Confronting Climate Change with Generative and Self-Supervised Machine Learning

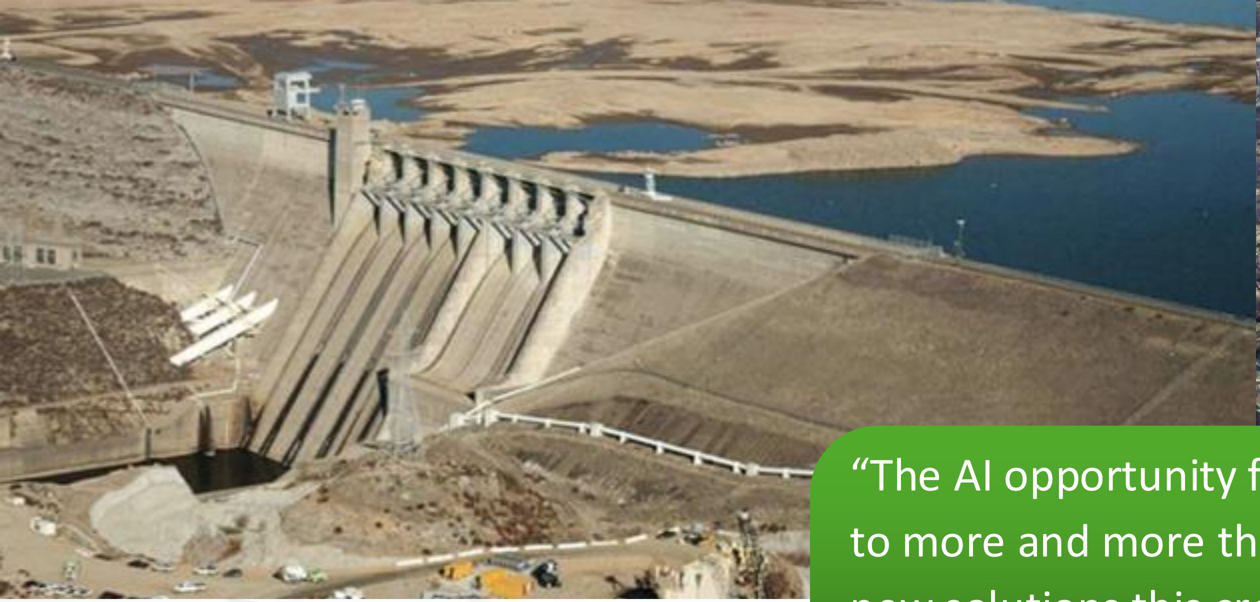
Claire Monteleoni

INRIA Paris

University of Colorado Boulder

Inria

Choose France™



“The AI opportunity for the Earth is significant. Today’s AI explosion will see us add AI to more and more things every year.... As we think about the gains, efficiencies and new solutions this creates for nations, business and for everyday life, we must also think about how to maximize the gains for society and our environment at large.”

– The World Economic Forum: Harnessing Artificial Intelligence for the Earth. 2018

Climate Informatics: using Machine Learning to address Climate Change

2008 Started research on Climate Informatics, with Gavin Schmidt, NASA

2010 "Tracking Climate Models" [Monteleoni et al., NASA CIDU, Best Application Paper Award]

2011 Launched International Workshop on Climate Informatics, New York Academy of Sciences

2012 Climate Informatics Workshop held at NCAR, Boulder, for next 7 years

2013 "Climate Informatics" book chapter [M et al., SAM]

2014 "Climate Change: Challenges for Machine Learning," [M & Banerjee, NeurIPS Tutorial]

2015 Launched Climate Informatics Hackathon, Paris and Boulder

2018 World Economic Forum recognizes Climate Informatics as key priority

2021 Computing Research for the Climate Crisis [Bliss, Bradley @ M, CCC white paper]

2022 First batch of articles published in Environmental Data Science, Cambridge University Press

2024 13th Conference on Climate Informatics, Turing Institute, London

2025 14th Conference on Climate Informatics, April 28-30, Rio de Janeiro, Brazil

Creating transdisciplinary research communities

Hackathons: data science challenges on science problems

Help students to collaborate with their peers in other fields

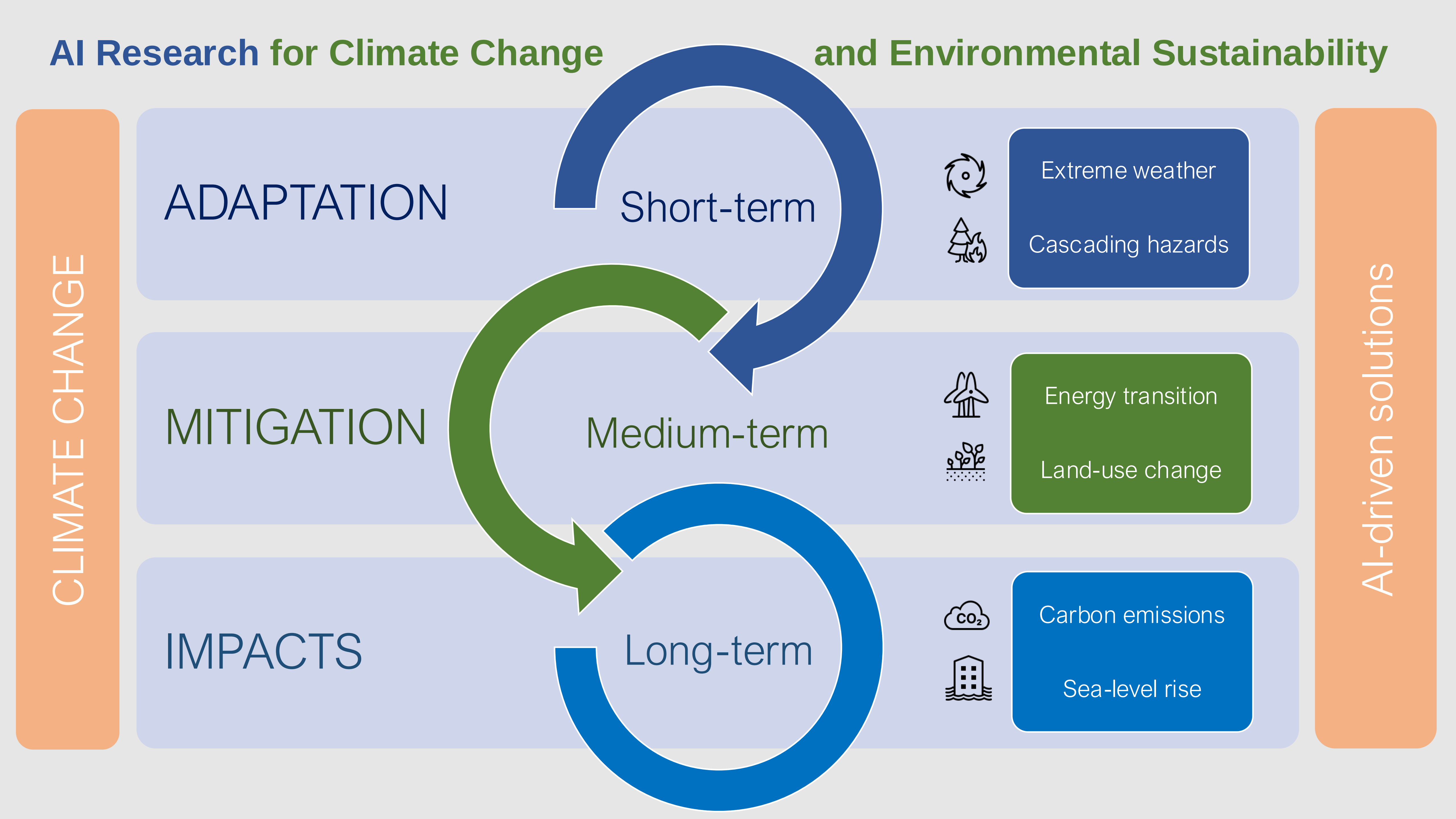
New publication venues that are truly cross-discipline

Events, including time for casual discussion

- Machine Learning for the Physics of Climate, Kavli Institute, UCSB

  → forthcoming paper in Nature Reviews Physics
- This ELLIIT Symposium and study period is a great example!

- **Call for Working Groups: Environmental Data Science Innovation and Inclusion Lab (ESIIL)**

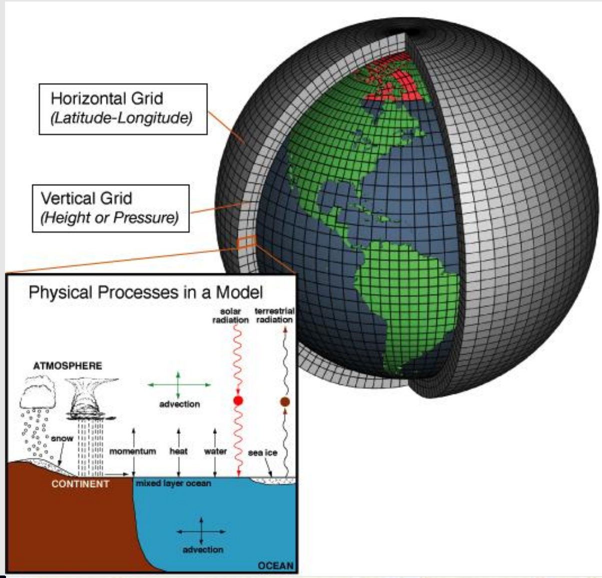

AI Research for Climate Change and Environmental Sustainability
CLIMATE CHANGE
ADAPTATION
Short-term
Extreme weather
Cascading hazards
MITIGATION
Medium-term
Energy transition
Land-use change
IMPACTS
Long-term
Carbon emissions
Sea-level rise
AI-driven solutions

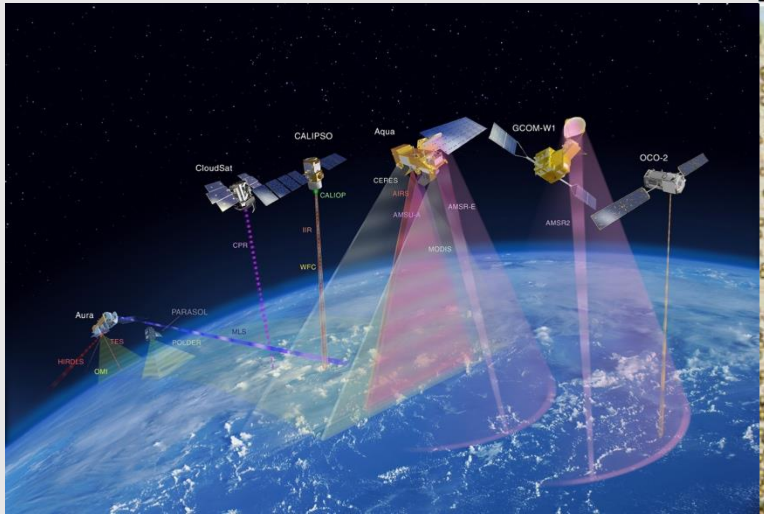
Approach: Exploit all available data

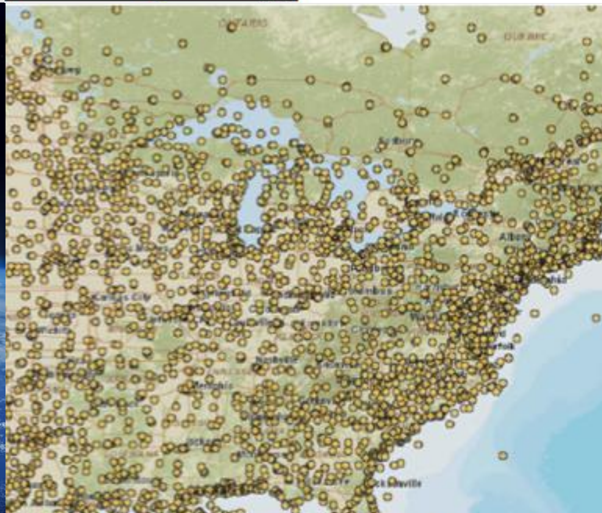
- Simulated data generated by physics-based models
  - Numerical Weather Prediction (NWP) models
  - General Circulation Models (GCM)
  - Regional Climate Models (RCM)

- Reanalysis data
  - Gridded data products from <u>data assimilation</u>: applies physical laws to observations

