

AI Research
for Climate Change and Environmental Sustainability

Claire Monteleoni

INRIA Paris

University of Colorado Boulder

Inria

Choose France™



December 2021: Boulder County, Colorado

- Snow drought conditions through fall and winter 2021 created dry land-cover
- 130-160 km/h winds, plus ignition, launched an uncontrollable “fire storm”
- Loss of 2 lives. 1000 homes and 20 businesses were destroyed, and more damaged





"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 (2018) "Harnessing Artificial Intelligence for the Earth"





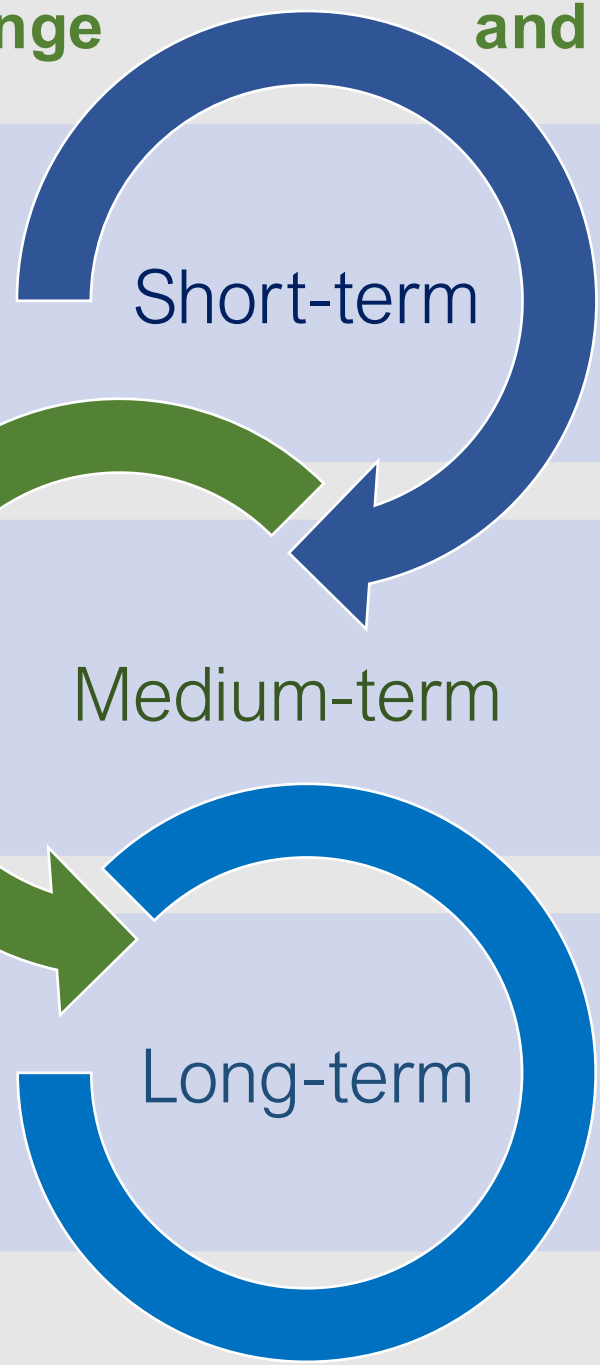
AI Research for Climate Change and Environmental Sustainability

CLIMATE CHANGE

ADAPTATION

MITIGATION

IMPACTS



Extreme weather

Cascading hazards

Energy transition

Land-use change

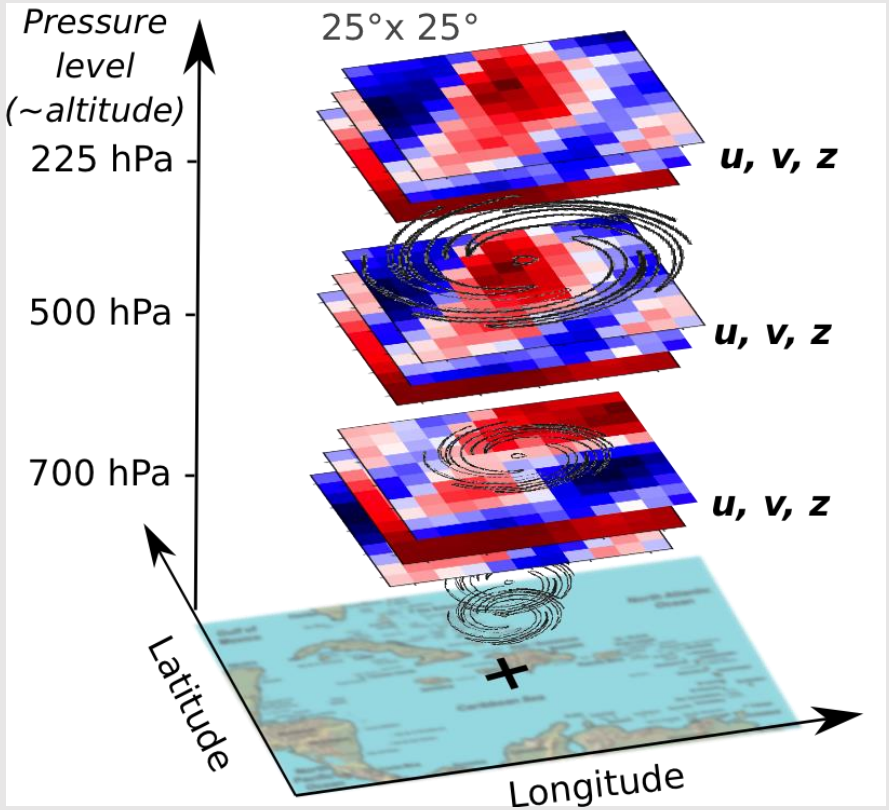
Carbon emissions

Sea-level rise

AI-driven solutions

ADAPTATION

AI for Extreme Weather and Cascading Hazards



[Giffard-Roisin et al., Frontiers 2020]

Hurricane track prediction

Forecasting Indian Summer Monsoon precipitation extremes

Avalanche detection

Generative AI for weather forecasting



MITIGATION

Reducing carbon emissions

Accelerate green energy transition

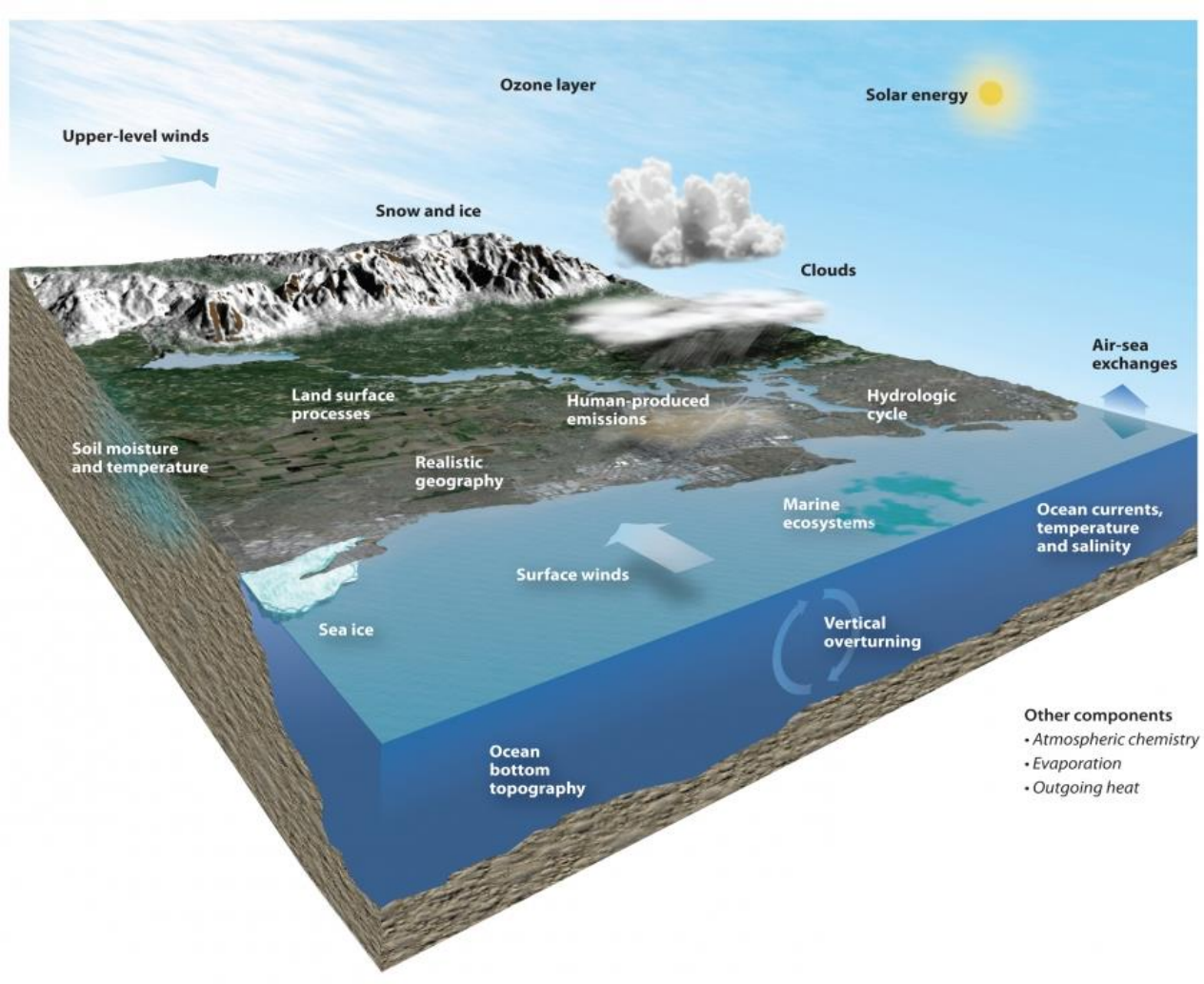
- AI-driven forecasting of solar, wind
- AI to downscale solar and wind data

Reduce compute for weather and climate modeling

- Once trained, AI is significantly faster at prediction than physical models

IMPACTS

AI for Understanding and Predicting Climate Change



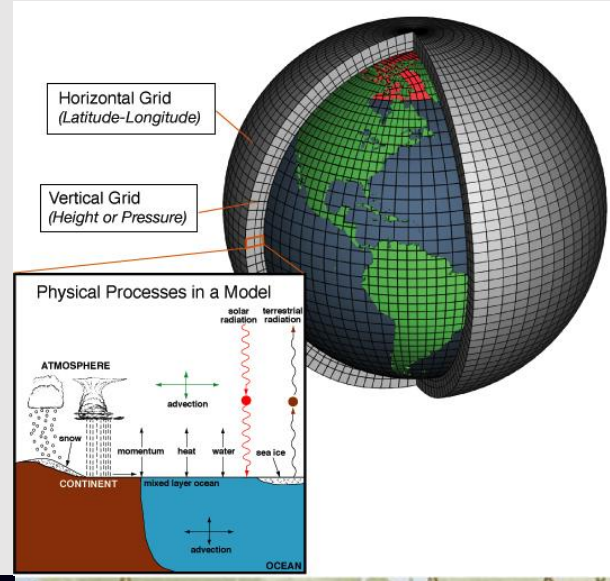
UCAR Science Education

Use AI to learn relations between IPCC simulations and observations

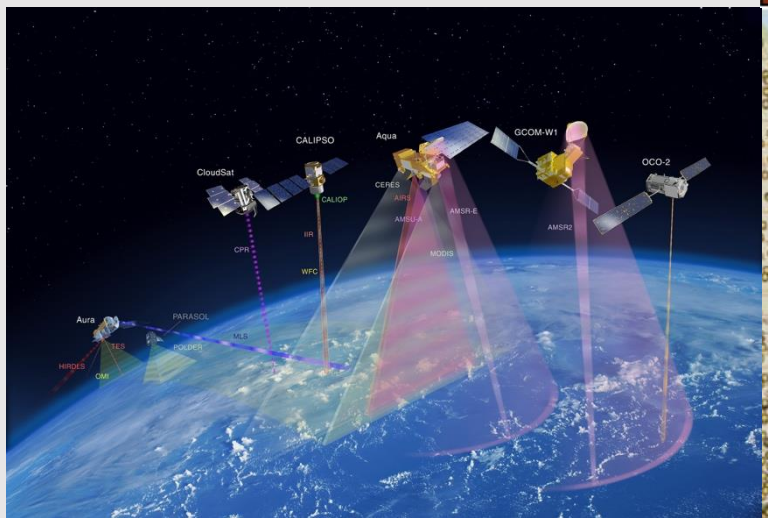
- Robustify climate model ensemble forecasts
- Projecting long-term sea-level rise
- Projecting long-term carbon emissions

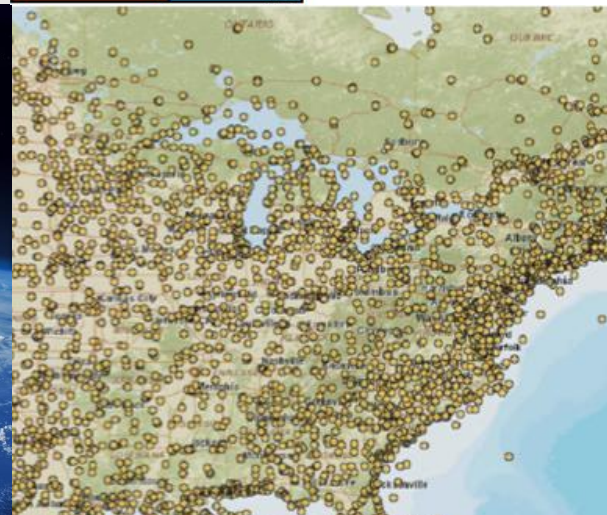
Approach: Exploit all available data

- 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 data assimilation: applies physical laws to observations





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

ML tasks underlying many of these problems

- ❑ Data fusion
- ❑ Forecasting
- ❑ Unsupervised learning
- ❑ Probabilistic domain alignment
- ❑ Downscaling and interpolation

AI Methods

- Unsupervised learning
- Generative deep learning

Outline for the rest of today

I. **Context:** Revolution in AI for Weather Forecasting

Implications for the study of Climate Change

II. **AI Methods:** What is self-supervised learning?

Applications in climate and weather

III. **AI Methods:** What is generative deep learning?

Applications in climate and weather

Part I: Context:
Revolution in AI for Weather Forecasting

The quiet revolution in weather forecasting

- Numerical weather prediction (NWP) has been steadily progressing over the past decades
- Great scientific achievement
  - Complex simulation problem being solved **routinely** in operational forecasting centers across the world
- Strong socio-economical value
  - Climate hazards mitigation, Logistics, Property loss prevention, …

Credit for first few slides: David Landry (INRIA → Google DeepMind)

doi:10.1038/nature14956

REVIEW

doi:10.1038/nature14956

The quiet revolution of numerical weather prediction

Peter Bauer[1], Alan Thorpe[1] & Gilbert Brunet[2]

Advances in numerical weather prediction represent a quiet revolution because they have resulted from a steady accumulation of scientific knowledge and technological advances over many years that, with only a few exceptions, have not been associated with the aura of fundamental physics breakthroughs. Nonetheless, the impact of numerical weather prediction is among the greatest of any area of physical science. As a computational problem, global weather prediction is comparable to the simulation of the human brain and of the evolution of the early Universe, and it is performed every day at major operational centres across the world.

At the turn of the twentieth century, Abbe[1] and Bjerknes[2] proposed that the laws of physics could be used to forecast the weather; they recognized that predicting the state of the atmosphere could be treated as an initial value problem of mathematical physics, wherein future weather is determined by integrating the governing partial differential equations, starting from the observed current weather. This proposition, even with the most optimistic interpretation of Newtonian determinism, is all the more audacious given that, at that time, there were few routine observations of the state of the atmosphere, no computers, and little understanding of whether the weather possesses any significant degree of predictability. But today, more than 100 years later, this paradigm translates into solving daily a system of nonlinear differential equations at about half a billion points per time step between the initial time and weeks to months ahead, and accounting for dynamic, thermodynamic, radiative and chemical processes working on scales from hundreds of metres to thousands of kilometres and from seconds to weeks.

A touchstone of scientific knowledge and understanding is the ability to predict accurately the outcome of an experiment. In meteorology, this translates into the accuracy of the weather forecast. In addition, today's numerical weather predictions also enable the forecaster to assess quantitatively the degree of confidence users should have in any particular forecast. This is a story of profound and fundamental scientific success built upon the application of the classical laws of physics. Clearly the success has required technological acumen as well as scientific advances and vision.

Accurate forecasts save lives, support emergency management and mitigation of impacts and prevent economic losses from high-impact weather, and they create substantial financial revenue—for example, in energy, agriculture, transport and recreational sectors. Their substantial benefits far outweigh the costs of investing in the essential scientific research, super-computing facilities and satellite and other observational programmes that are needed to produce such forecasts[3].

These scientific and technological developments have led to increasing weather forecast skill over the past 40 years. Importantly, this skill can be objectively and quantitatively assessed, as every day we compare the forecast with what actually occurs. For example, forecast skill in the range from 3 to 10 days ahead has been increasing by about one day per decade: today's 6-day forecast is as accurate as the 5-day forecast ten years ago, as shown in Fig. 1. Predictive skill in the Northern and Southern hemispheres is almost equal today, thanks to the effective use of observational information from satellite data providing global coverage.

More visible to society, however, are extreme events. The unusual path and intensification of hurricane Sandy in October 2012 was predicted 8 days ahead, the 2010 Russian heat-wave and the 2013 US cold spell were forecast with 1–2 weeks lead time, and tropical sea surface temperature variability following the El Niño/Southern Oscillation phenomenon can be predicted 3–4 months ahead. Weather and climate prediction skill are intimately linked, because accurate climate prediction needs a good representation of weather phenomena and their statistics, as the underlying physical laws apply to all prediction time ranges.

This Review explains the fundamental scientific basis of numerical weather prediction (NWP) before highlighting three areas from which the largest benefit in predictive skill has been obtained in the past—physical process representation, ensemble forecasting and model initialization. These are also the areas that present the most challenging science questions in the next decade, but the vision of running

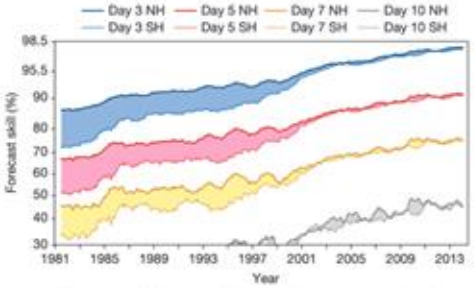
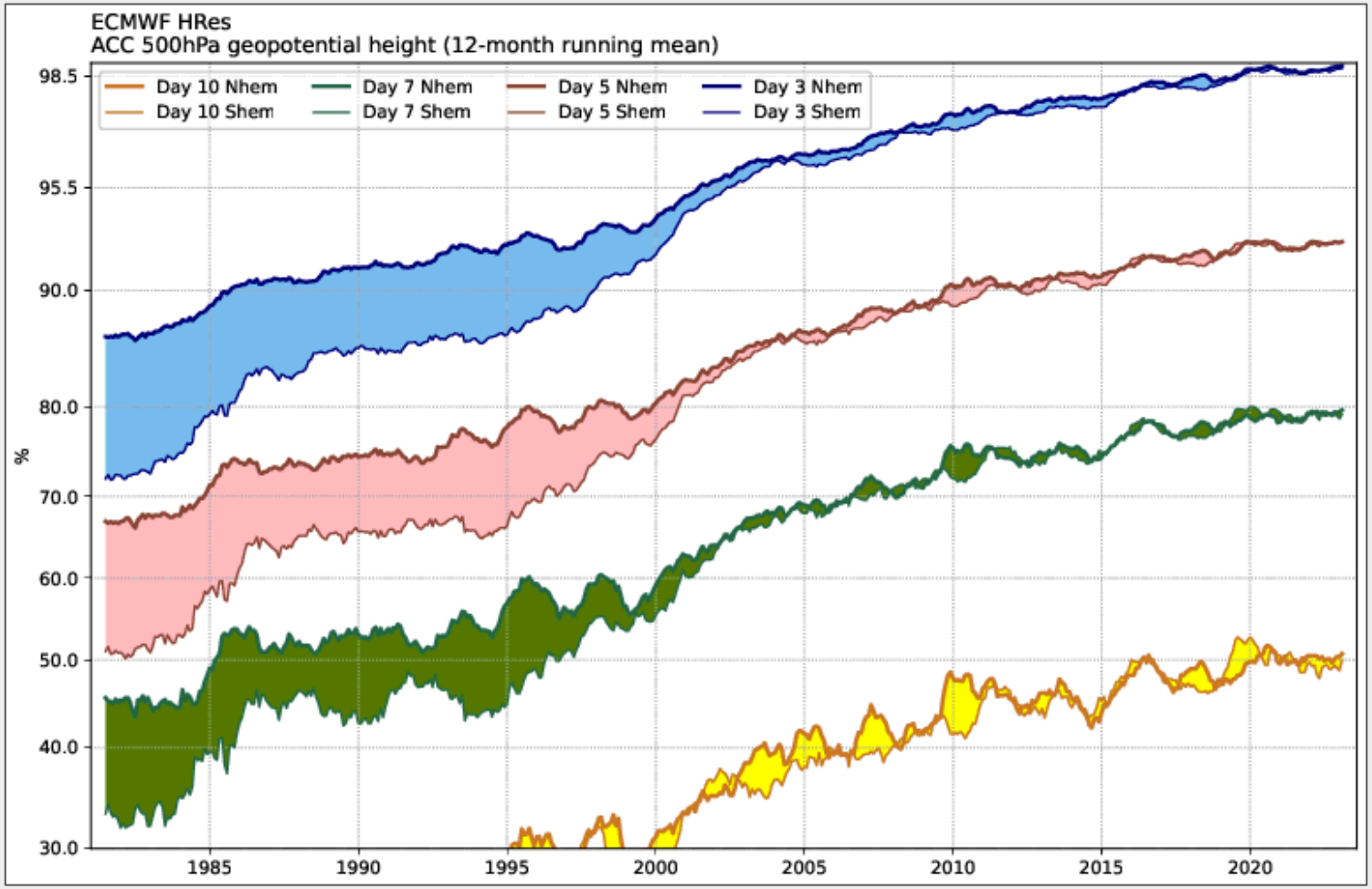


Figure 1 | A measure of forecast skill at three-, five-, seven- and ten-day ranges, computed over the extra-tropical northern and southern hemispheres. Forecast skill is the correlation between the forecasts and the verifying analysis of the height of the 500-hPa level, expressed as the anomaly with respect to the climatological height. Values greater than 60% indicate useful forecasts, while those greater than 80% represent a high degree of accuracy. The convergence of the curves for Northern Hemisphere (NH) and Southern Hemisphere (SH) after 1999 indicates the breakthrough in exploiting satellite data through the use of variational data[100].

[1]European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK. [2]Environment Canada, Trans-Canada Highway Dorval, Québec H9P 1J3, Canada.

©2015 Macmillan Publishers Limited. All rights reserved

The quiet revolution in weather forecasting



ECMWF Global Forecast performance over 4 decades

https://charts.ecmwf.int/products/plwww_m_hr_ccaf_adrian_ts?single_product=latest

First AI-NWP models

- Starting in 2022, multiple different research teams have used deep neural networks to perform medium-range weather forecasts
- Some of these models' skill became ~equivalent to the leading NWP models



arXiv:2202.07575v1 [physics.ao-ph] 15 Feb 2022

Forecasting Global Weather with Graph Neural Networks

Ryan Keisler
rkeisler@gmail.com

Abstract

We present a data-driven approach for forecasting glo
neural networks. The system learns to step forward the
state by six hours, and multiple steps are chained tog
forecasts going out several days into the future. The un
on reanalysis data from ERA5 or forecast data from G
metrics such as $Z500$ (geopotential height) and $T850$ (te
previous data-driven approaches and is comparable to o
physical models from GFS and ECMWF, at least whe
scales and when using reanalysis initial conditions. W
connecting this data-driven model to live, operational fo

1 Introduction

Numerical weather prediction (NWP), as part of the broader weath
and positive impact on society. Decades of steady improvem
observational data, better modeling techniques, and more com
increasingly accurate weather forecasts and growing adoption of

While statistical techniques have been used within NWP for deca
these models continue to be based on the physical principles go
More recently, spurred on by advancements in machine learnin
interest in statistical, data-driven techniques for weather forecasti
to improve upon an already extremely successful NWP program t
forecasts, faster forecasts, or more forecasts, i.e. larger ensemble
for using ML to advance our scientific understanding of the under
et al., 2020].

