Generative Data Assimilation of Sparse Weather Station Observations at Kilometer Scales

Peter Manshausen, Yair Cohen, Peter Harrington, Jaideep Pathak, Michael Pritchard, Piyush Garg, Morteza Mardani, Karthik Kashinath, Simon Byrne, and Noah Brenowitz

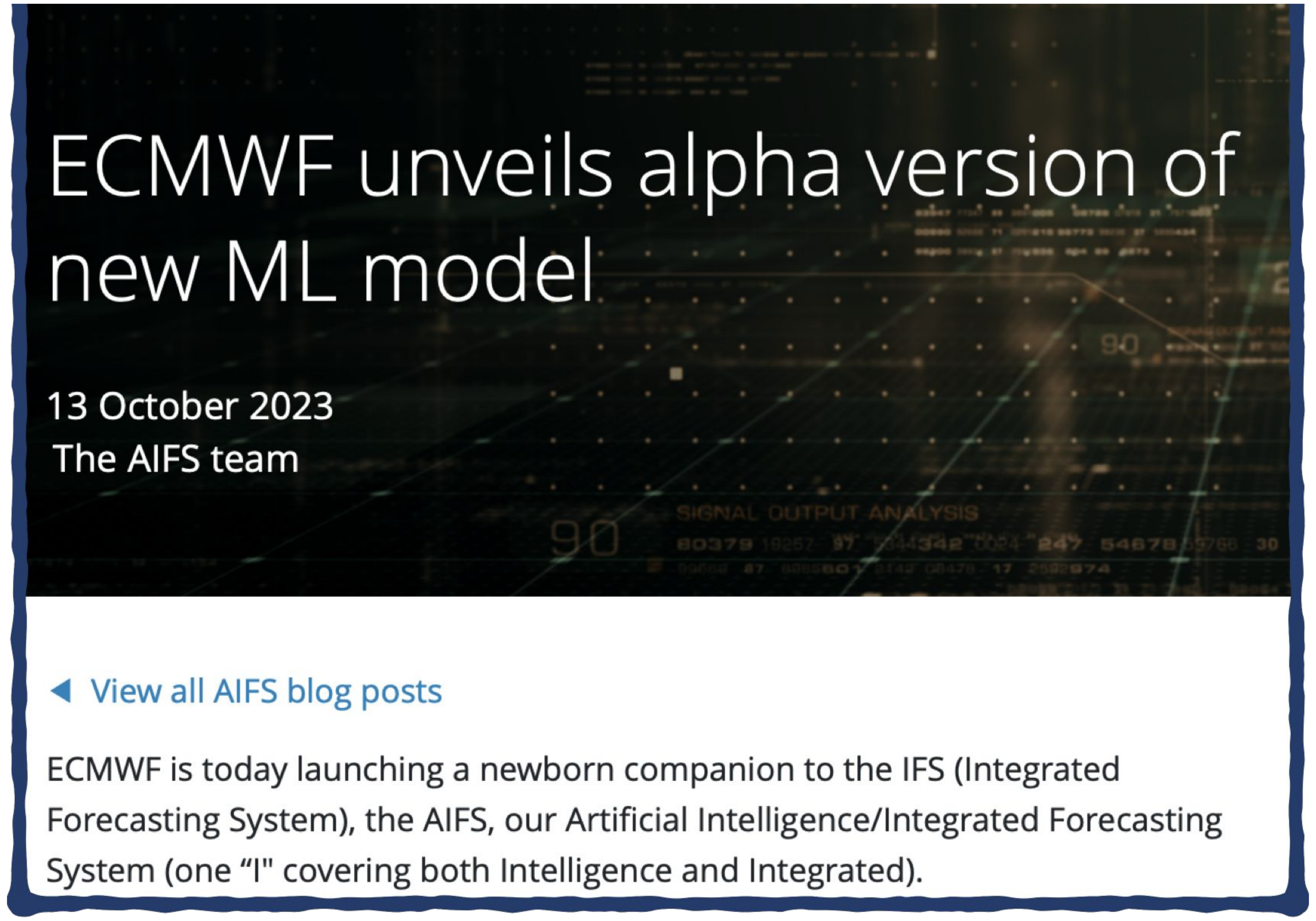
22 July 2025

CLIVAR summit, Boulder, CO

2023 Was a Milestone Year for Autoregressive AI Weather Prediction

Global 25-km AI weather forecasting has exited its infancy w/ large foundation models

- Several AI/ML weather models are now as accurate or better than state-of-the-art numerical weather prediction at global 25km resolution.
- AI weather models offer massive speedups of over10,000x and huge ensemble sizes
- Common configuration
  - 0.25 deg data (1440x721)
  - 6 – 12 hours



ECMWF unveils alpha version of new ML model

13 October 2023
The AIFS team

◄ View all AIFS blog posts

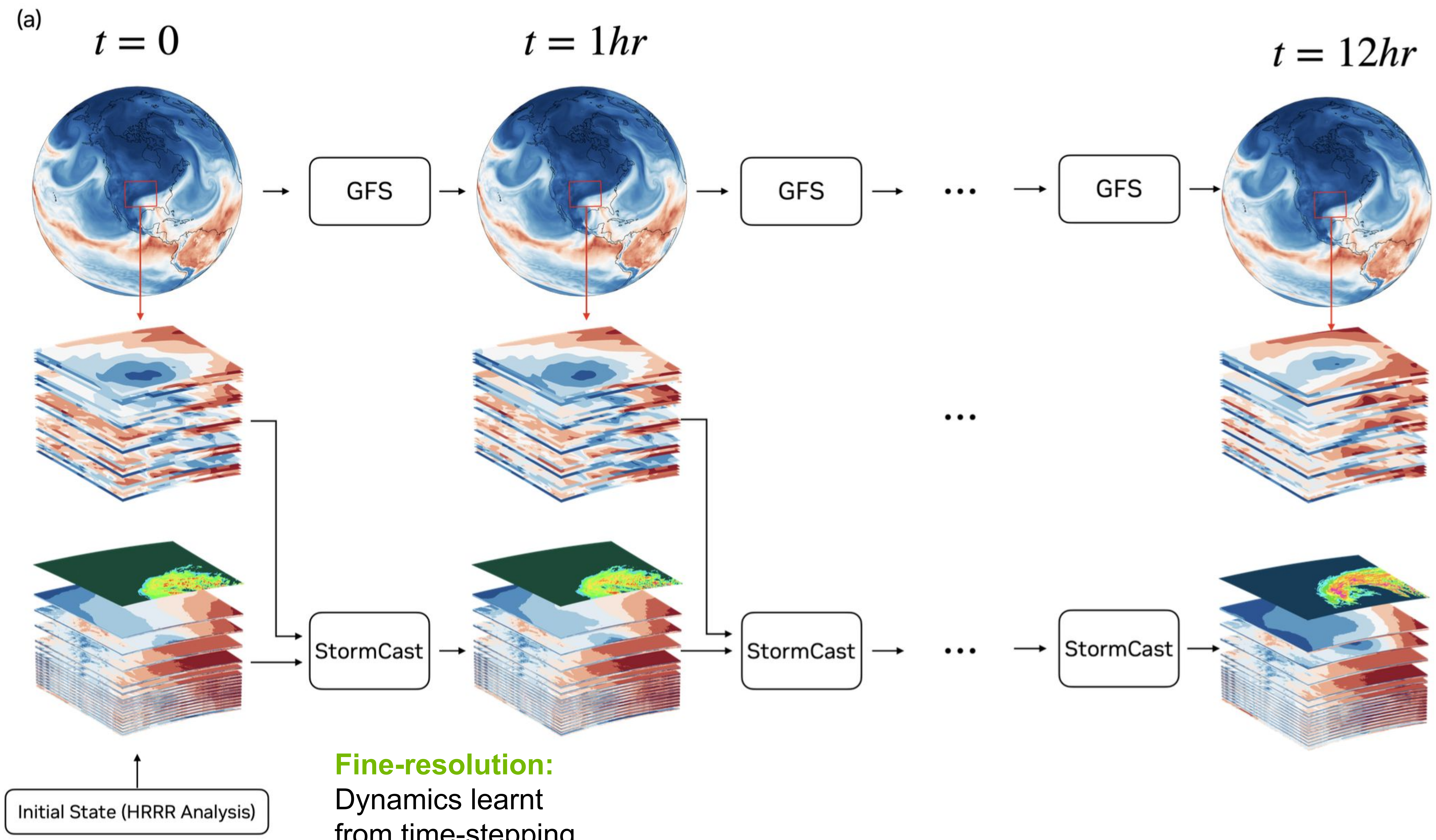
ECMWF is today launching a newborn companion to the IFS (Integrated Forecasting System), the AIFS, our Artificial Intelligence/Integrated Forecasting System (one “I" covering both Intelligence and Integrated).