- Observation data
  - Satellite remote sensing data
  - In-situ data





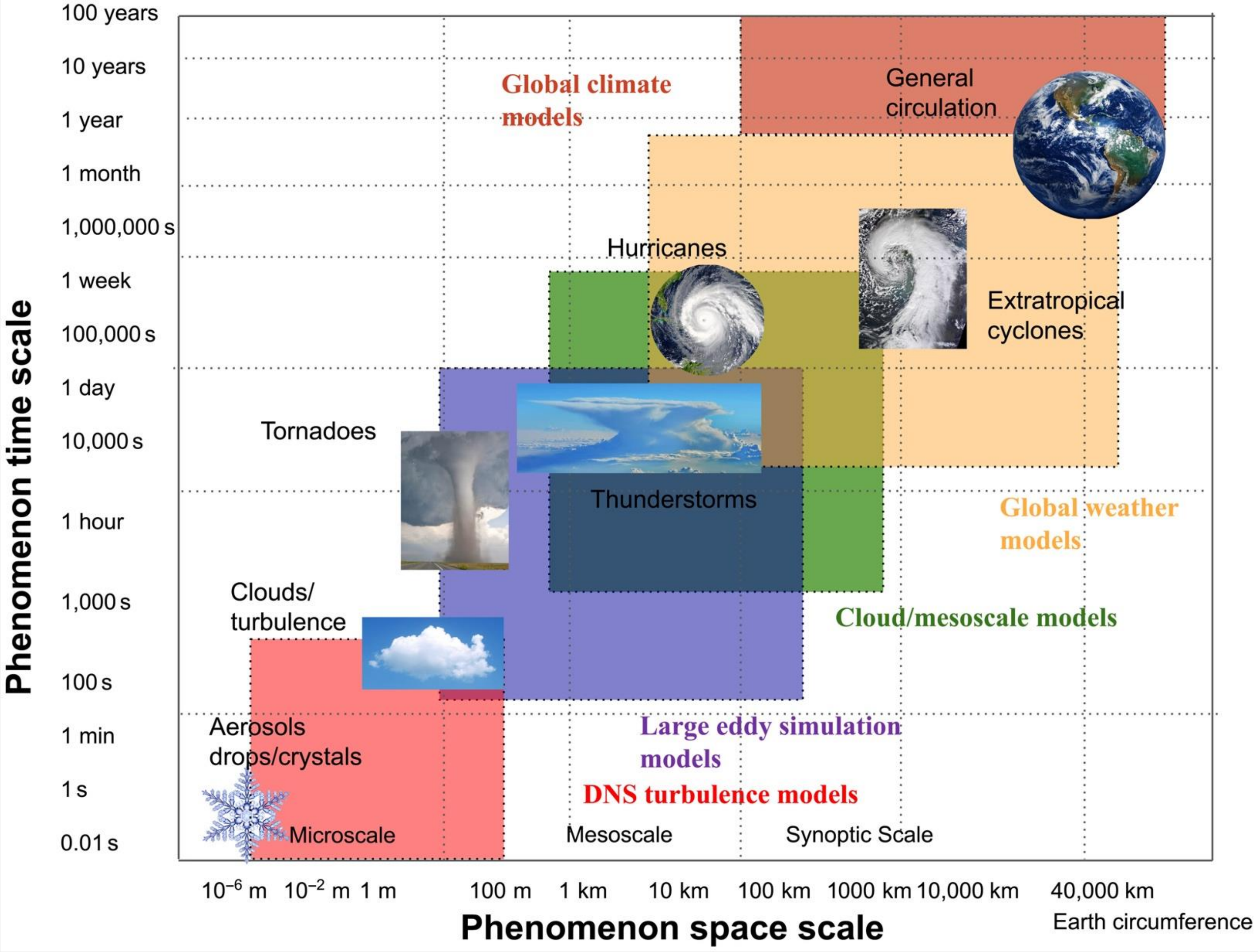


Downscaling climate model simulations

Global climate model simulations are coarser scale (in space and time) than needed for multiple tasks in:

- Climate change adaptation
- Climate change mitigation
- Projecting long-term impacts

Approach: Use generative AI to downscale climate model data to relevant scales



[Gettelman, et al., Science Advances, 2022]

AI Methods

- Semi-supervised, unsupervised, self-supervised learning
  - New methods for downscaling (super-resolution), interpolation of geospatial data
  - New pretext tasks for self-supervised learning, e.g., STINT [Harilal et al., 2024]
  - Regularization via multi-tasking over variables, lead-times
- Generative AI
  - VAE, Normalizing Flows
  - Diffusion models
  - Develop new generative downscaling methods, e.g., [Groenke et al., 2020]
- Learning under non-stationarity
  - Learn level of non-stationarity over time and space

AI Methods

- Semi-supervised, unsupervised, self-supervised learning
  - New methods for downscaling (super-resolution), interpolation of geospatial data
  - New pretext tasks for self-supervised learning, e.g., STINT [Harilal et al., 2024]
  - Regularization via multi-tasking over variables, lead-times

- Generative AI
  - VAE, Normalizing Flows
  - Diffusion models
  - Develop new generative downscaling methods, e.g., [Groenke et al., 2020]

- Learning under non-stationarity
  - Learn level of non-stationarity over time and space

Outline

What is self-supervised learning?

- A pretext task for temporal interpolation of geospatial data

What is generative deep learning?

- Normalizing flows for downscaling geospatial data

Implications for Climate Data Equity

Outline

What is self-supervised learning?

A pretext task for temporal interpolation of geospatial data

What is generative deep learning?

Normalizing flows for downscaling geospatial data

Implications for Climate Data Equity

Unsupervised Deep Learning

- Supervised DL. Prediction loss is a function of the label, y, and the network's output on input x.

Network output

$$f_W(x) = \hat{y}$$

Loss function

$$\mathcal{L}(\hat{y}, y)$$

- Unsupervised DL. Prediction loss is only a function of x, and the network's output on input x. There is no label, y.

Network output

$$f_W(x) = \hat{x}$$

Loss function

$$\mathcal{L}(\hat{x}, x)$$

Self-Supervised Approach to Unsupervised learning

Self-supervised learning

A state-of-the-art approach to (deep) unsupervised learning

Design a pretext task:

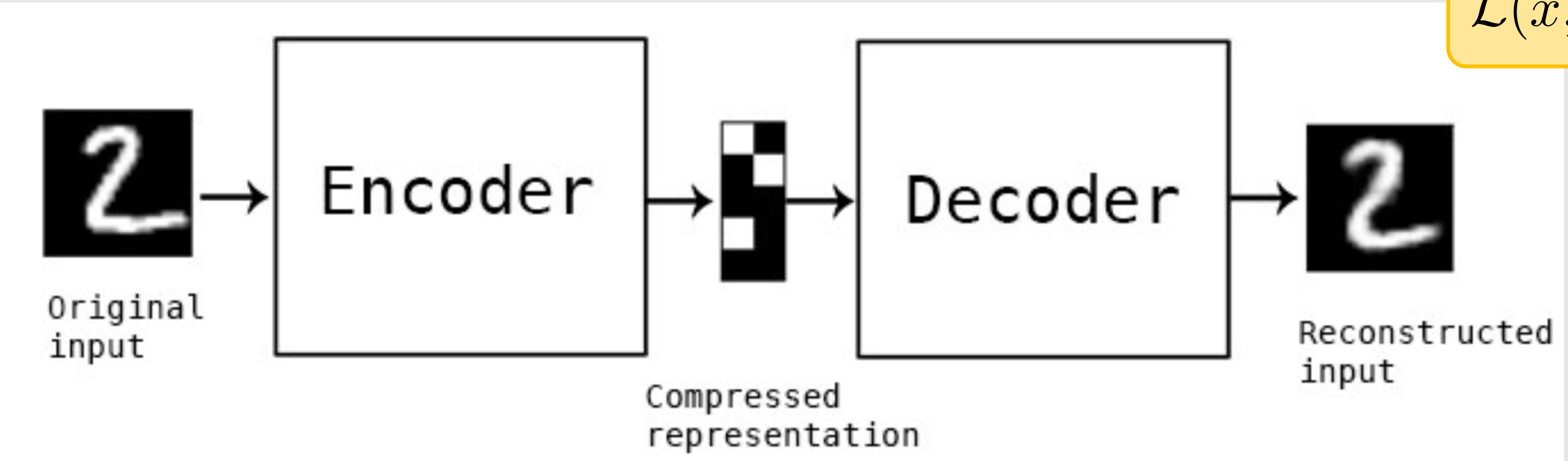
- Design a supervised learning task using only the available data.
- Train a model on this task such that,
- the learned features (or the learned posterior over a feature space) will be useful for another (down-stream) task.

Pretext Task: Example

Classic example of a pretext task: Autoencoder

- Train a neural network in an unsupervised way
  - Use the unlabeled data both as input, and to evaluate the output
- After training, the bottleneck layer will be a compact representation of the input distribution

$$\mathcal{L}(\hat{x}, x)$$



Outline

What is self-supervised learning?

A pretext task for temporal interpolation of geospatial data

What is generative deep learning?

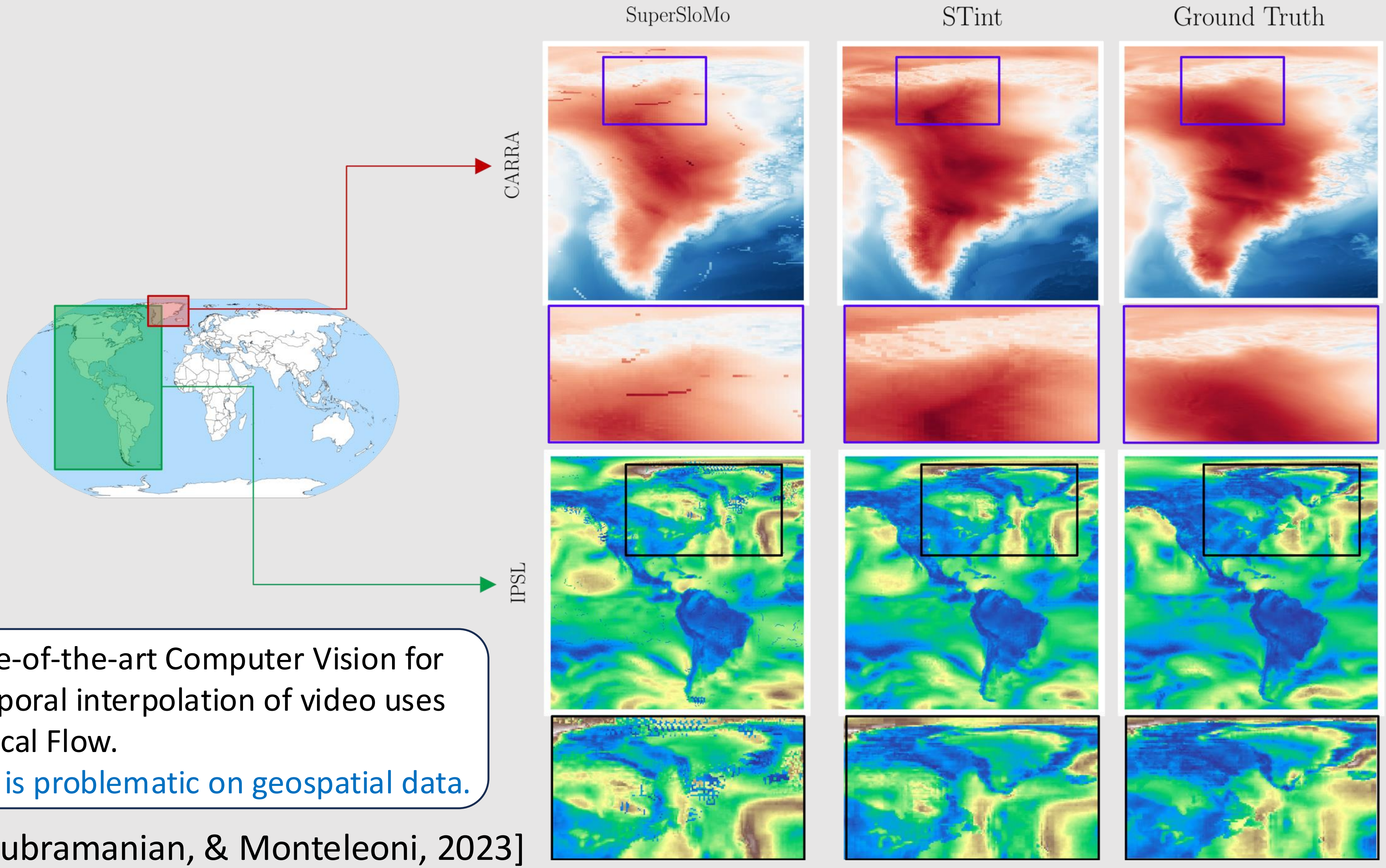
Normalizing flows for downscaling geospatial data

Implications for Climate Data Equity

Why we can’t just use existing AI algorithms

- Climate change applications involve geospatial data evolving with time
  - Observation data that has been gridded over the globe using data assimilation
  - Simulations output by physics-driven models (NWP, GCM, RCM)
- These are tensors of real-values over latitude, longitude, time,
  and possibly over multiple climatological variables
- Computer Vision algorithms for “spatiotemporal data,” rely on properties of **video data** that **do not generalize well to geospatial data**
  - e.g., depth, edges, and “objects”
  - vs. ephemeral patterns in fluids

STINT: Self-supervised Temporal Interpolation



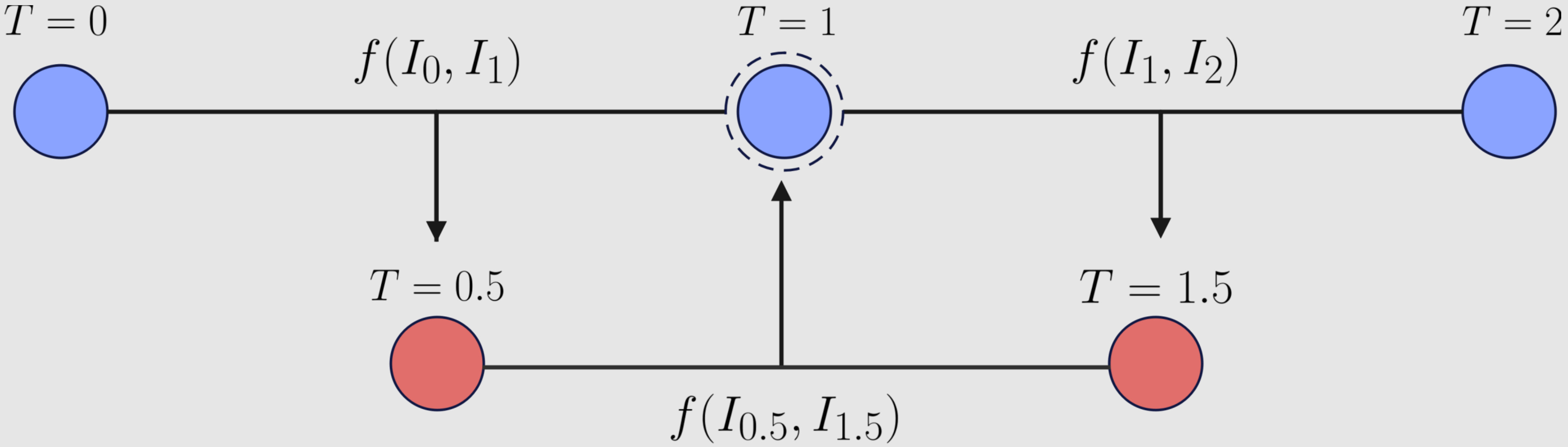
State-of-the-art Computer Vision for temporal interpolation of video uses Optical Flow.
This is problematic on geospatial data.