There is currently a very active hub of research at the intersection
areas include: faster or more accurate solving of relevant PDEs [
2021, Um et al., 2021, Li et al., 2021, Brandstetter et al., 2022
2021, Maulik et al., 2021], subgrid parameterization [Brenowitz a
2021, Yuval et al., 2021, Meyer et al., 2021], nowcasting [Sen
2019, Espeholt et al., 2021, Klocek et al., 2021, Ravuri et al., 20
[Watt-Meyer et al., 2021], replacing global NWP models [Dueben

[physics.ao-ph] 22 Feb 2022

FOURCASTNET: A GLOBAL DATA-DRIVEN HIGH-RESOLUTION WEATHER MODEL USING ADAPTIVE FOURIER NEURAL OPERATORS

A PREPRINT

Jaideep Pathak
NVIDIA Corporation
Santa Clara, CA 95051

Shashank Subramanian
Lawrence Berkeley
National Laboratory
Berkeley, CA 94720

Peter Harrington
Lawrence Berkeley
National Laboratory
Berkeley, CA 94720

Sanjeev Raja
University of Michigan
Ann Arbor, MI 48109

Ashesh Chattopadhyay
Rice University
Houston, TX 77005

Morteza Mardani
NVIDIA Corporation
Santa Clara, CA 95051

Thorsten Kurth
NVIDIA Corporation
Santa Clara, CA 95051

David Hall
NVIDIA Corporation
Santa Clara, CA 95051

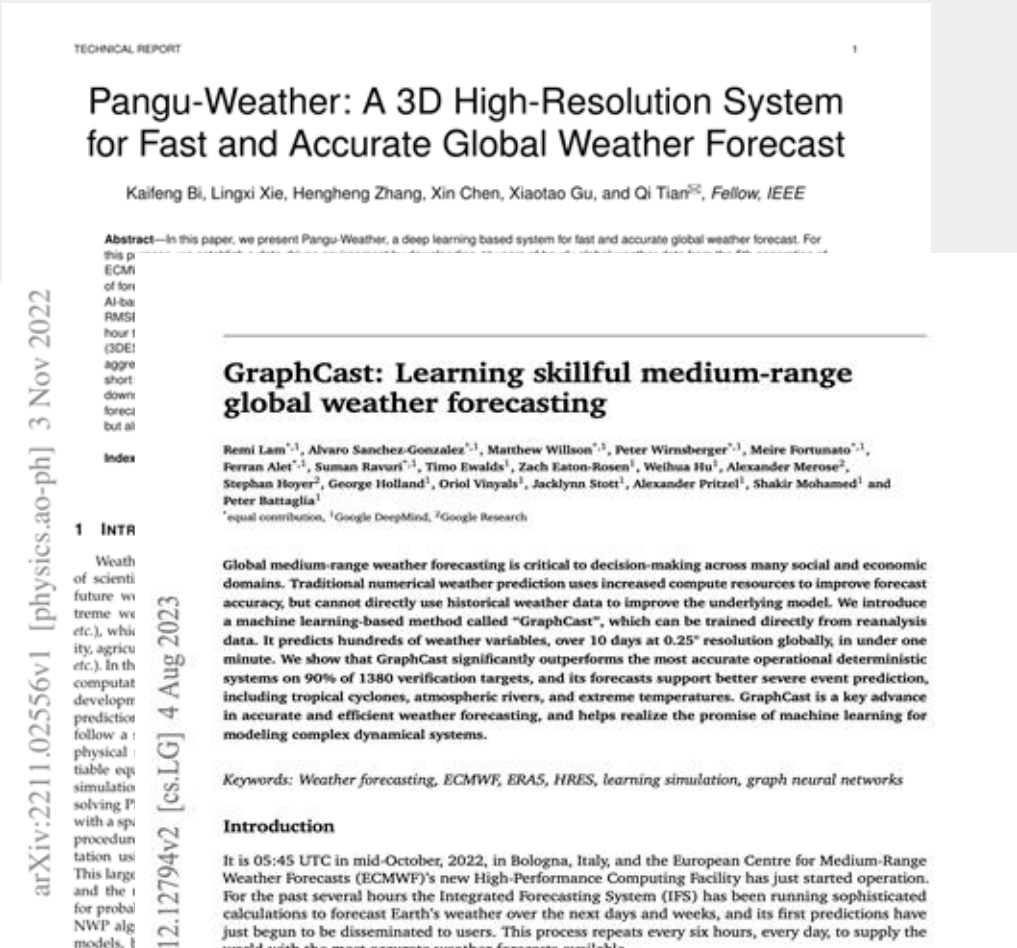
Zongyi Li
California Institute of Technology
Pasadena, CA 91125
NVIDIA Corporation
Santa Clara, CA 95051

Kamyar Azizzadenesheli
Purdue University
West Lafayette, IN 47907

Pedram Hassanzadeh
Rice University
Houston, TX 77005

Karthik Kashinath
NVIDIA Corporation
Santa Clara, CA 95051

Animashree Anandkumar
California Institute of Technology
Pasadena, CA 91125
NVIDIA Corporation



arXiv:2211.02556v1 [physics.ao-ph] 3 Nov 2022

Pangu-Weather: A 3D High-Resolution System for Fast and Accurate Global Weather Forecast

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian✉, *Fellow, IEEE*

Abstract—In this paper, we present Pangu-Weather, a deep learning based system for fast and accurate global weather forecast. For

arXiv:2212.12794v2 [cs.LG] 4 Aug 2023

GraphCast: Learning skillful medium-range global weather forecasting

Remi Lam*,†,1, Alvaro Sanchez-Gonzalez*,1, Matthew Willson*,1, Peter Wirnsberger*,1, Meire Fortunato*,1, Ferran Alet*,1, Suman Ravuri*,1, Timo Ewalds1, Zach Eaton-Rosen1, Weihua Hu1, Alexander Merose2, Stephan Hoyer2, George Holland1, Oriol Vinyals1, Jacklynn Stott1, Alexander Pritzel1, Shakir Mohamed1 and Peter Battaglia1
*equal contribution, †Google DeepMind, ²Google Research

Global medium-range weather forecasting is critical to decision-making across many social and economic domains. Traditional numerical weather prediction uses increased compute resources to improve forecast accuracy, but cannot directly use historical weather data to improve the underlying model. We introduce a machine learning-based method called "GraphCast", which can be trained directly from reanalysis data. It predicts hundreds of weather variables, over 10 days at 0.25° resolution globally, in under one minute. We show that GraphCast significantly outperforms the most accurate operational deterministic systems on 90% of 1380 verification targets, and its forecasts support better severe event prediction, including tropical cyclones, atmospheric rivers, and extreme temperatures. GraphCast is a key advance in accurate and efficient weather forecasting, and helps realize the promise of machine learning for modeling complex dynamical systems.

Keywords: Weather forecasting, ECMWF, ERA5, HRES, learning simulation, graph neural networks

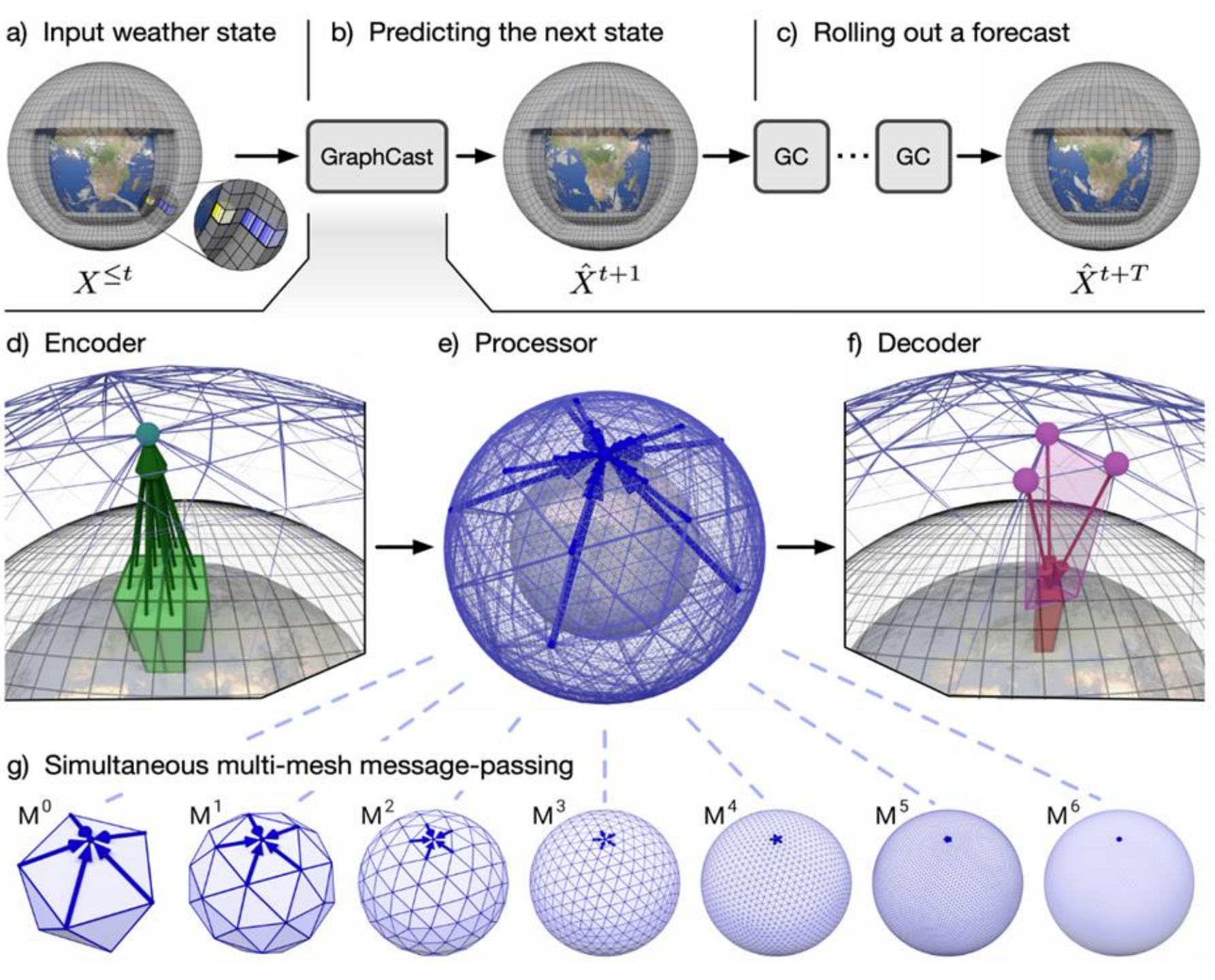
Introduction

It is 05:45 UTC in mid-October, 2022, in Bologna, Italy, and the European Centre for Medium-Range Weather Forecasts (ECMWF)'s new High-Performance Computing Facility has just started operation. For the past several hours the Integrated Forecasting System (IFS) has been running sophisticated calculations to forecast Earth's weather over the next days and weeks, and its first predictions have just begun to be disseminated to users. This process repeats every six hours, every day, to supply the

Computational savings of these first AI-NWP models

Model	**Training**	**Inference**
Pangu-Weather	16 days on 192 V100 GPUs	24 hours forecast 1xV100 1.4 seconds on
GraphCast	21 days on 32 Cloud TPU V4 devices (~128 GPUs)	10 days forecast 1 Cloud TPU V4 < 60 seconds
NWP (ECMWF HRES) ***Higher resolution**	-	10 days forecast 1/3/6 hours 11 664 cores HPC cluster

GraphCast architecture



doi:10.48550/arXiv.2212.12794

AI for Weather Forecasting: A New Paradigm

Data Assimilation → ERA5 Dataset → Simulation with Neural Network → P(Forecast)

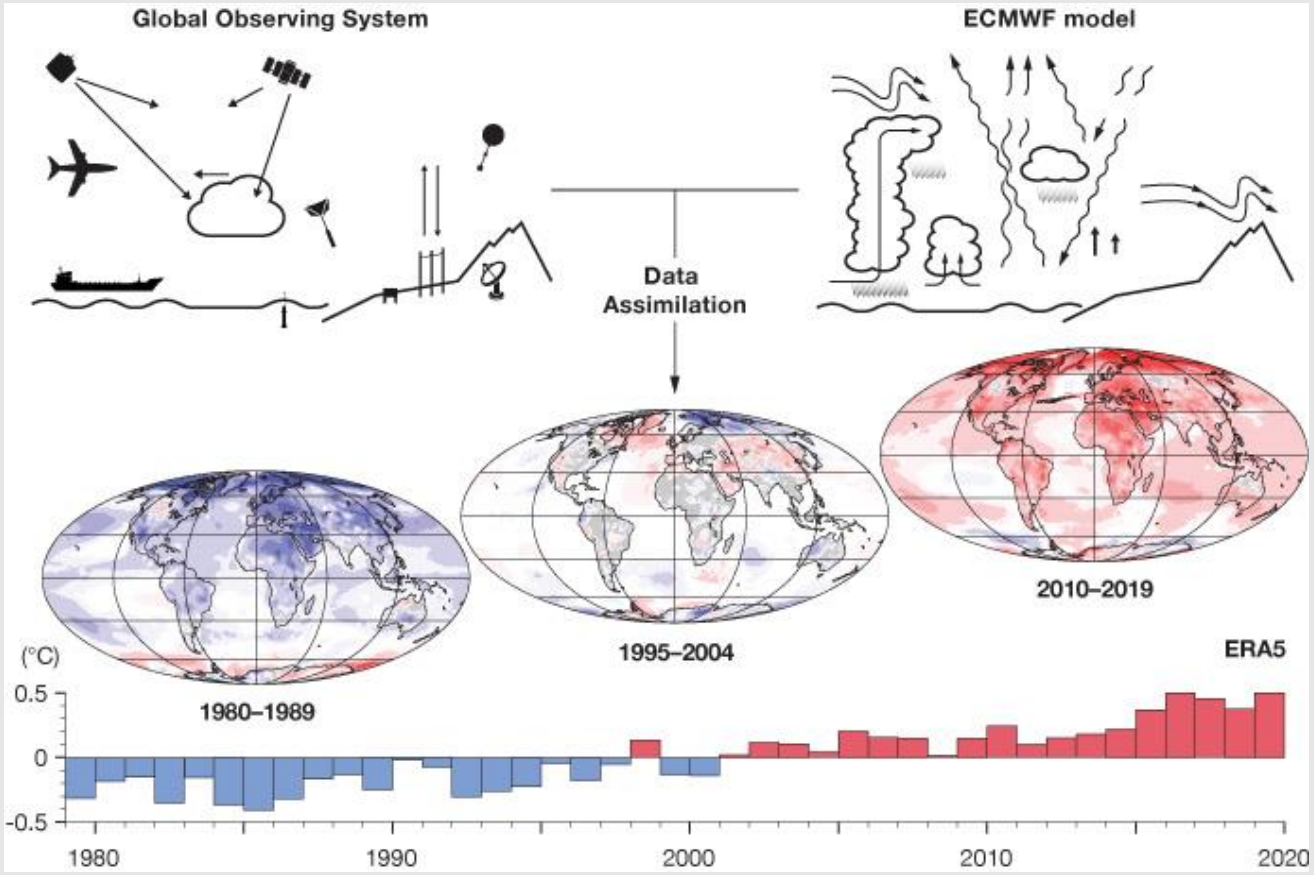
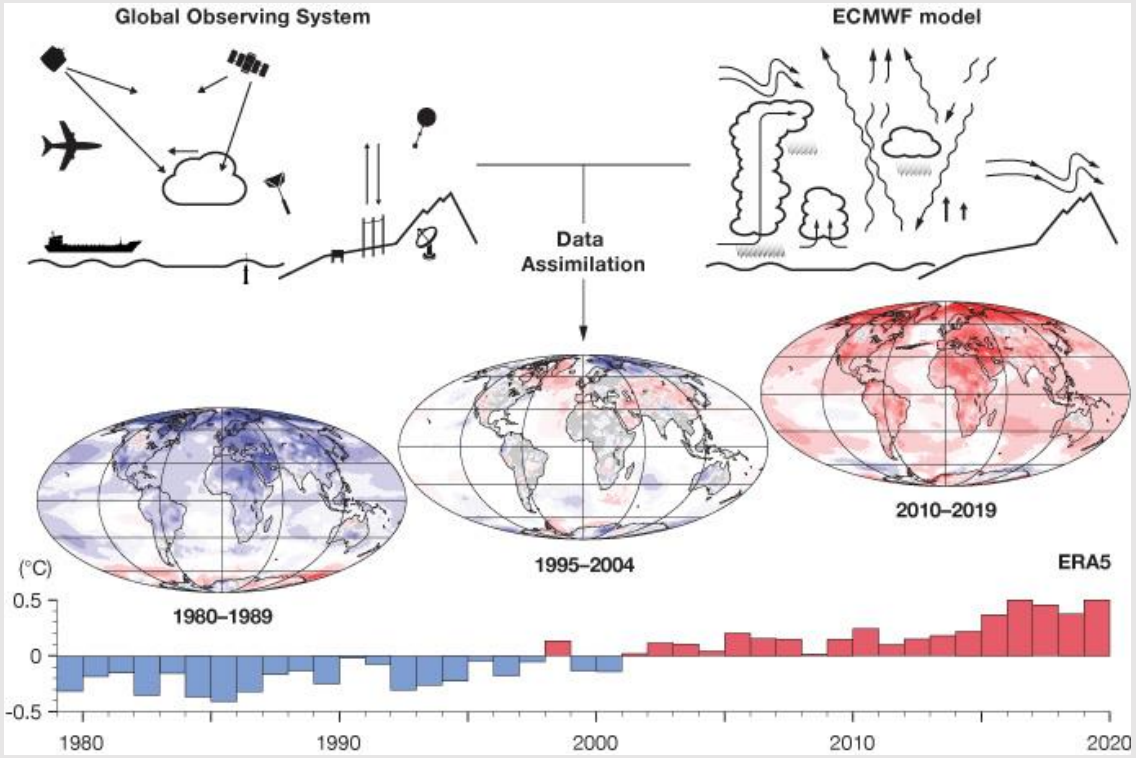


Figure: European Center for Medium Range Weather Forecasting (ECMWF)

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

Training task: auto-regression: forecasting 6-24 hours ahead

- Rolled-out to forecast 7-10 days ahead

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

Rollouts

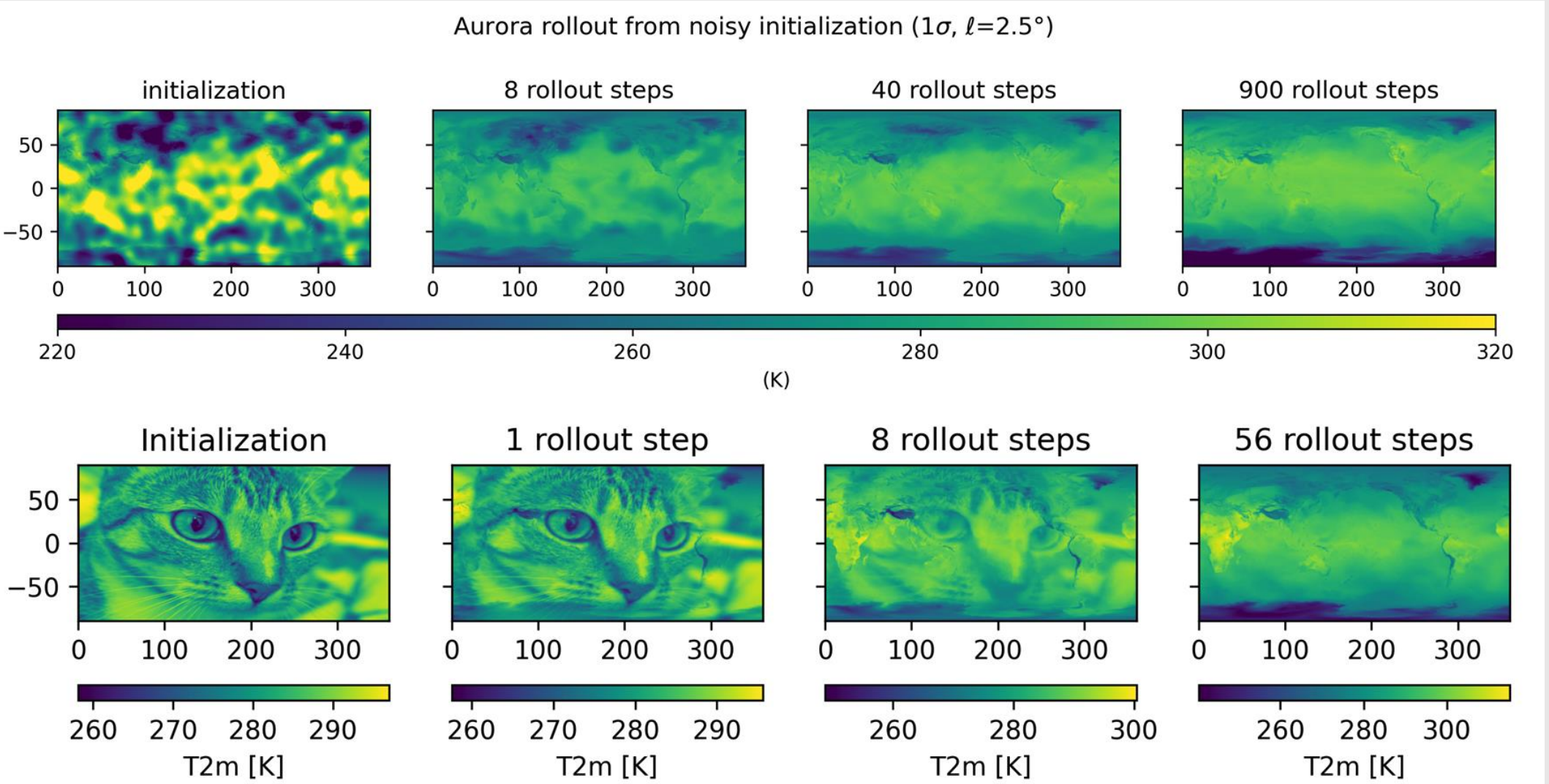


Figure A.9: 2-meter temperature predicted at different time stamps when (top) inputs are perturbed with an additive Gaussian random field with coefficient of variation σ^2 (where σ is the standard deviation of the physical variable) and correlation length of 10 pixels (equivalent to $2.5°$), (bottom) surface and atmospheric variables are initialized with a cat picture.

Lehmann et al., (2026) ”Can AI Weather Models Predict Beyond Two Weeks?” arXiv:2605.30184

AI for Weather Forecasting: A New Paradigm

Time

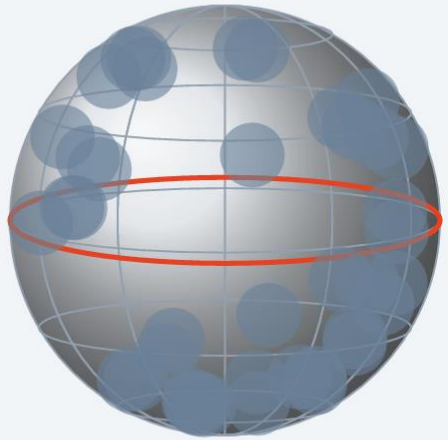
45+ years

1979 — 2024

Hourly

~400,000 timesteps

Space



$721 \times 1440 \times 240$

0.25° resolution

37 levels × 6 vars + 18 surface

Memory

per atmospheric state:

0.25° — 0.94 GB

×6.25

0.1° — 5.9 GB

ECMWF's ERA5 Reanalysis Dataset ≈ 13 Petabytes

ArchesWeatherGen:
Frugal AI weather generator
[Couairon et al., *Science Advances*, 2026]

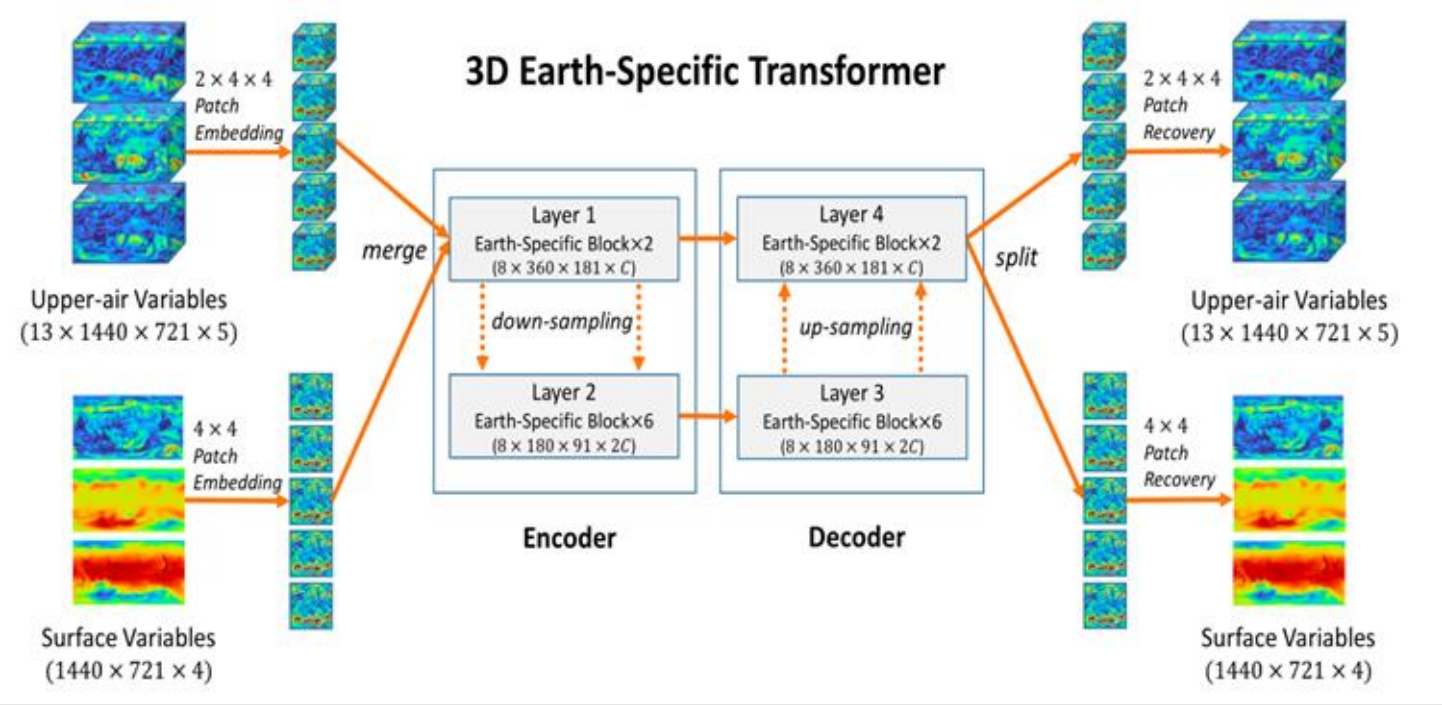
Figure: Emmanuel de Bézenac (INRIA)

