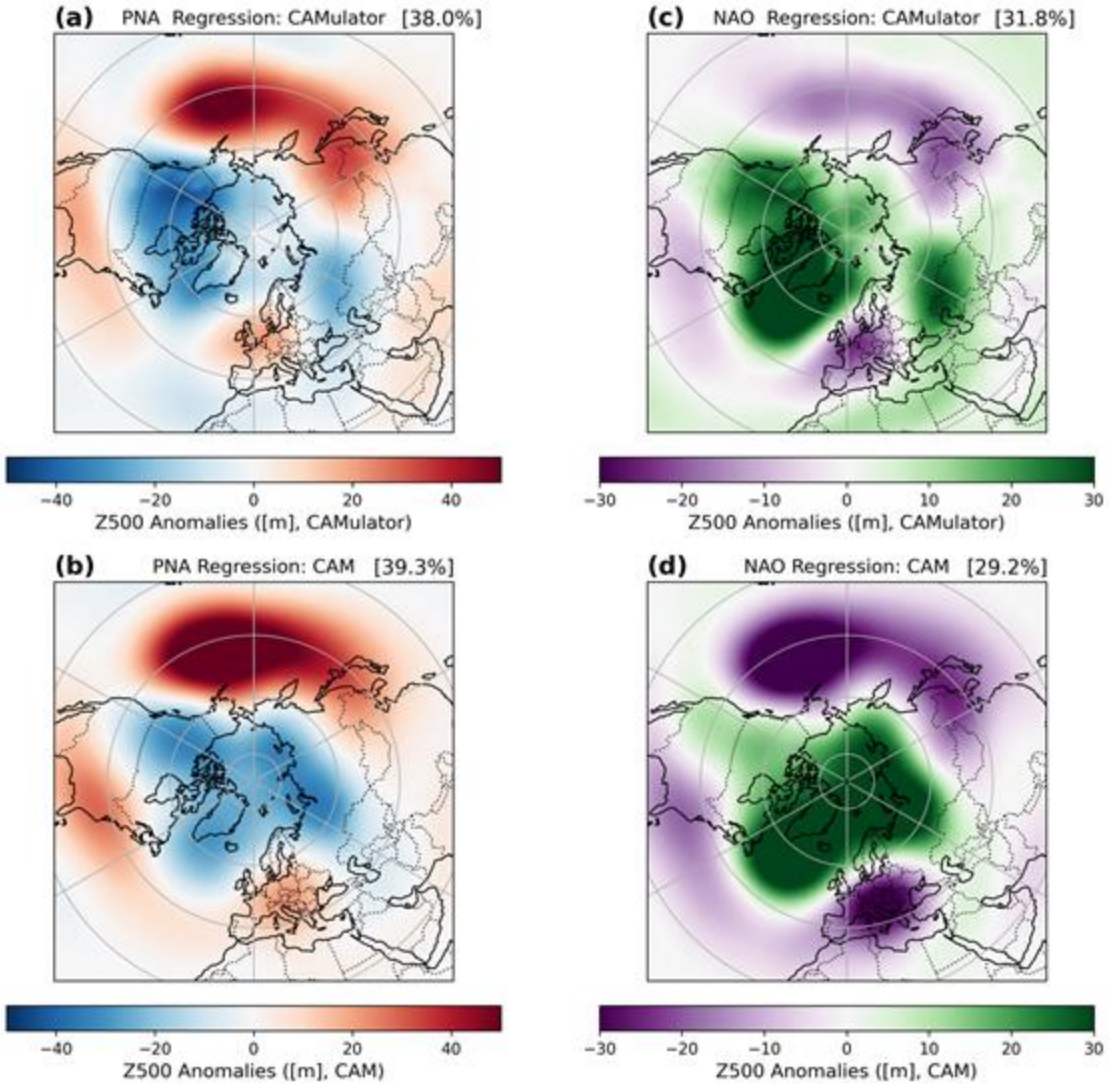


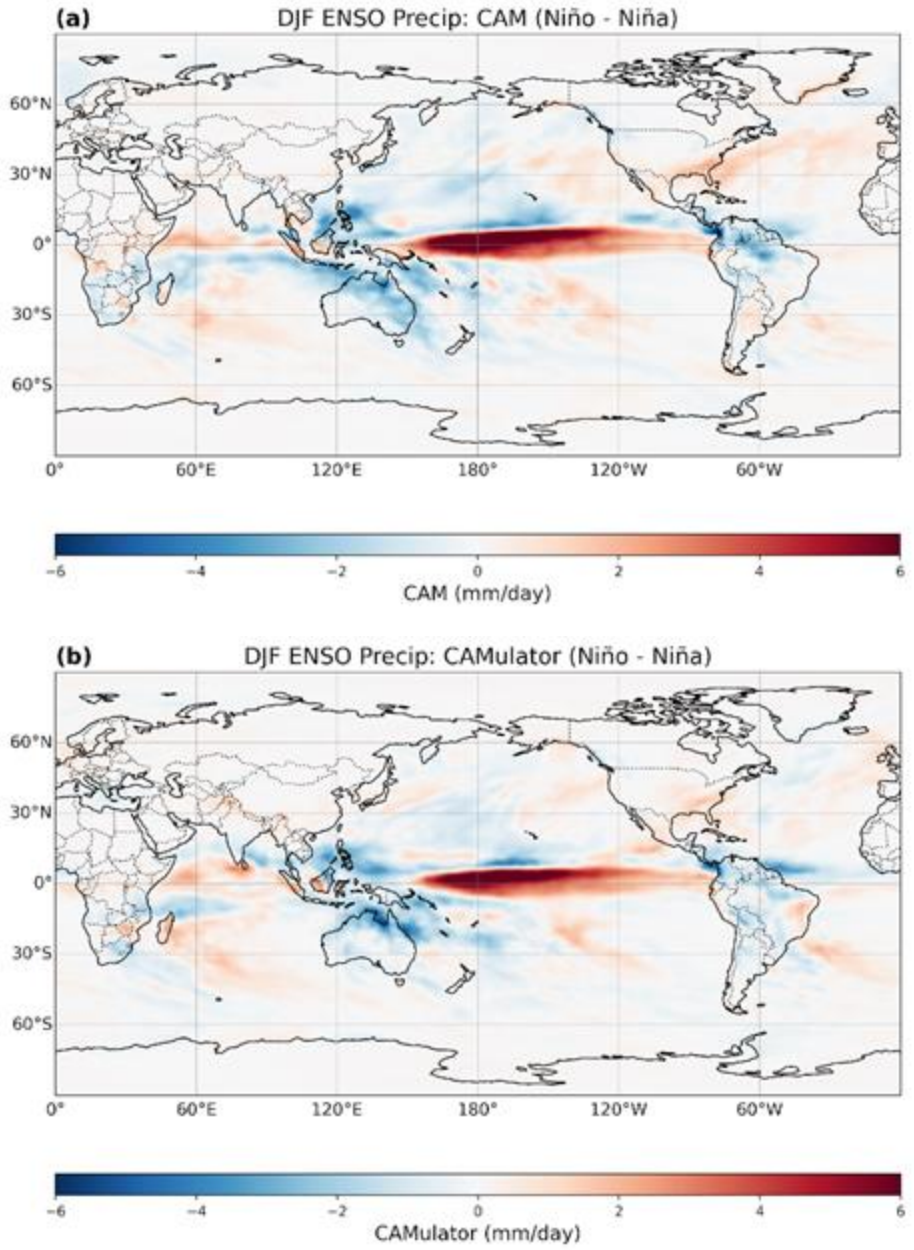




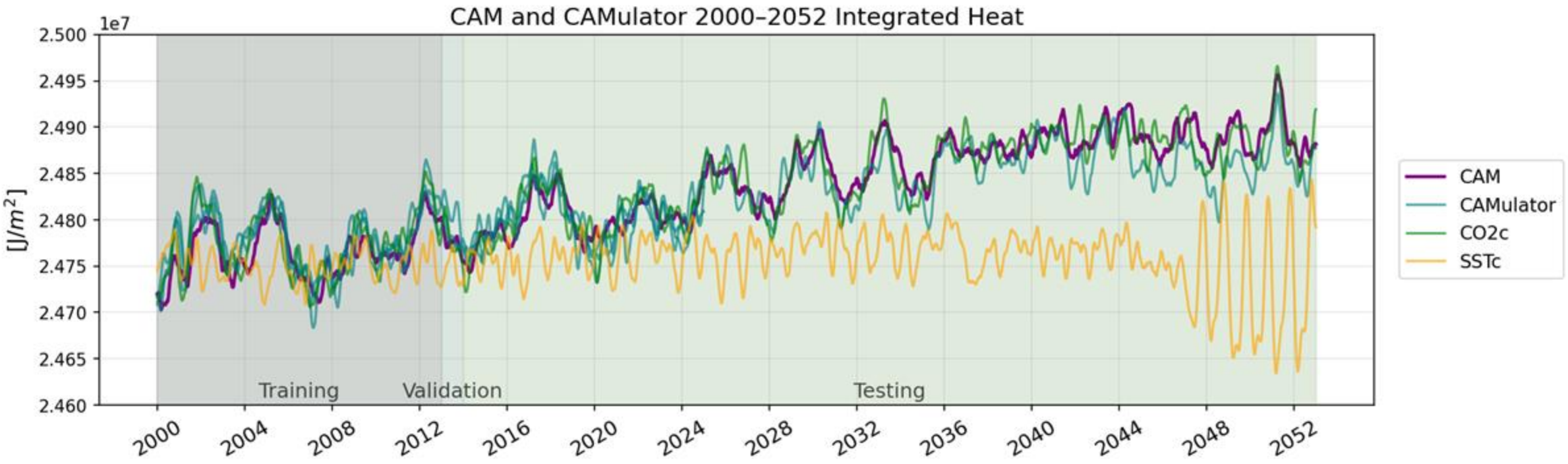


CAMulator Teleconnections





CAMulator SST and CO2 Forcing Experiments



Ongoing Challenges

- **Process Evaluation**: how can we validate the degree to which AI weather/climate models capture particular processes more systematically?
- **Fast versus slow timescale processes**: ML models tend to overemphasize fast processes over slow processes. How to learn both effectively?
  - Stratosphere
  - Deep ocean
  - Vertical mixing
- **Learning causal structure of forced responses**: if two forcings are happening simultaneously, how to learn to disentangle their effects?
  - CO2 vs SST

Summary

- Choices of relevant data and post-model global physics constraints result in both improved accuracy and structural representations of weather events
- Performance improvements with physical constraints extend out to climate timescales and can capture teleconnections and longer term trends
- Email: dgagne@ucar.edu
- CREDIT source: https://github.com/NCAR/miles-credit
- Links to CREDIT papers: https://miles.ucar.edu/projects/credit/