[Harilal, Hodge, Subramanian, & Monteleoni, 2023]

A pretext task for temporal downscaling

STINT: Self-supervised Temporal Interpolation for Geospatial Data

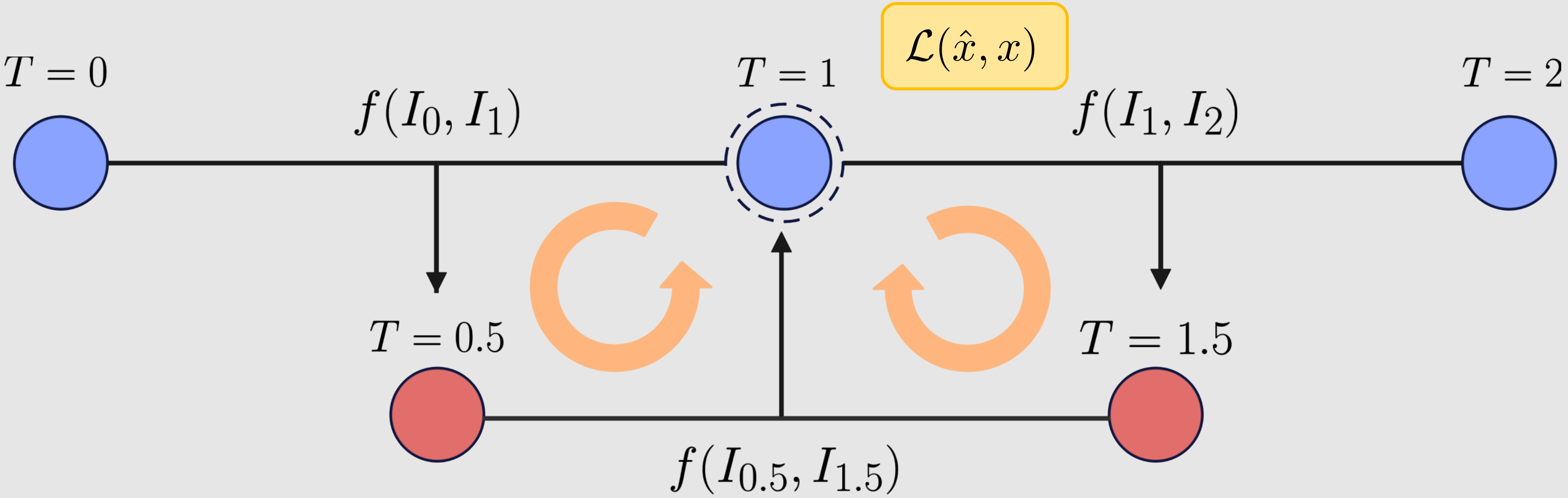
[Harilal, Hodge, Subramanian, & Monteleoni, 2023]



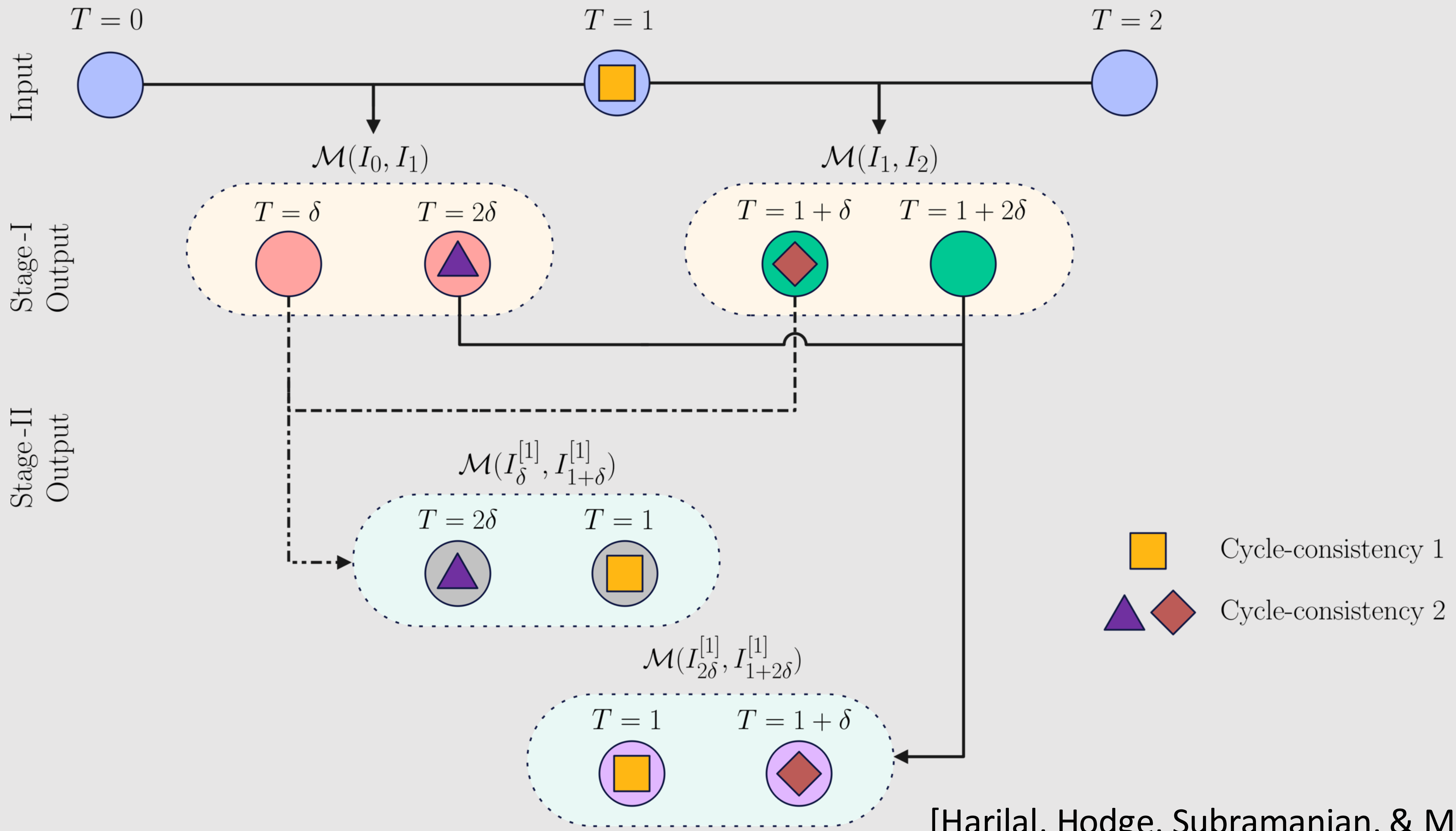
A pretext task for temporal interpolation

STINT: Self-supervised Temporal Interpolation for Geospatial Data

[Harilal, Hodge, Subramanian, & Monteleoni, 2023]

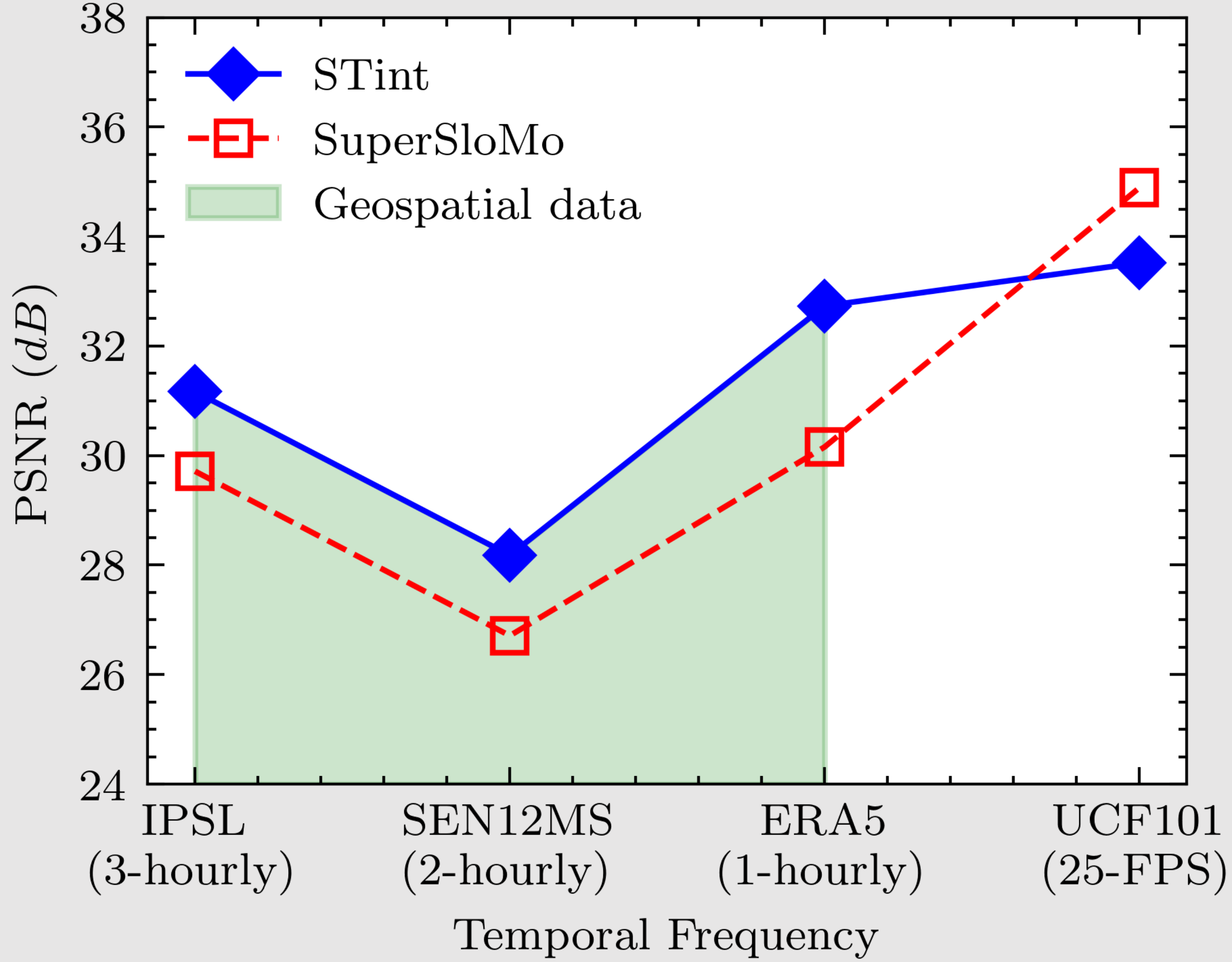


STINT: Self-supervised Temporal Interpolation



[Harilal, Hodge, Subramanian, & Monteleoni, 2023]

STINT: Self-supervised Temporal Interpolation



	$\frac{MSE}{Capacity}$ (↓)	PSNR (↑)	SSIM (↑)
ERA5 Solar			
Baseline	0.3086	25.238	0.623
SuperSloMo	0.0907	30.157	0.733
Proposed	**0.0561**	**32.731**	**0.792**
IPSL Wind			
Baseline	0.6206	24.097	0.619
SuperSloMo	0.4150	29.713	0.681
Proposed	**0.2904**	**31.167**	**0.713**
CARRA Temperature			
Baseline	0.5319	27.832	0.667
SuperSloMo	0.1604	30.276	0.724
Proposed	**0.0975**	**31.908**	**0.775**

[Harilal, Hodge, Subramanian, & Monteleoni, 2023]