ArchesWeather



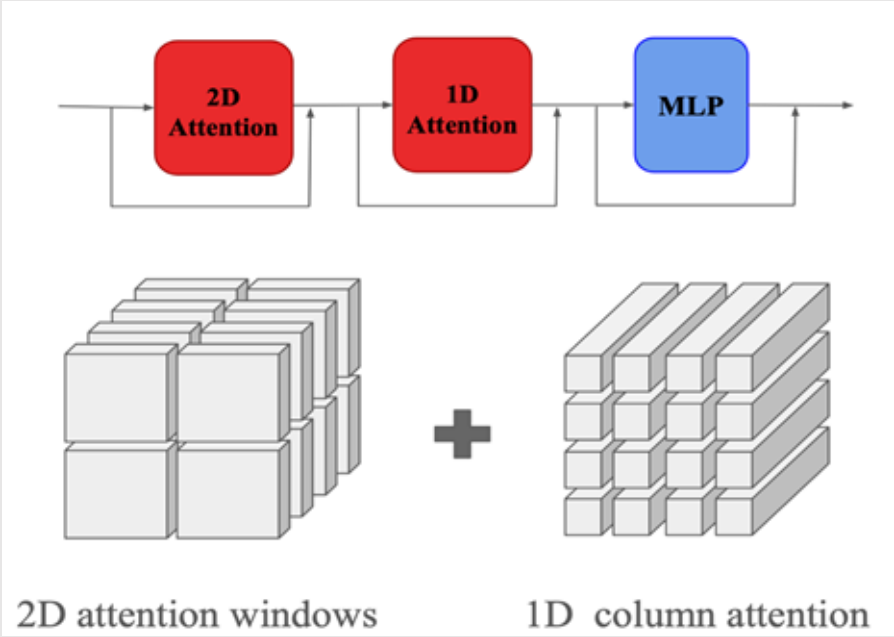
[Couairon et al., ArchesWeather: an efficient AI weather model at 1.5º resolution ICML 2024 workshop on Earth System Modeling]

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



New in ArchesWeather:

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



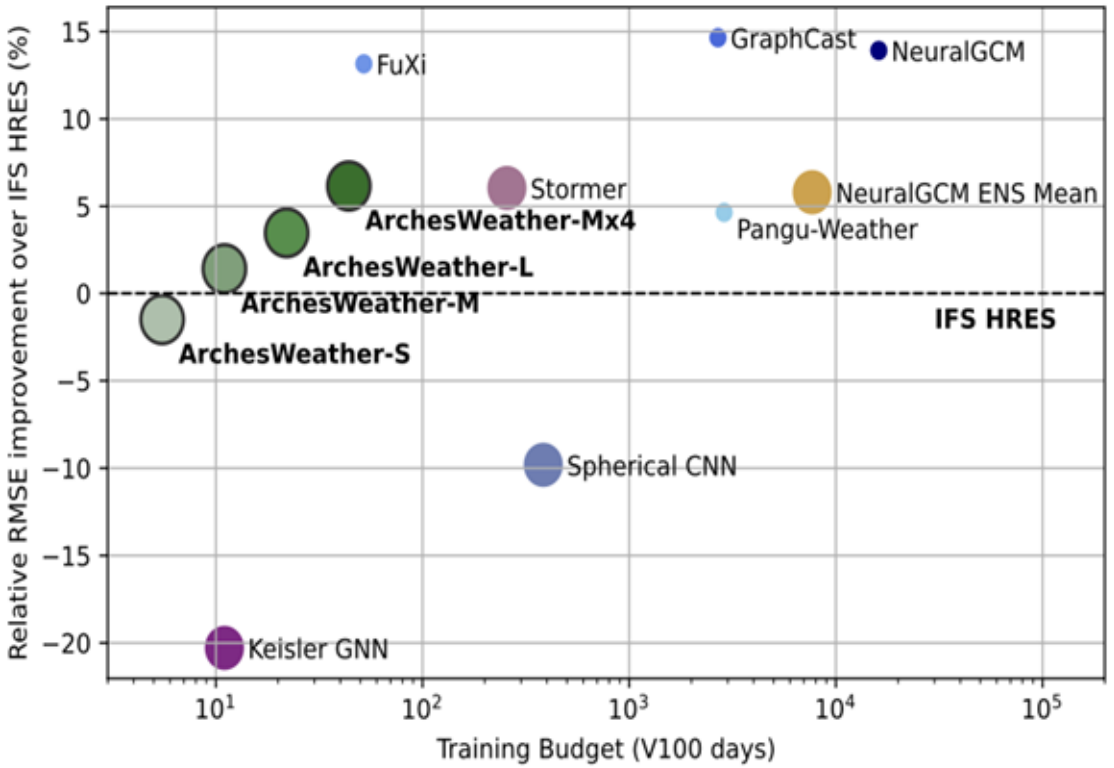
latitude + longitude

altitude

ArchesWeather

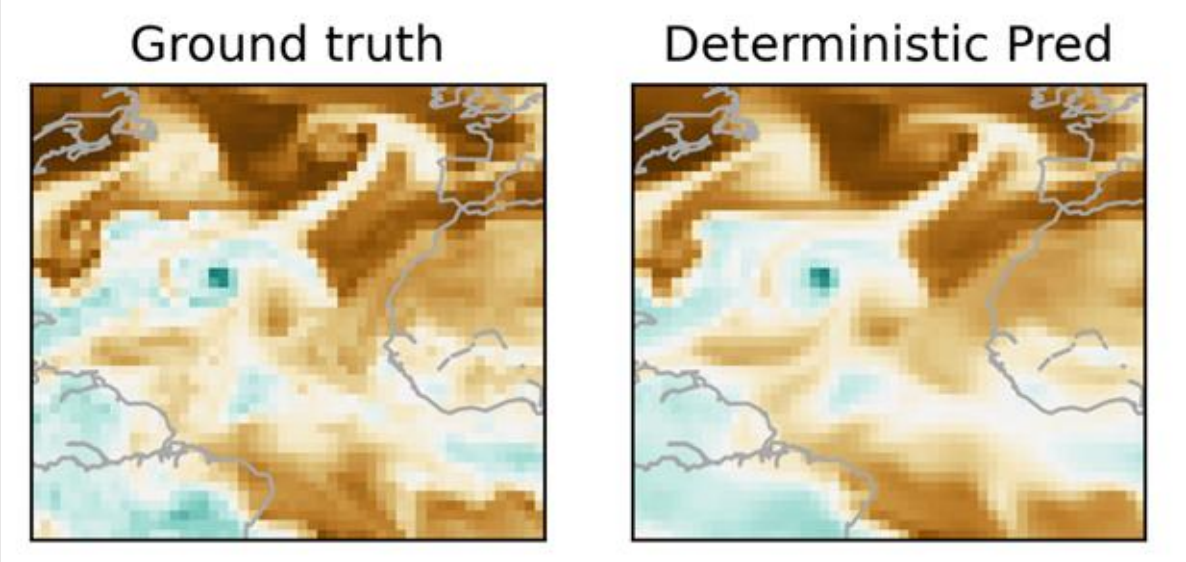
	Res.	Cost	Z500 m^2/s^2	T850 K	Q700 g/kg	U850 m/s	V850 m/s	T2m K	SP Pa	U10m m/s	V10m m/s
IFS HRES			42.30	0.625	0.556	1.186	1.206	0.513	60.16	0.833	0.872
Pangu-Weather	0.25°	2880	44.31	0.620	0.538	1.166	1.191	0.570	55.14	0.728	0.759
NeuralGCM	0.25°	16128	37.94	0.547	0.488	1.050	1.071	N/A	N/A	N/A	N/A
FuXi	0.25°	52	40.08	0.548	N/A	1.034	1.055	0.532	49.23	0.660	0.688
GraphCast	0.25°	2688	39.78	0.519	0.474	1.000	1.02	0.511	48.72	0.655	0.683
Keisler	1°	11	66.87	0.816	0.658	1.584	1.626	N/A	N/A	N/A	N/A
SphericalCNN	1.4°	384	54.43	0.738	0.591	1.439	1.471	N/A	N/A	N/A	N/A
Stormer	1.4°	256	45.12	**0.607**	0.527	**1.138**	**1.156**	0.570	**53.77**	**0.726**	**0.760**
NeuralGCM ENS (50)	1.4°	7680	**43.99**	0.658	0.540	1.239	1.256	N/A	N/A	N/A	N/A
ArchesWeather-S	1.5°	9	48.92	0.650	0.542	1.289	1.325	0.562	60.30	0.833	0.873
ArchesWeather-M	1.5°	18	44.96	0.615	**0.527**	1.219	1.251	**0.539**	56.22	0.783	0.819
ArchesWeather-Mx4	1.5°	36	**41.93**	**0.593**	**0.513**	**1.172**	**1.203**	**0.517**	**52.22**	**0.749**	**0.783**

Performance tradeoff



RMSE, 24 hour lead time

Why transition to **ensemble generation**…



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

ArchesWeatherGen

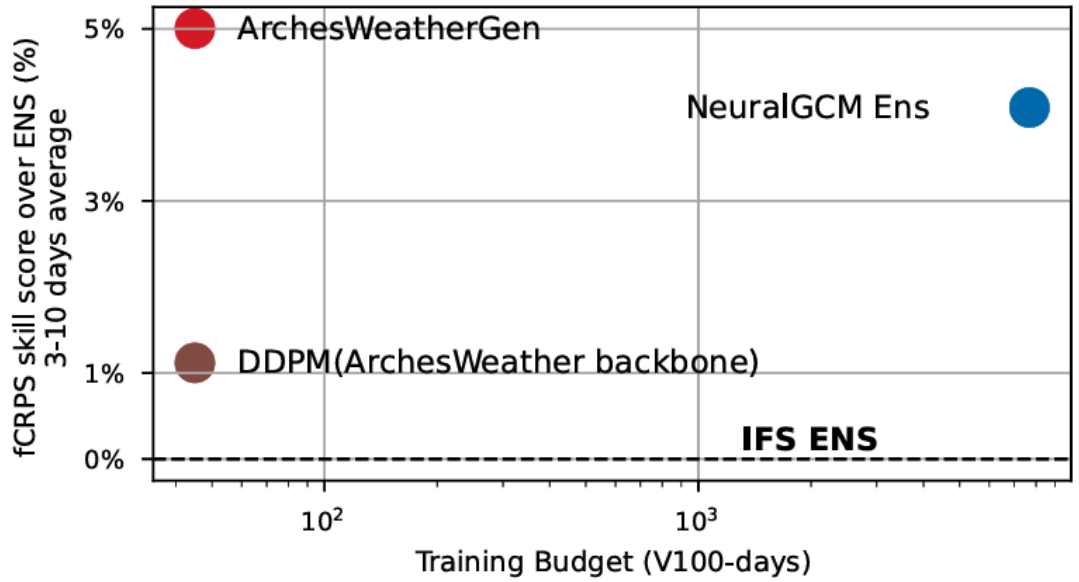
[Couairon et al., *Science Advances*, 2026]

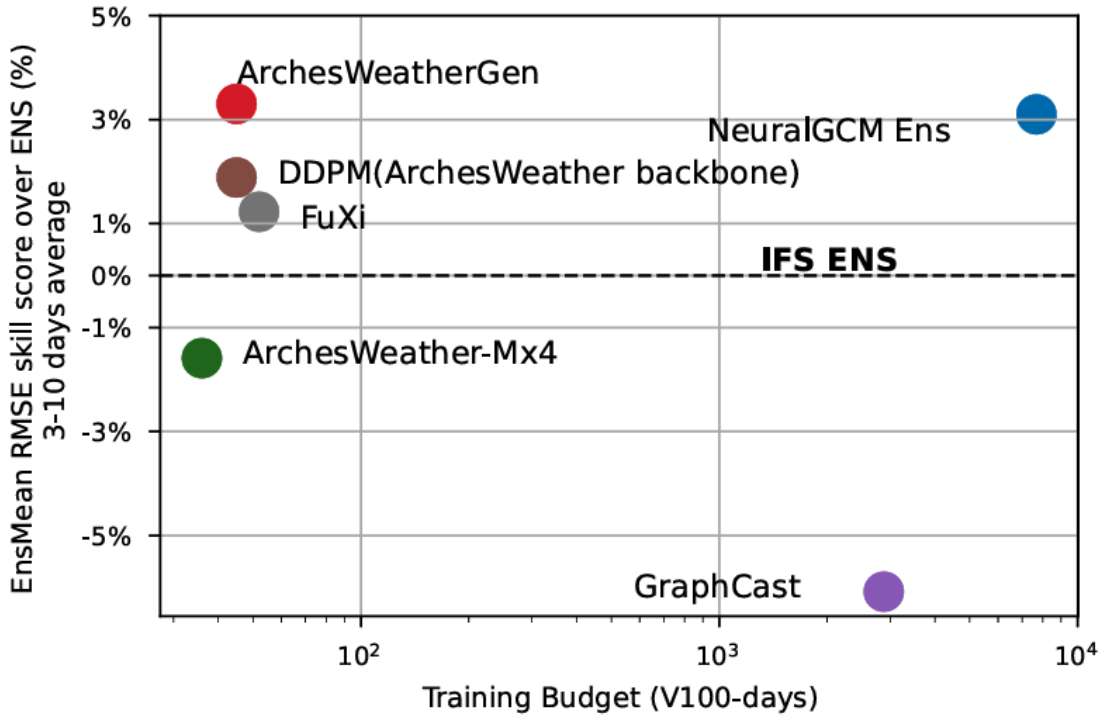
Backbone: ArchesWeather (variant of Swin U-Transformer) 44M parameters

Trained with Flow Matching

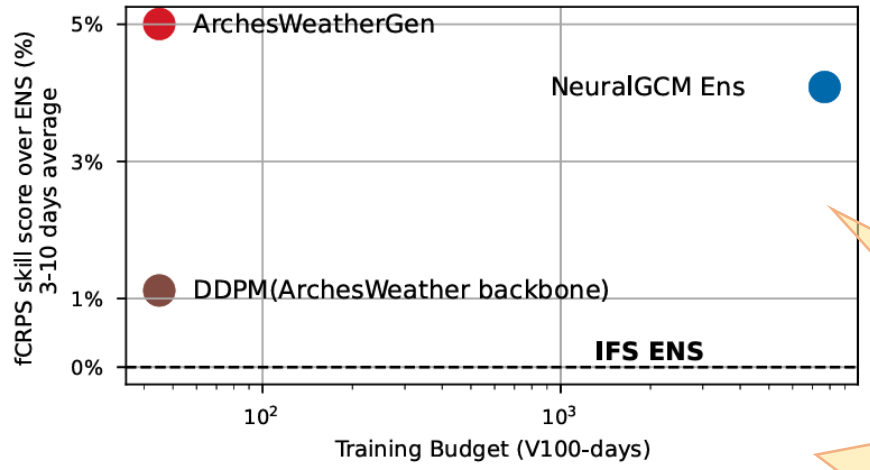
Inference done with 25 sampling steps per 24 hours

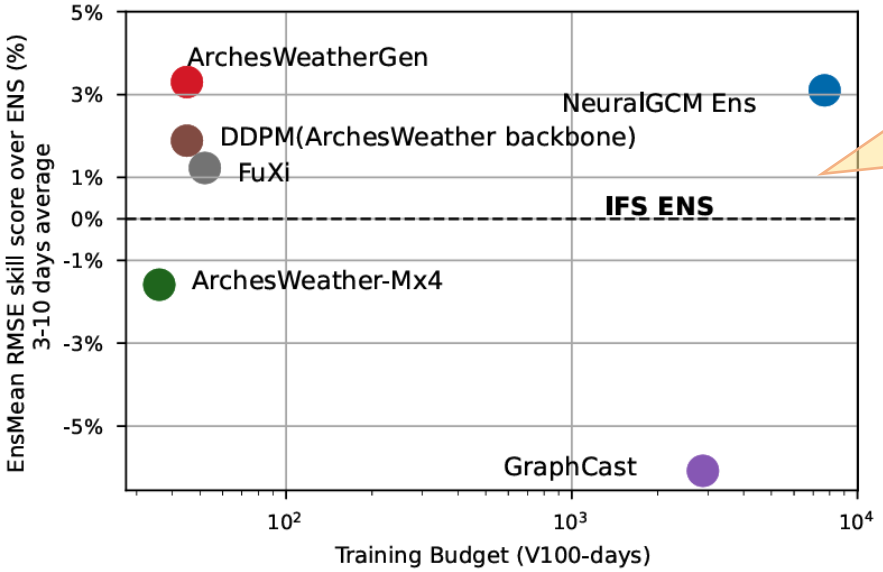
Roll-outs done via one sample each 24 hours. Input this output into trained model and repeat.

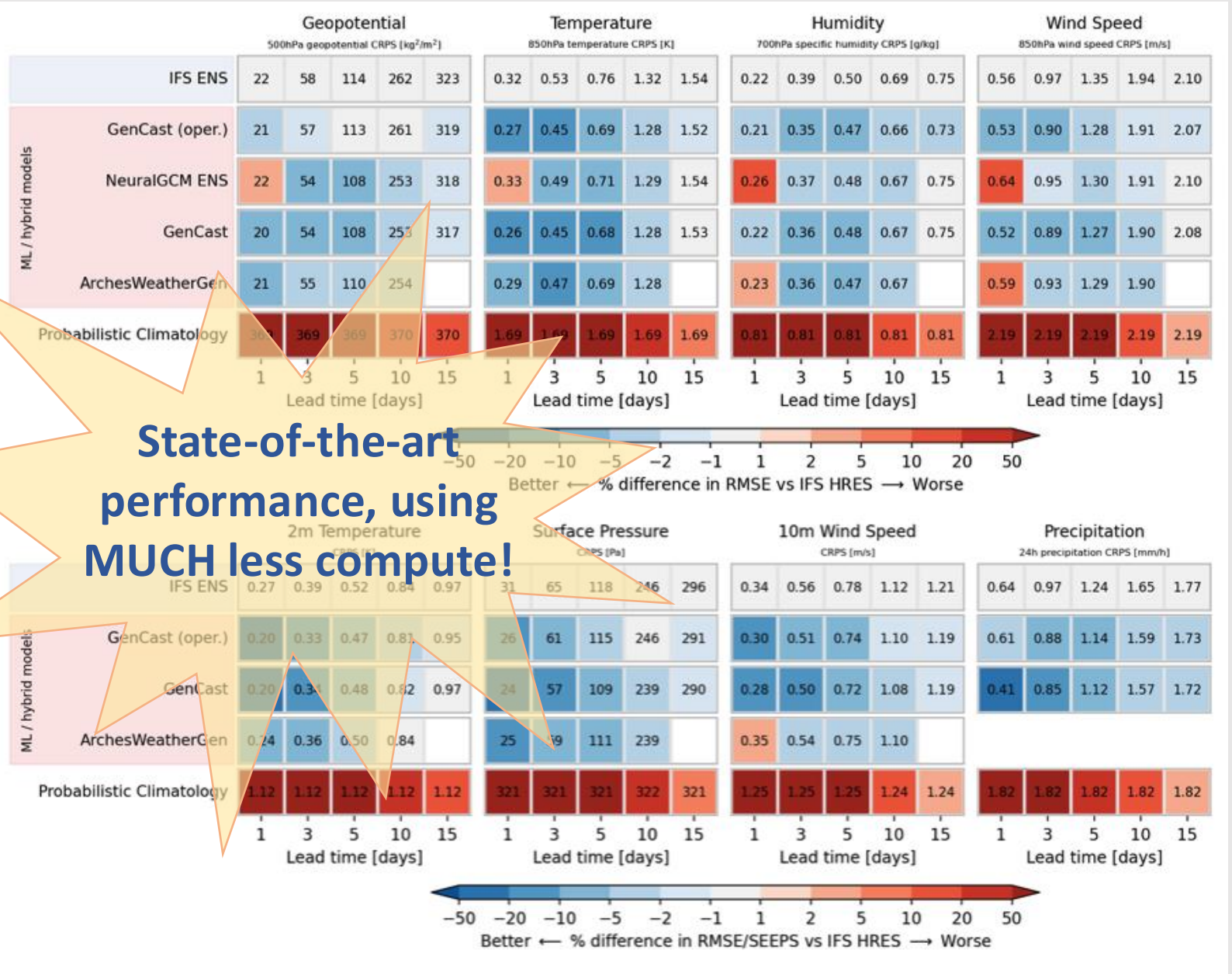




ArchesWeatherGen





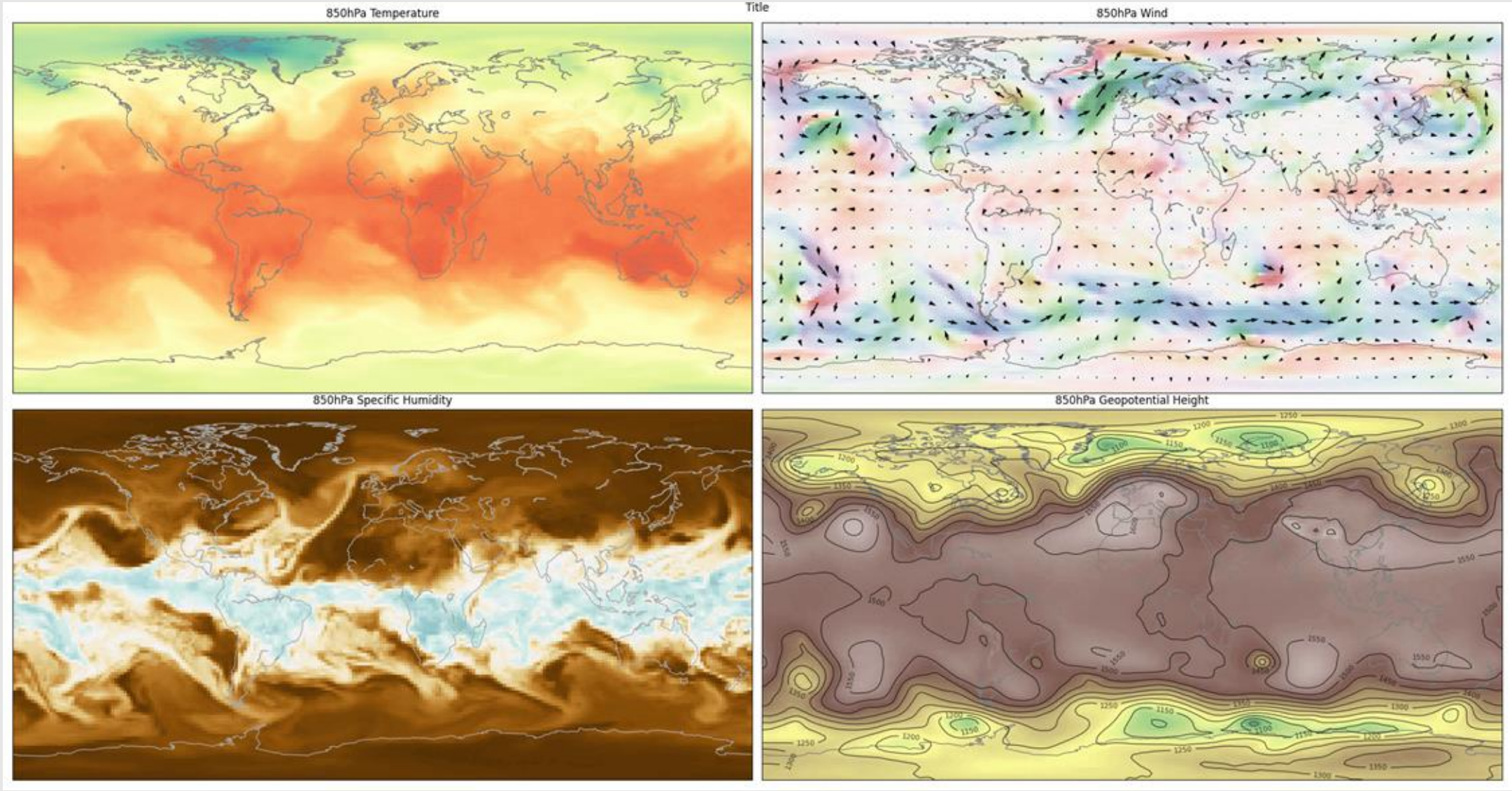


State-of-the-art performance, using MUCH less compute!