And Also for Regional Forecasting

A Multi-Scale Inference Setup

Coarse-resolution:
ML model is made aware of a prior, 25-km global forecast.

Fine-resolution:
Is initialized with radar-assimilating HRRR...



Fine-resolution:
Dynamics learnt from time-stepping km-scale state.

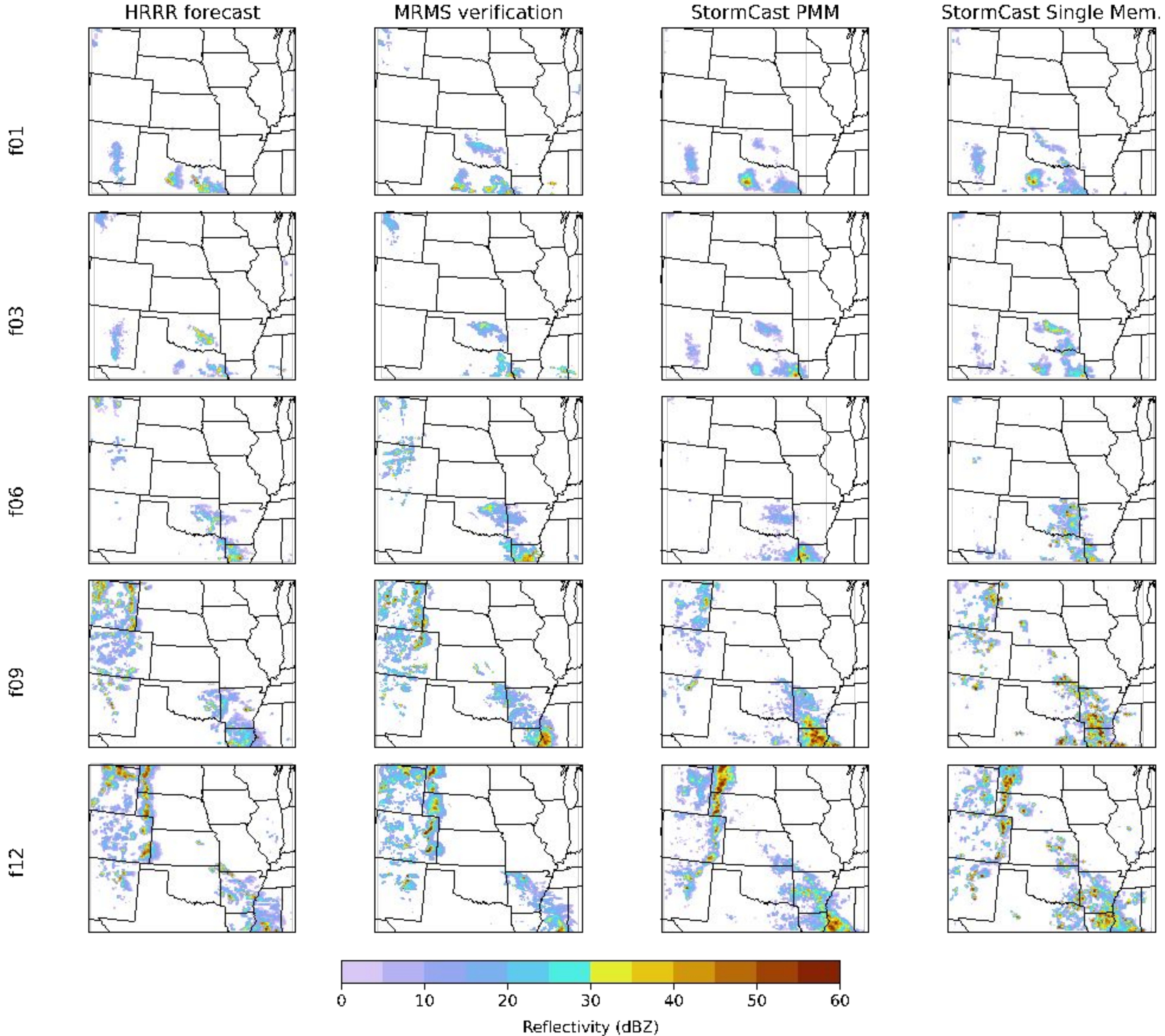
Lead time:
1 hour

3 hours

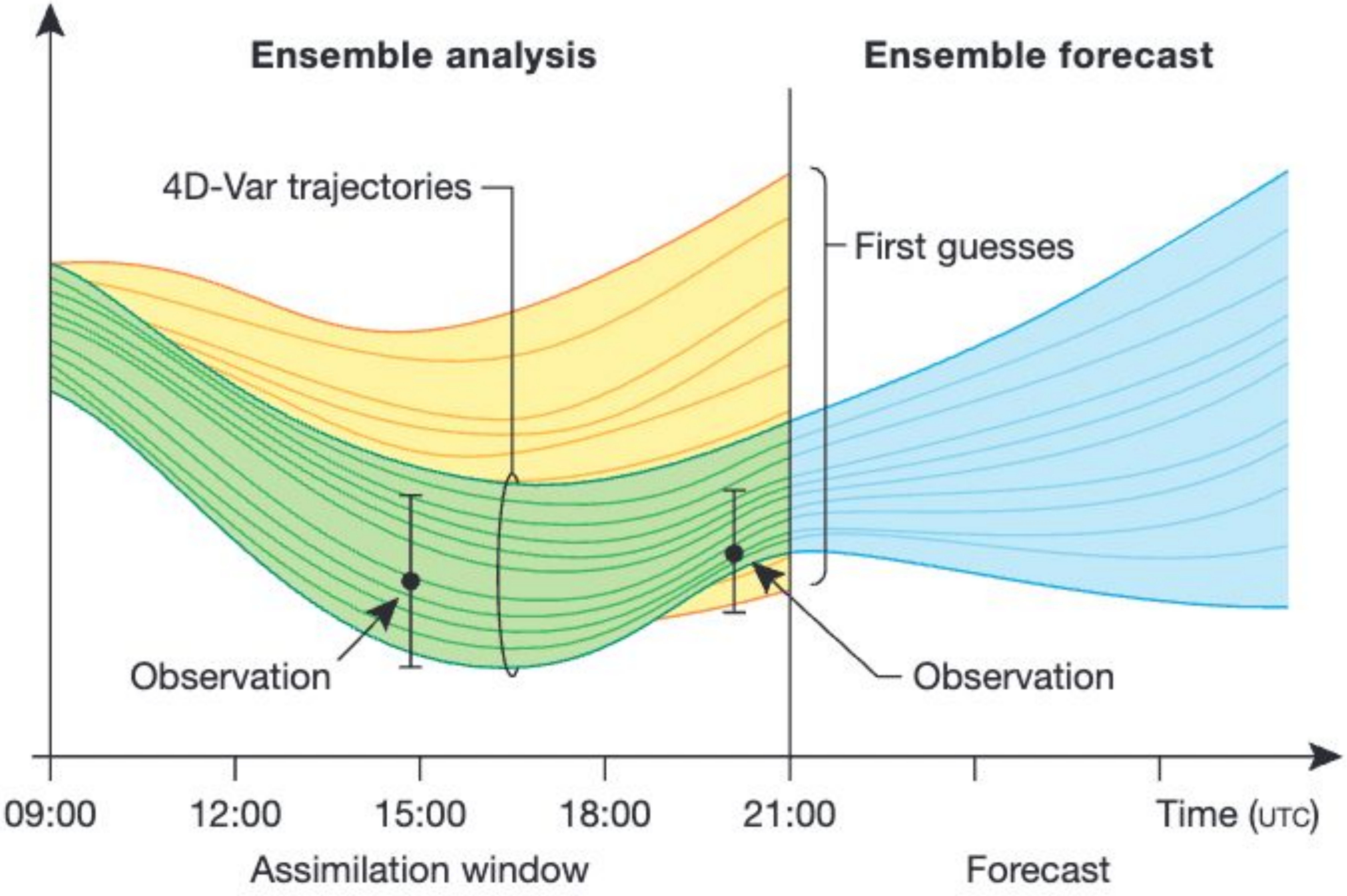
6 hours

9 hours

12 hours



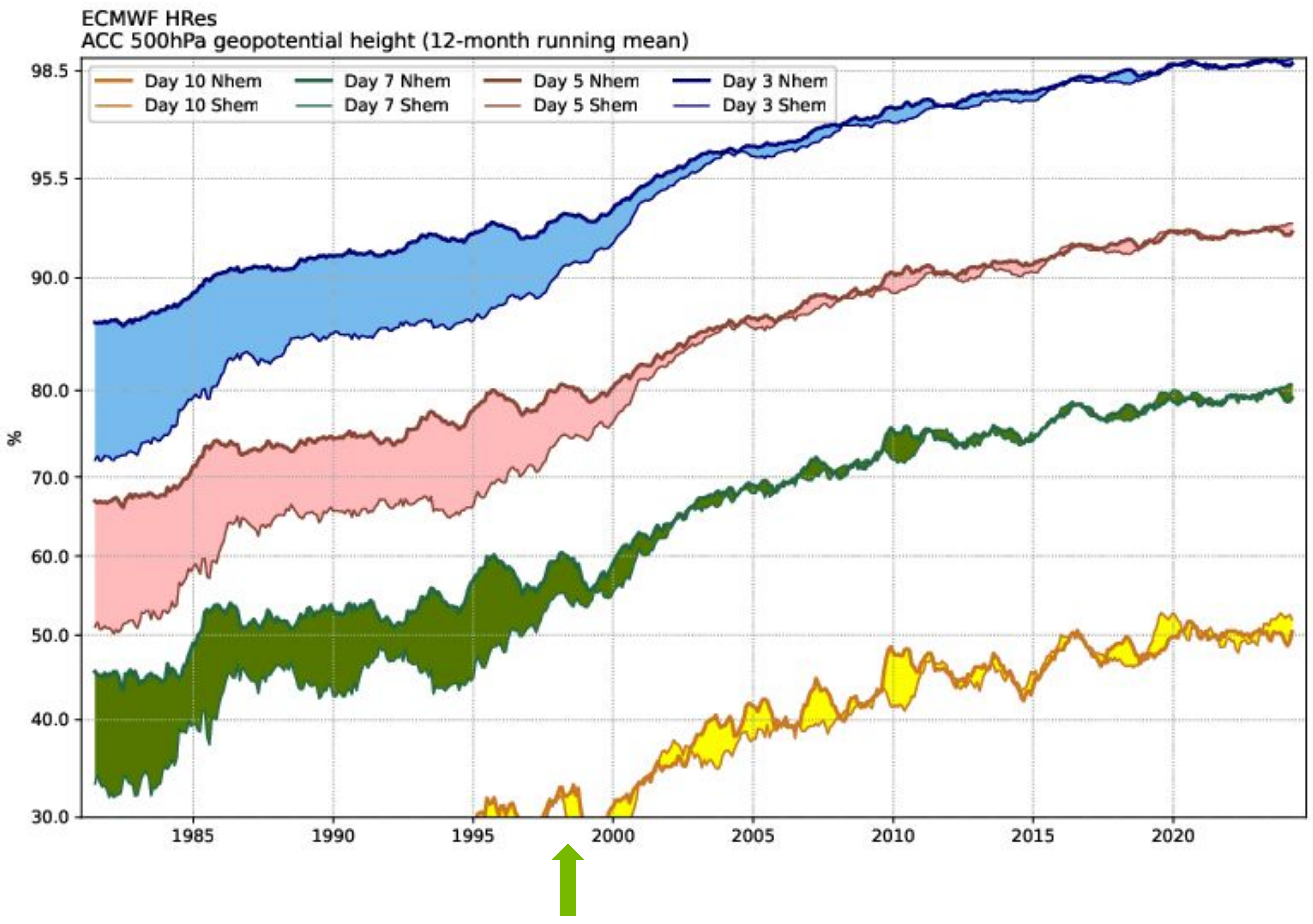
The Missing Piece: Data Assimilation



Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. Nature, 525(7567), 47–55. https://doi.org/10.1038/nature14956

NVIDIA

Large skill improvements from DA have happened before



4DVar put
into operations

Data Assimilation Is Very Similar to Inpainting

Song, Y., et. al. (2020). Score-based generative modeling through stochastic differential equations

The Magic of Diffusion Models

You can add conditioning without retraining the model

Bayes rule of the score function:

$$\begin{aligned}\nabla_{\mathrm{x}} \log p(x|y) &= \nabla_x \log \frac{p(y|x)p(x)}{p(y)} \\ &= \nabla_x[\log p(x) + \log p(y|x) - \cancel{\log p(y)}] \\ &= \nabla_x \log p(x) + \nabla_x \log p(y|x) \\ &\approx s_\theta(x, \sigma) + \nabla_x \log p(y|x)\end{aligned}$$