Outline

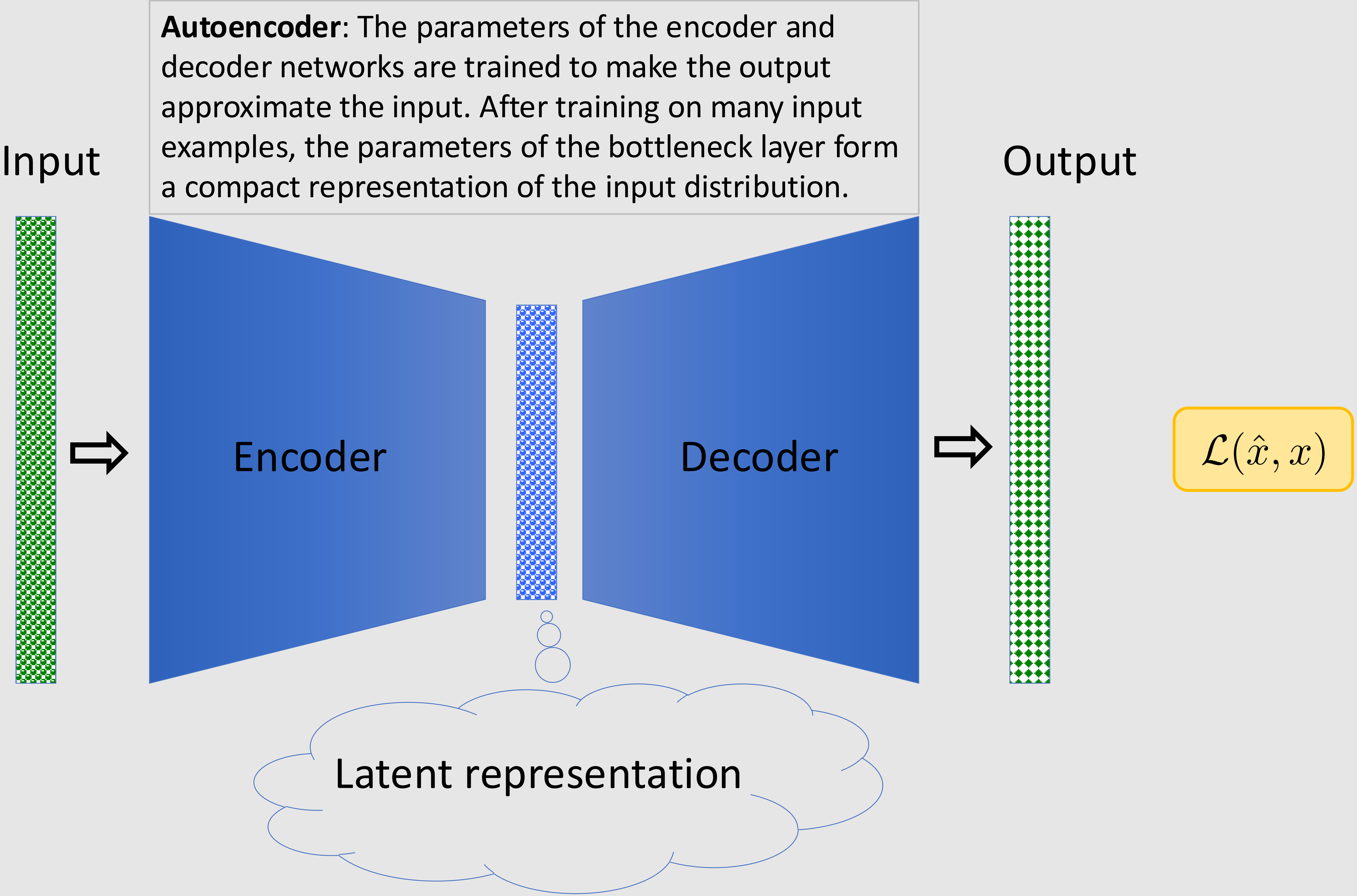
What is self-supervised learning?

A pretext task for temporal interpolation of geospatial data

What is generative deep learning?

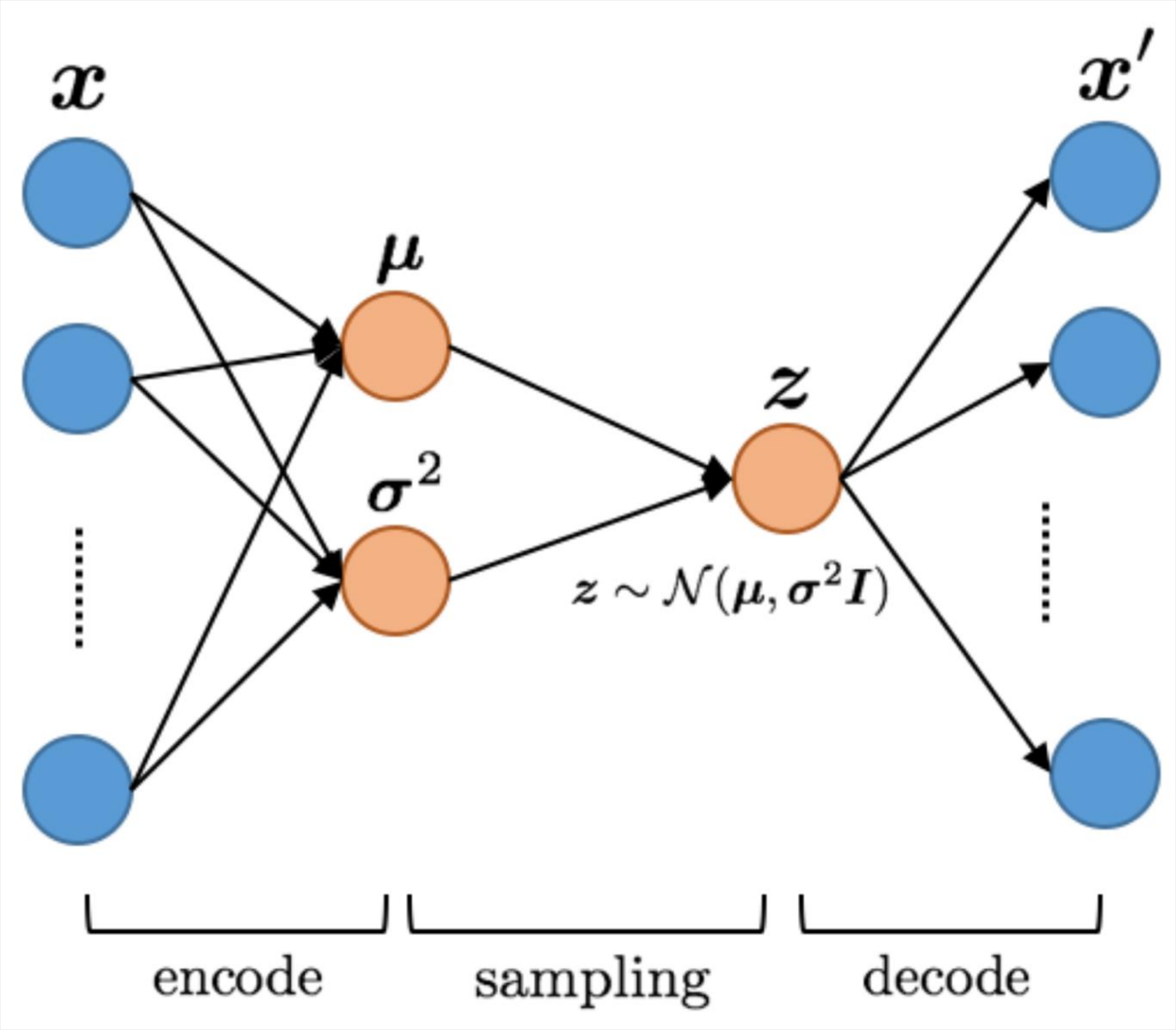
Normalizing flows for downscaling geospatial data

Implications for Climate Data Equity

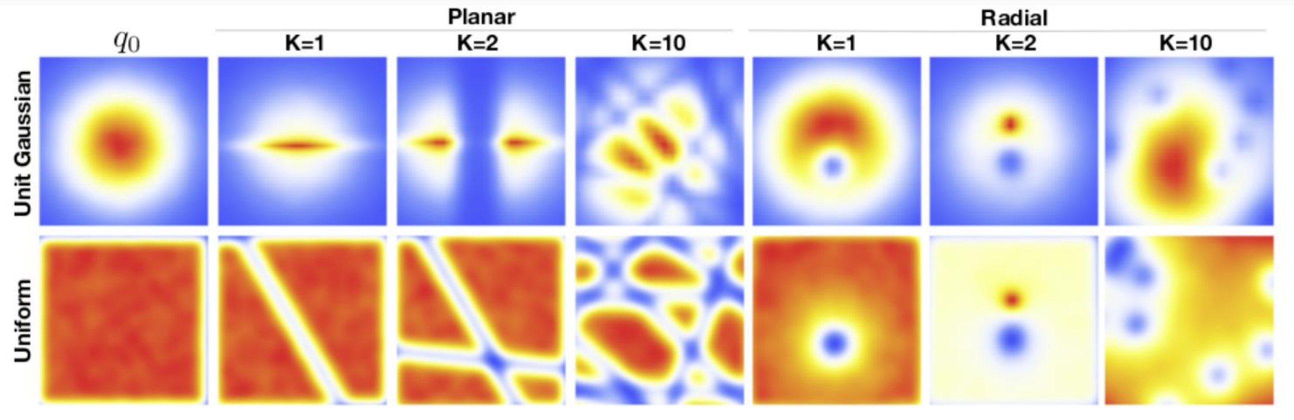

Autoencoder: The parameters of the encoder and decoder networks are trained to make the output approximate the input. After training on many input examples, the parameters of the bottleneck layer form a compact representation of the input distribution.
Input
Encoder
Decoder
Output
$\mathcal{L}(\hat{x}, x)$
Latent representation

Variational Autoencoder (VAE)

Learn a distribution over latent representations, instead of a single encoding



Normalizing Flows

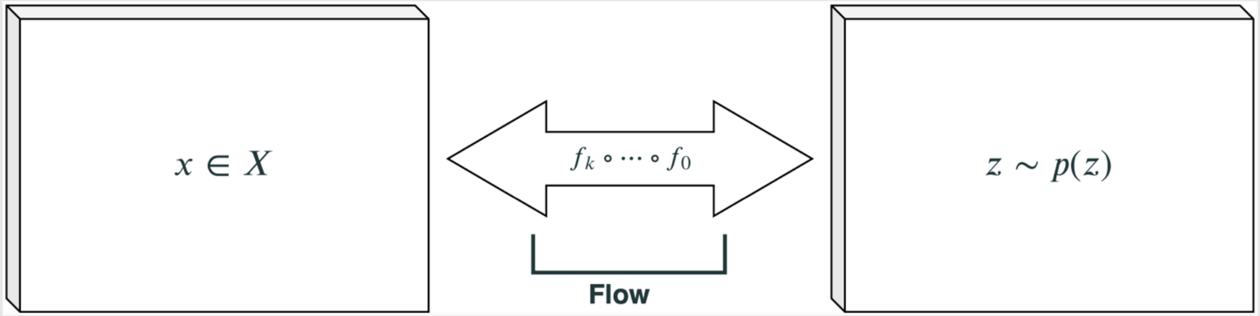


[Rezende & Mohamed, ICML 2015]

Can be viewed as extension of VAE beyond Gaussian assumption on latent space

Learn a series of invertible transformations, $\{f_i\}$, from a simple prior on latent space, Z, to allow for more informative distributions on the latent space:

$$z_k = f_k \circ f_{k-1} \circ \cdots \circ f_1(z_0)$$



Outline

What is self-supervised learning?

A pretext task for temporal interpolation of geospatial data

What is generative deep learning?

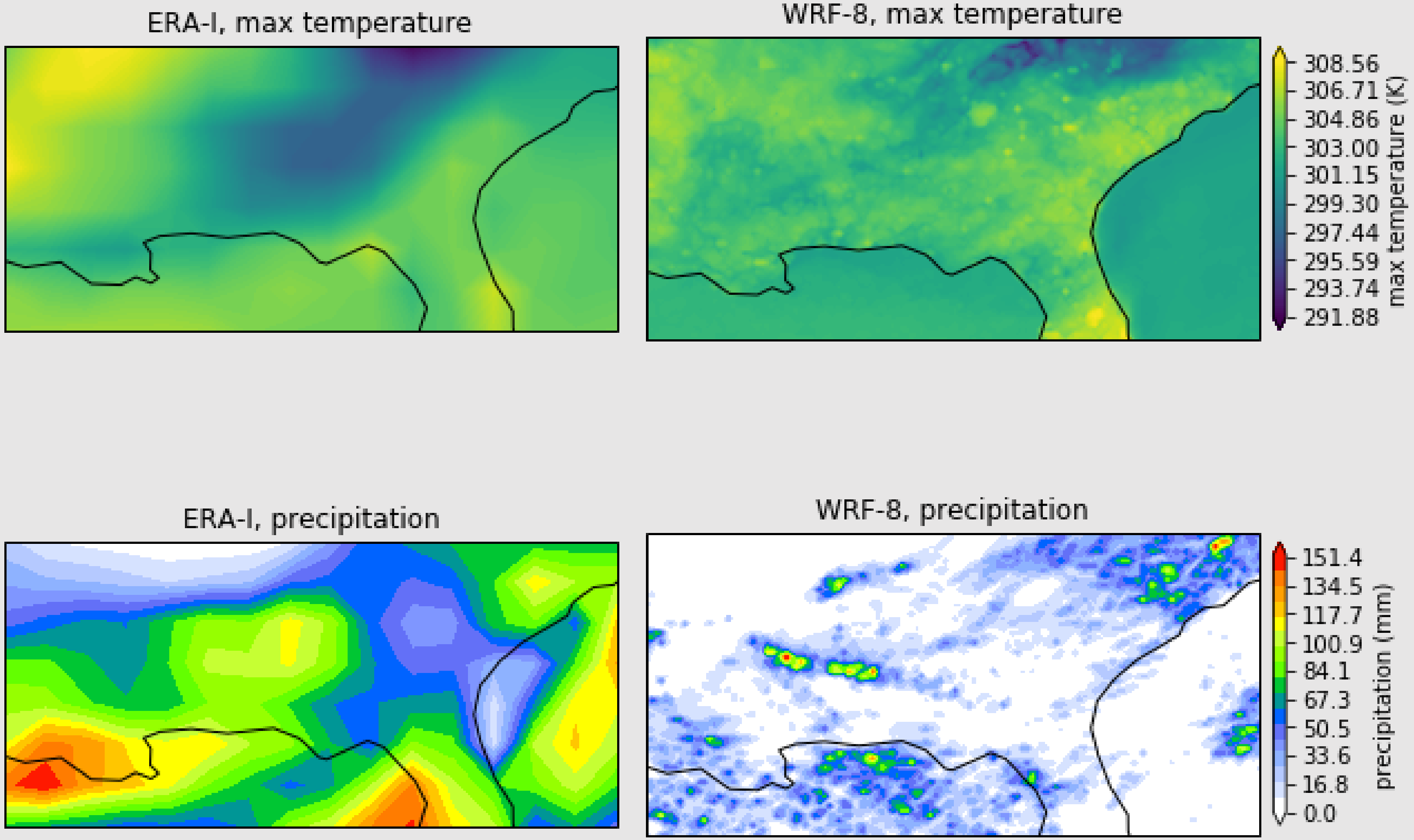
Normalizing flows for downscaling geospatial data