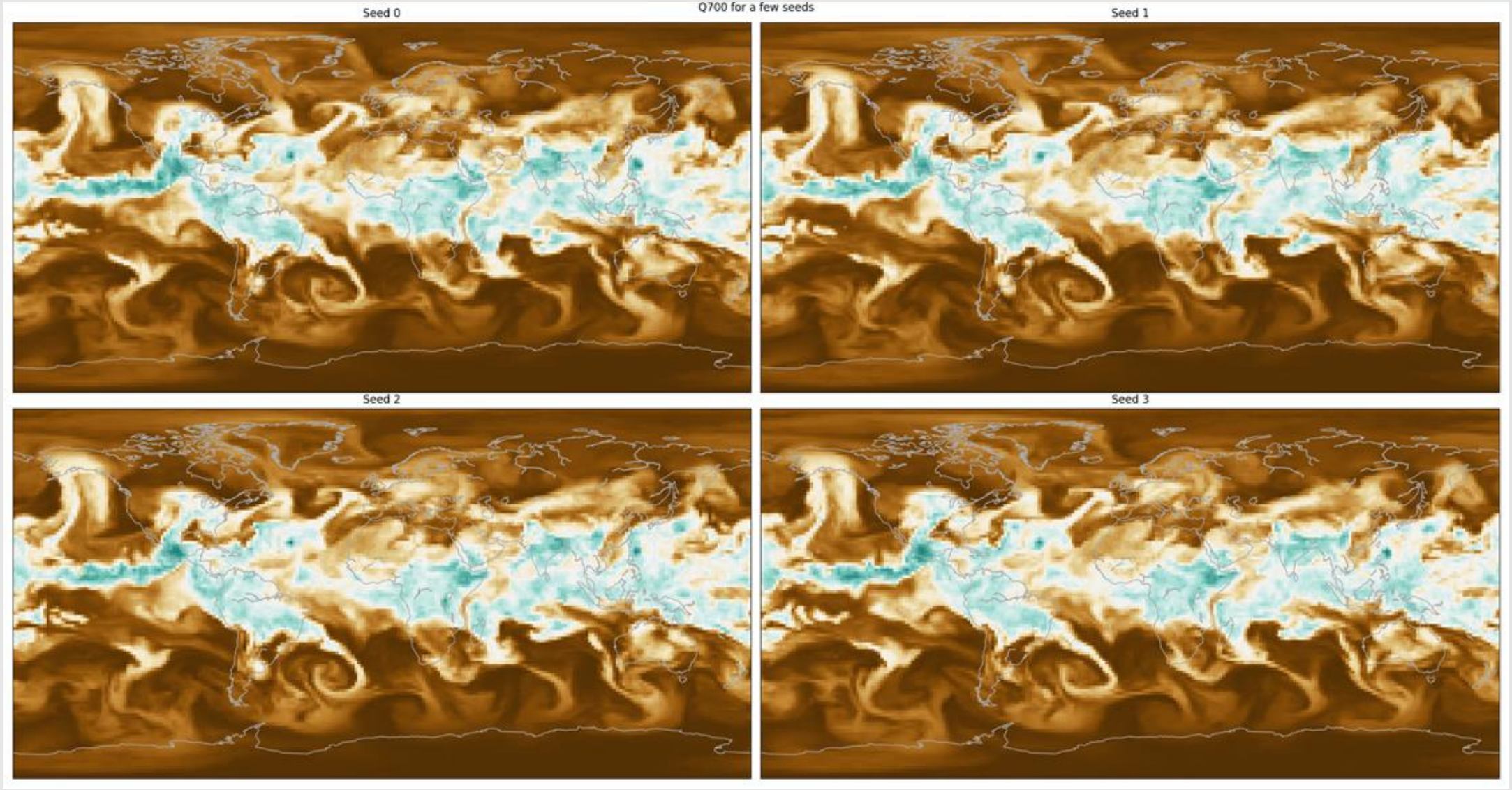
[Couairon, Singh, Charantonis, Lessig, Monteleoni, *Science Advances*, 2026]

Score cards on Google's Weather Bench 2

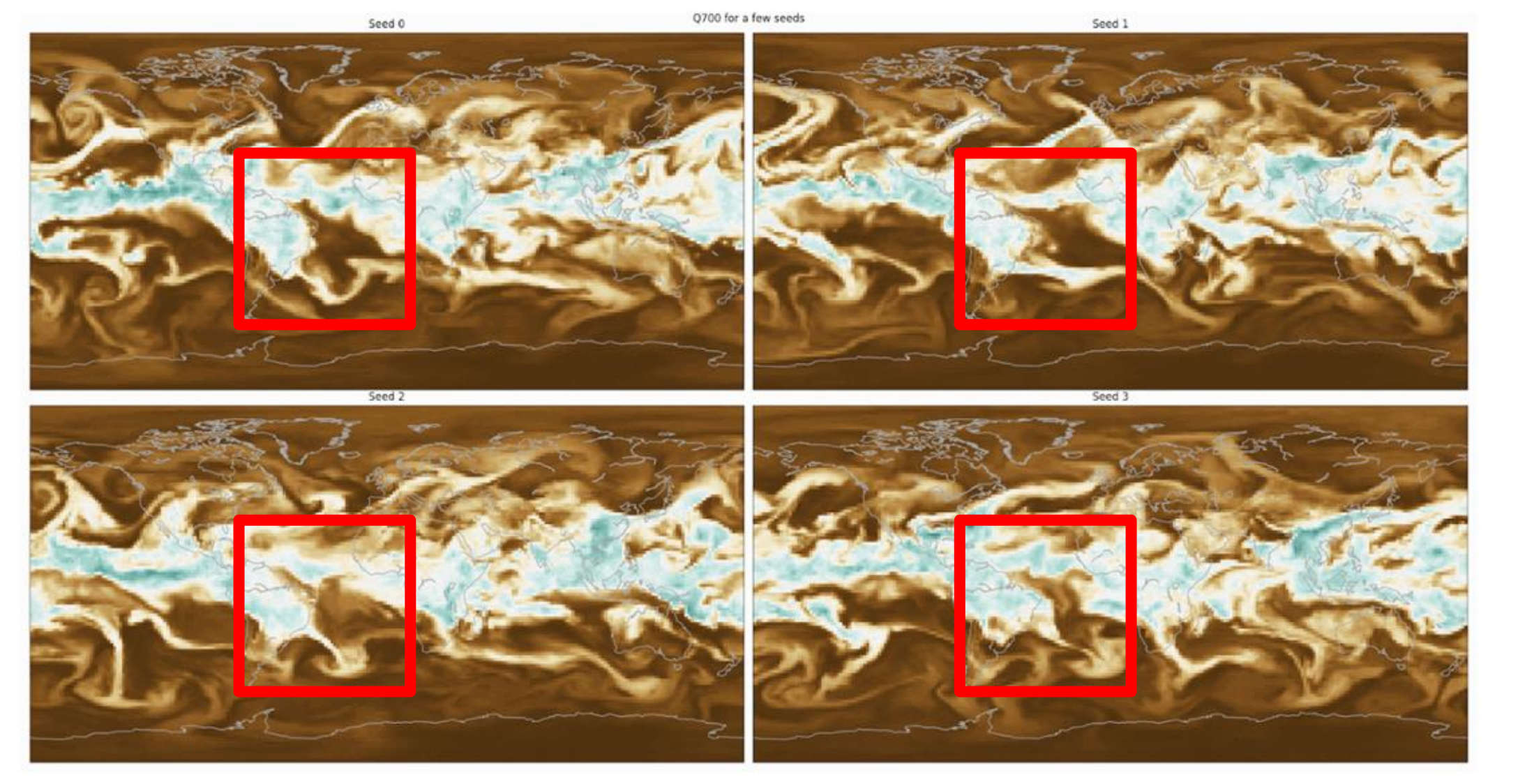
A sample trajectory



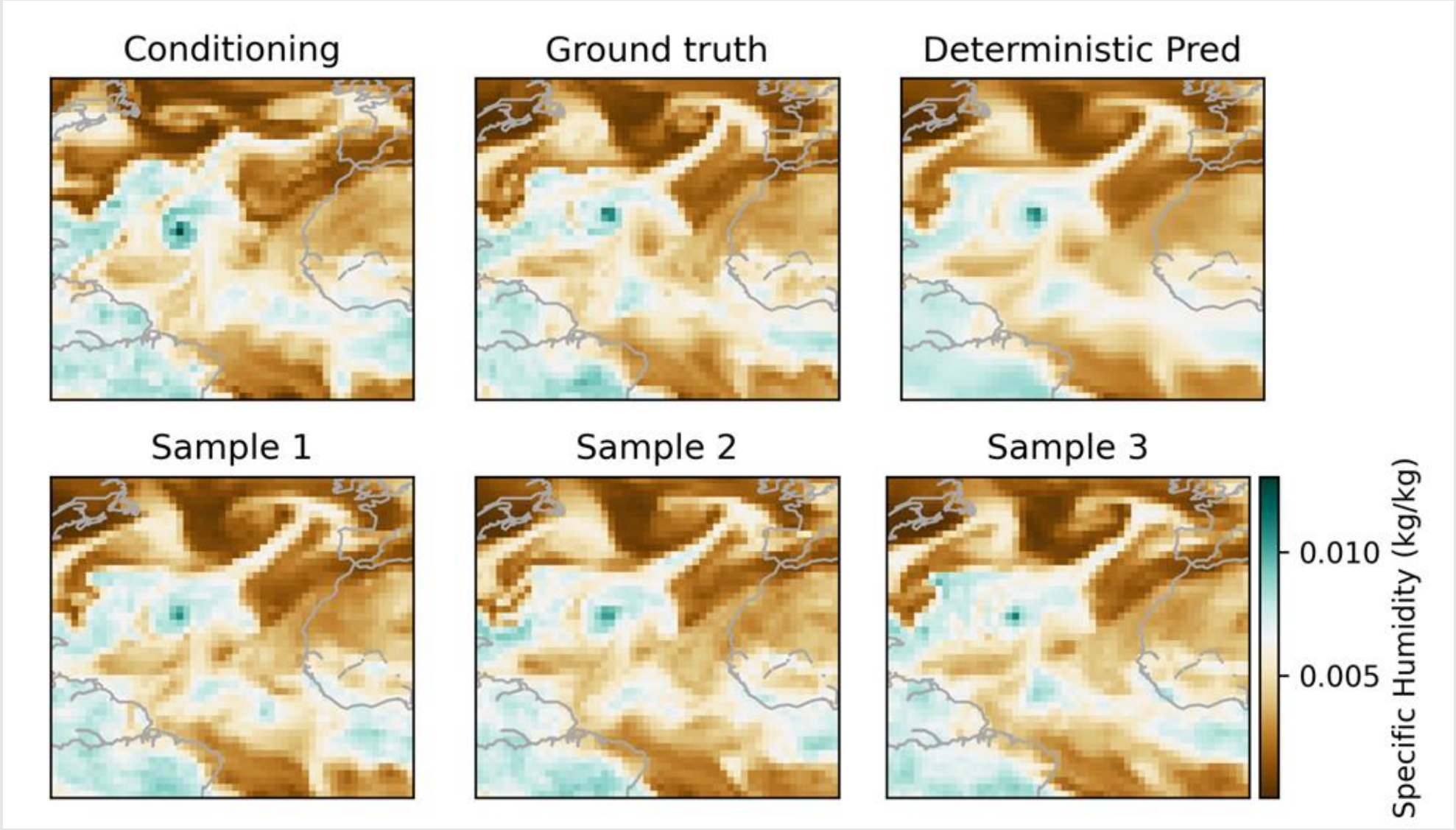
Different seeds - Q700



Ensemble member diversity; lead-time 10 days



Case study: Lorenzo - Q700 -init date 25 sept 2019



ArchesClimate: AI climate model emulation



Ongoing PhD work of Graham Clyne
Primary supervisor: Anastase Charantonis

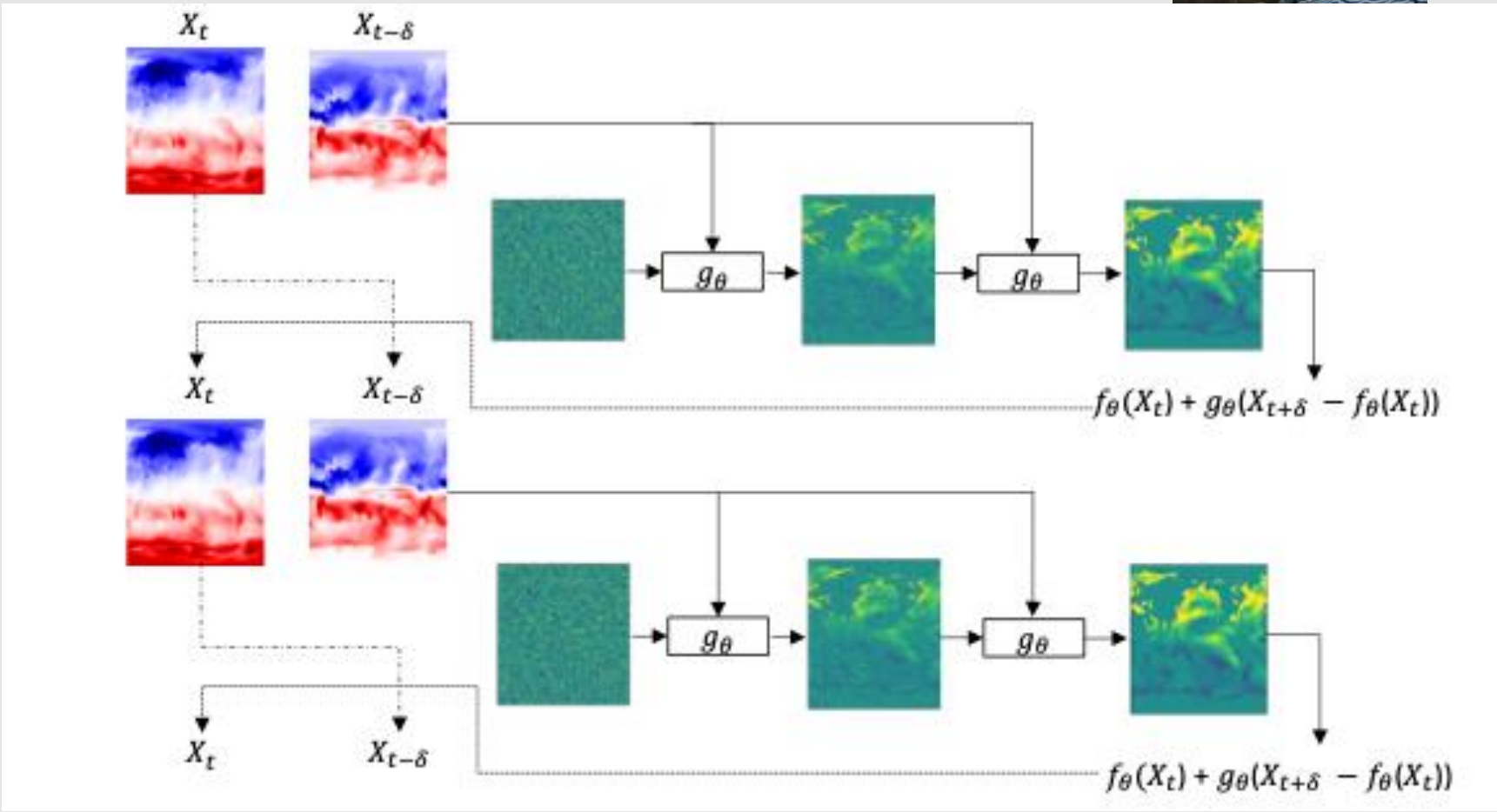
Pipeline: Modified ArchesWeatherGen:

- No column-wise 1D attention
- Include forcings

Training data: Climate model simulations (from IPSL DCPP)

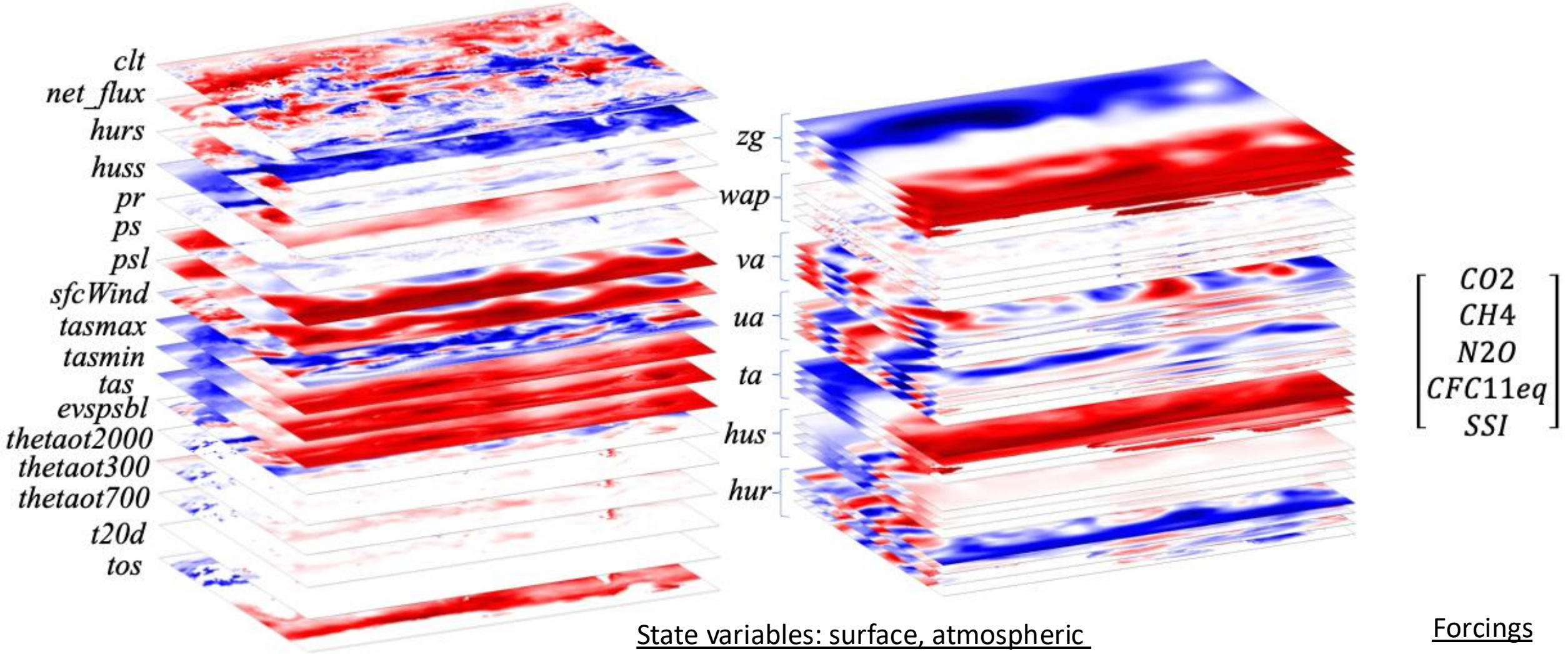
Training: Predict one month ahead

Roll-outs done via one sample each month. Input this output into trained model and repeat.



Clyne et al., *Preliminary results presented at EGU 2025 and Climate Informatics 2025*

