AI/ML Applications for Downscaling to Convection Permitting and Catchment Resolving Scales

Ethan Gutmann
Interim Director - Hydrometeorological Applications Program, NSF NCAR RAL
Kyle Sha, Andrew Newman, Rachel McCrary, Seth McGinnis,
Tracy Hertneky, David John Gagne and many others

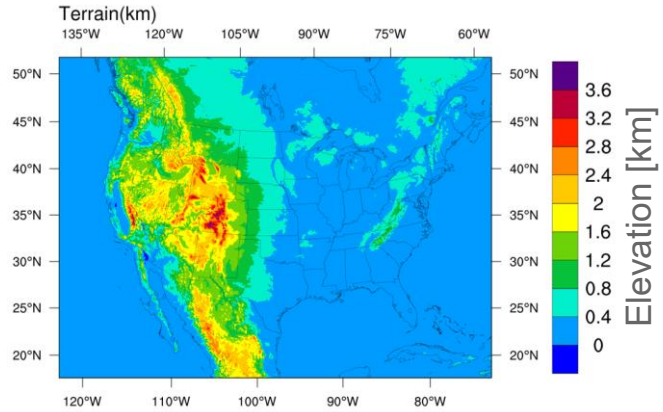
Sept 14, 2025

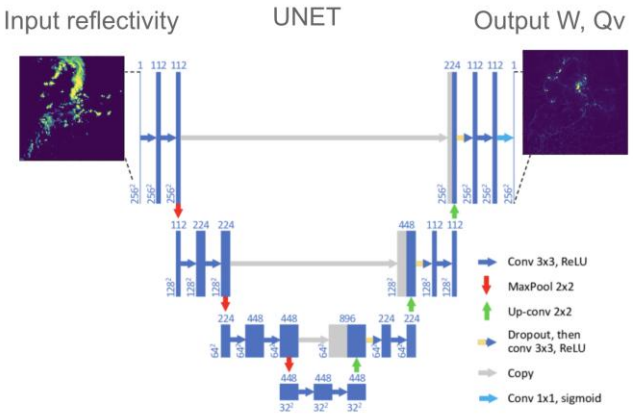
This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

Overview

- Catchment scales require finer resolution information than global models can provide.
- Regional MLWP model based on CISL MILES CREDIT framework
  - For regional climate “dynamical” downscaling
  - Leveraging CONUS404 (40-years 4km 3D output) training dataset
- Development of Generalized AI framework for application to new catchments without retraining
  - Simultaneous training over many regions in N. America and S. America
- Current model is not useful yet, but extremely promising
  - 10,000x faster than WRF
  - Physically interpretable