Implications for Climate Data Equity

Normalizing Flows: Application to Spatial Downscaling

[Groenke, Madaus, & Monteleoni, Climate Informatics 2020]

ERA: reanalysis data, 1° resolution; WRF: numerical weather model prediction, $\frac{1}{8}$° resolution



Downscaling as Domain Alignment

- Domain alignment task: given random variables X, Y, learn a mapping f: X → Y such that, for any $x_i \in X$ and $y_i \in Y$,

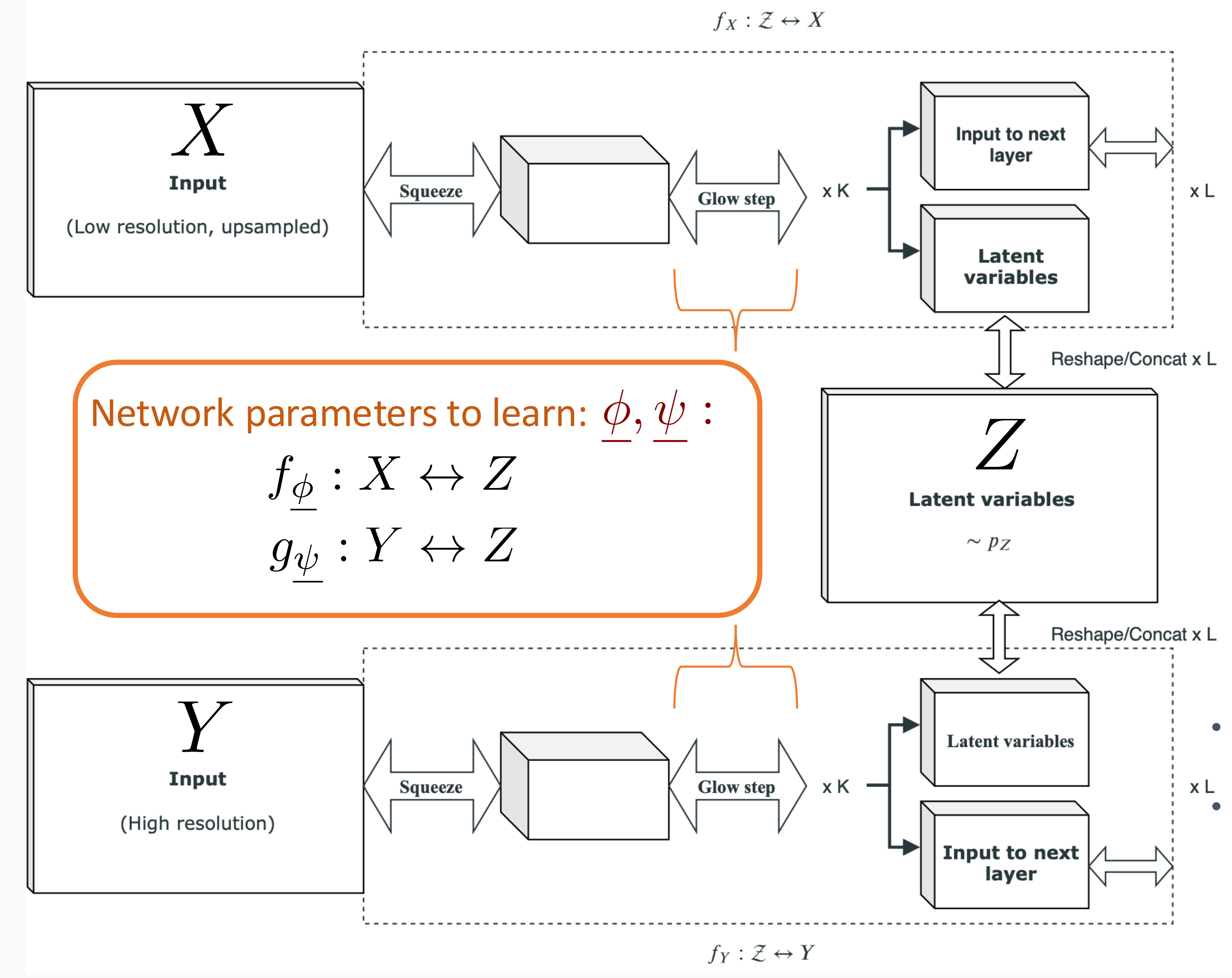
$$f(x_i) \sim P_Y \qquad\qquad f^{-1}(y_i) \sim P_X$$

- Downscaling as domain alignment
  - Given i.i.d. samples at low resolution (X) and high-resolution (Y)
  - Learn the joint PDF over X and Y by assuming conditional independence over a shared latent space Z,

$$P_{XY}(x,y) = \int_{z \in Z} P_{XYZ}(x,y,z)dz \;= \int_{z \in Z} P(x|z)P(y|z)P_Z(z)dz$$

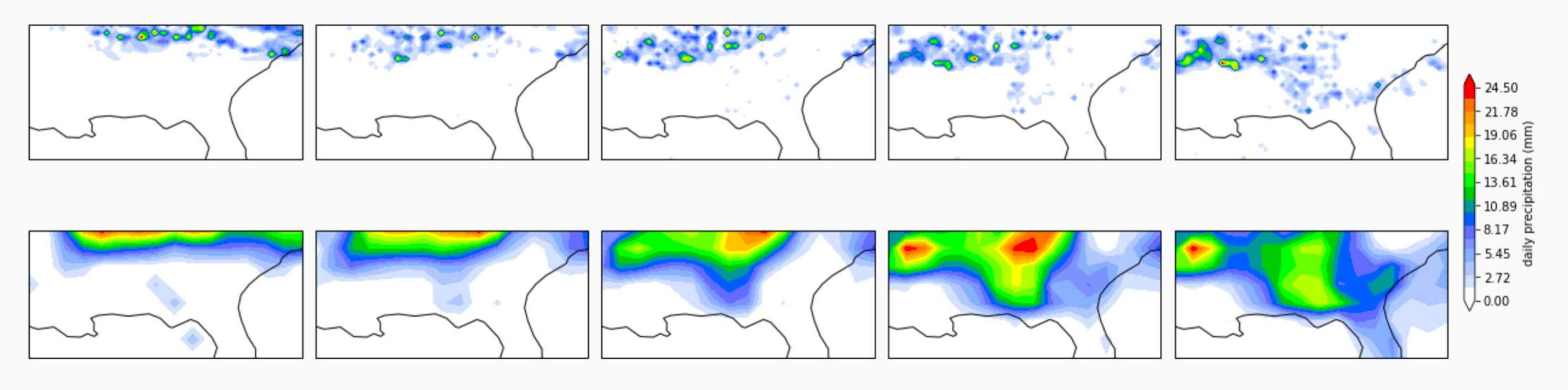
  - Model $P(x|z), P(y|z)$ using AlignFlow [Grover et al. 2020]
    - Starting with a simple prior on P_Z, learn normalizing flows
    - No pairing between x and y examples needed!

ClimAlign architecture



- Architecture follows AlignFlow [Grover et al., 2020]
- Normalizing flow: Glow [Kingma & Dhariwal, 2018]

ClimAlign: Unsupervised, generative downscaling



General downscaling technique via domain alignment with normalizing flows
[AlignFlow: Grover et al., AAAI 2020][Glow: Kingma & Dhariwal, NeurIPS 2018]

- **Unsupervised**: do not need paired maps at low and high resolution
- **Generative**: can sample from posterior over latent representation OR sample conditioned on a low (or high!) resolution map
- **Intepretable**, e.g., via interpolation

[Groenke, Madeus, & Monteleoni, Climate Informatics 2020]

Summary & Outlook

- A pretext task for temporal downscaling of geospatial data
  - Works best when input data is spatially aligned
- Normalizing flows for spatial downscaling of geospatial data
  - Does not require temporal alignment of the coarse and fine scale data
  - Works best when data is spatially aligned

- Is there one pretext task for downscaling in both space and time?
  - Does it provide features that are useful for other downstream tasks?

Other generative DL projects

- Landry, D., Charantonis, A., & Monteleoni, C. (2024). Leveraging deterministic weather forecasts for in-situ probabilistic temperature predictions via deep learning. Monthly Weather Review.

- Generative downscaling for solar and wind energy planning

- Ensemble generation via climate model emulation with diffusion training

Outline

What is self-supervised learning?

A pretext task for temporal interpolation of geospatial data

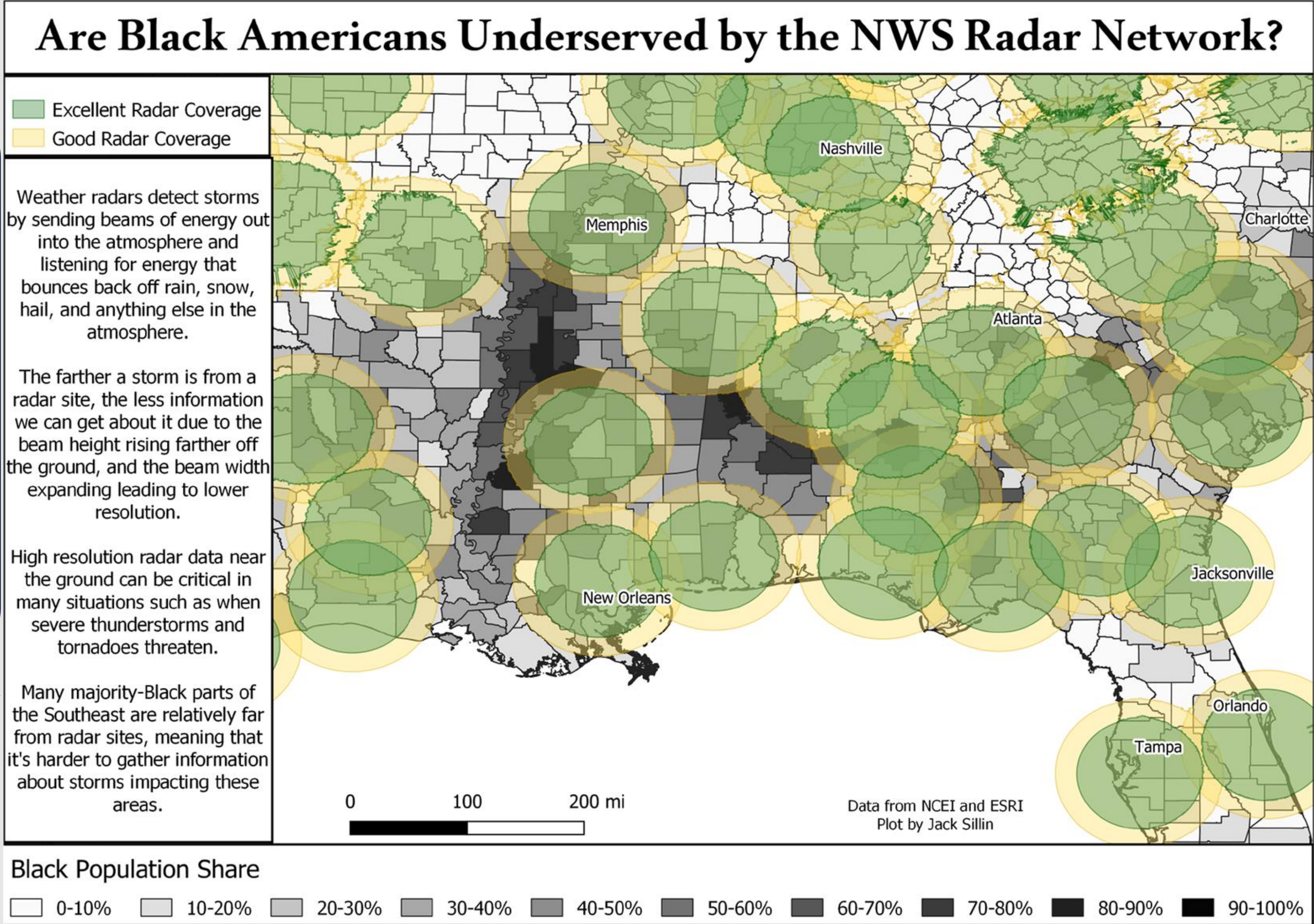
What is generative deep learning?

Normalizing flows for downscaling geospatial data

Implications for Climate Data Equity

"Many majority-Black parts of the Southeast [USA] are relatively far from radar sites, meaning that it's harder to gather information about storms impacting these areas."