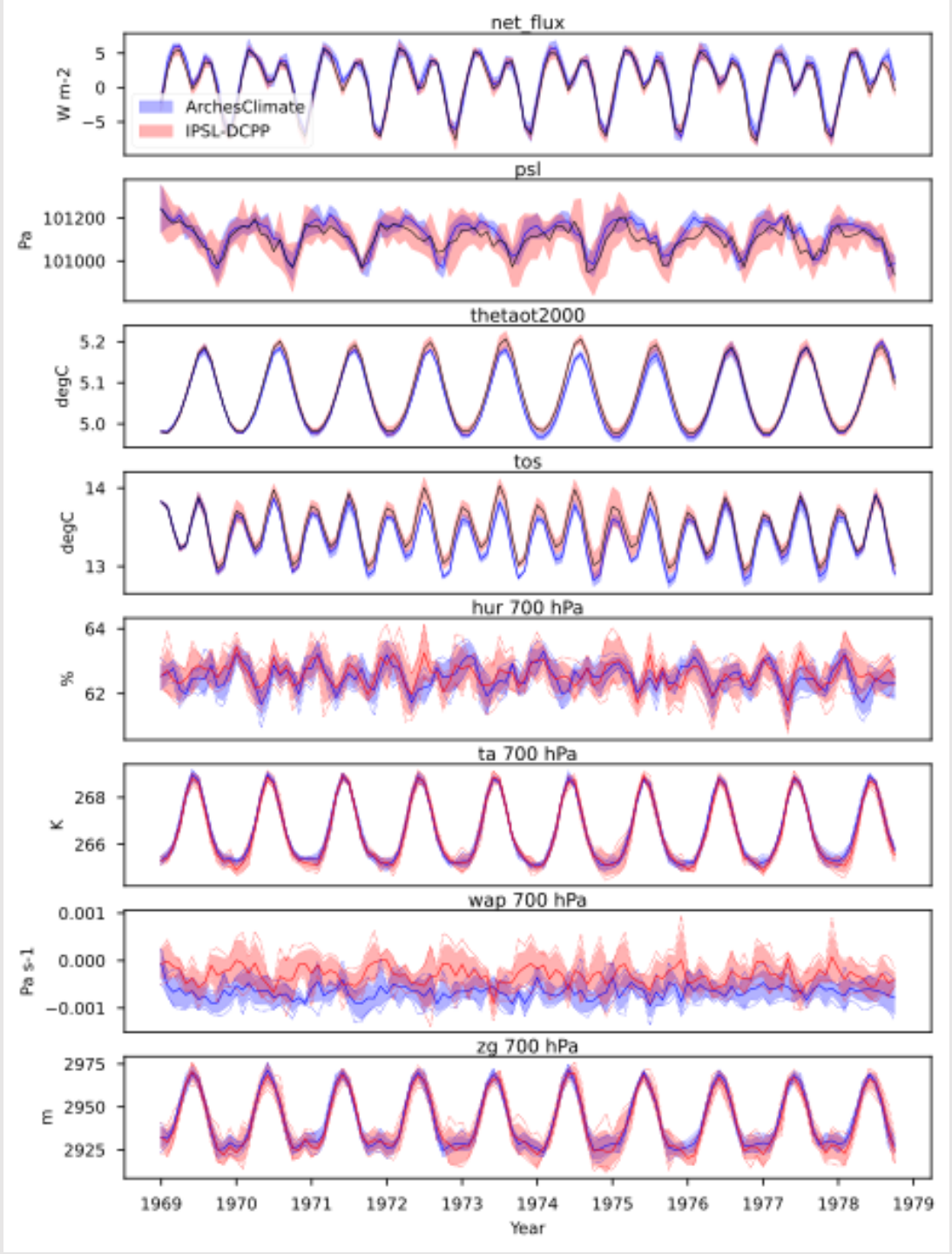
Input variables and forcings



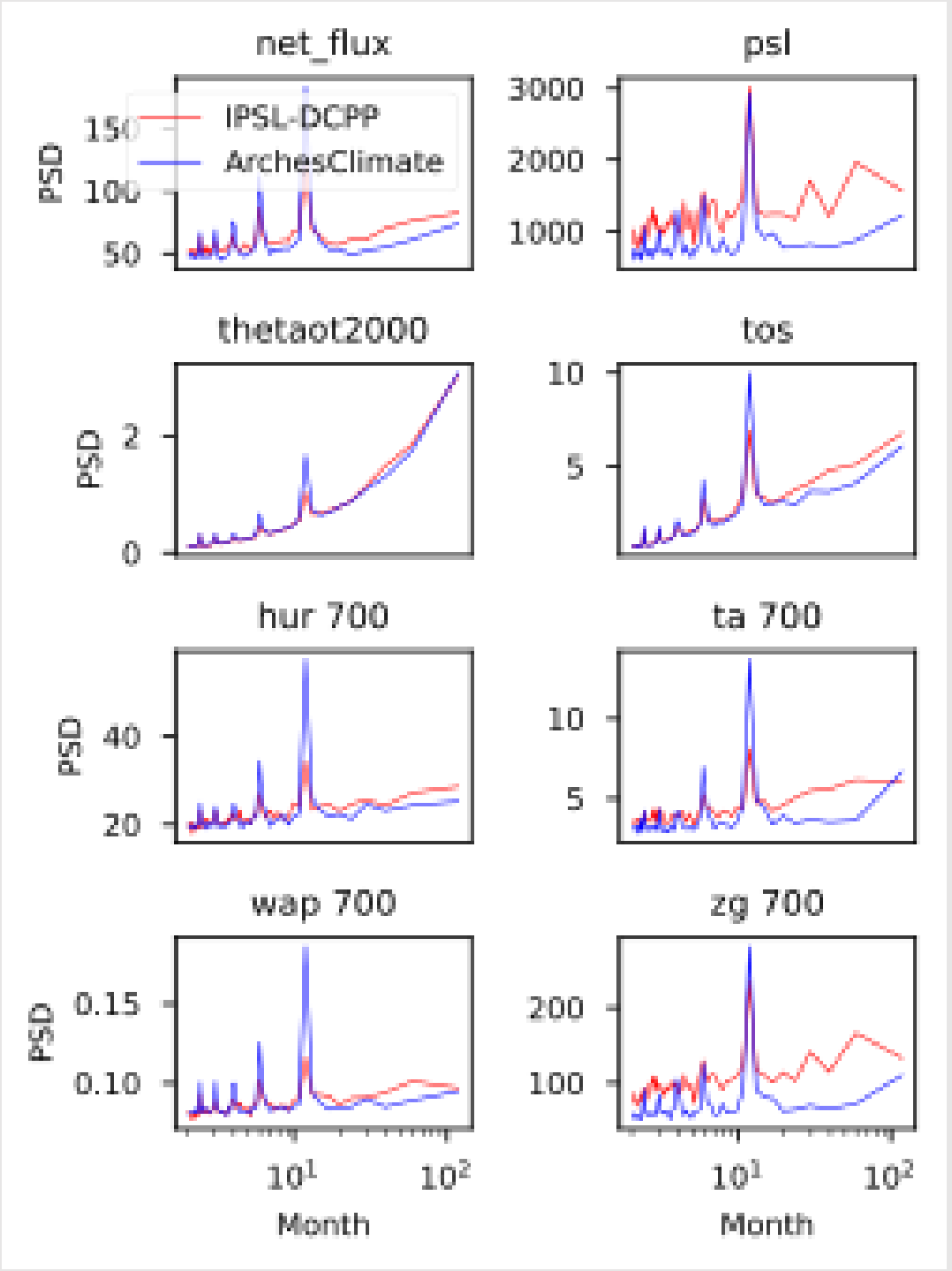
State variables: surface, atmospheric

Forcings

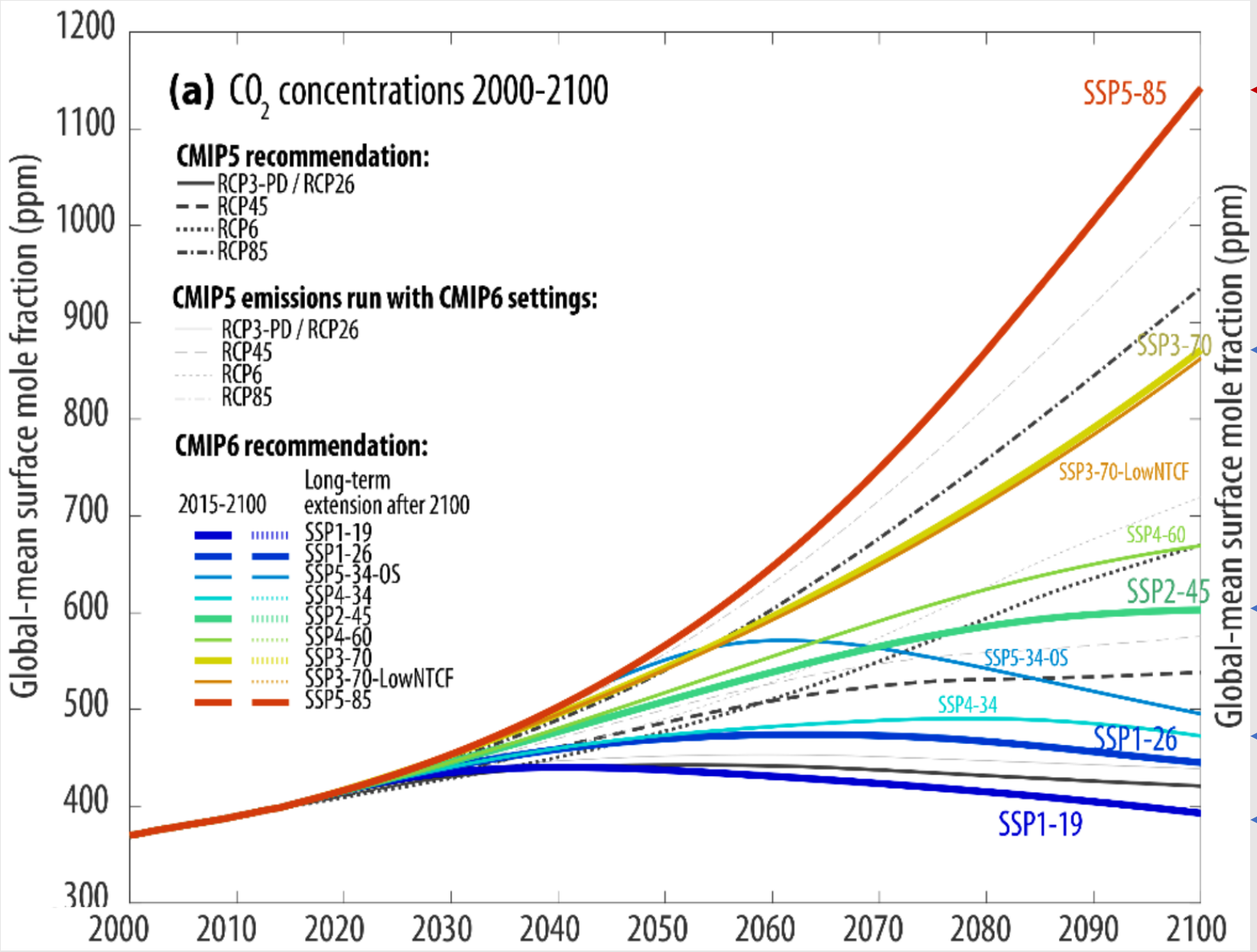
Decadal rollouts

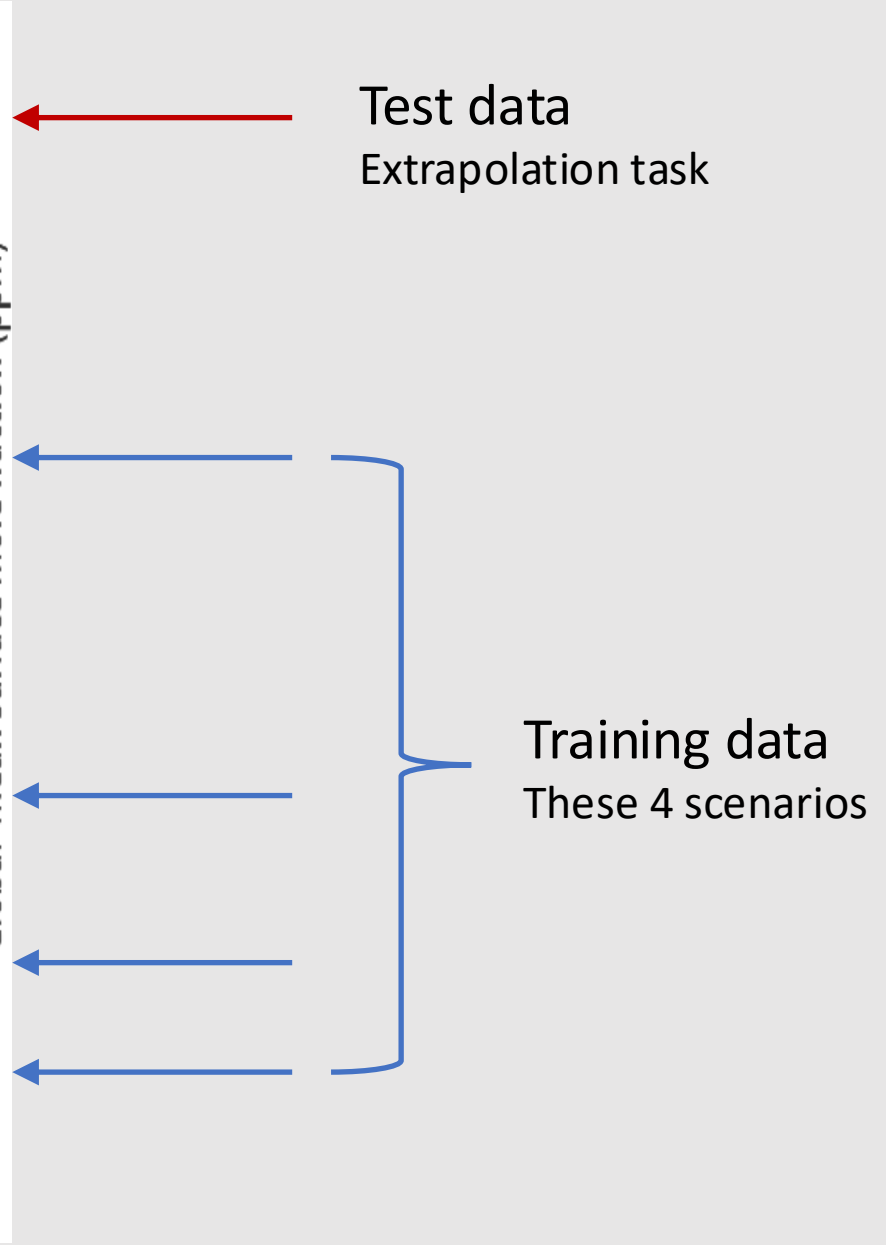


Preliminary results



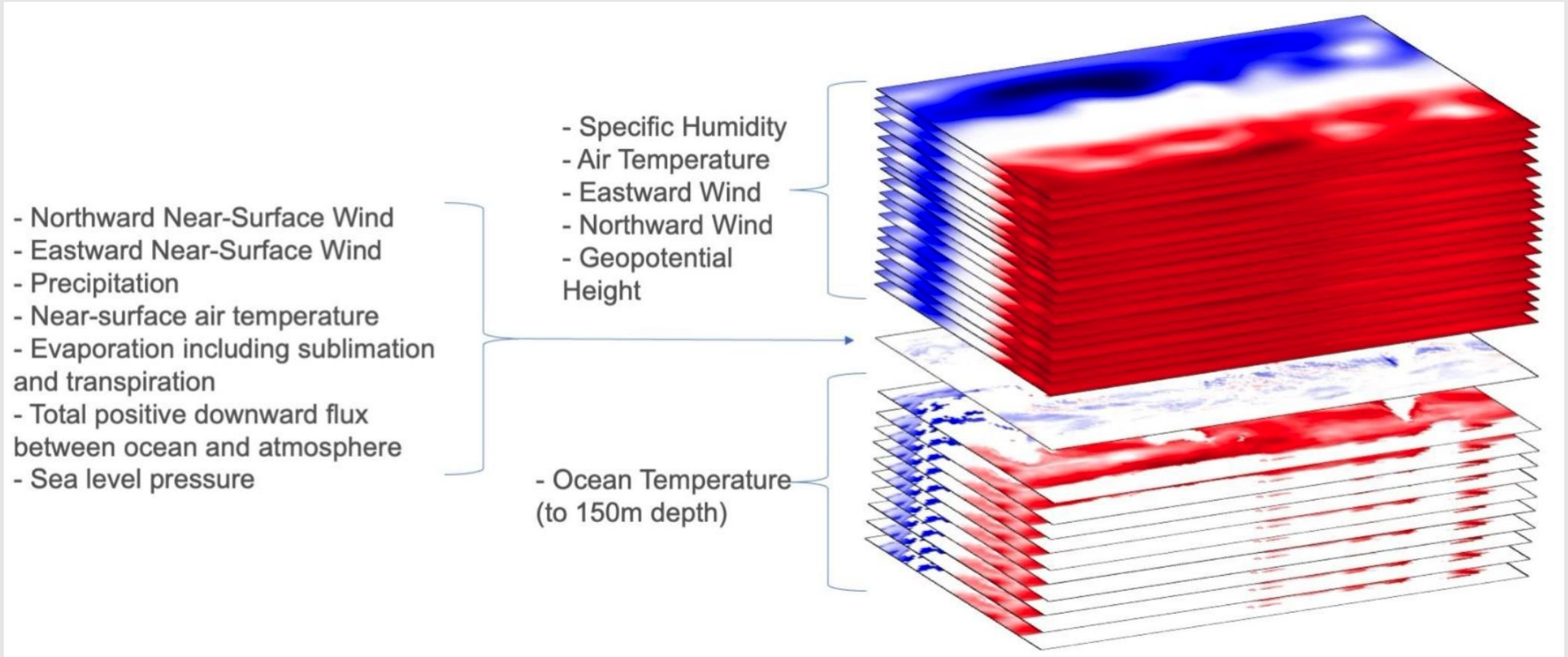
Extrapolating climate model scenarios





Interpolation and extrapolation in climate model scenario space

Clyne et al., *in preparation*



Scenario Extrapolation

Matching temporal patterns of El Niño-Southern Oscillation

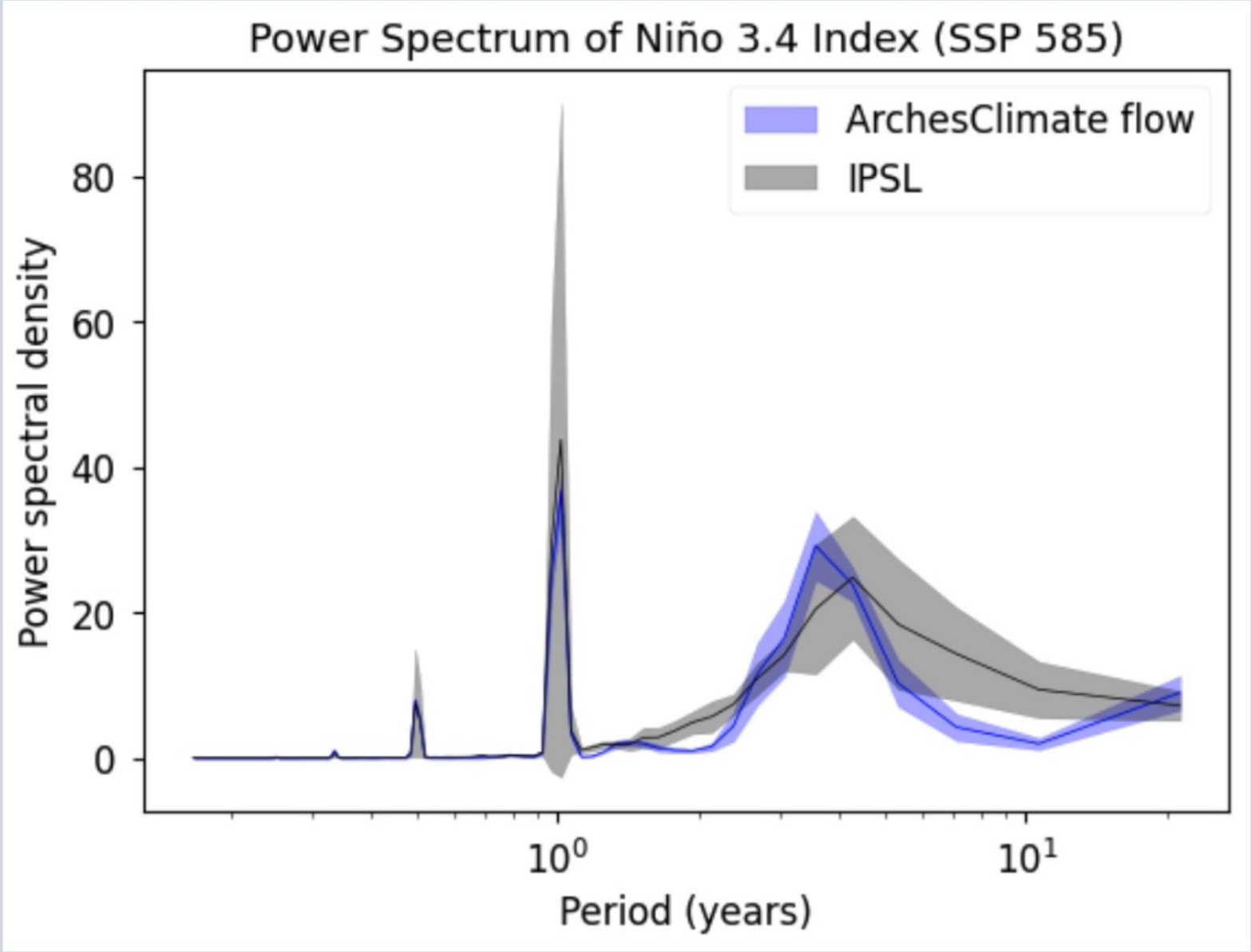


Figure 9. Power spectra for El Niño 3.4 index for SSP5-8.5.

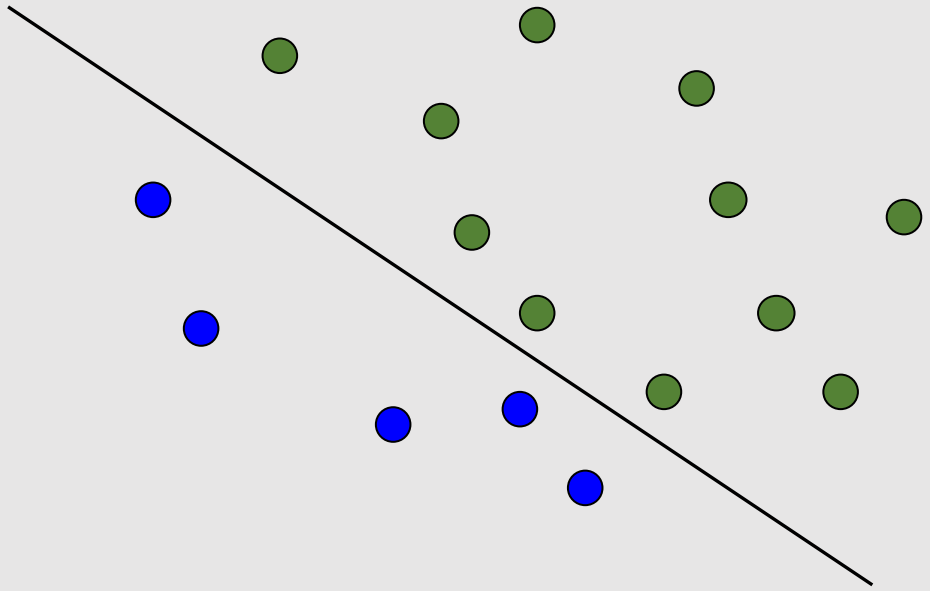
Clyne et al., *in preparation*

Part II: AI Methods:
Self-supervised learning

Supervised Learning

Given labeled data points, find a good classification rule.

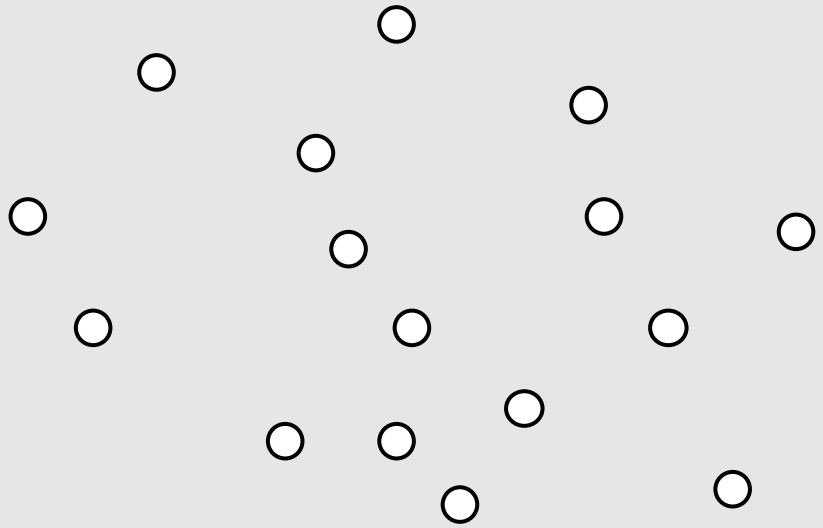
- Describes the data
- Generalizes well



Unsupervised Learning

What can be done without any labels?

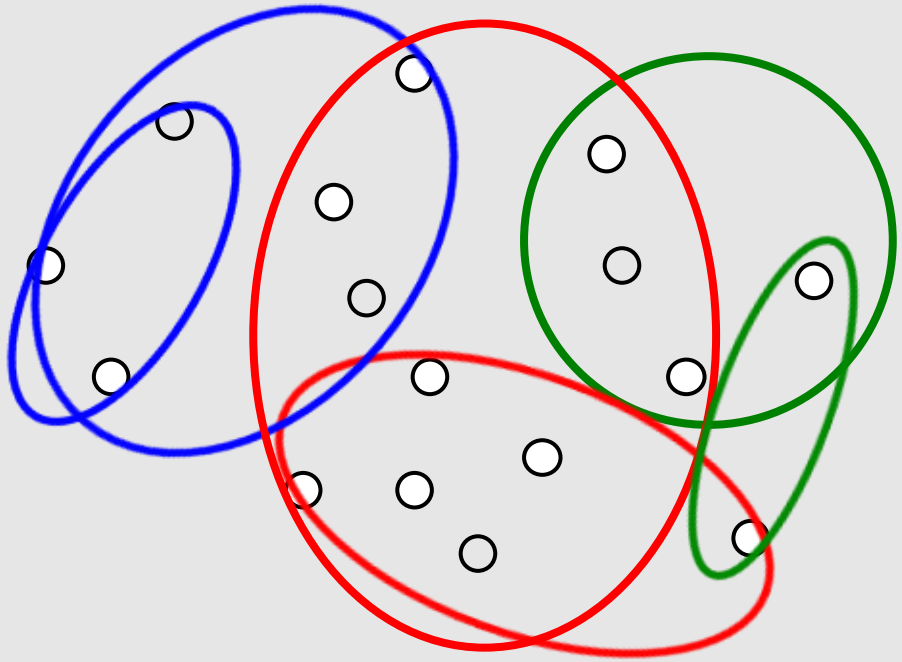
Unsupervised learning



Unsupervised Learning

What can be done without any labels?

Unsupervised learning



Example: Clustering

Learning from “unlabeled" data

- Unsupervised learning tasks
  1. Clustering
  2. Embedding (dimensionality reduction, including non-linear methods)
  3. Interpolation
  4. Learning a generative model

- There are deep learning (DL) approaches to all of these tasks

- There are non-DL approaches to all of these tasks

- There are self-supervised DL approaches to all of these tasks

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:

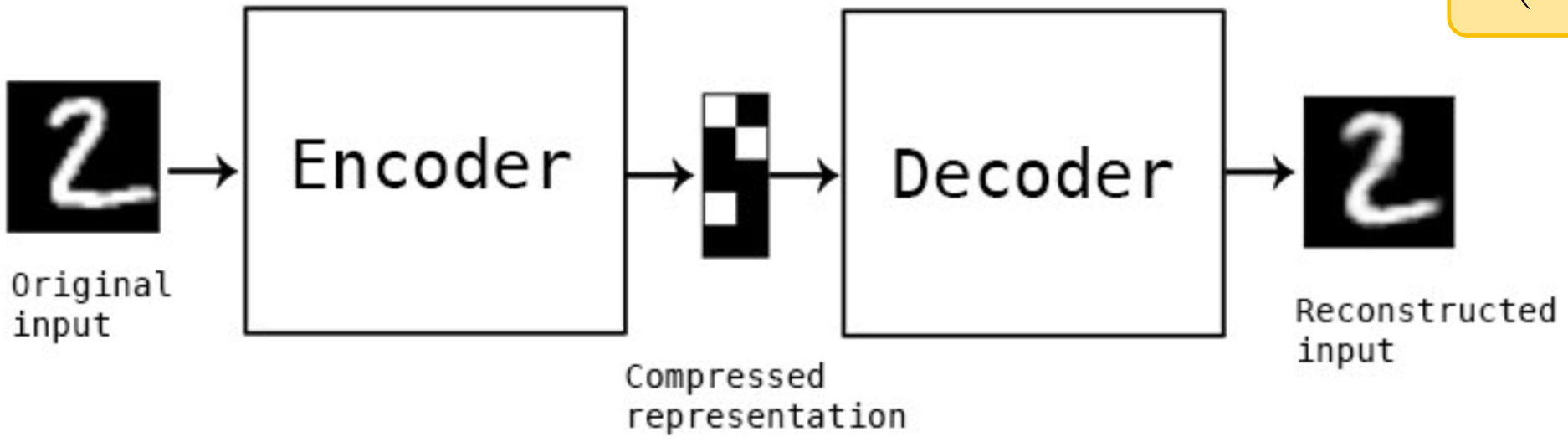
- Design a supervised learning task using only the available data.
- Train a model on this task such that,
- the learned features (or the learned posterior over a feature space) will be useful for another (down-stream) task.

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

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



Self-supervised learning

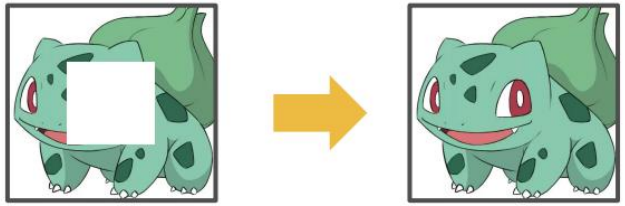
- Design a pretext task that will cause useful features to be learned.

Vision Pretext Tasks: Masked Prediction

- **Denoising autoencoder** (Vincent et al. 2008)
  - Add noise = Randomly mask some pixels
  - Only reconstruction loss



- **Context autoencoder** (Pathak et al. 2016)
  - Mask a random region in the image
  - Reconstruction loss + adversarial loss



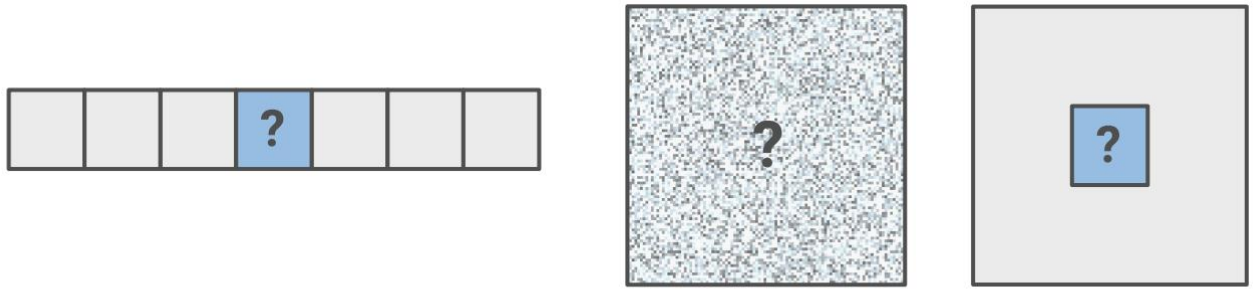
Lilian Weng, Jong Wook Kim NeurIPS 2021 Tutorial

Self-supervised learning

- Design a pretext task that will cause useful features to be learned.

Self-Prediction: Masked Generation

We mask a random portion of information and pretend it is missing, irrespective of the natural sequence. The model learns to predict the missing portion given other unmasked information.



Examples:

- Masked language modeling (BERT)
- Images with masked patch (denoising autoencoder, context autoencoder, colorization)

Lilian Weng, Jong Wook Kim NeurIPS 2021 Tutorial

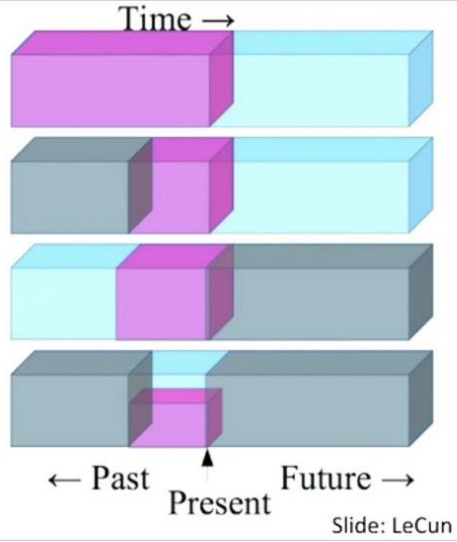
Self-supervised learning

- Design a pretext task that will cause useful features to be learned.

Self-Prediction

Self-prediction construct prediction tasks within every individual data sample: to predict a part of the data from the rest while pretending we don't know that part.

- Predict any part of the input from any other part.
- Predict the future from the past.
- Predict the future from the recent past.
- Predict the past from the present.
- Predict the top from the bottom.
- Predict the occluded from the visible
- Pretend there is a part of the input you don't know and predict that.



(Famous illustration from Yann LeCun)

Lilian Weng, Jong Wook Kim NeurIPS 2021 Tutorial

Self-supervised learning: Contrastive learning

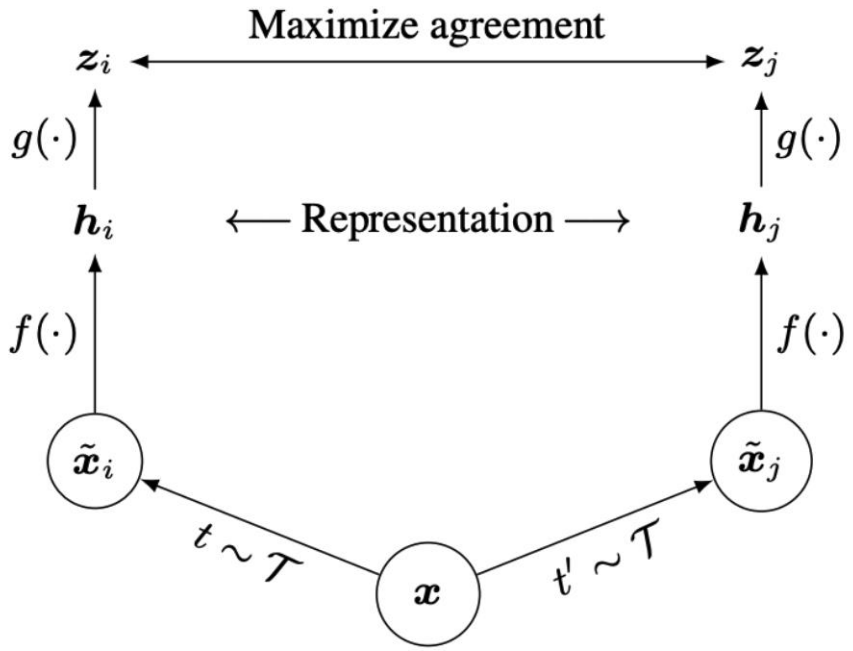


Figure 2. A simple framework for contrastive learning of visual representations. Two separate data augmentation operators are sampled from the same family of augmentations ($t \sim \mathcal{T}$ and $t' \sim \mathcal{T}$) and applied to each data example to obtain two correlated views. A base encoder network $f(\cdot)$ and a projection head $g(\cdot)$ are trained to maximize agreement using a contrastive loss. After training is completed, we throw away the projection head $g(\cdot)$ and use encoder $f(\cdot)$ and representation $\boldsymbol{h}$ for downstream tasks.