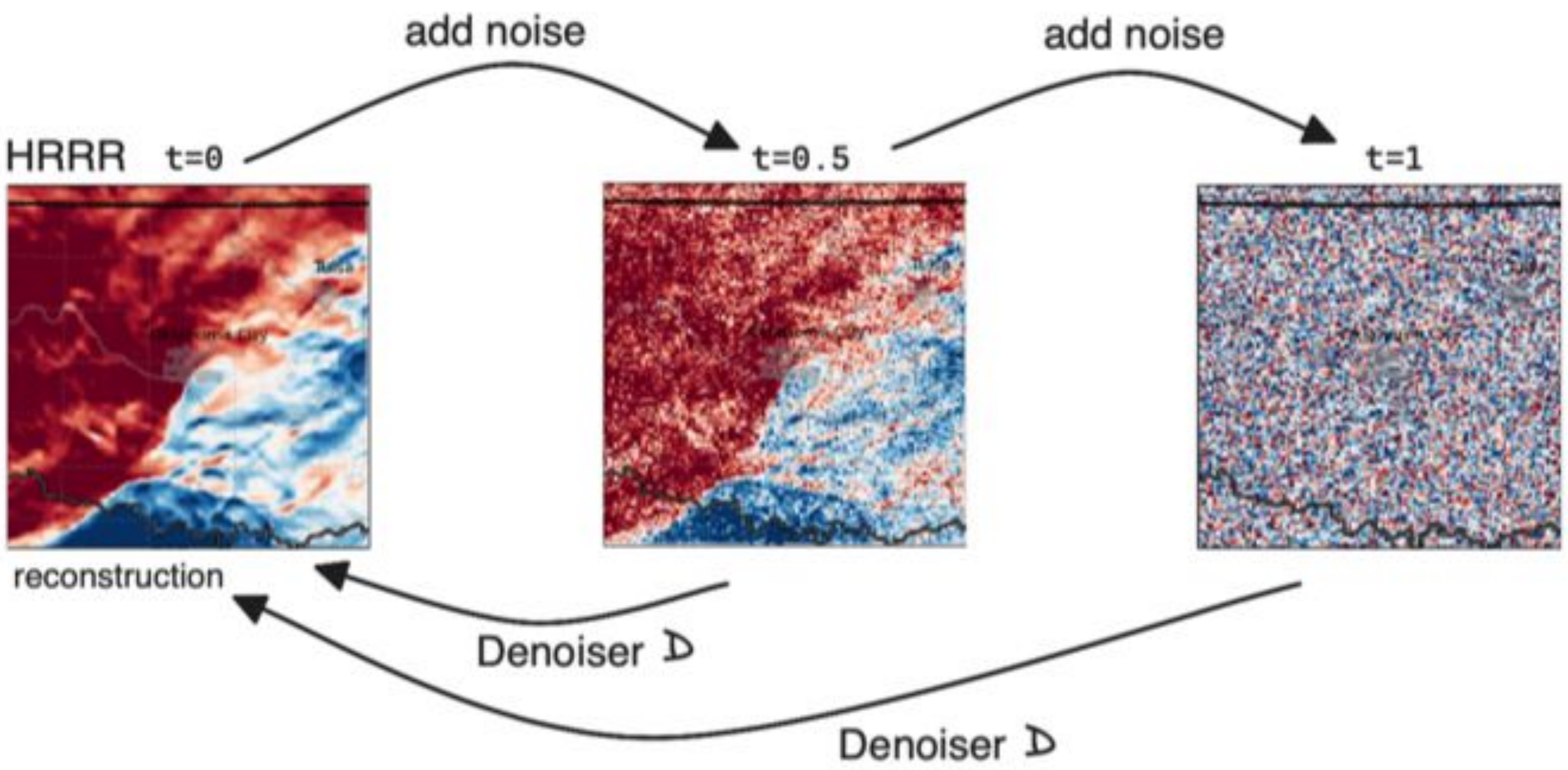
Observation operator:

$$\nabla_x \log p(y|x) = \nabla_{\mathrm{x}} \log N(\cdot\,|Ax, \sigma) = \sigma^{-2} A'(Ax - y)$$

Rozet, F., & Louppe, G. (2023). Score-based Data Assimilation. *arXiv [Cs.LG]*. arXiv. https://arxiv.org/abs/2306.10574

Step 1: Emulating the Km-Scale Models with Diffusion Models

Forward diffusion process: $d\boldsymbol{x} = d\boldsymbol{W} \rightarrow \boldsymbol{x}(t) = \boldsymbol{x}(0) + \boldsymbol{W}(t)$



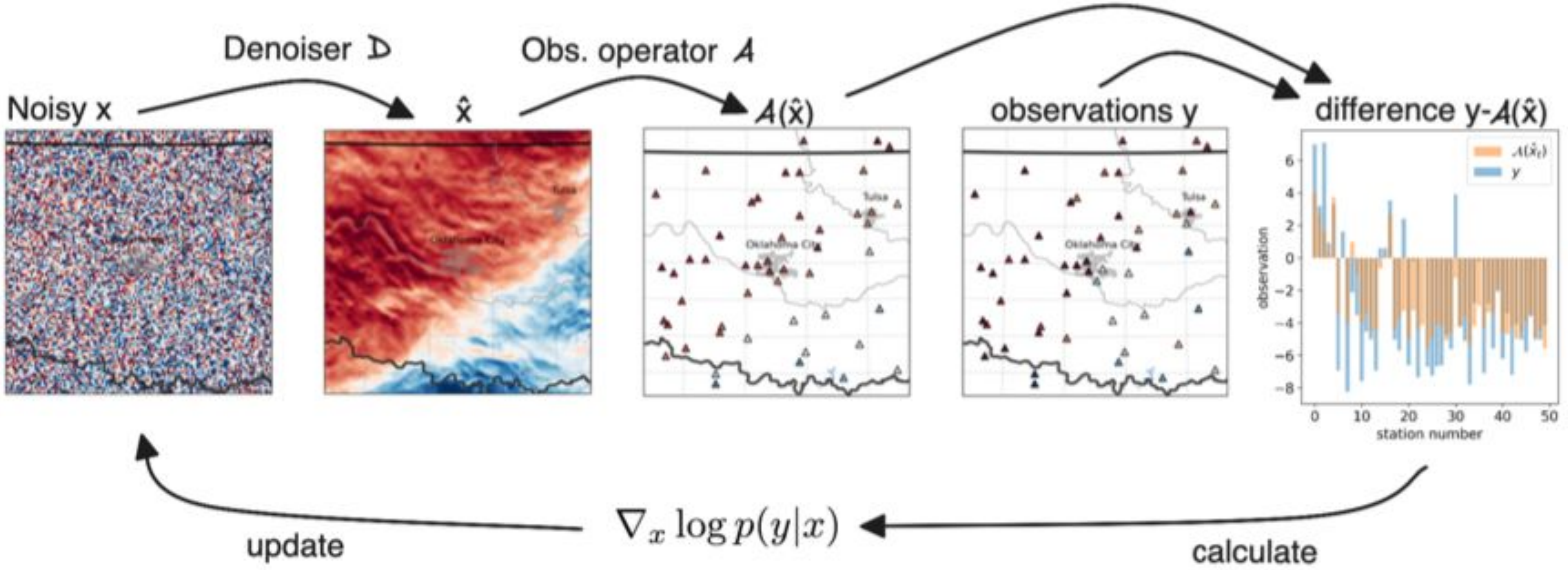
Backwards process:

$$\frac{d\boldsymbol{x}}{dt} = -t\nabla_x \log\big(p(\boldsymbol{x}; c, t)\big) = -\frac{(D(x, \boldsymbol{c}, t) - x)}{t}$$

Architecture: Elucidated Diffusion (Karras et al. 2022)

Step 2: Data Assimilation

Generation conditioned on weather gauge observations



Rozet, F., & Louppe, G. (2023). Score-based Data Assimilation. arXiv [cs.LG]. arXiv. https://arxiv.org/abs/2306.10574

The Setup

- Central OK testbed
- Prior: Training data for prior: HRRR analysis from 2018-2021
  - Total precipitation over one hour (tp) - log transformed
  - U10m, v10m
- Surface Observations from NOAA Integrate Surface Databas
  - Observation operator $y \sim N(interp(x), \sigma)$
- Model: Standard Unet architecture
- Elucidated diffusions (Karras et al.) diffusion framework:
  - https://github.com/NVlabs/edm
- Cost:
  - Training: 64 GPU hours (A100)
  - Inference: 10 seconds per sample (RTX 6000 Ada)