Credit: Jack Sillin, in [McGovern et al., Environmental Data Science, 2022]



AI for Climate Data Equity

- Train models in high-data regions and apply them in low-data regions
  - Can evaluate them against supervised learning models in high-data regions
  - Can fine-tune them using the limited data in the low-data regions
- Contribution to climate data equity
  - Local scales (e.g. legacy of environmental injustice in USA)
  - Global scales:
    - Global North historically emitted more carbon; Meanwhile there's typically more data there
    - Global South is suffering the most severe effects of the resulting warming

Climate and Machine Learning Boulder (CLIMB)

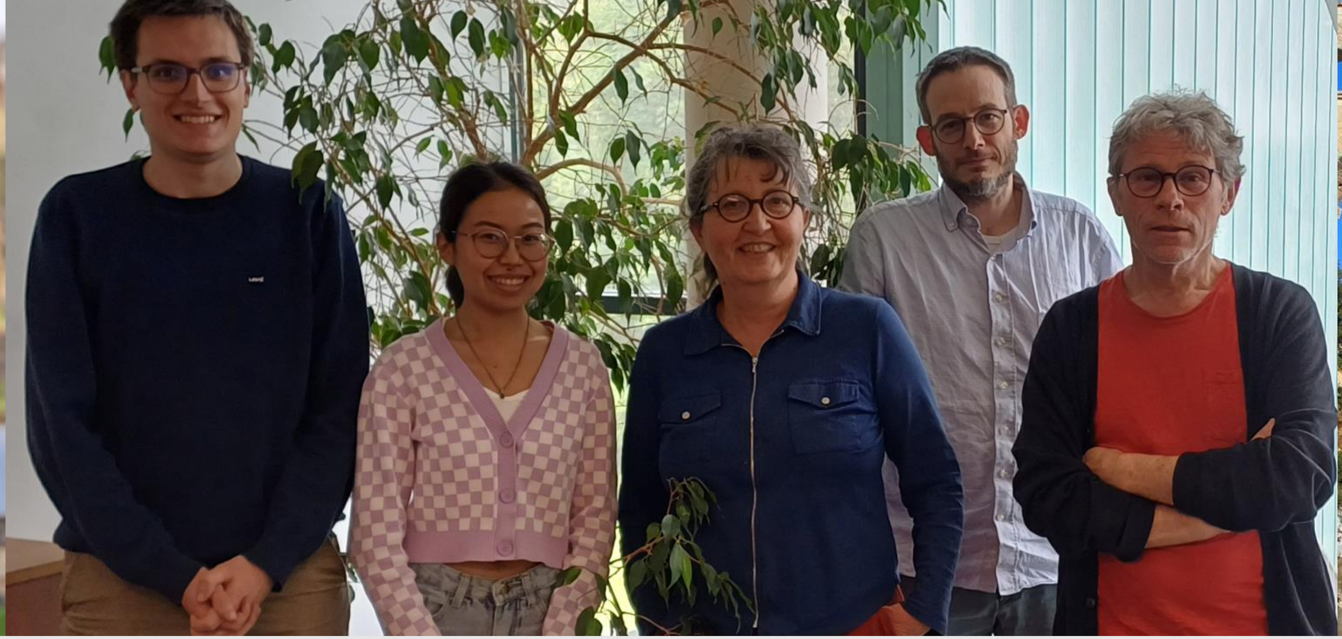
Boulder
NSF
Climate Change AI
futurearth
Research. Innovation. Sustainability.

Thank you!

And many thanks to:

Anastase Charantonis, *INRIA Paris*
Guillaume Couairon, *INRIA Paris*
Graham Clyne, *INRIA Paris*
Brian Groenke, *Alfred Wegener Institute*
Nidhin Harilal, *University of Colorado Boulder*
David Landry, *INRIA Paris*
Christian Lessig, *ECMWF*

eDF
METEO FRANCE
Choose France™
NASA
Inria


AI Research for Climate Change and Environmental Sustainability (ARCHES)

Environmental Data Science Innovation & Inclusion Lab

A national accelerator linking data, discovery, & decisions

ESIIL

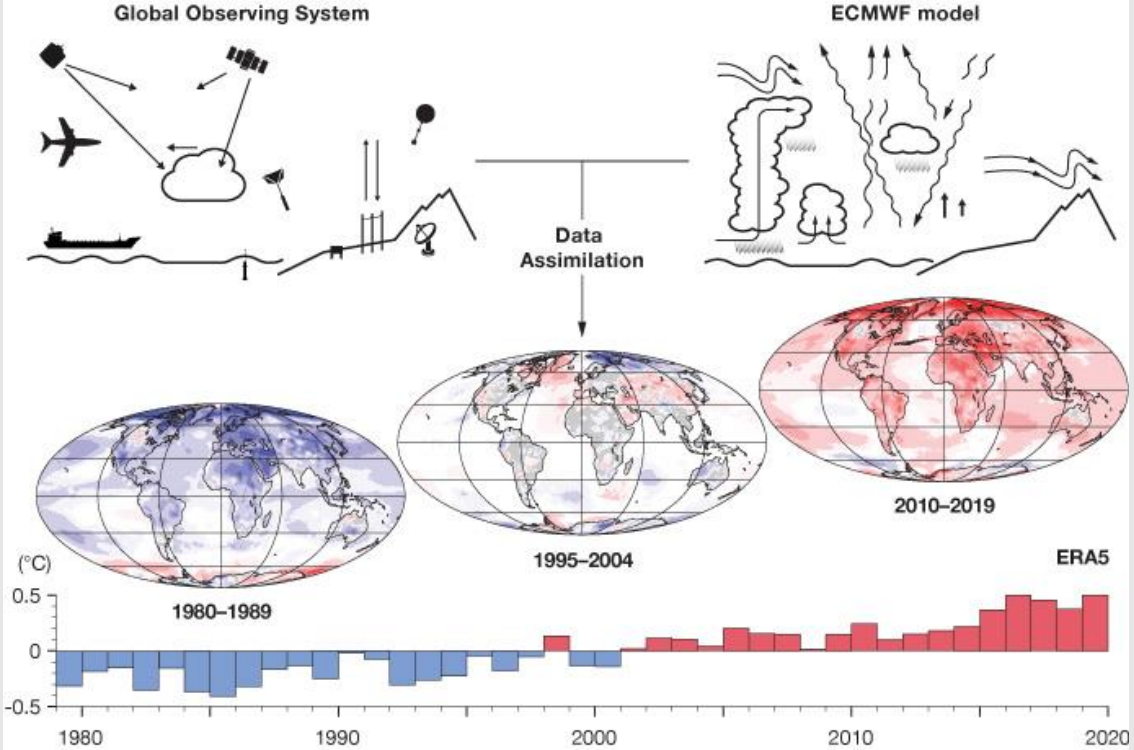
NSF's newest data synthesis center,
hosted by the University of Colorado Boulder & CIRES,
with key partners CyVerse & the University of Oslo

Bonus slides

Revolution in AI for weather forecasting

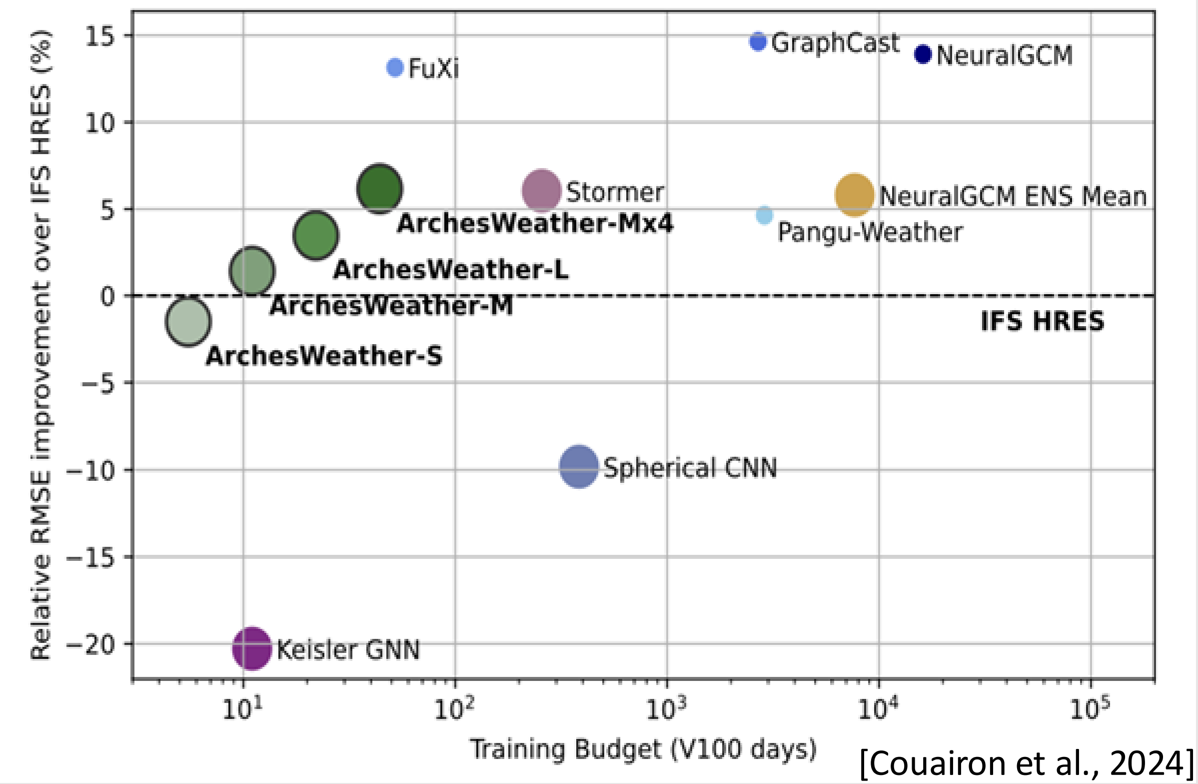
Since 2022, a variety of deep learning models have shown weather forecasting performance comparable or **BETTER** than numerical weather prediction (NWP), the previous SOTA.

- Training data: ERA5, a reanalysis data set produced by data assimilation



[European Center for Medium Range Weather Forecasting (ECMWF) website]

- Training task: auto-regression: forecasting 6-24 hours ahead
- Rolled-out to forecast 7-10 days ahead

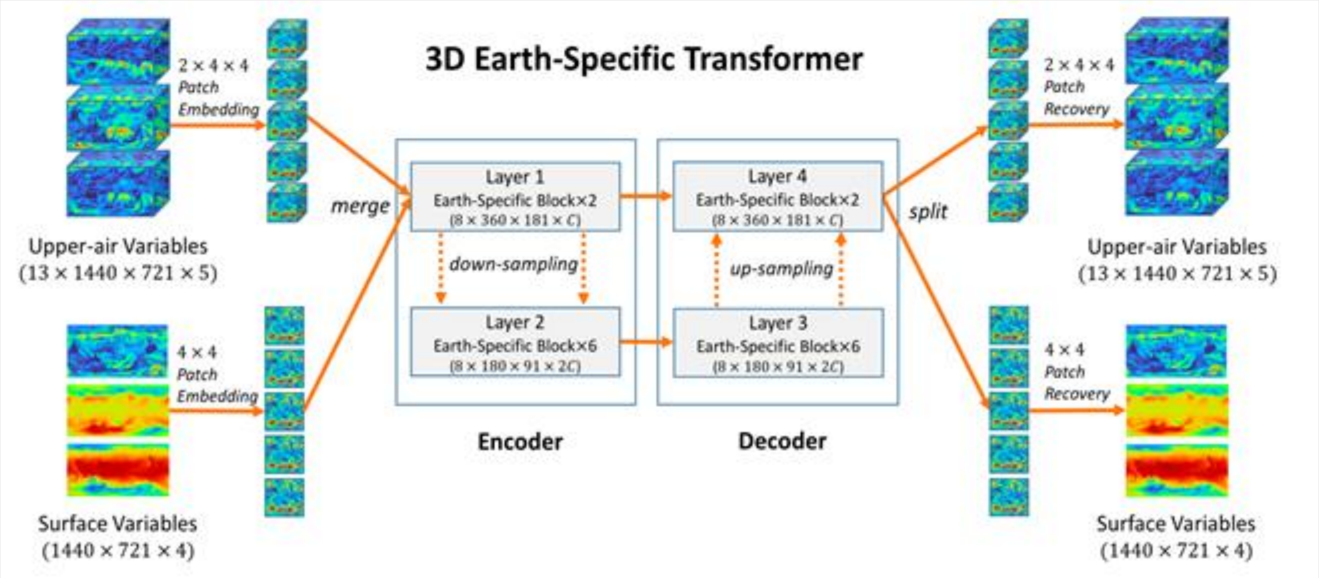


[Couairon et al., 2024]

Lighter-weight AI weather forecasting

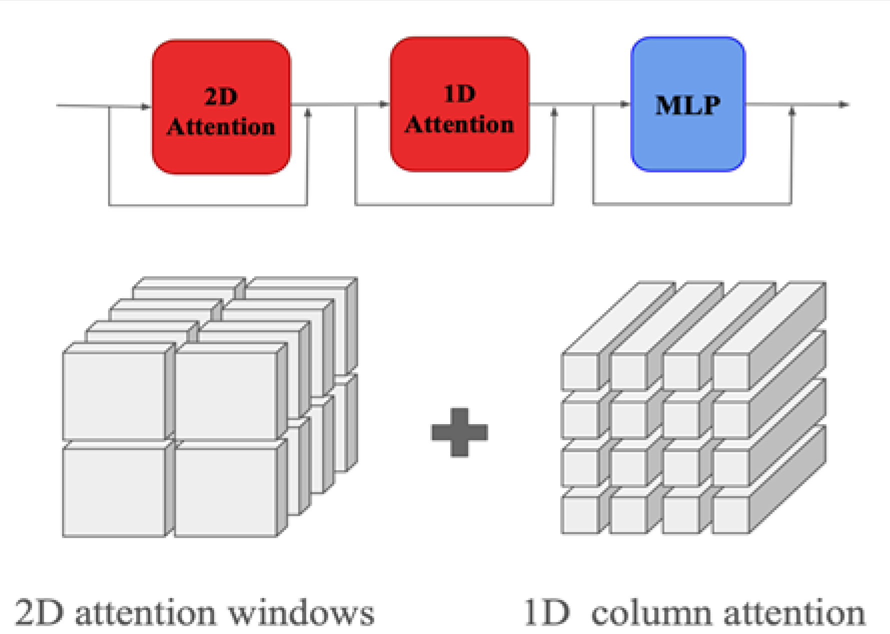
[Couairon et al., ArchesWeather: an efficient AI weather model at 1.5º resolution, ICML 2024 workshop on Earth System Modeling]