Pretext task:

1. Augment input x twice: $\tilde{x}_i, \tilde{x}_j$
2. Learn a mapping , $g(f(\cdot))$, by minimizing a loss function to bring $g(f(\tilde{x}_i)), g(f(\tilde{x}_j))$ close.

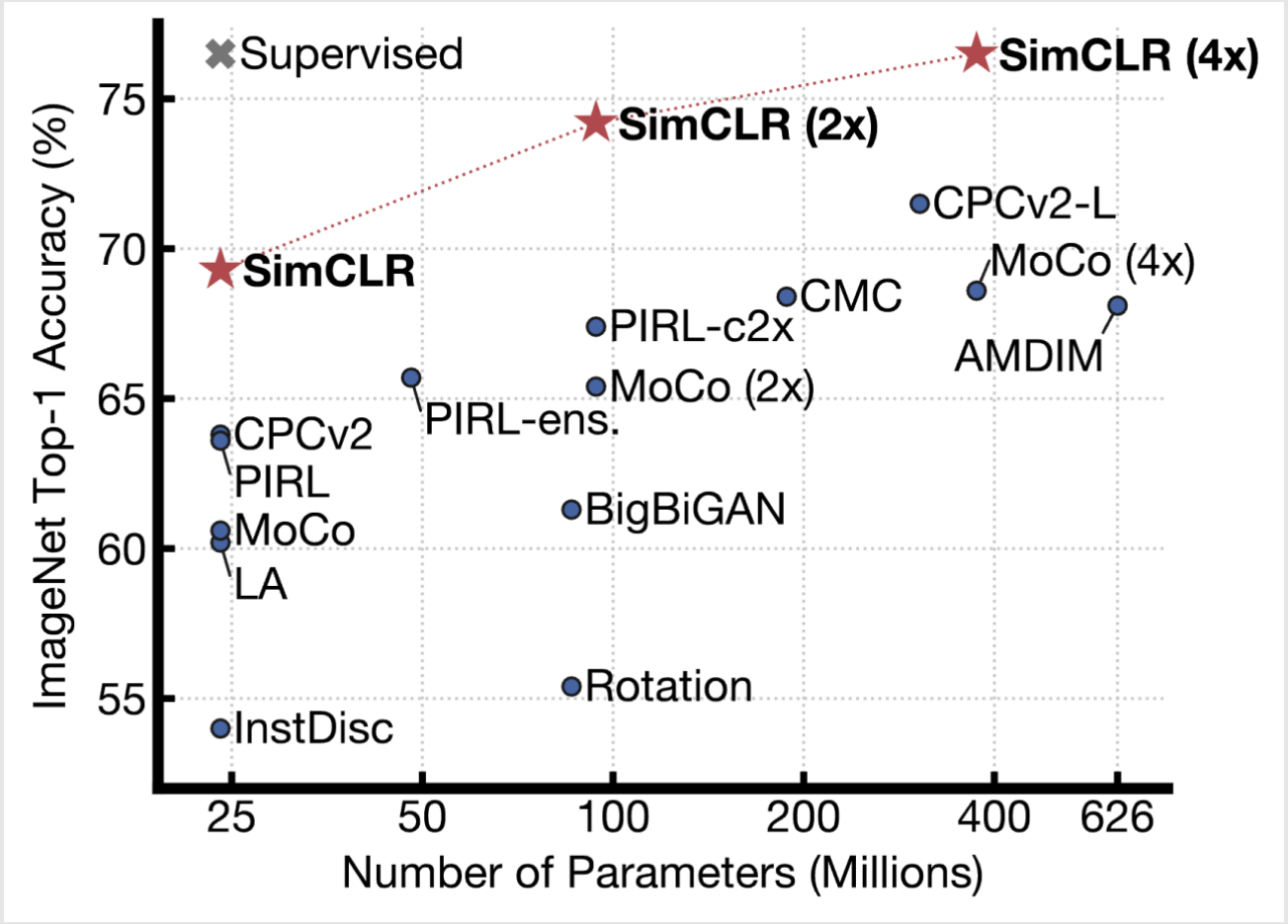
$$\mathcal{L}(\hat{x}, x)$$

$$\hat{x} := \{g(f(\tilde{x}_i)), g(f(\tilde{x}_j))\}$$

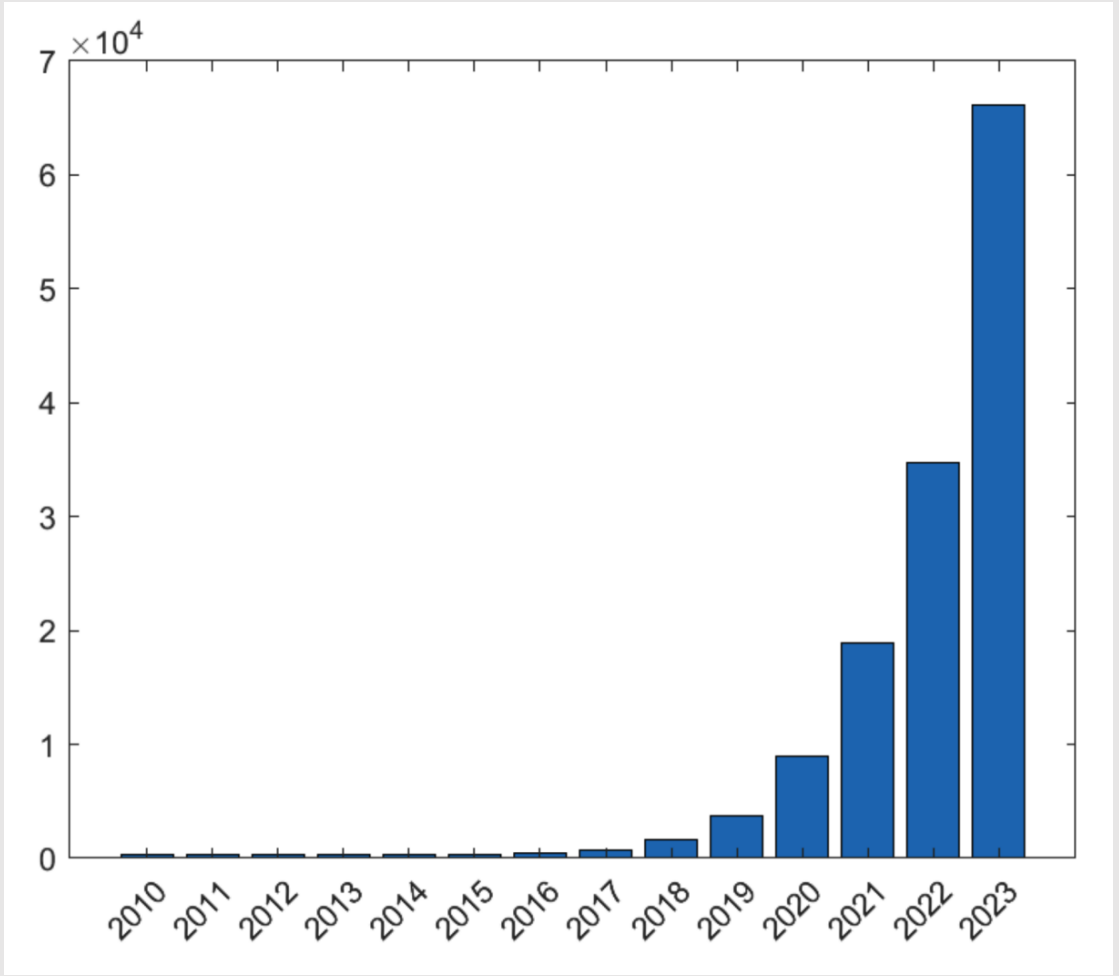
$$\tilde{x} := aug_t(x), t \sim \mathcal{T}$$

SimCLR : [Chen et al., *ICML* 2020]

Self-supervised learning can outperform supervised



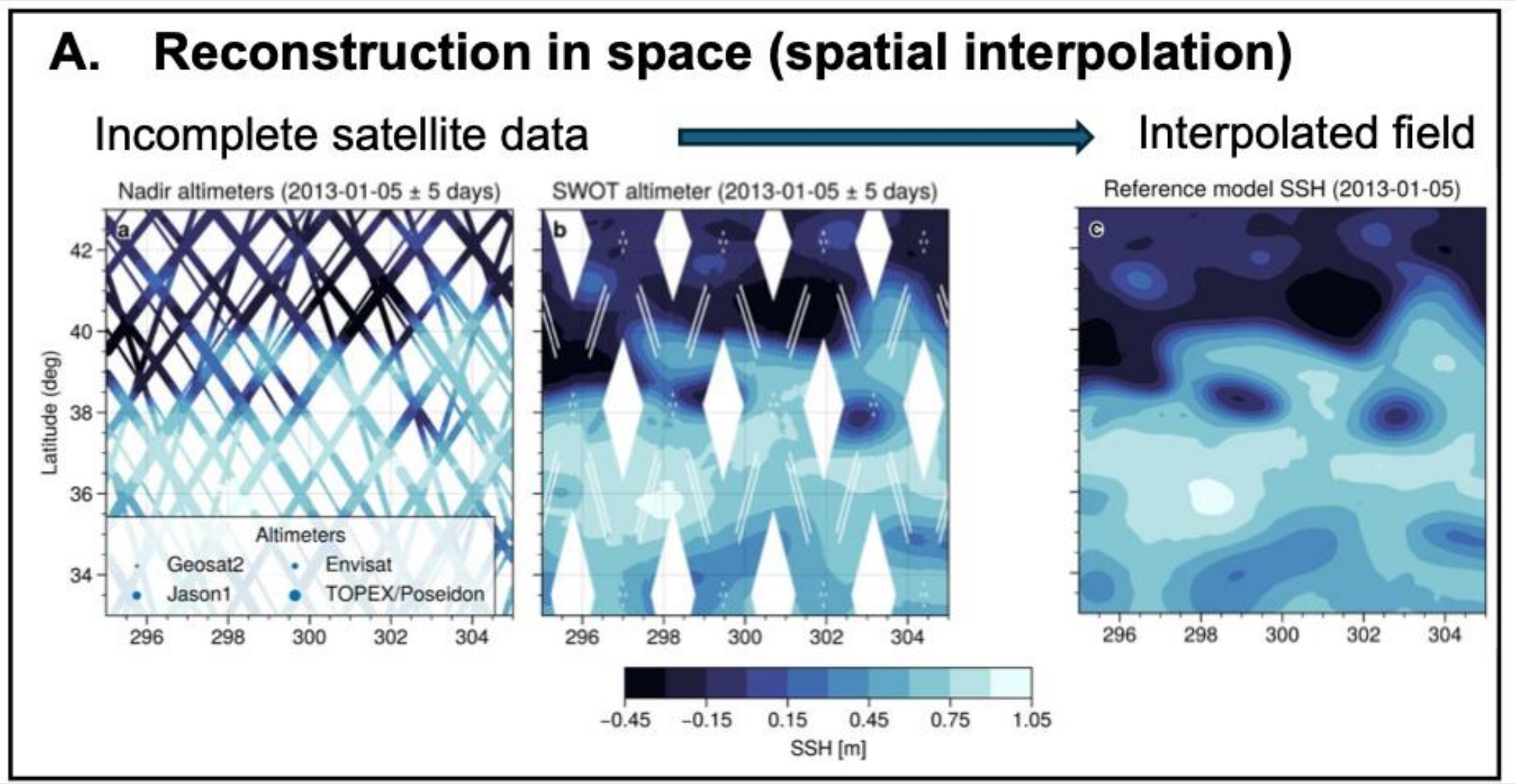
Comparison of self-supervised methods



Publications on self-supervised learning

SimCLR : [Chen et al., *ICML* 2020]

Self-supervised learning for your applications



[Bracco et al., Machine Learning for the Physics of Climate. *Nature Reviews Physics*, 2024]

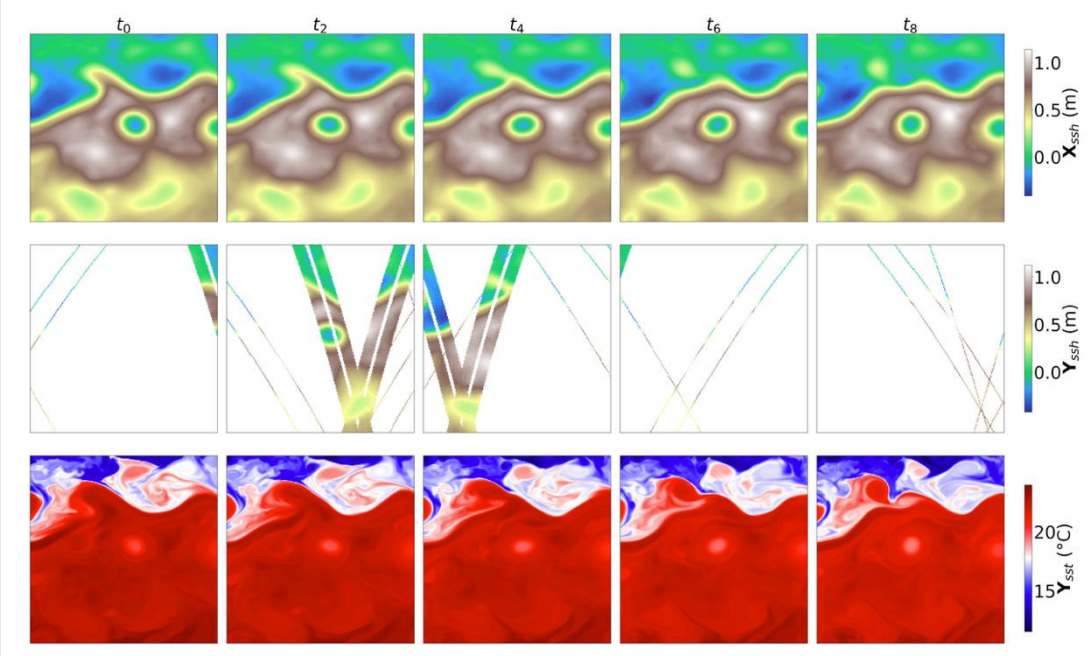
Unsupervised spatiotemporal interpolation

MUSTI: Multi-modal Unsupervised Spatio-Temporal Interpolation of satellite ocean altimetry maps. [Archambault et al. VISAPP 2023]

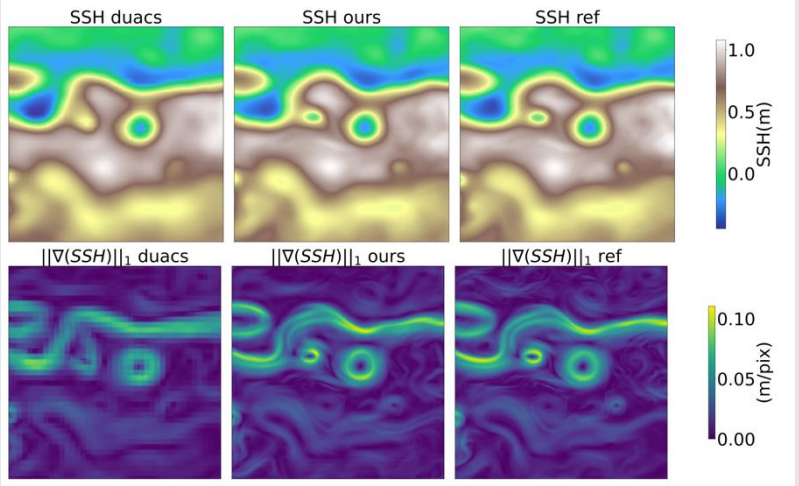
Ground truth SSH fields (not used in training)

SSH data along satellite tracks

SST fields from physical model



Training Data



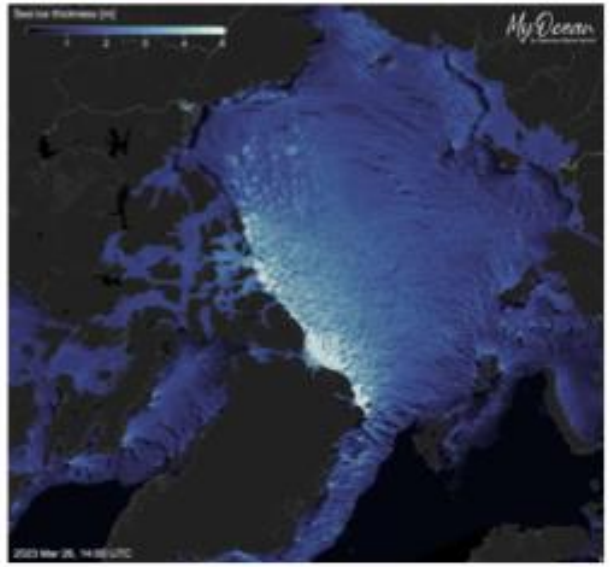
Results

Self-supervised learning for your applications

C. Reconstruction in scales (downscaling)

Low resolution field → High resolution field





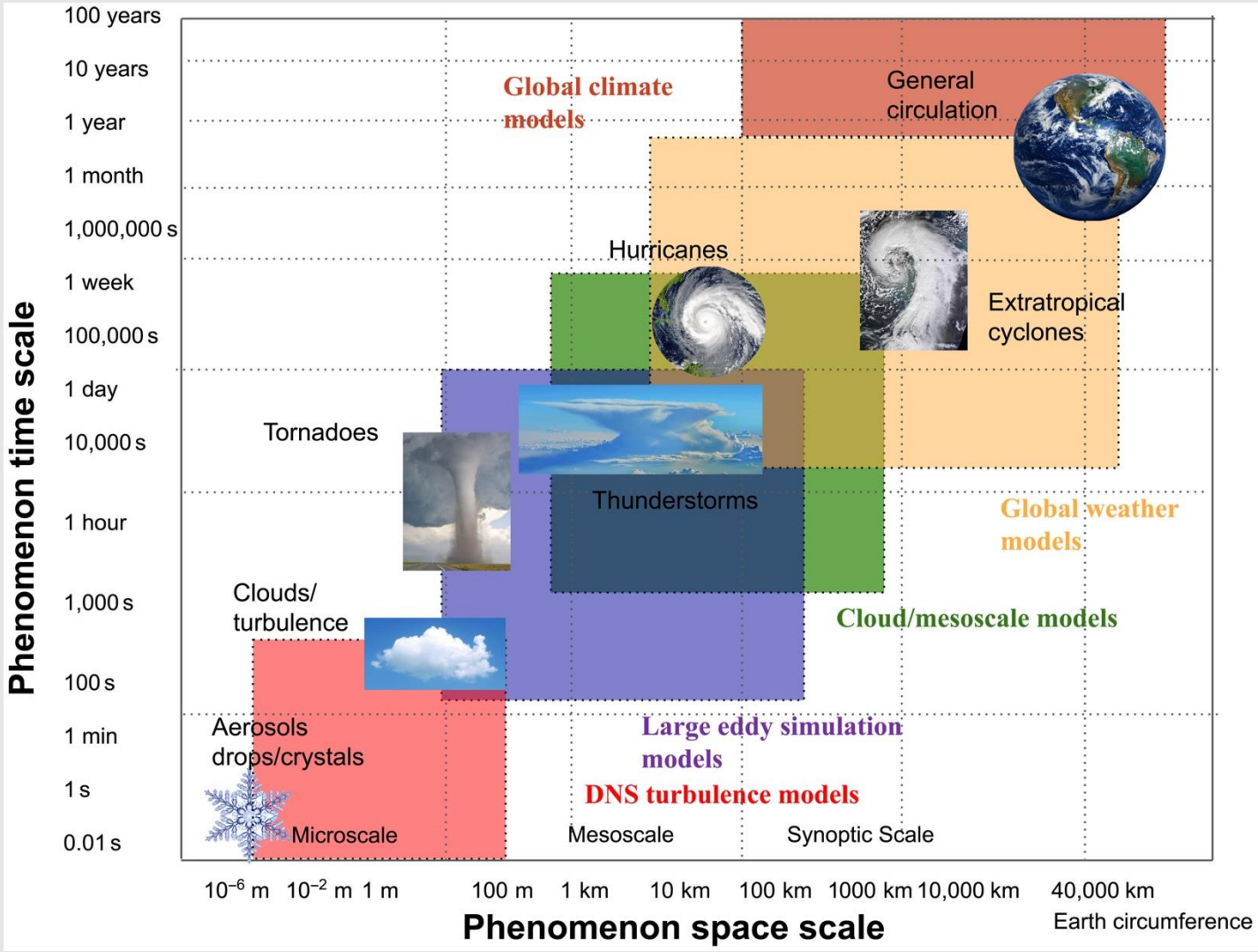
[Bracco et al., Machine Learning for the Physics of Climate. *Nature Reviews Physics*, 2024]

Downscaling climate model simulations

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

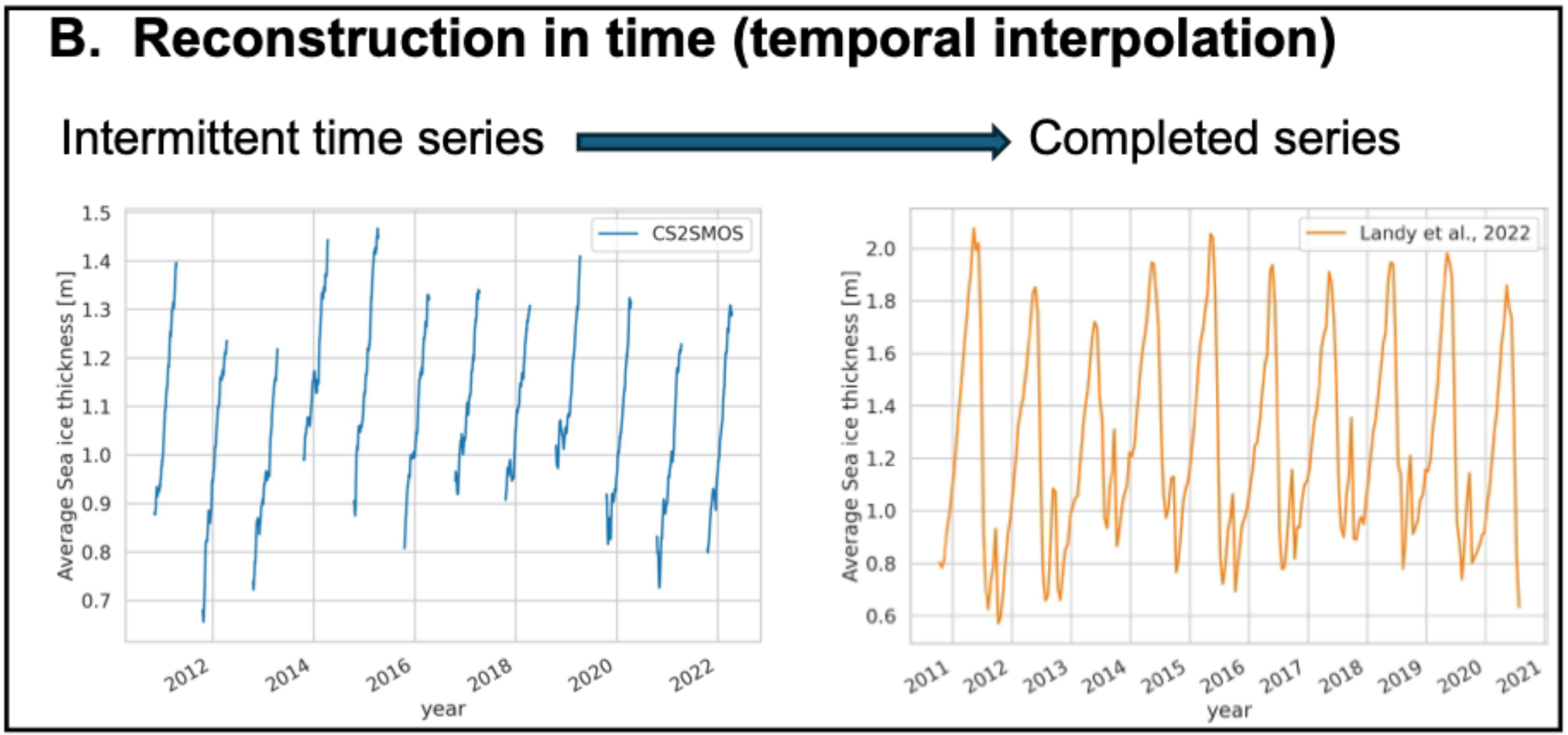
- Climate change adaptation
- Climate change mitigation
- Projecting long-term impacts

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



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

Self-supervised learning for your applications



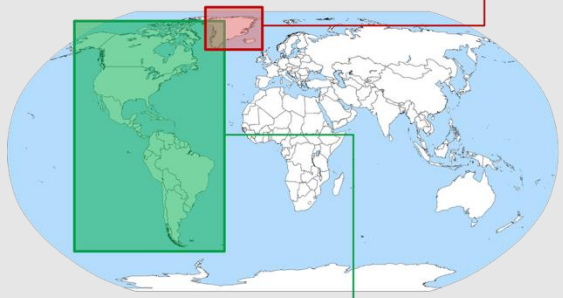
[Bracco et al., Machine Learning for the Physics of Climate. *Nature Reviews Physics*, 2024]

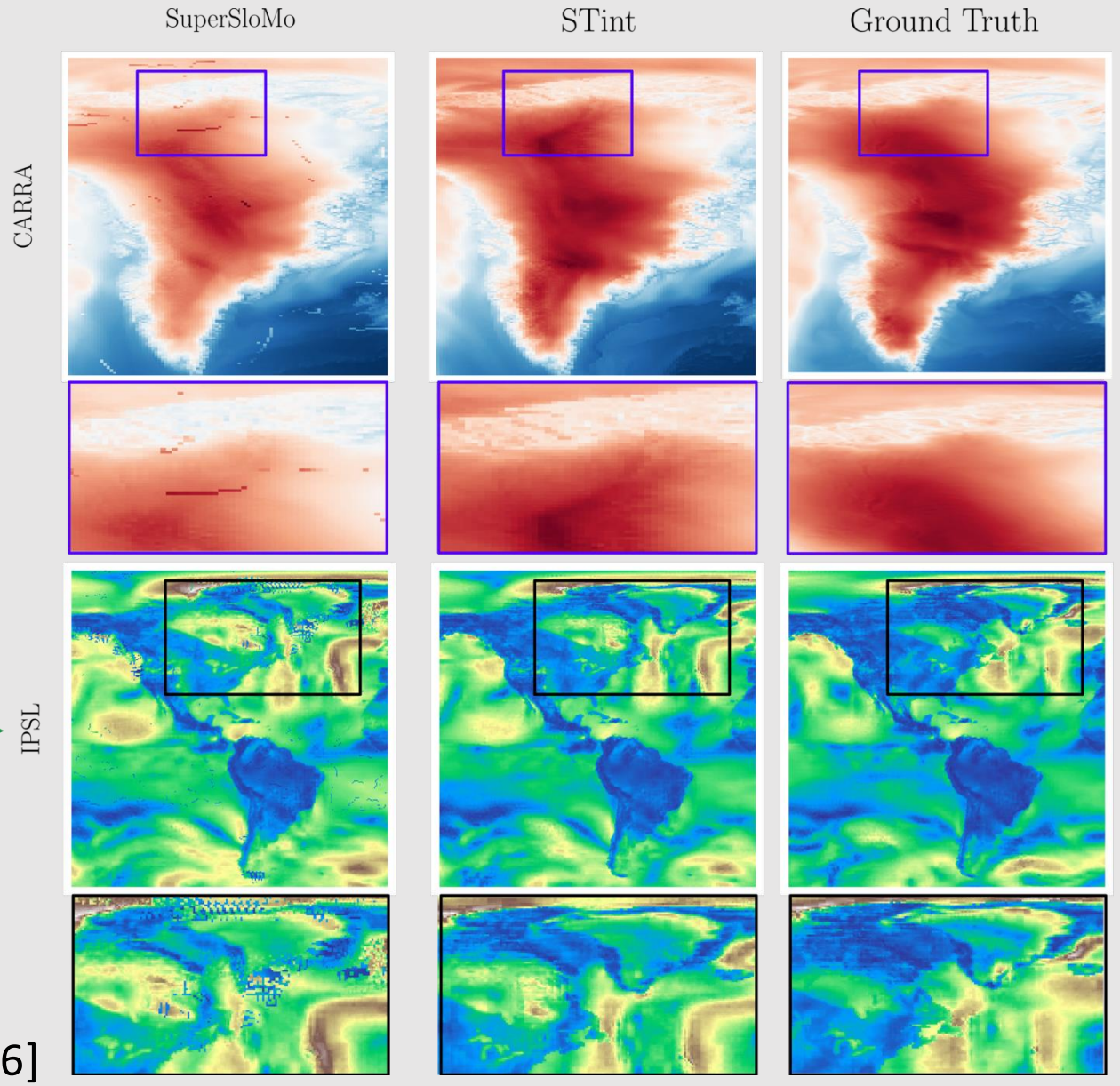
Example: Temporal downscaling (interpolation) of geospatial data

Why can’t we just use existing AI methods?