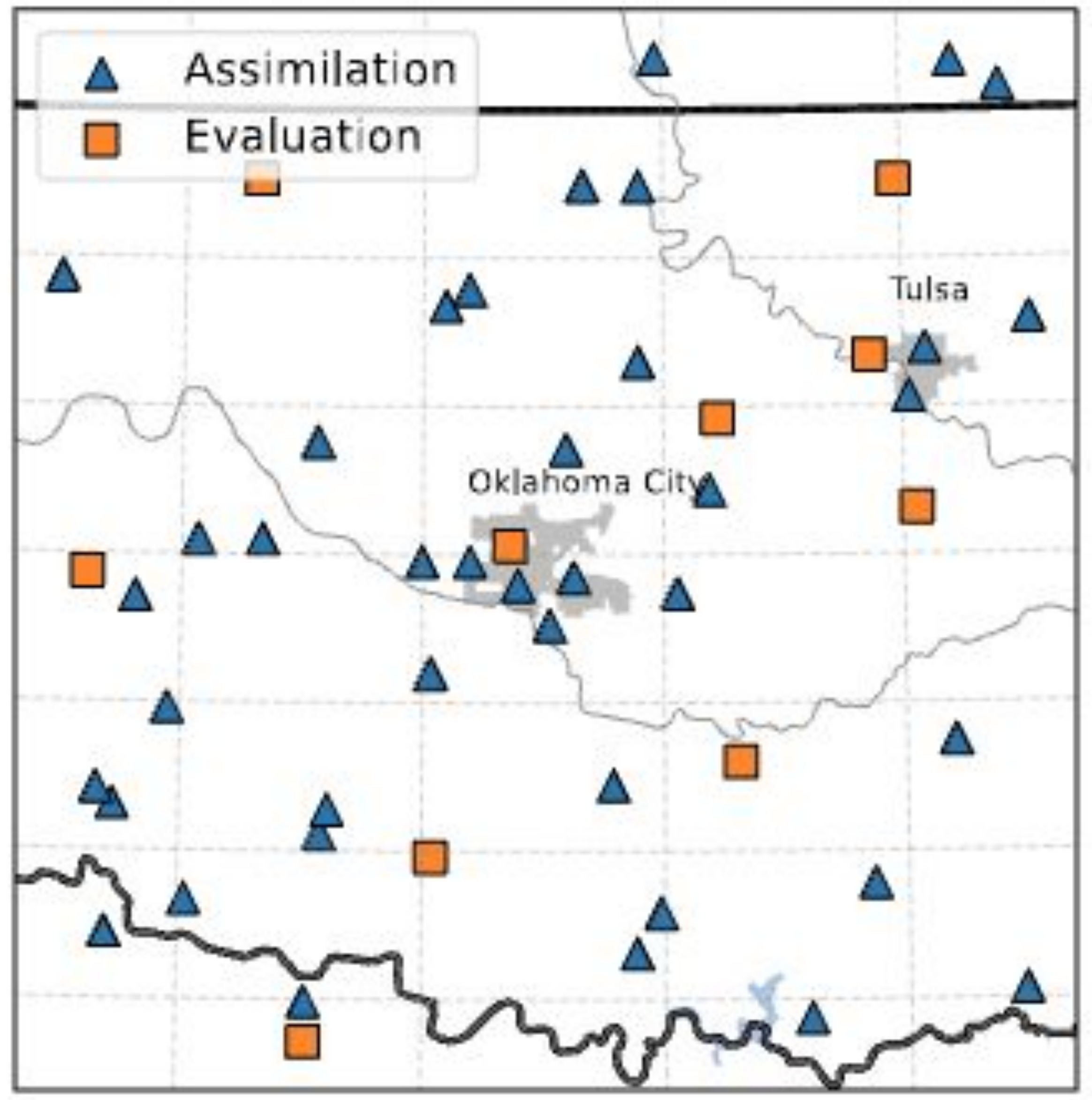
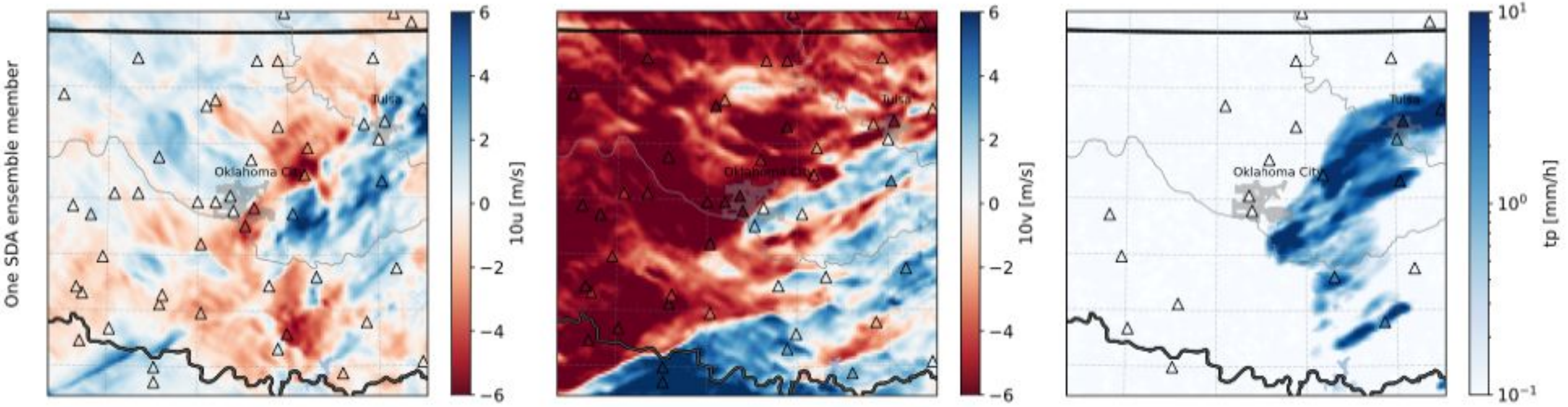


Figure A2. For evaluation we leave out ten randomly selected stations.