Pangu-Weather: A 3D High-Resolution System for Fast and Accurate Global Weather Forecast, Bi et al., Nature 2023



<u>New in ArchesWeather:</u>

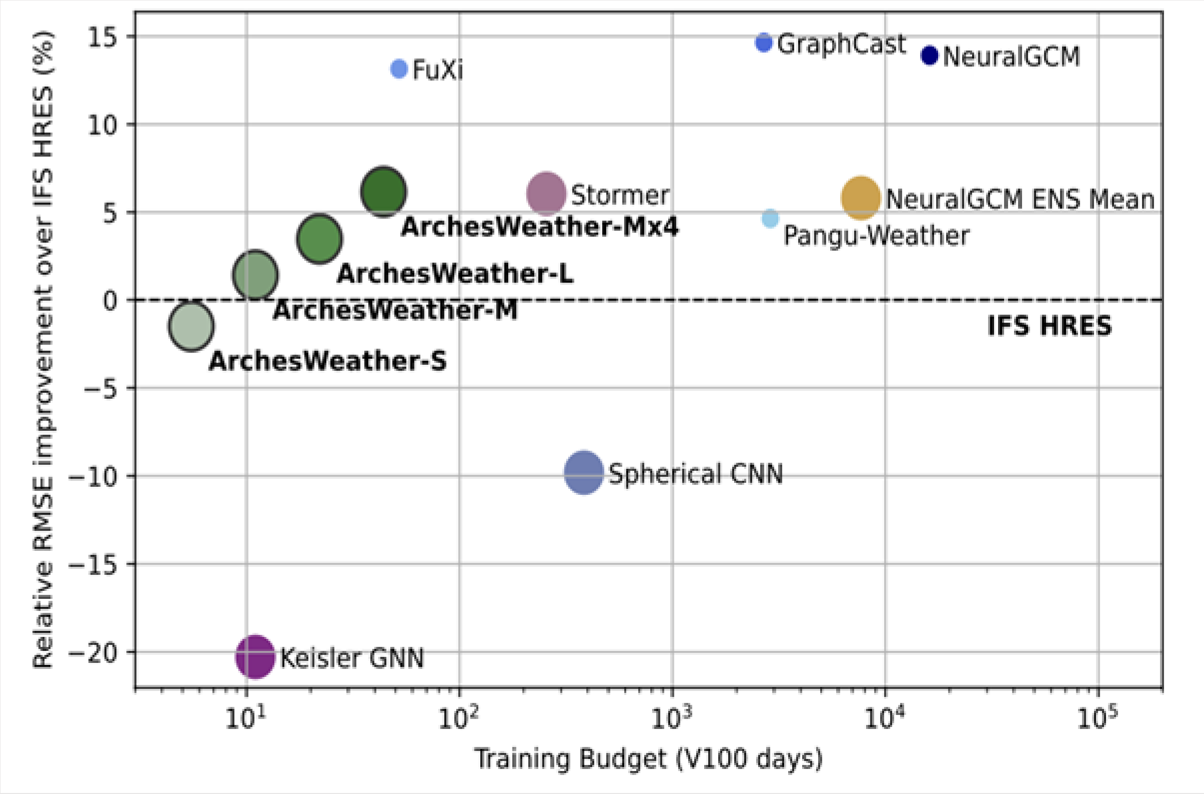
- Train at courser data resolution
- Replace 3D attention with:

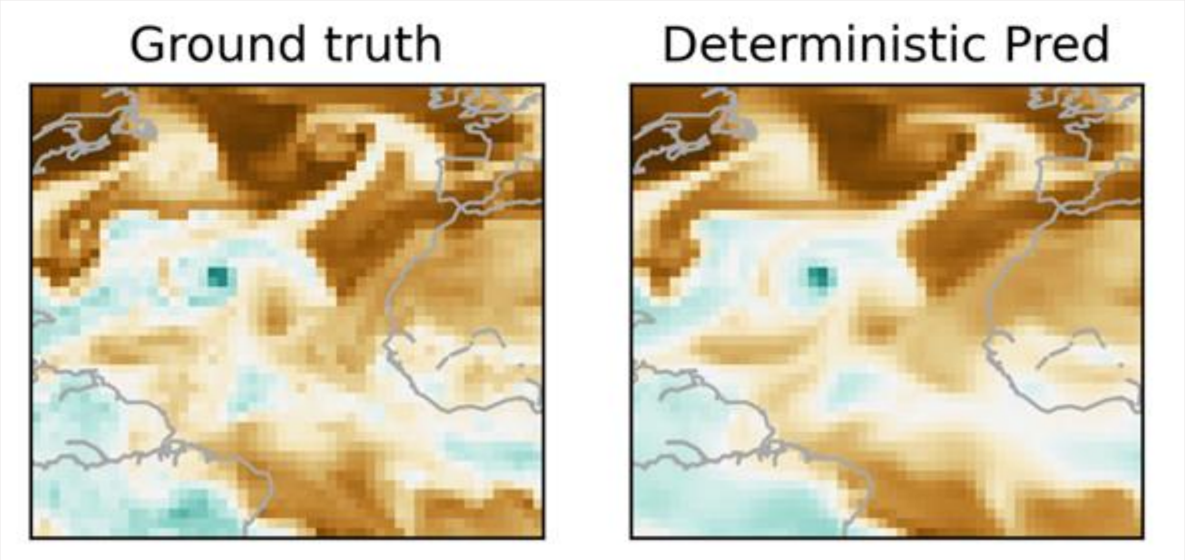


latitude + longitude

altitude

Lighter-weight AI weather forecasting





24h lead time Q700 forecast
init date: 28 sept 2019

- Deterministic prediction shows unrealistic smoothing
- Try ensemble generation via diffusion training
- Goal: each sampled member should be more physical

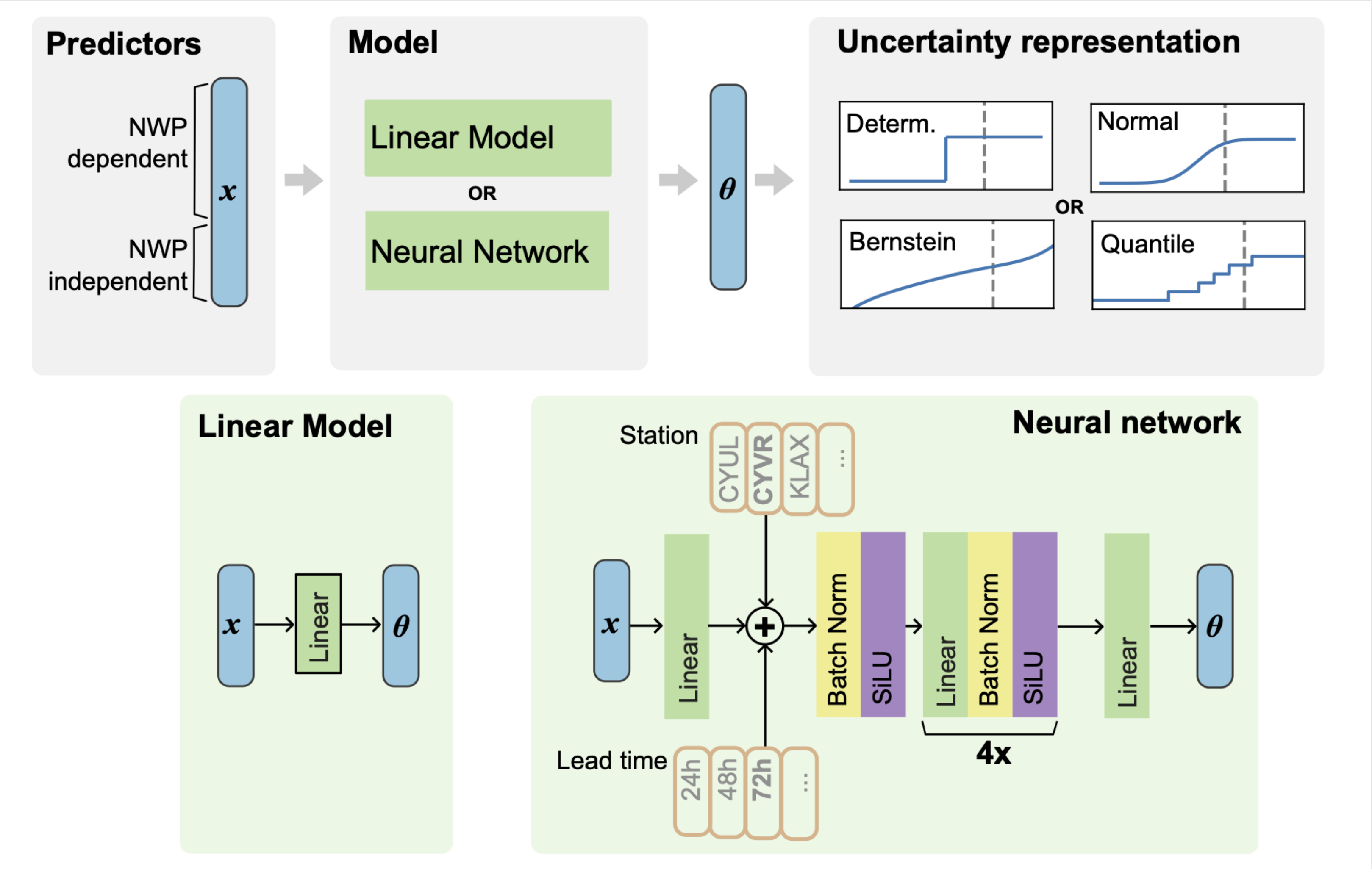
Generative AI for weather and climate

Ensemble forecast generation

- Multivariate emulation of kilometer-scale numerical weather predictions with generative adversarial networks : a proof-of-concept. C. Brochet, L. Raynaud, N. Thome, M. Plu et C. Rambour, *Artificial Intelligence for the Earth Systems, 2023*
- GenCast: Diffusion-based ensemble forecasting for medium-range weather, Price et al., *2023*
- Leveraging deterministic weather forecasts for in-situ probabilistic temperature predictions via deep learning. David Landry, Anastase Charantonis, and Claire Monteleoni. *Monthly Weather Review, 2024*
- ArchesWeather: an efficient AI weather model at 1.5º resolution. Guillaume Couairon, Anastase Charantonis, Christan Lessig, and Claire Monteleoni. *In preparation. Preliminary results in ICML 2024 workshop on Earth System Modeling*
- Diffusion-based ensemble generation for emulating climate models at decadal time scales. Graham Clyne, Guillaume Couairon, Anastase Charantonis, Guillaume Gastinau, Juliette Mignot, and Claire Monteleoni. *Work in progress*

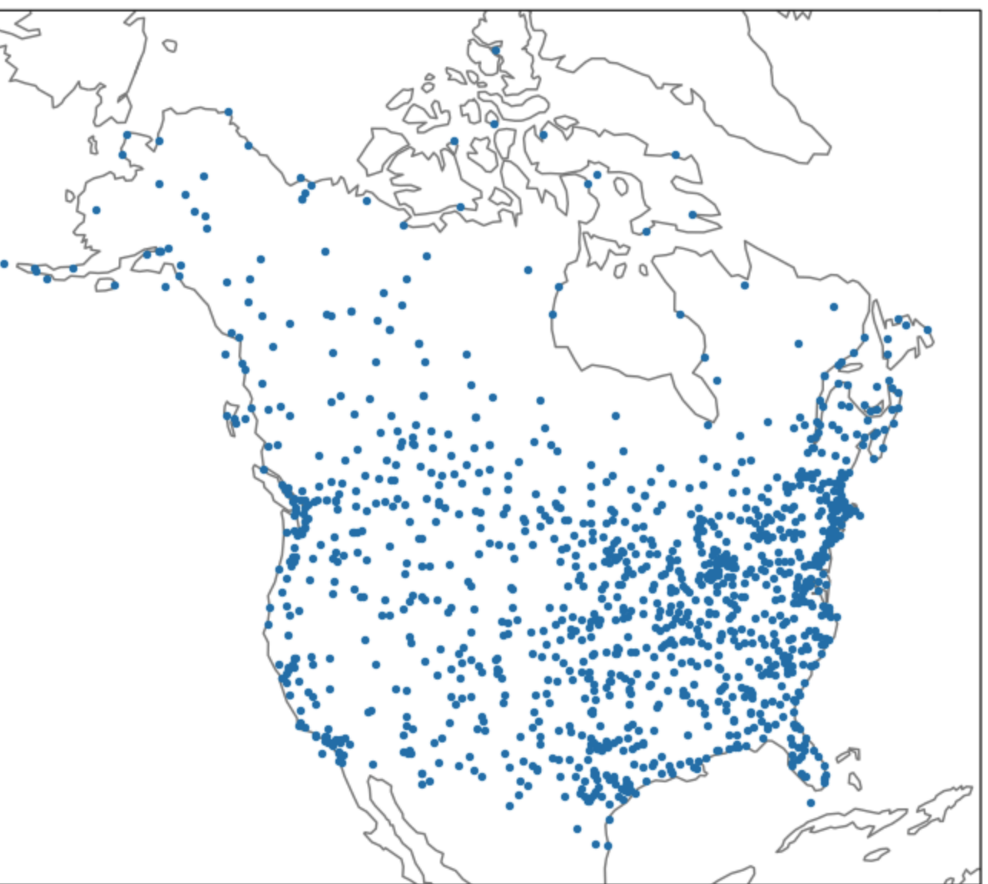
Probabilistic ensemble generation from a single forecast