- 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, 2026]

Limitations of previous work: Bicubic Interpolation

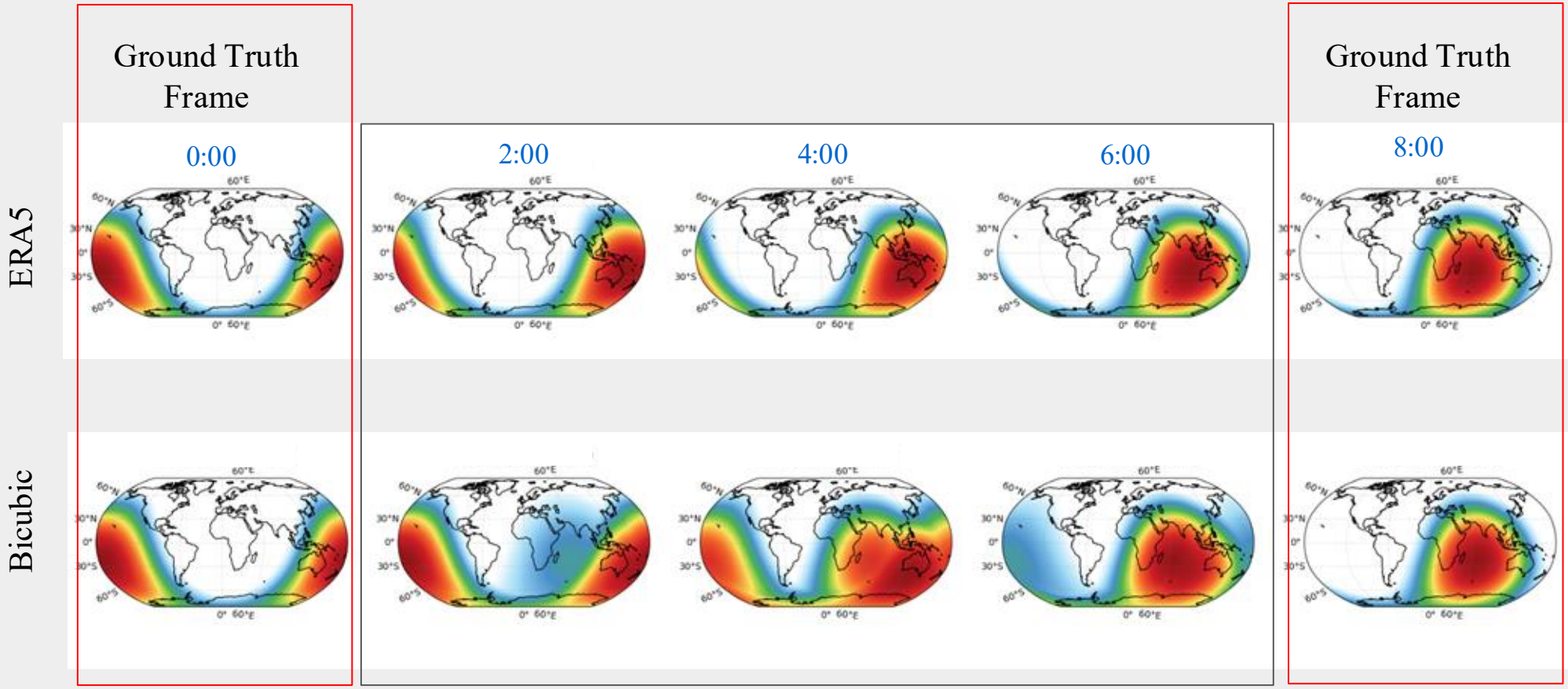
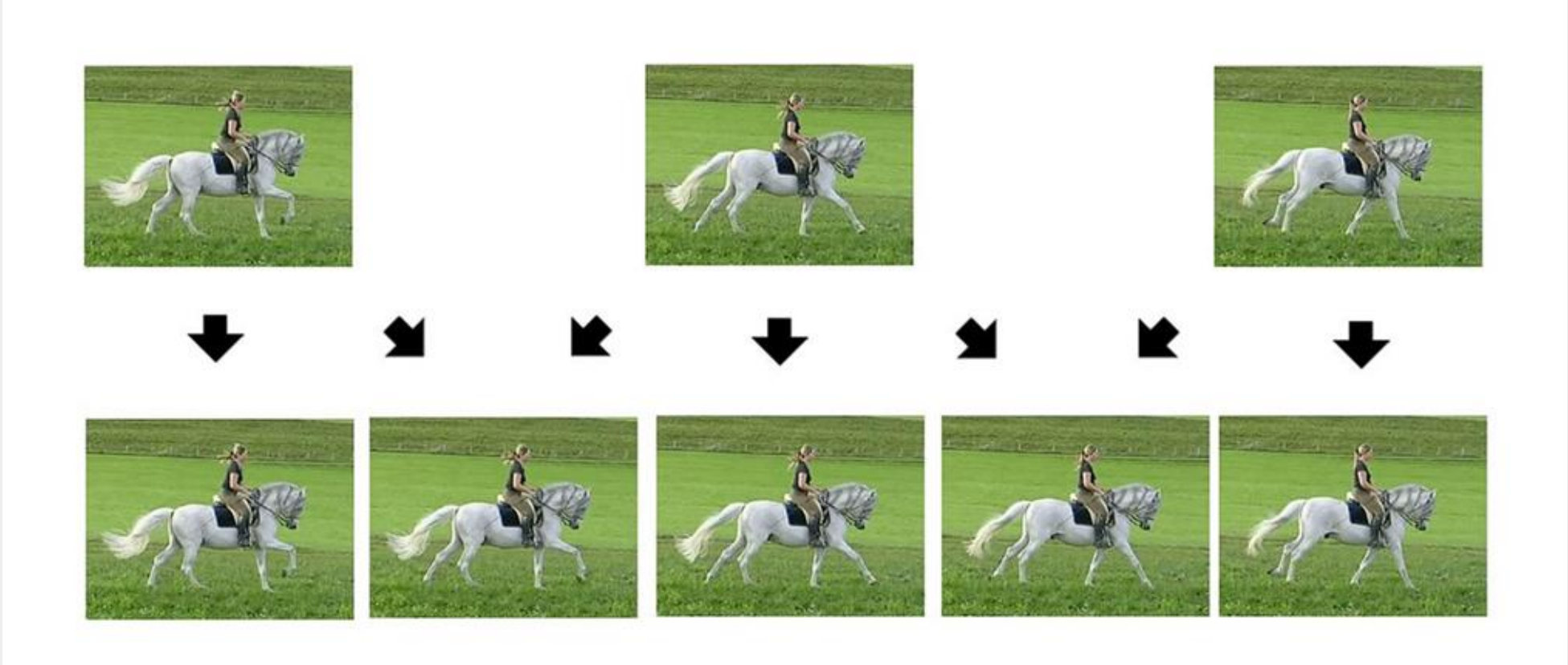


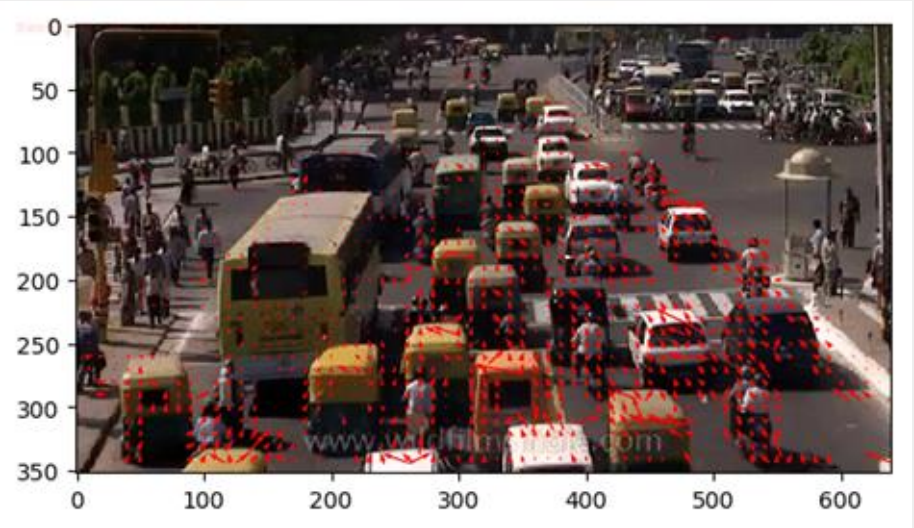
Figure. *Temporal bicubic interpolation of ERA5 solar irradiance fields*

Video Interpolation in Computer Vision

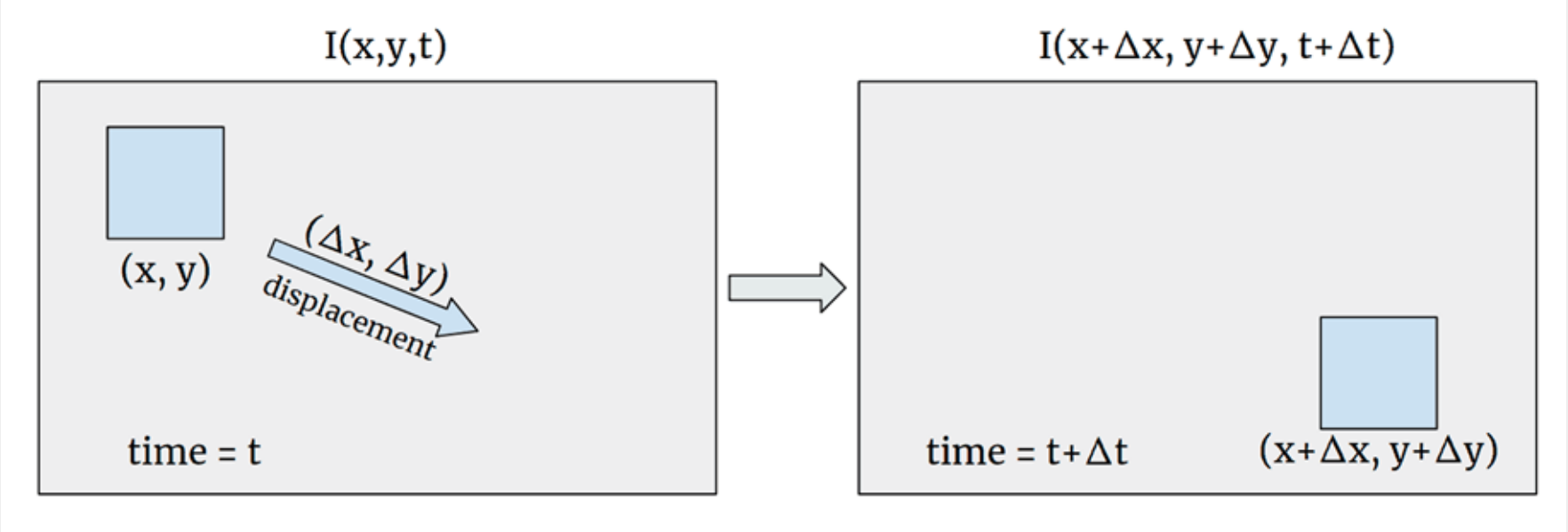


From Computer Vision: Optical Flow

- Optical flow: pixel-wise motion field between two images (e.g. video frames).
- It encodes the “perceived” 2D projection of the physical movement of points relative to the observer.



(source: Nvidia OF SDK)



From Computer Vision: Optical Flow

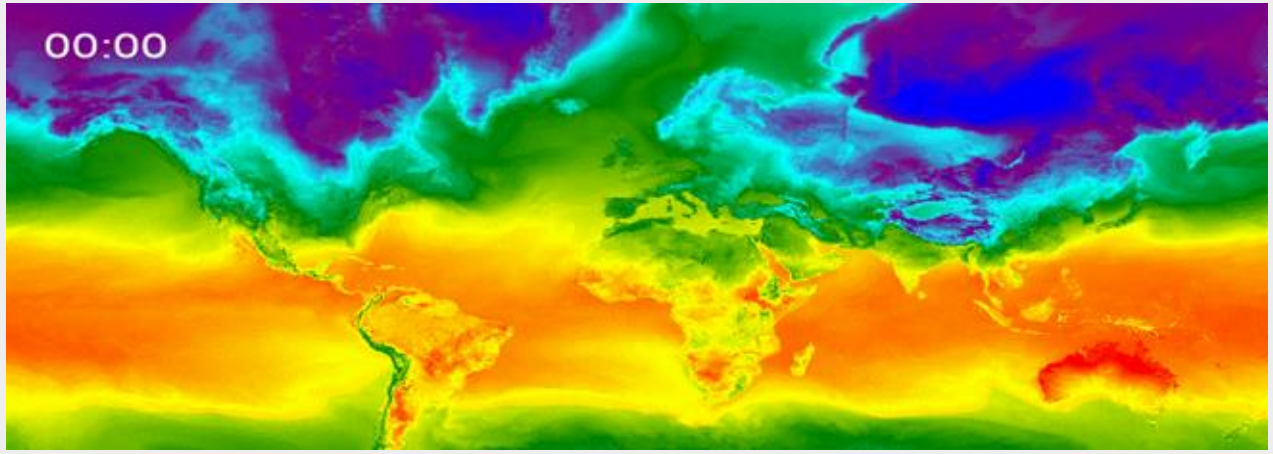
Assumptions:

- Small movement (Δt).
- No/Small changes in reflected light.
- No/Small changes in scale.
- No occlusion or disocclusion.



Not applicable for geospatial data!

- No concept of objects or depth.
- Entities can form/deform arbitrarily.
- Much coarser temporal resolution.



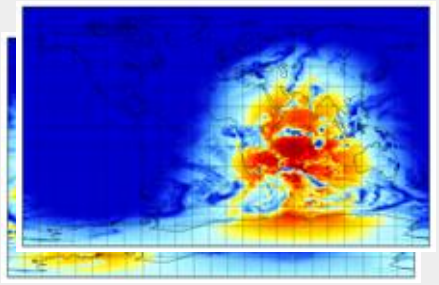
NOAA GFS Hourly Temperature

Goal: Enhancing Temporal Resolution

- The task is Temporal interpolation
- But we <u>don't have 1-hourly data</u> to train on

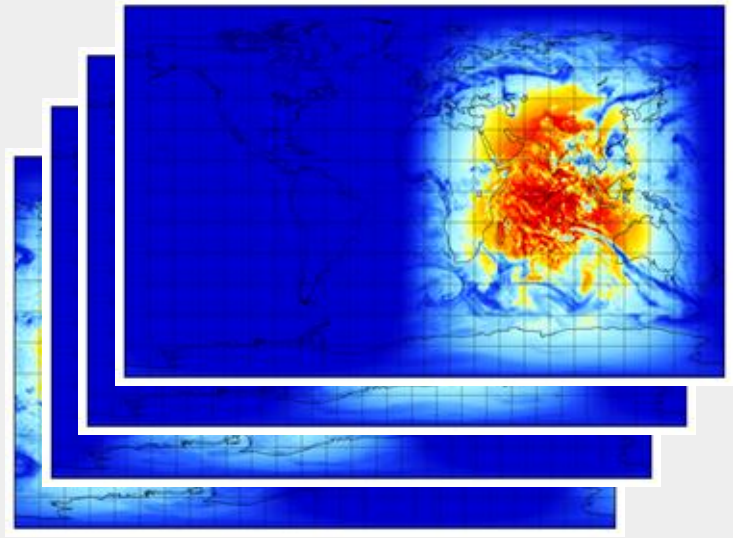
→ need unsupervised (**Self-supervised**) techniques.

CMIP6-IPSL



3-hourly

?



1-hourly

Self-supervised Learning recap

- Use pretext/surrogate/pseudo tasks in a supervised fashion.
  - Automatically **generated** labels.

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

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

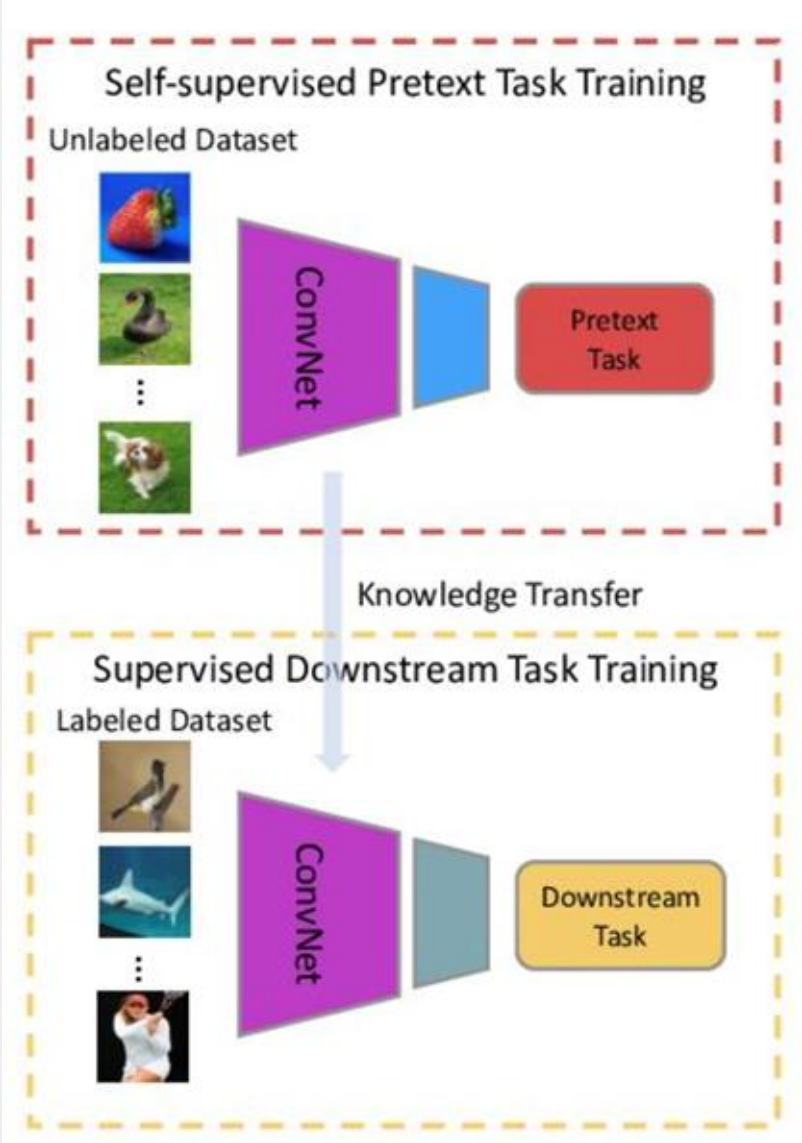
Supervised learning

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

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

Self-supervised learning

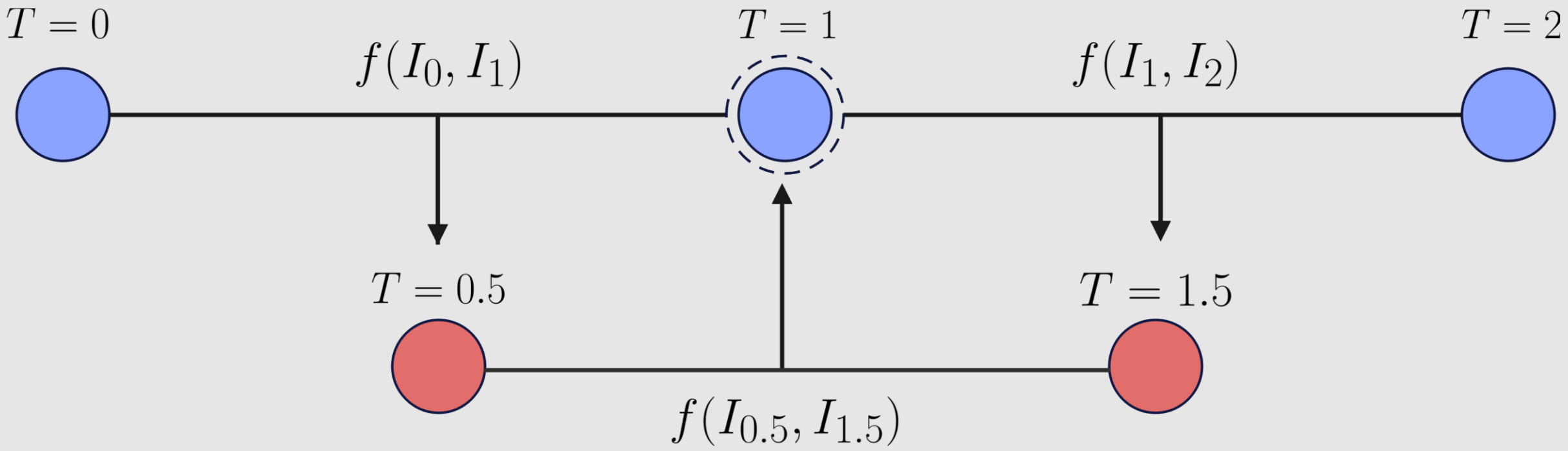
- Downstream task: retrieval, supervised or semi-supervised classification etc.