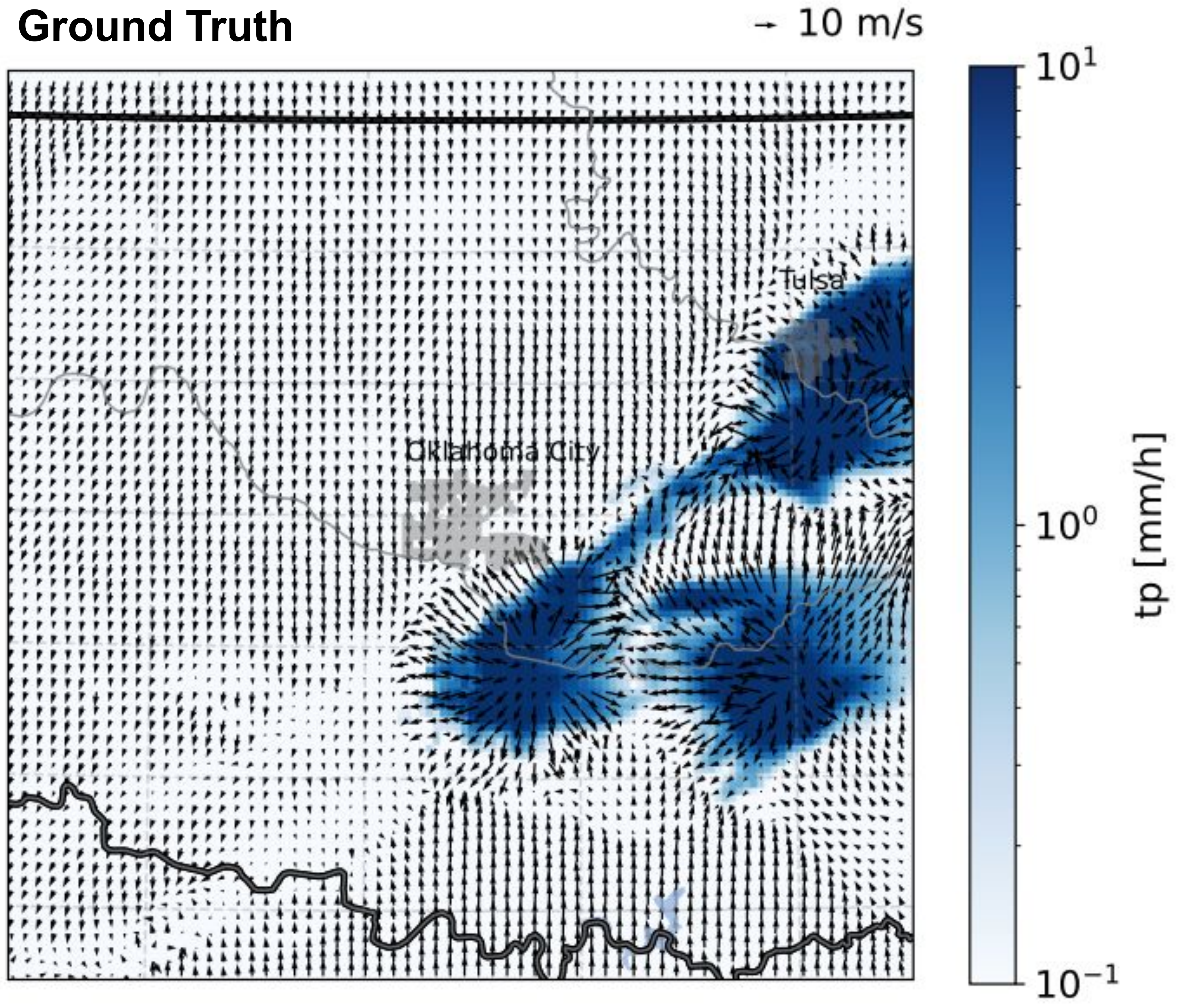
State Estimation

Data source: NOAA weather stations

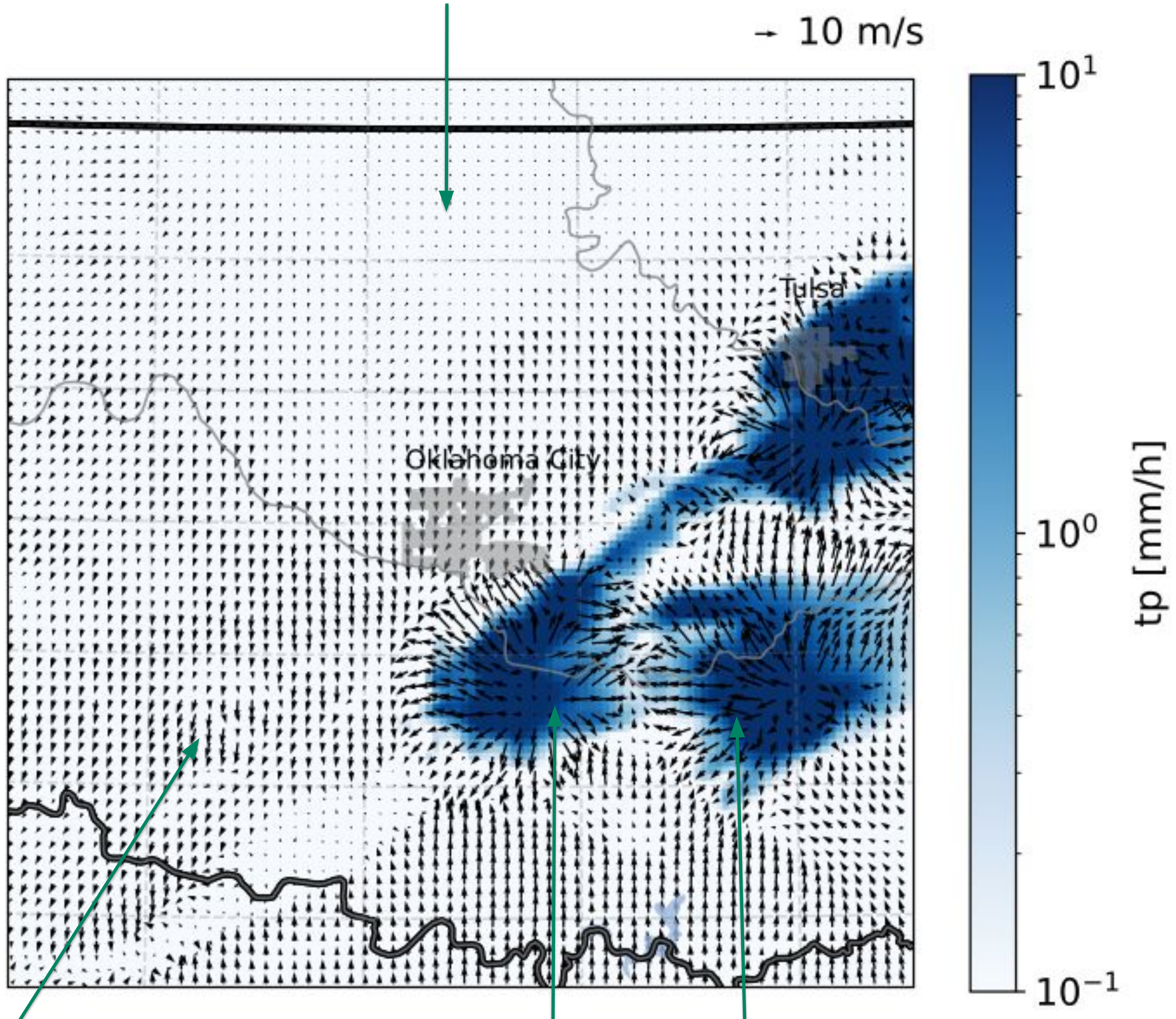


Learnt Multivariate Physics Subject to Data Constraints

Exercise: Generate northward wind component from eastward & rainfall rate



Nothing to constrain



Flow channeling
associated
with topography

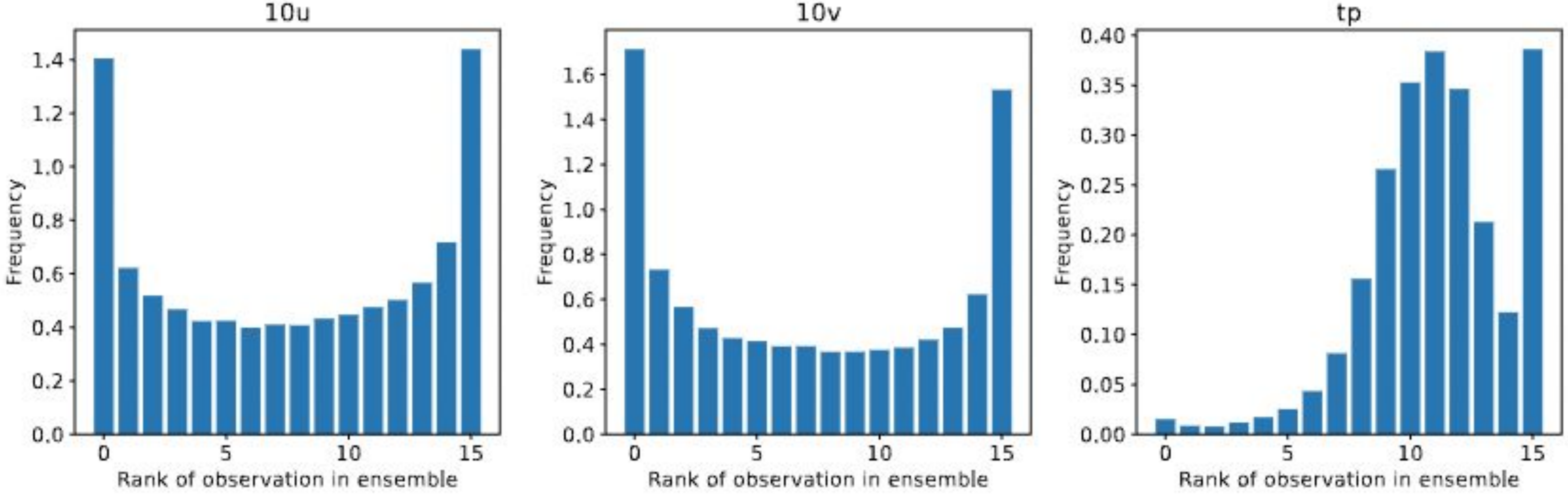
Cold pool divergence
associated with
evaporating rainfall.