[Landry, Charantonis & Monteleoni. *Monthly Weather Review, 2024*]



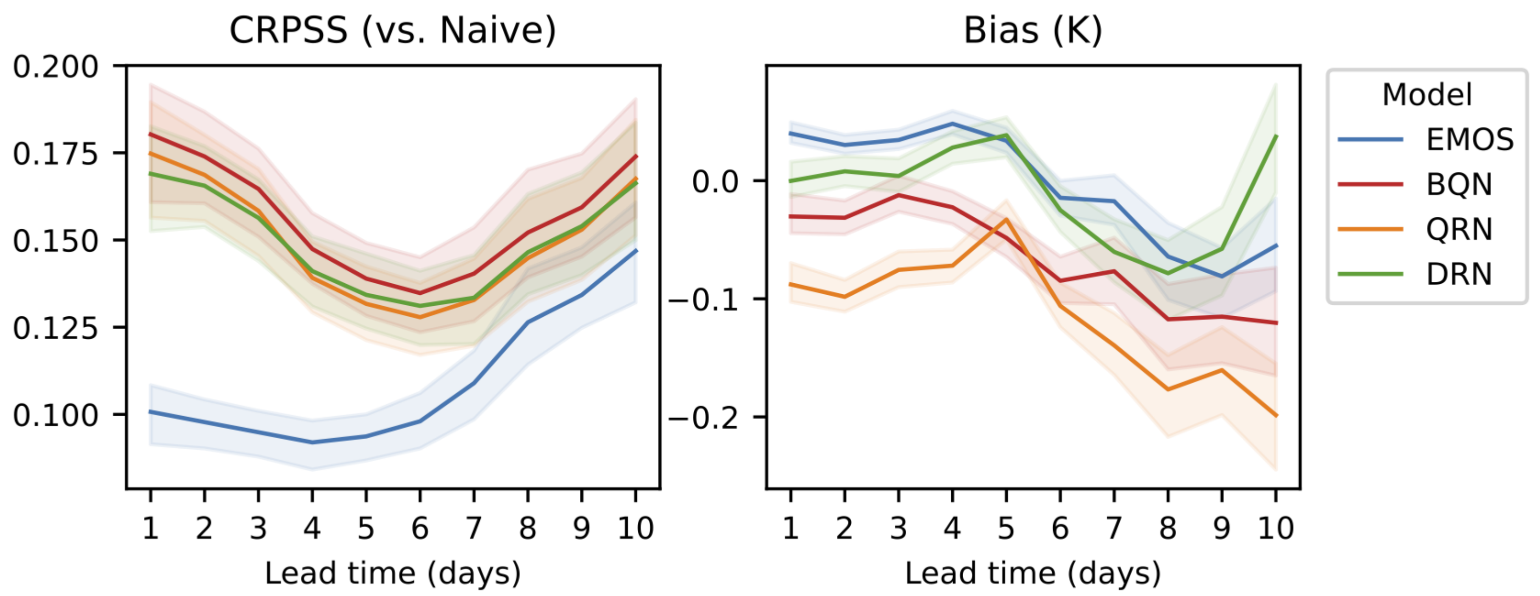
Probabilistic ensemble generation from a single forecast [Landry, Charantonis & Monteleoni. *Monthly Weather Review, 2024*]

Data from METAR: 1066 weather stations



Probabilistic prediction skill score

Mean forecast error



ArchesWeather with generative training

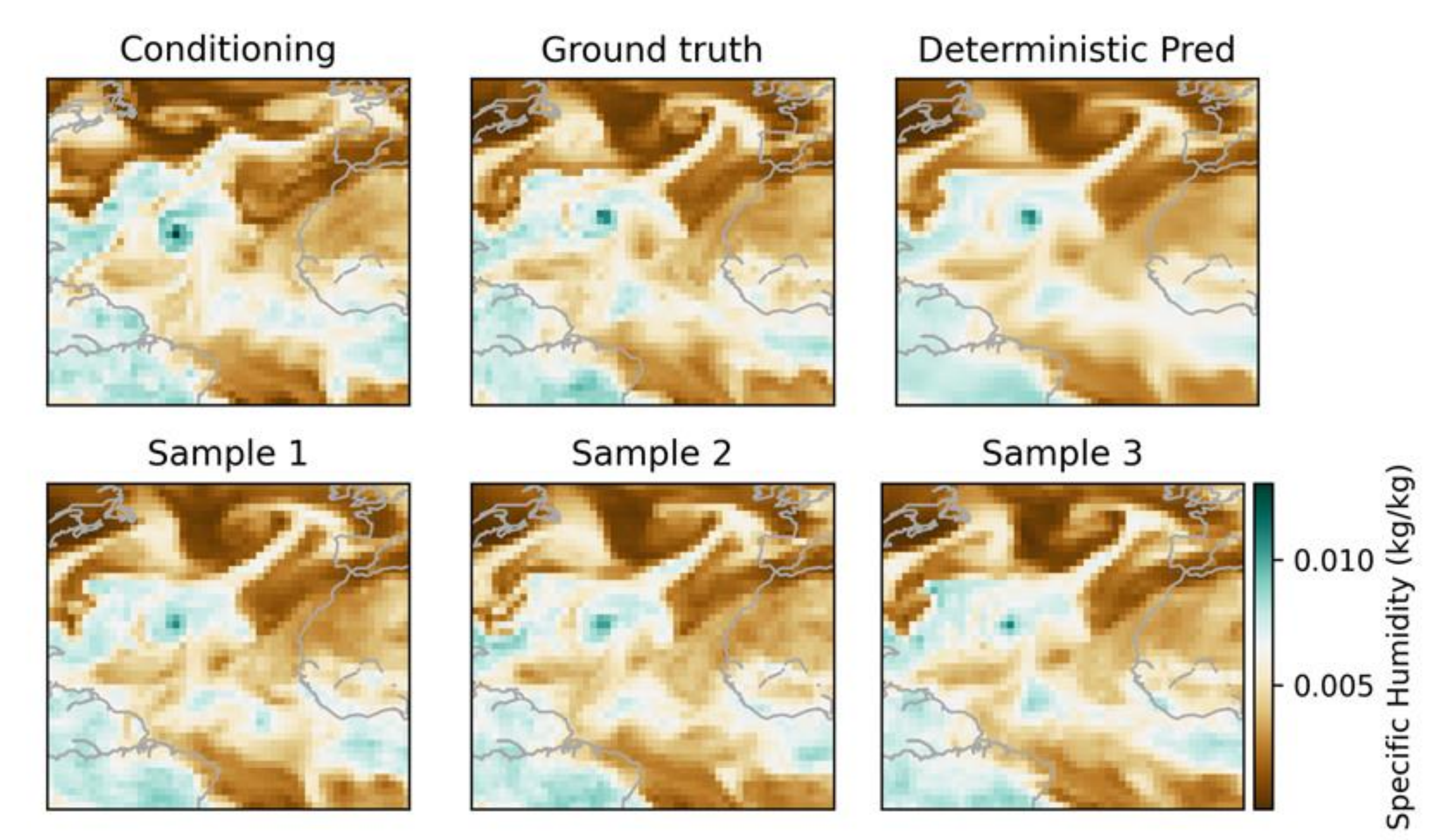
[Couairon et al., ArchesWeather: an efficient AI weather model at 1.5º resolution, manuscript]

<u>Backbone</u>: ArchesWeather variant of Swin U-Transformer; 44M parameters

<u>Trained</u> with Flow Matching

<u>Inference</u> done with 25 sampling steps per 24 hours

<u>Roll-outs</u> done via one sample each 24 hours. Input this output into trained model and repeat.

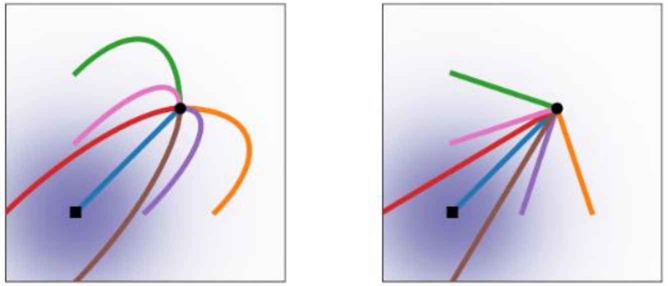


Flow Matching vs. Diffusion

FLOW MATCHING FOR GENERATIVE MODELING

Yaron Lipman[1,2] **Ricky T. Q. Chen**[1] **Heli Ben-Hamu**[2] **Maximilian Nickel**[1] **Matt Le**[1]
[1]Meta AI (FAIR) [2]Weizmann Institute of Science

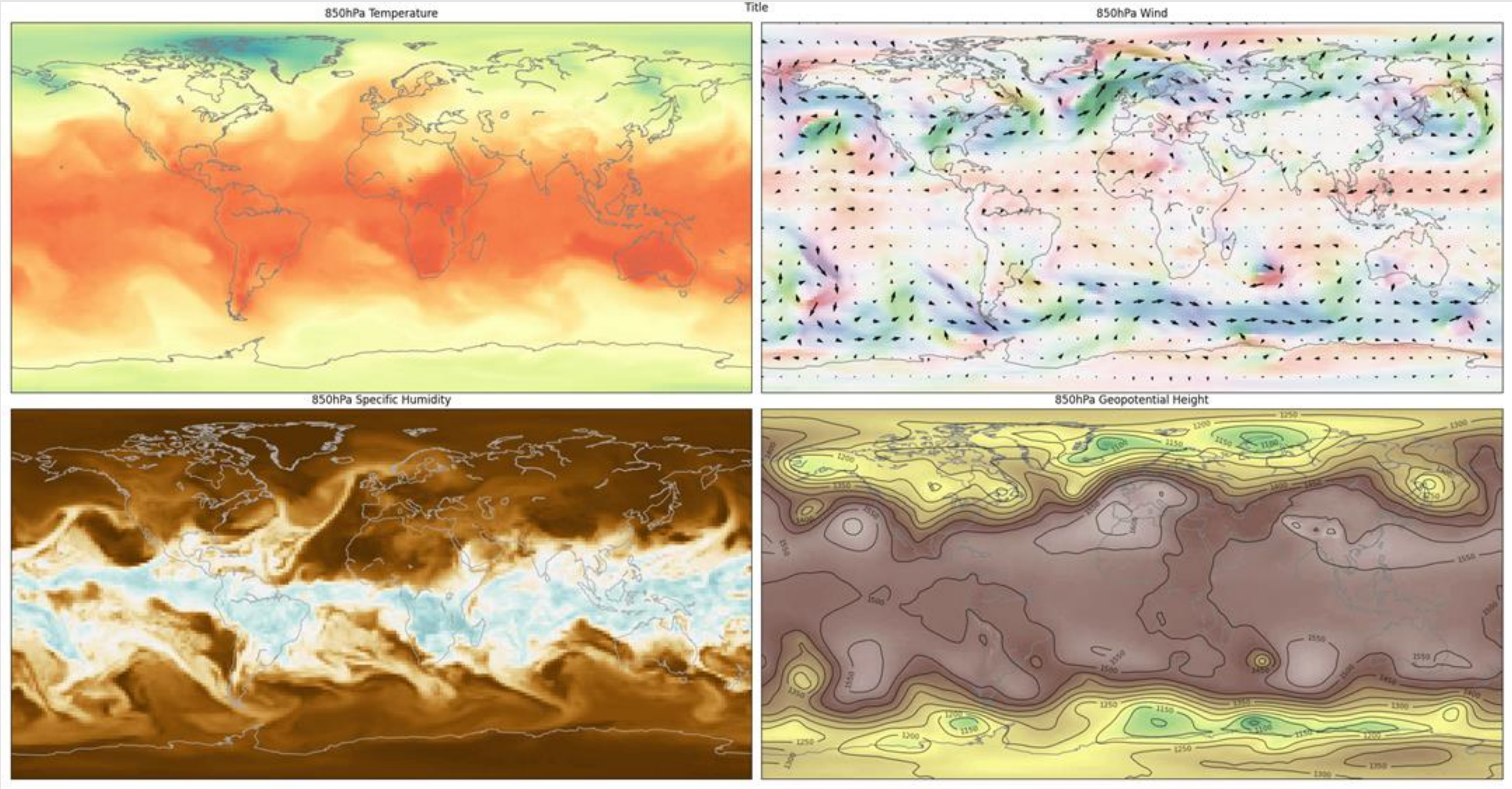
<u>Intuition</u>: Make straighter paths between data samples and noise samples



Diffusion OT

<u>Results:</u> Straighter paths in the probability flows, allowing us to sample in fewer diffusion steps
→ Speed-ups

A sample trajectory



Ensemble member diversity; lead-time 10 days