Regional domain

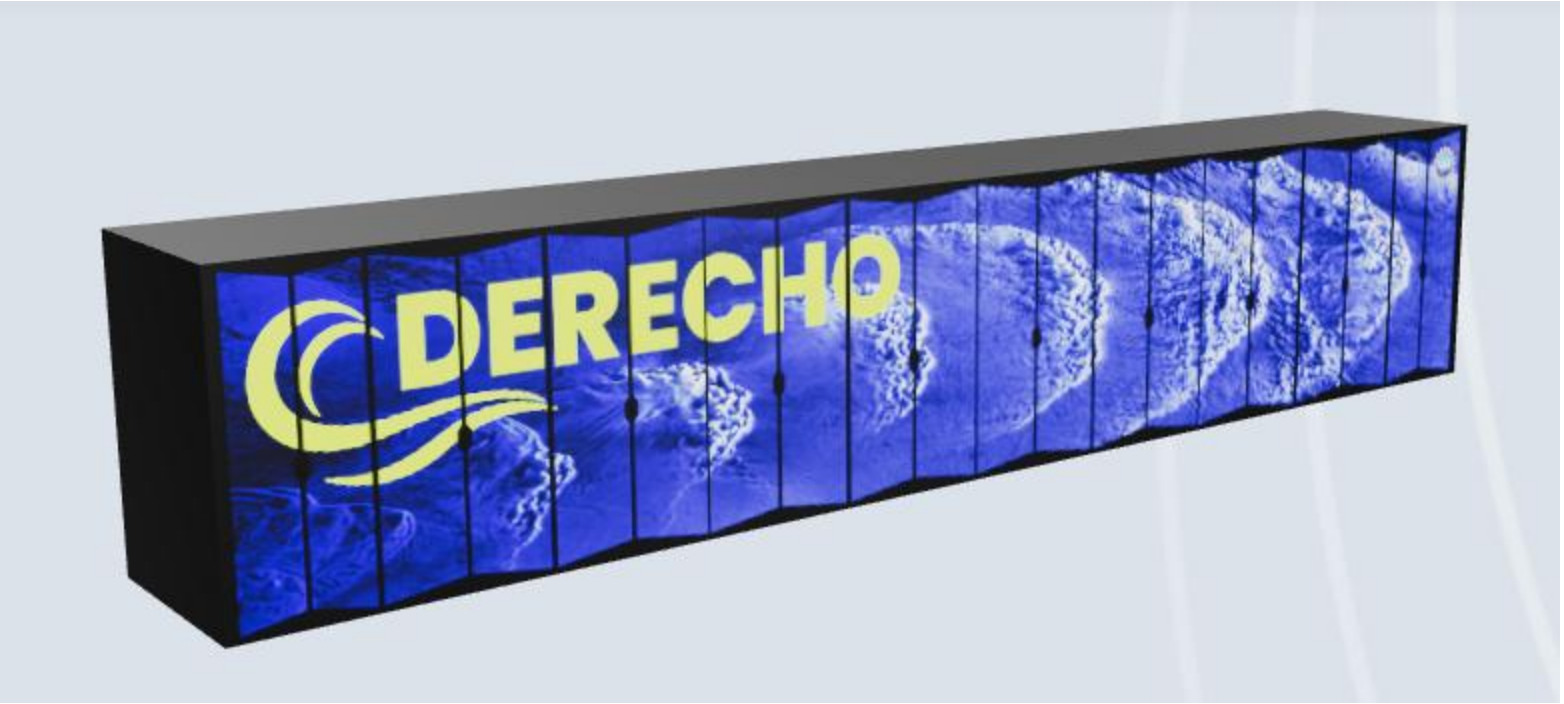




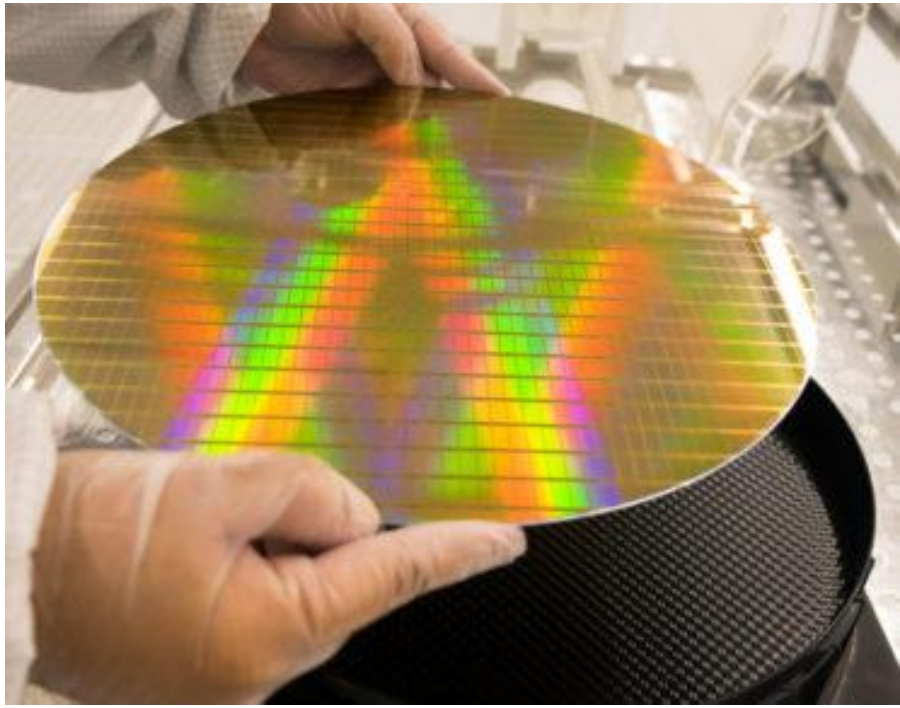
UNETs, ViT, SWIN…

Keeping up with Compute

- GPU computing is hard
- Next Generation accelerators are MUCH harder



Derecho only ~3% of nodes have GPUs because most code still can not run on GPUs 15 years later

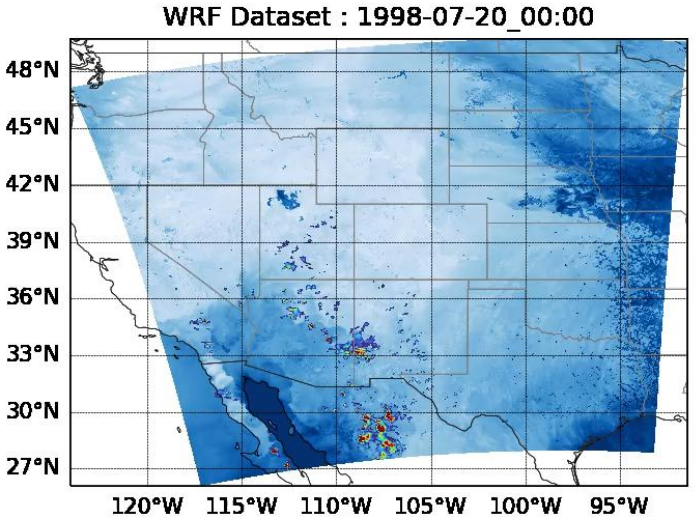


ML Models already work on Wafer Scale Engine (one chip is nearly Cheyenne's compute)