Accuracy Better than HRRR for Surface Wind Estimation

Ensemble predictions measured relative to held-out weather station observations

More work needed on calibration

Rank histograms show the predictions are under-dispersive or biased



Km-scale DA — Conclusions

- The diffusion-based data assimilation performs better than HRRR for winds and comparable for precipitation
- Model trained in purely generative mode, no need to retrain. (*toward foundation modeling*)
- Especially promising since we haven't yet used:
  - Any remote sensing
  - Any background state
- **A vision**: AI models trained from simulation data, then initialized using DA
- Try it out yourself: Code available on NVIDIA PhysicsNeMo (see link in preprint)

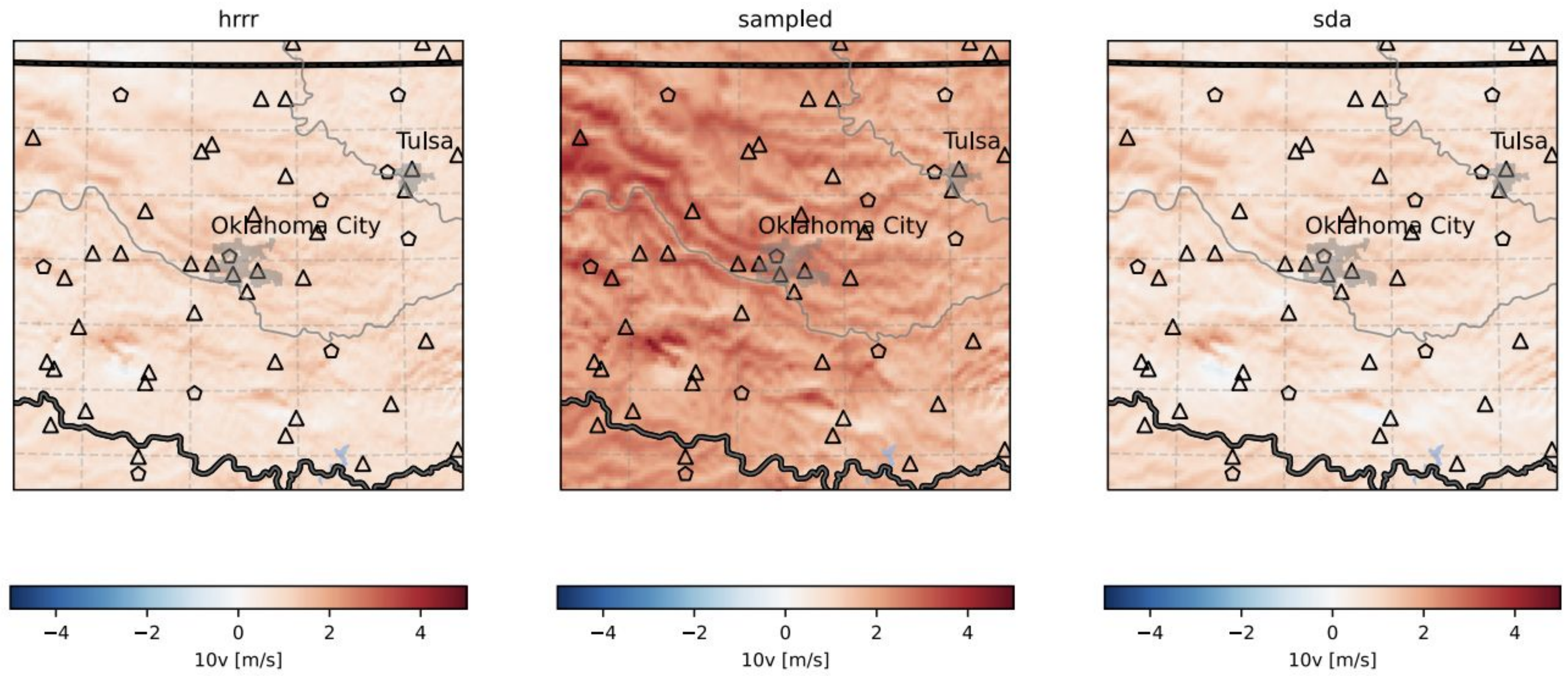
Manshausen, P., Cohen, Y., Pathak, J., Pritchard, M., Garg, P., Mardani, M., Kashinath, K., Byrne, S., & Brenowitz, N. (2024). Generative Data Assimilation of Sparse Weather Station Observations at Kilometer Scales. In *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/2406.16947