Regional MLWP model inputs and training data

CONUS 404 training variables:

- **Prognostic upper air**: P, U, V, T, Q_tot (15 sigma levels)
- **Prognostic surface**: MSLP, SP, T2, TD2, U10, V10, PWAT
- **Diagnostic**: Precip, REFL, TCC, OLR, SWDOWN, GLW
- **Static**: HGT, land-sea mask, LU, VAR



ERA5 “forcing” boundary conditions:

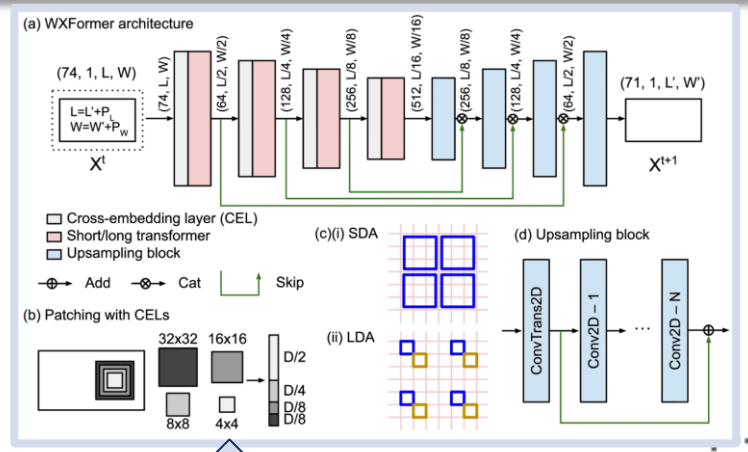
- **Upper air boundary conditions**: U, V, T, Q (11 pressure levels)
- **Surface boundary conditions**: MSLP, T2, U10, V10

C404 Derived variables:

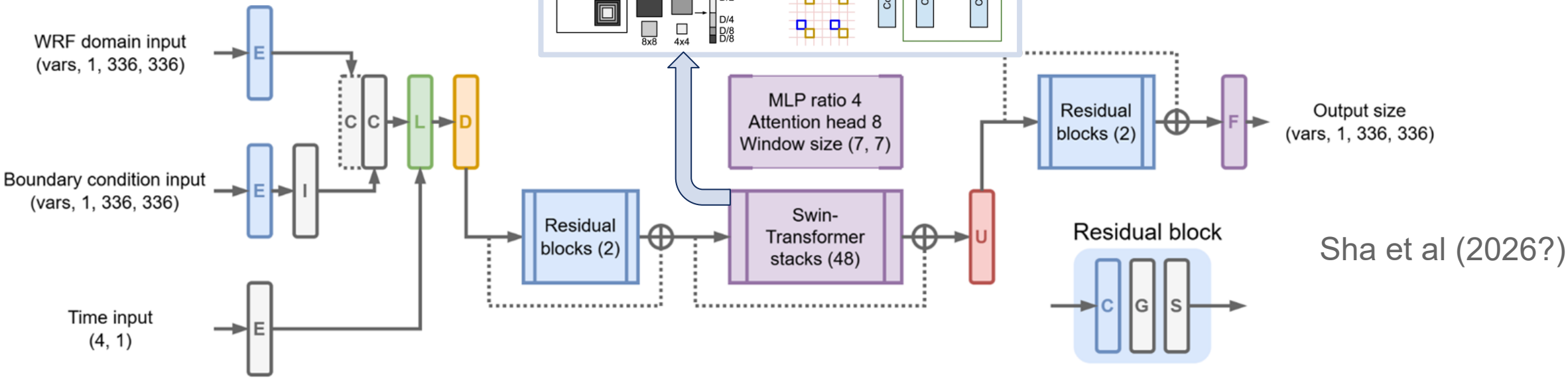
- Q_tot, MSLP, TCC
- The derivation is based on the 50 full levels (before level subsetting)

Variables selection is crucial, tested : Soil variables, 3D-W, others with mixed success

Regional CREDIT : model implementation

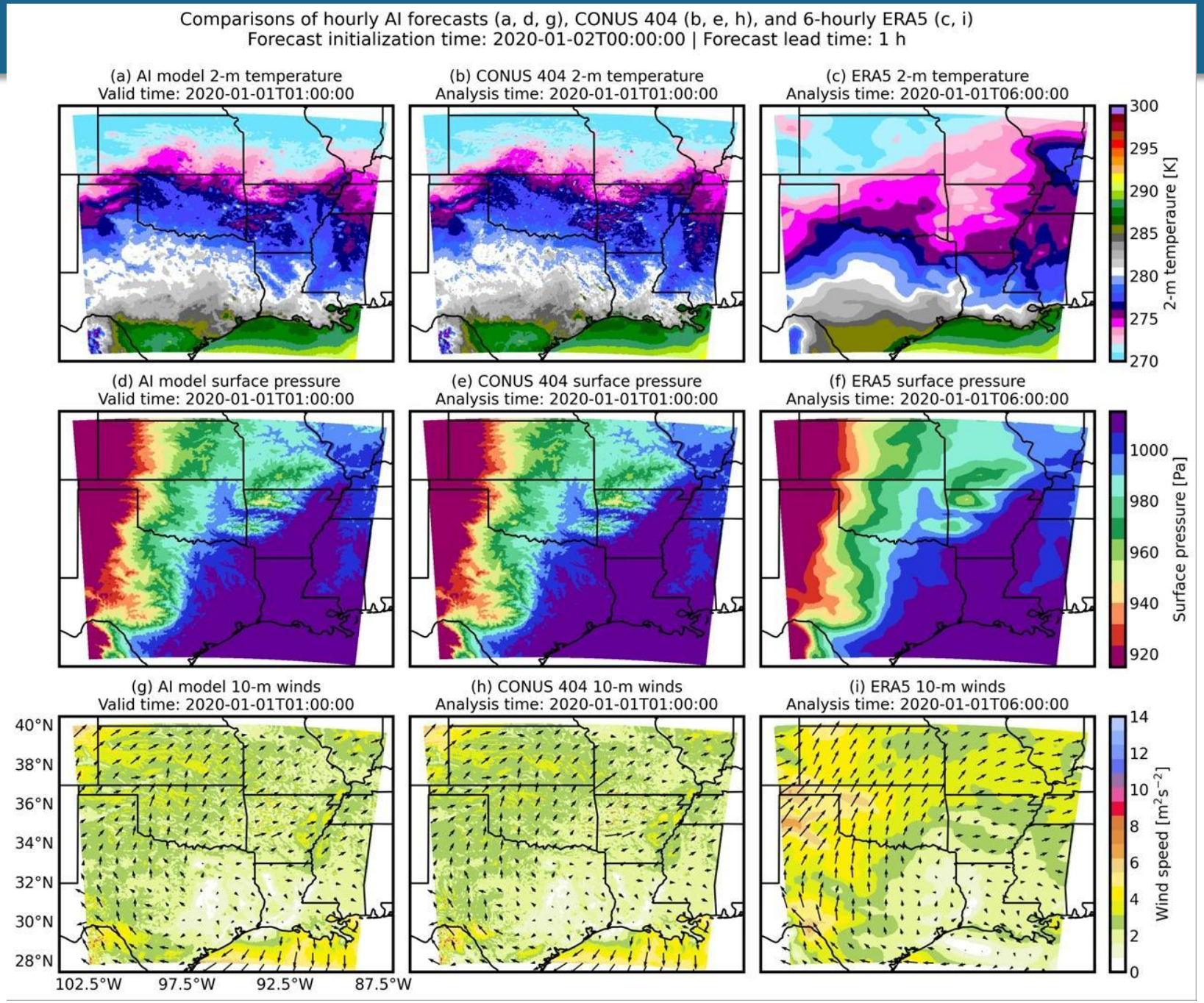


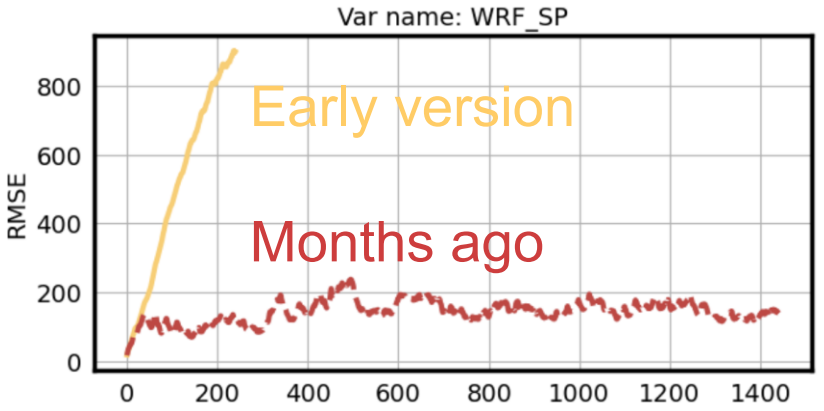
Schreck et al (2025)

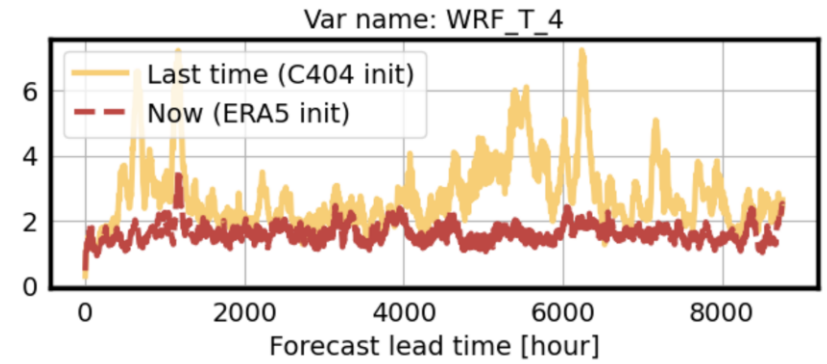


Sha et al (2026?)