NVIDIA

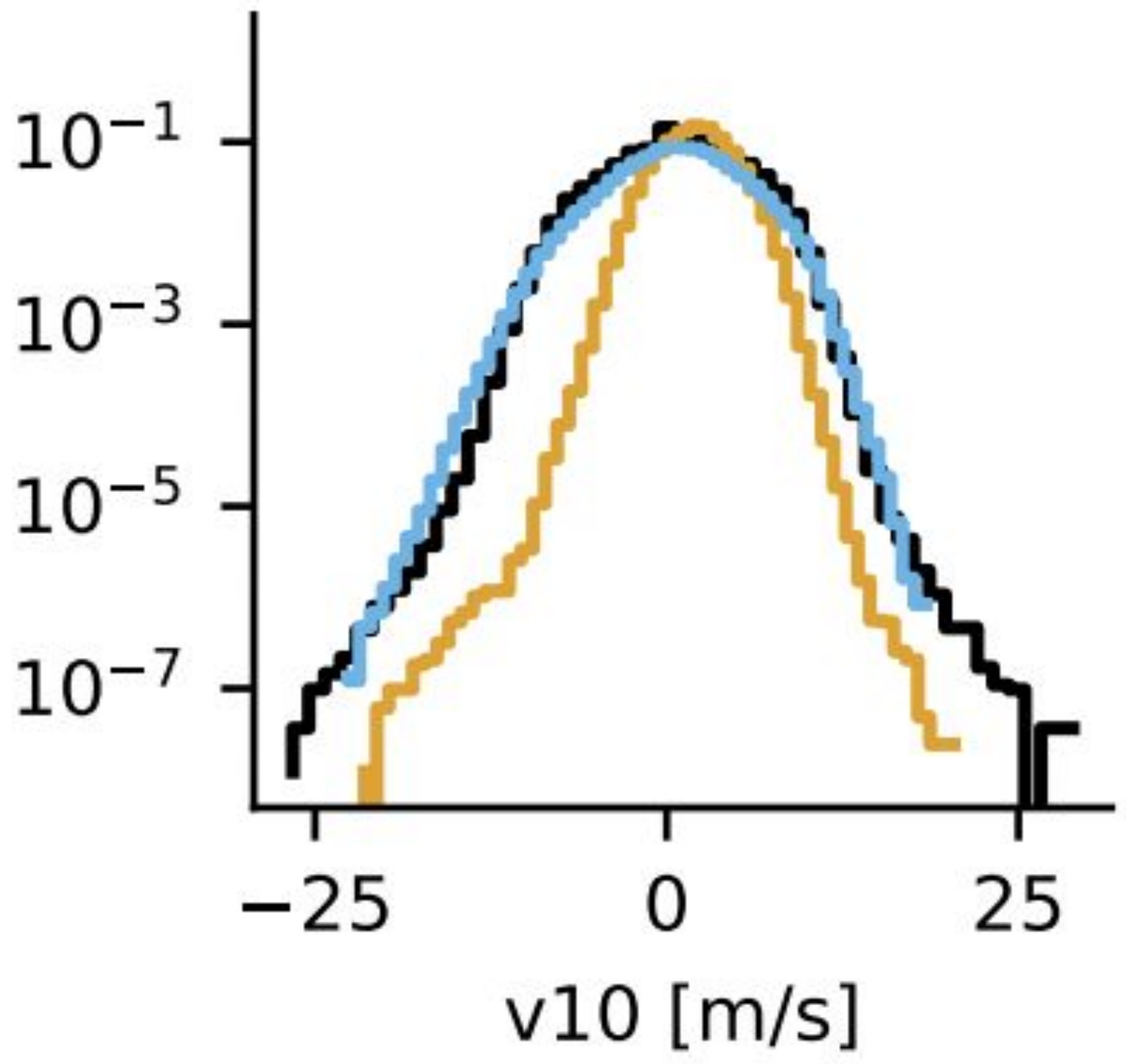
Unconditional samples from the prior

Meridional wind speed

Time mean

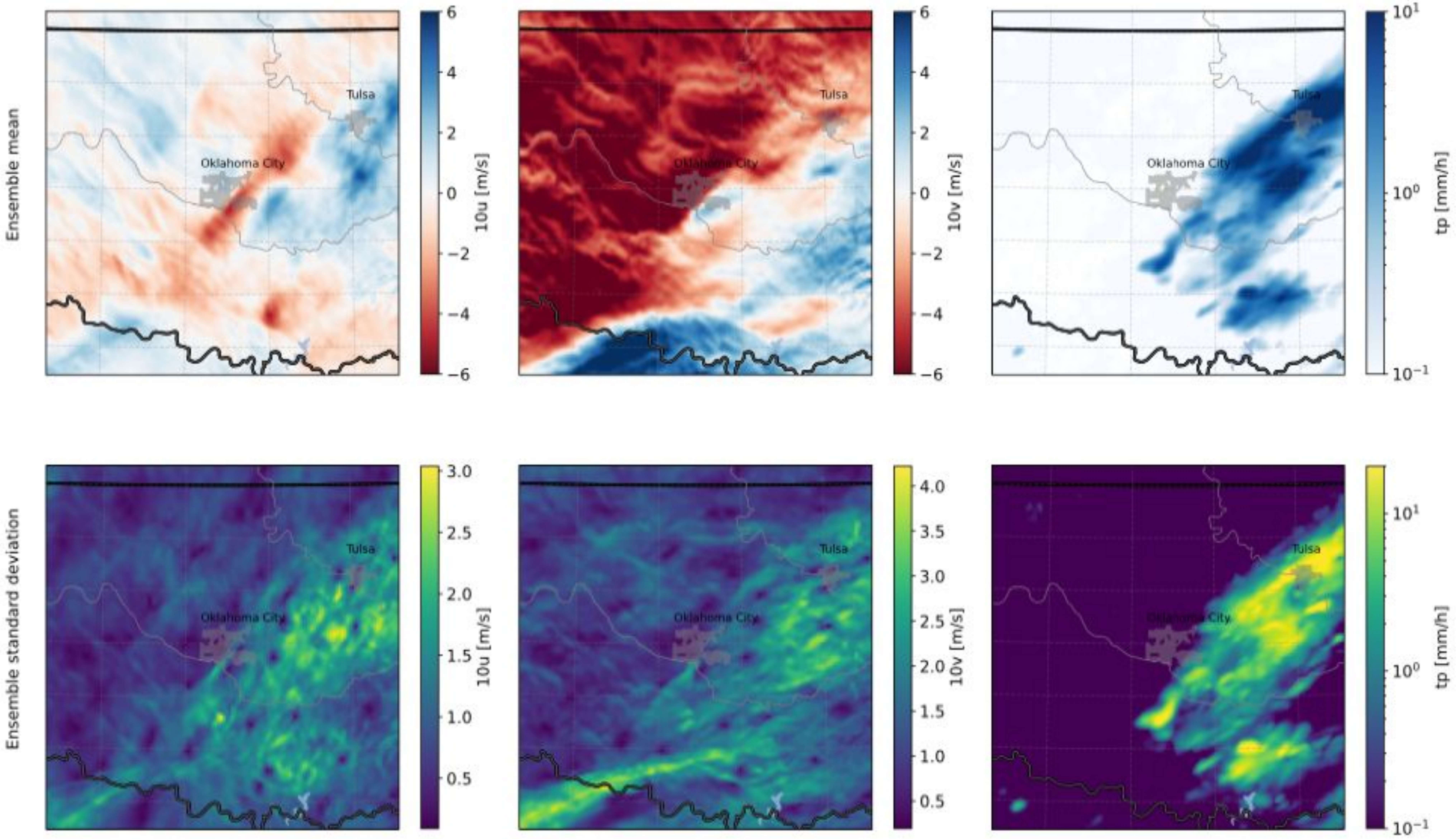


Histogram



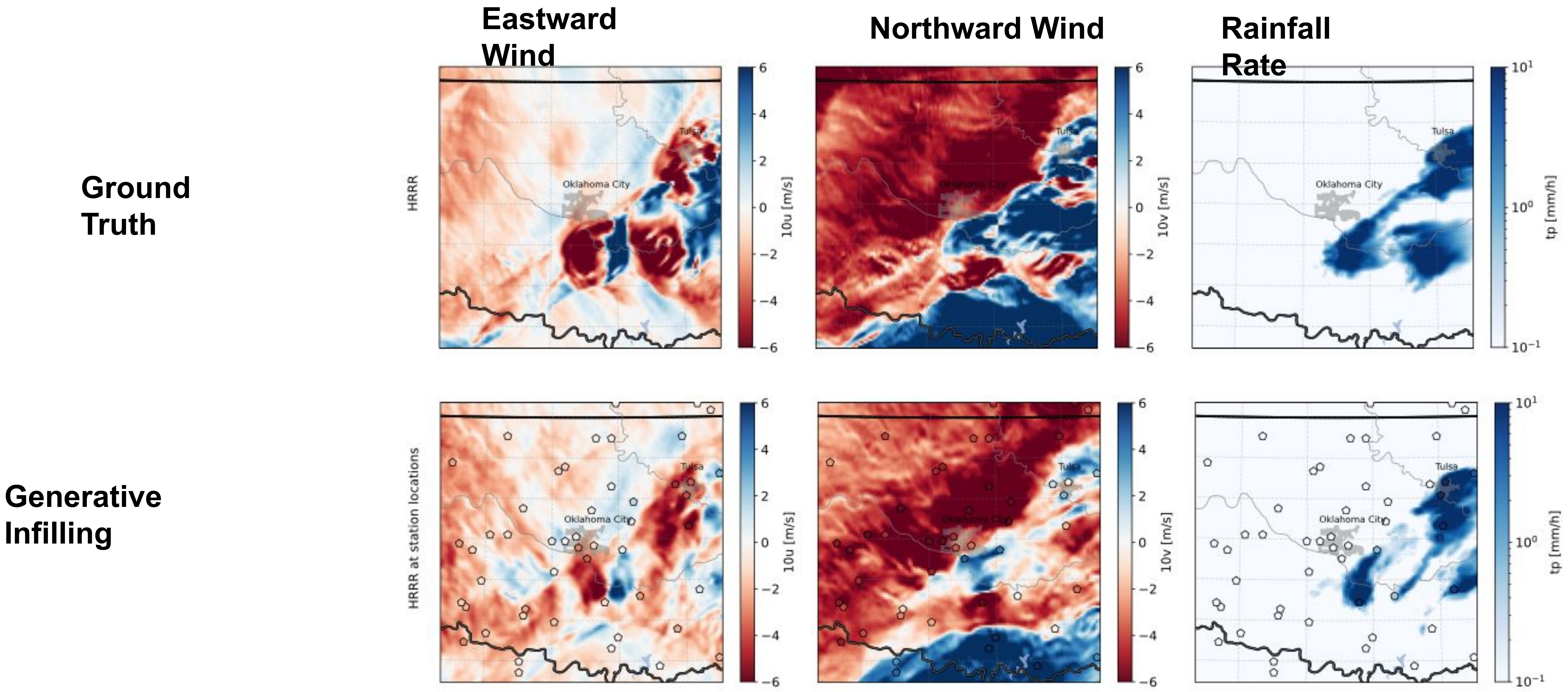
Ensemble State Estimation

Mean and spread from 20-member ensembles



Generative Infilling from Sparse Wind and Precipitation Data

Generation conditioned on weather gauge observations



On Comparing Gridded Data to Observations

A philosophical challenge

- Station observations may not represent the average atmospheric state on a 2km grid
- Yet, regional models and QPE products are evaluated directly against observations. See *James et. al. (2022). HRRR Part II. Forecast Perforamance*
- Here we assumed that no bias exists in the prior and assumed $y = x + \epsilon$.