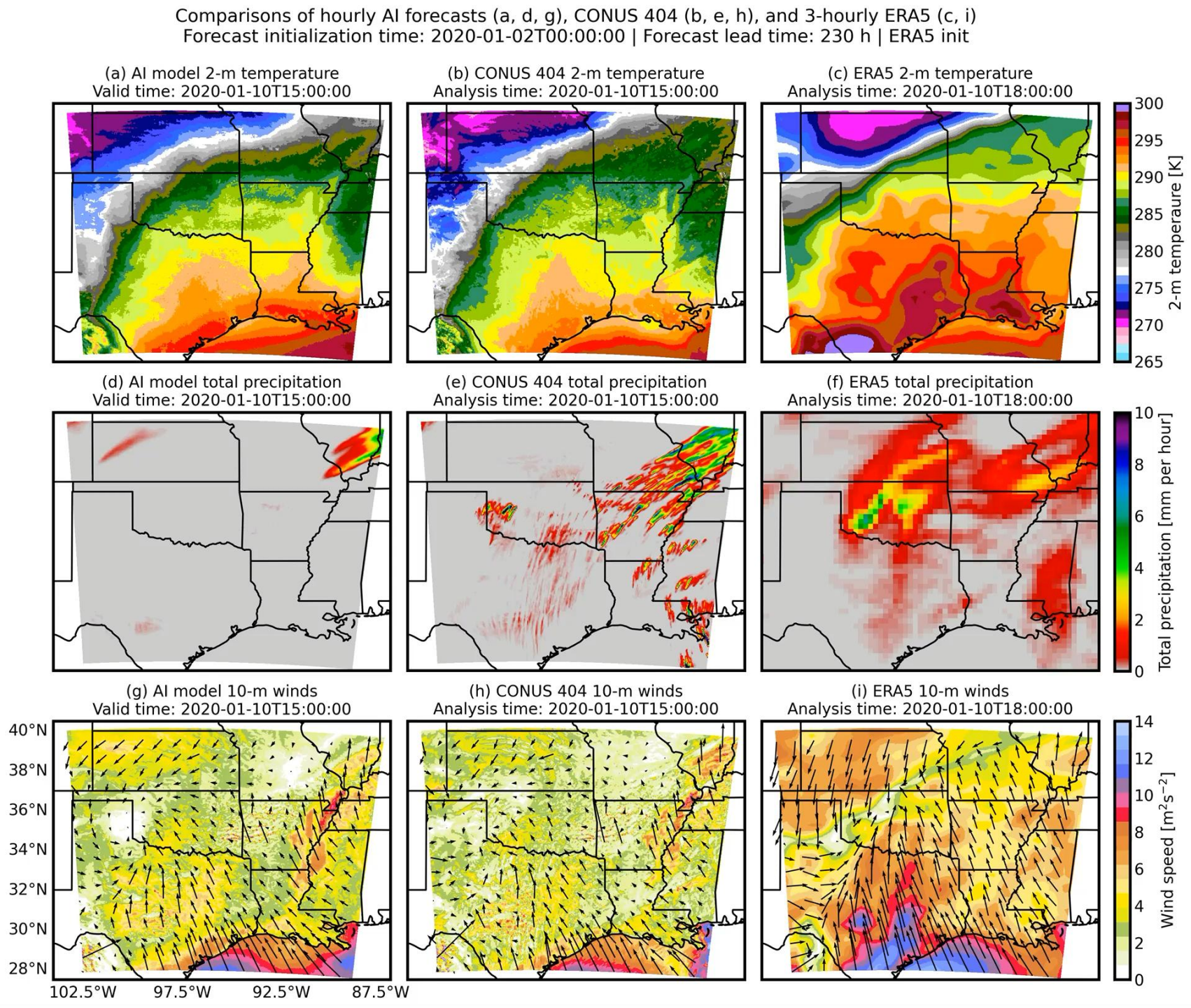
- C C Tensor concatenation.
- ⊕ Tensor add.
- S SiLU activation function.
- G Group normalization, number of groups 32.
- I Bilinear interpolation from (22, 28) to (84, 84).
- E Hour-of-day and day-of-year embedding using sine and cosine functions.
- L Feature-wise Linear Modulation (FiLM).
- C Conv2D, kernel size (3, 3), valid padding.
- E Cube Embedding. Conv3D, kernel size and stride (2, 4, 4).
- D Down-sampling layer. Conv2D, kernel size (3, 3), stride (2, 2).
- U Up-sampling layer. ConvTranspose2D, kernel size (2, 2), stride (2, 2).
- F Fully connected layer with tensor reshape. Input size 1536, output size 4x4xVar.

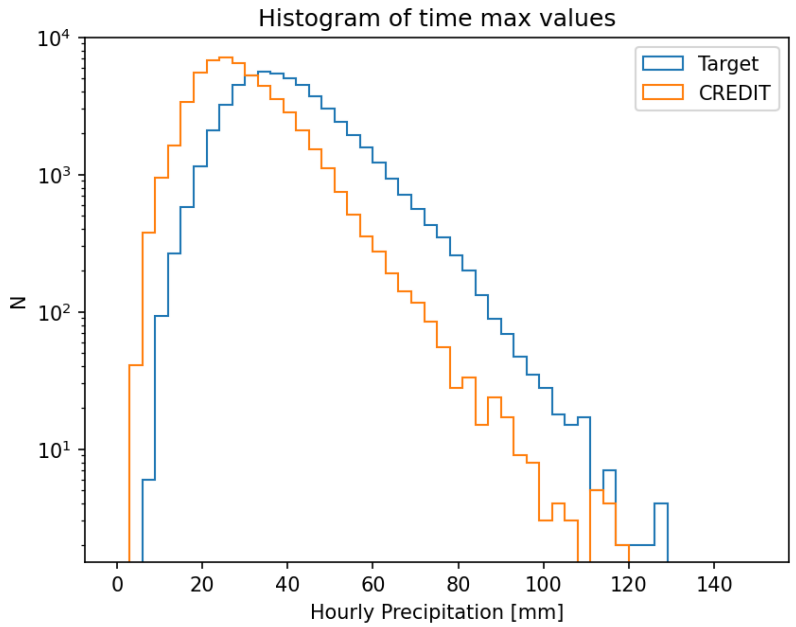




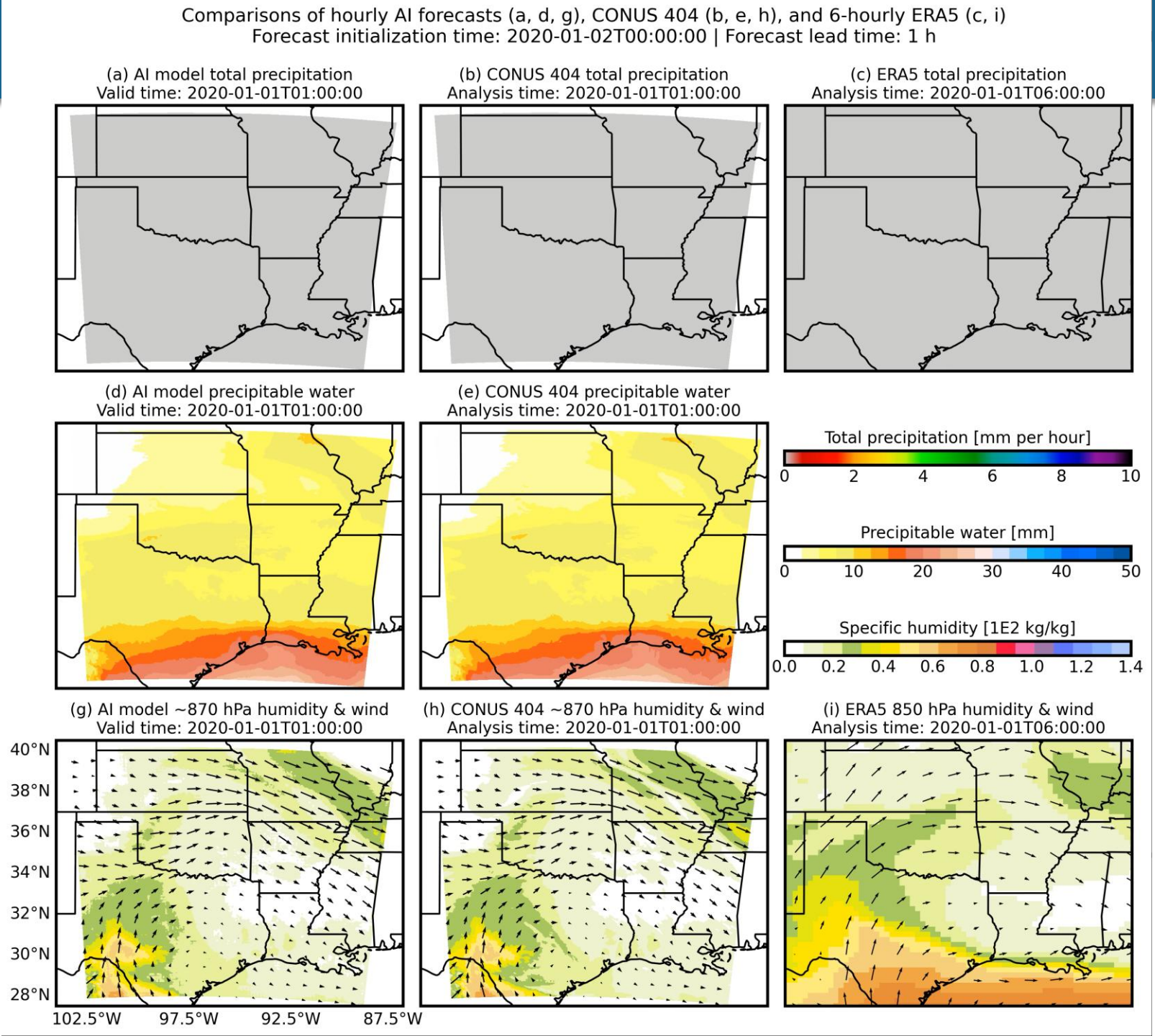


Now stable to 1-year+



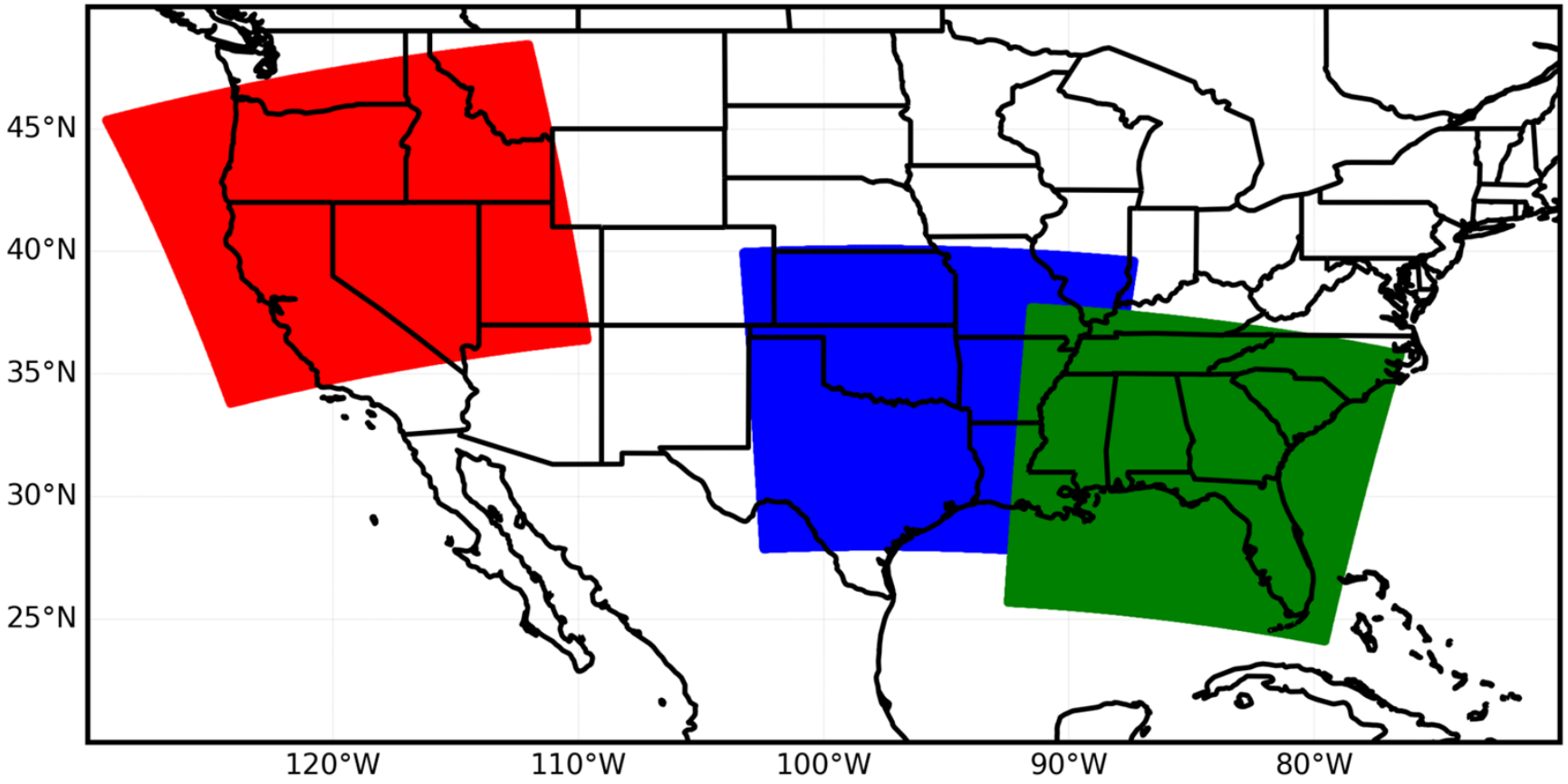


Able to produce hourly rainfall rates of 120mm/hr (for US that's high)
Still biased low, working on adding diffusion model for precipitation


Comparisons of hourly AI forecasts (a, d, g), CONUS 404 (b, e, h), and 6-hourly ERA5 (c, i)
Forecast initialization time: 2020-01-02T00:00:00 | Forecast lead time: 1 h
(a) AI model total precipitation
Valid time: 2020-01-01T01:00:00
(b) CONUS 404 total precipitation
Analysis time: 2020-01-01T01:00:00
(c) ERA5 total precipitation
Analysis time: 2020-01-01T06:00:00
(d) AI model precipitable water
Valid time: 2020-01-01T01:00:00
(e) CONUS 404 precipitable water
Analysis time: 2020-01-01T01:00:00
Total precipitation [mm per hour]
0 2 4 6 8 10
Precipitable water [mm]
0 10 20 30 40 50
Specific humidity [1E2 kg/kg]
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4
(g) AI model ~870 hPa humidity & wind
Valid time: 2020-01-01T01:00:00
(h) CONUS 404 ~870 hPa humidity & wind
Analysis time: 2020-01-01T01:00:00
(i) ERA5 850 hPa humidity & wind
Analysis time: 2020-01-01T06:00:00
40°N
38°N
36°N
34°N
32°N
30°N
28°N
102.5°W
97.5°W
92.5°W
87.5°W

Domain info

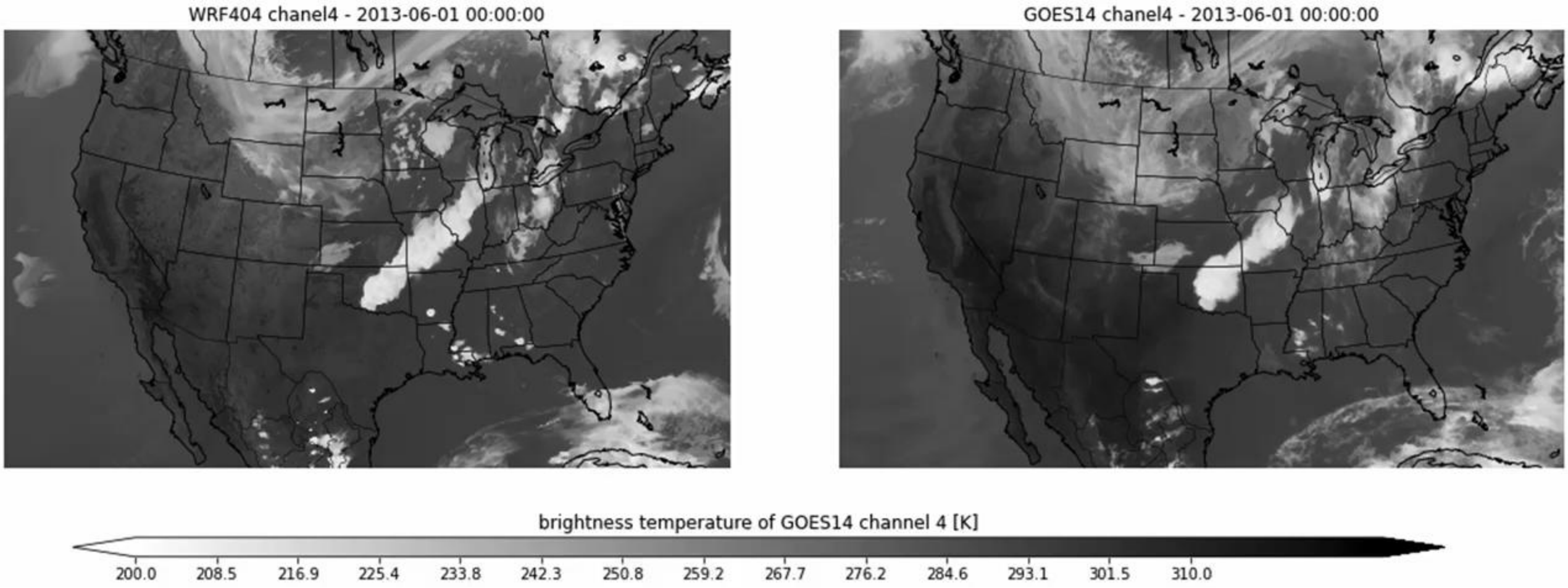
- Train AI models in the CONUS 404 **Great Plain** domain, then add the **Pacific Northwest, Southeastern** and many more domains. Test on regions / times left out of training.



Great Plain results only so far, adding other domains now.

Aim to understand transferability and generalizability of CP-MLWP

The Next Frontier : Training on Observations



Animation from Andreas Prein

Summary

- Advanced regional 4-km Convection Permitting AI/ML NWP system
  - Free running rollouts to >1 year stable with multi-step training for regional climate simulations
  - Some instability with different variables (e.g. soils) requiring new approaches
  - Major advances to come training on observations, and improving AI-DA
  - Adding new domains S. America? Europe? Alaska? …
  - Enables Ensemble Forecasts and RCM downscaling
- Current model missing most extreme precipitation
  - Adding stochastic precipitation sub-model
- Ongoing work is training over mountains and training for generalizability
- Future work will apply it to GCMs over INARCH basins