A pretext task for temporal interpolation

STINT: Self-supervised Temporal Interpolation for Geospatial Data

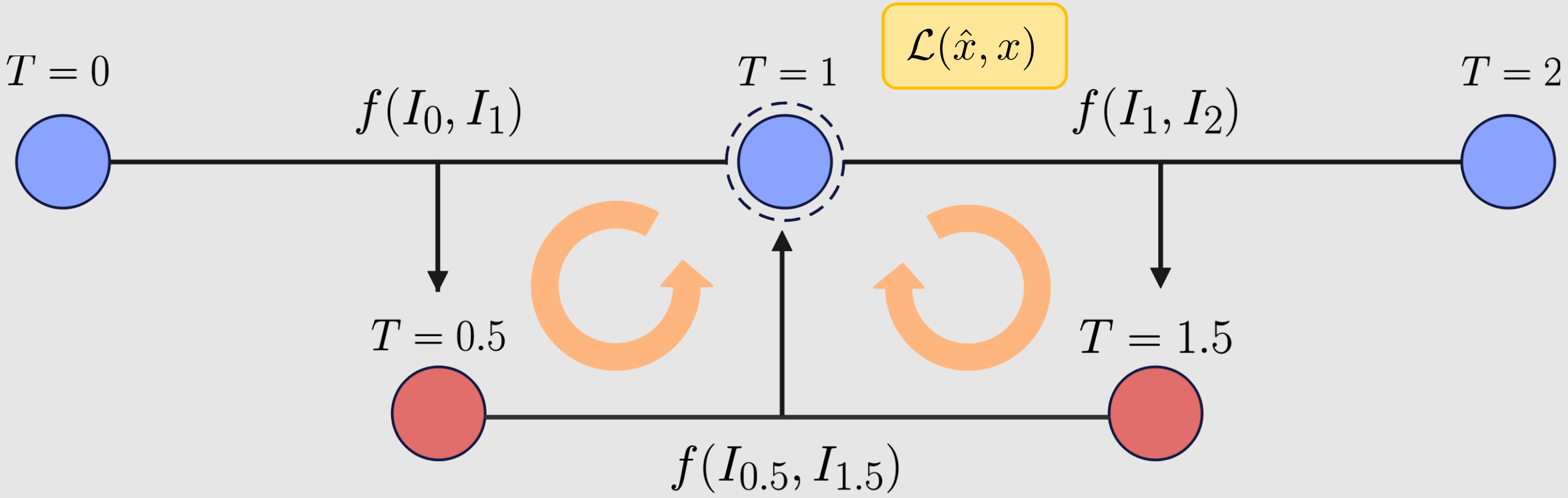
[Harilal, Hodge, Subramanian, & Monteleoni, Climate Informatics 2026]



A pretext task for temporal interpolation

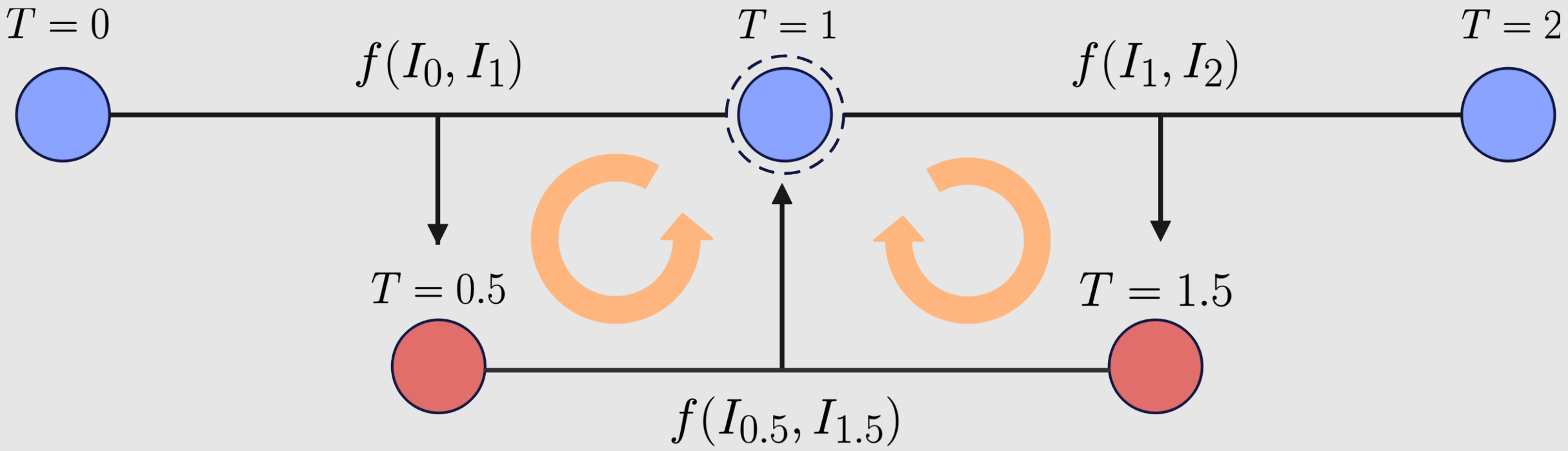
STINT: Self-supervised Temporal Interpolation for Geospatial Data

[Harilal, Hodge, Subramanian, & Monteleoni, Climate Informatics 2026]



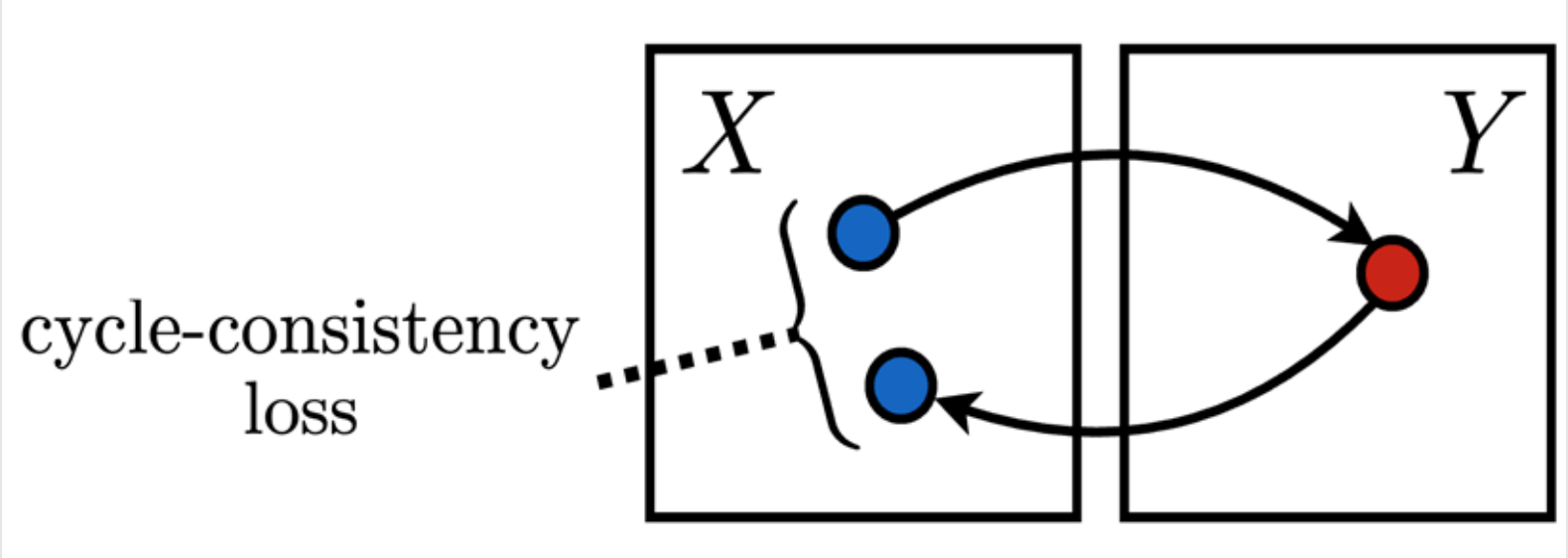
A pretext task for temporal interpolation

$$\mathcal{L}(\hat{x}, x) = \sum_{i=1}^{n} \mathcal{L}(T_i, \hat{T}_i) \qquad \hat{T}_i := f(f(T_{i-1}, T_i), f(T_i, T_{i+1}))$$

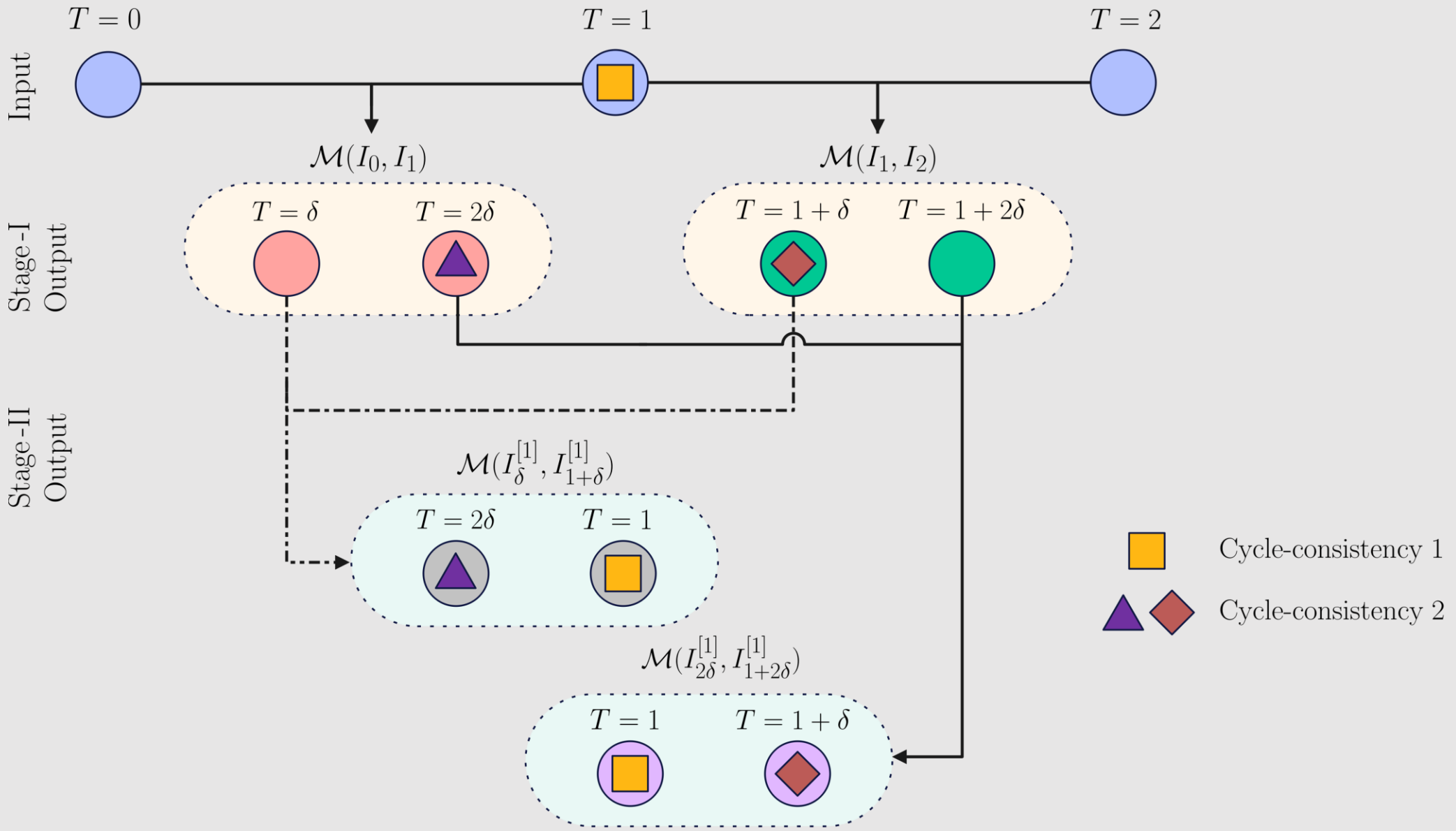


Cycle-consistency loss

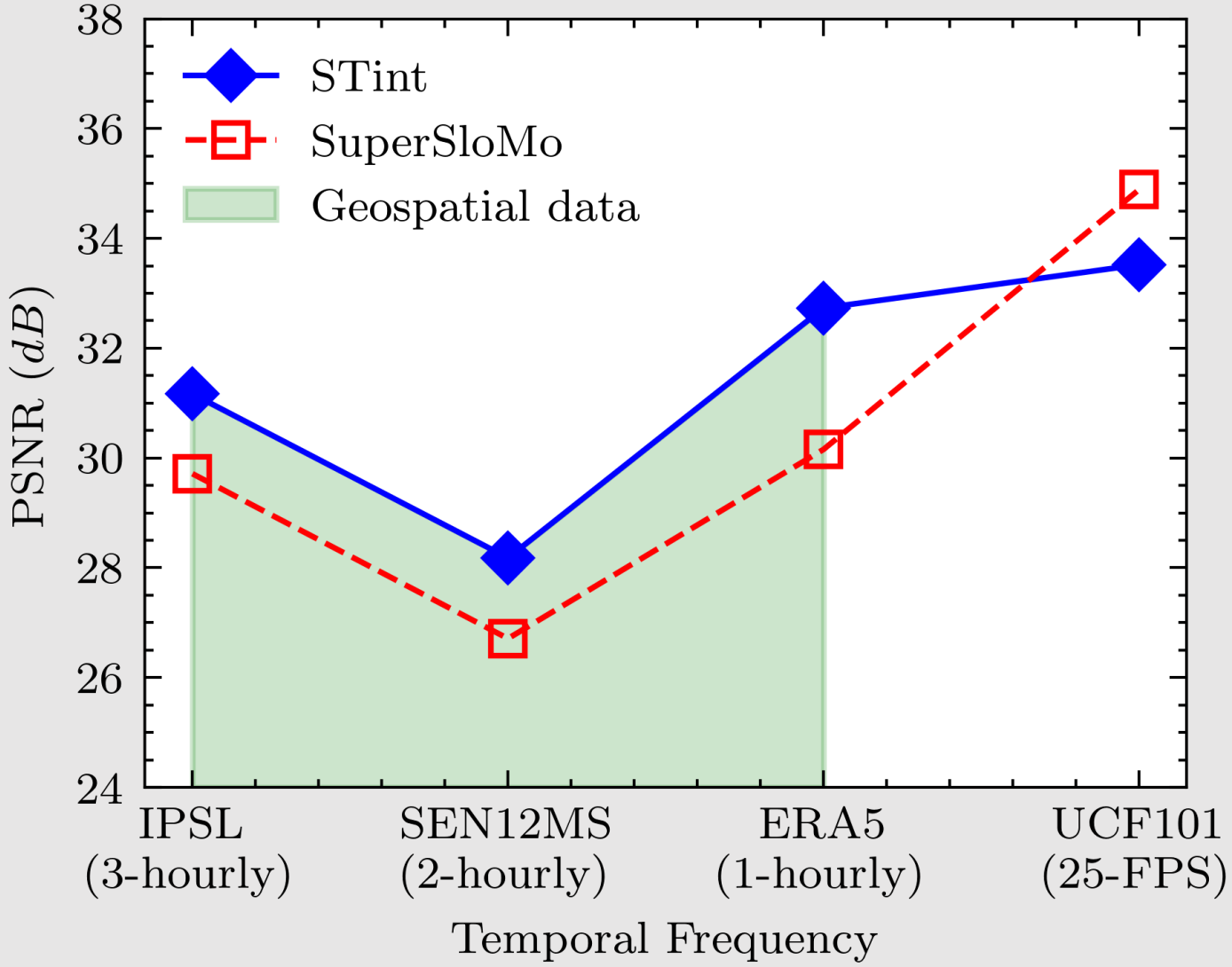
$$\mathcal{L}(\hat{x}, x) = \sum_{i=1}^{n} \mathcal{L}(T_i, \hat{T}_i) \qquad \hat{T}_i := f(f(T_{i-1}, T_i), f(T_i, T_{i+1}))$$



STINT: Self-supervised Temporal Interpolation



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**

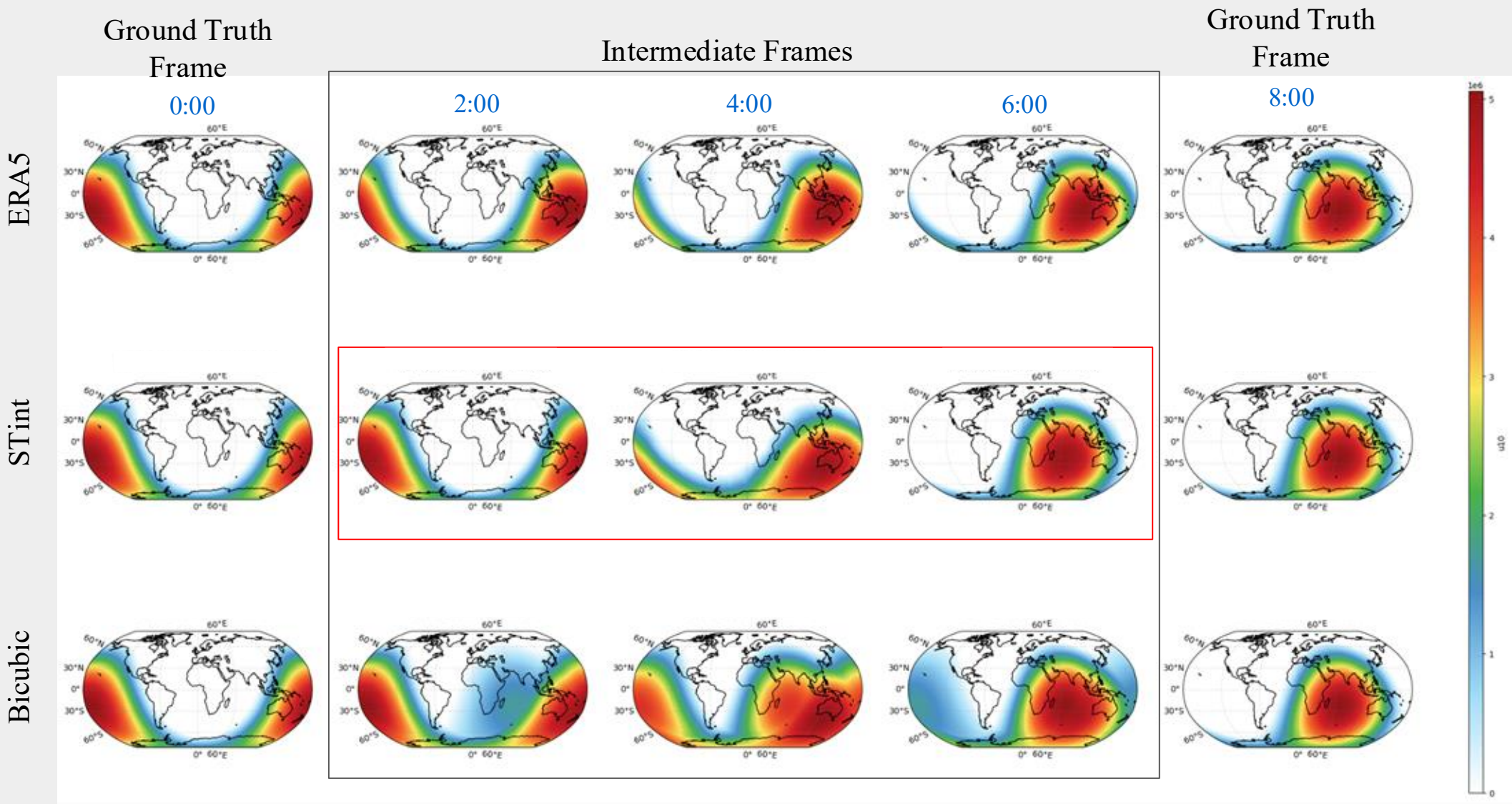


Figure. *Model trained on ERA5 Solar Irradiance fields and interpolation (3 unseen frames) between 0:00 and 8:00.*

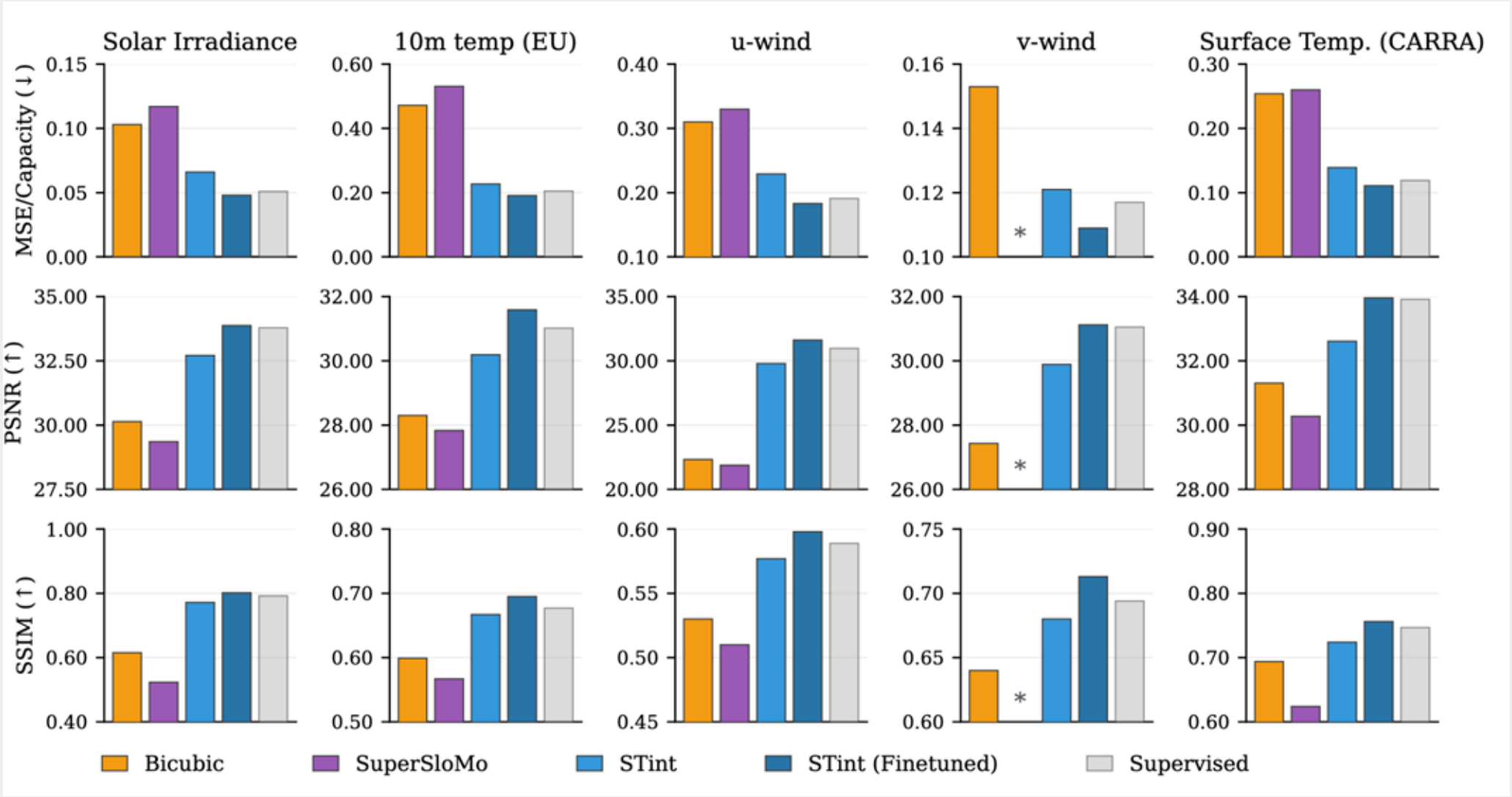
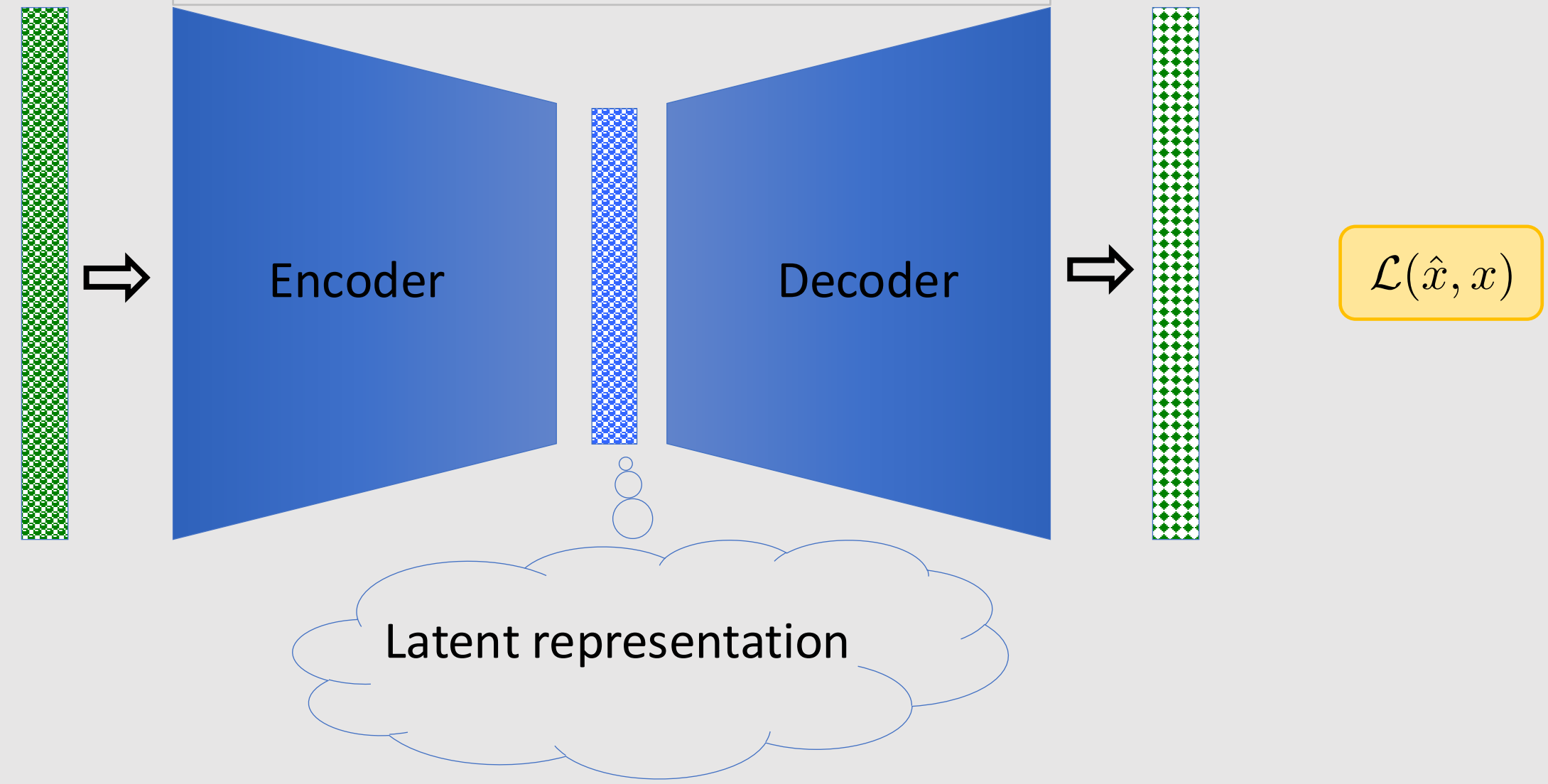


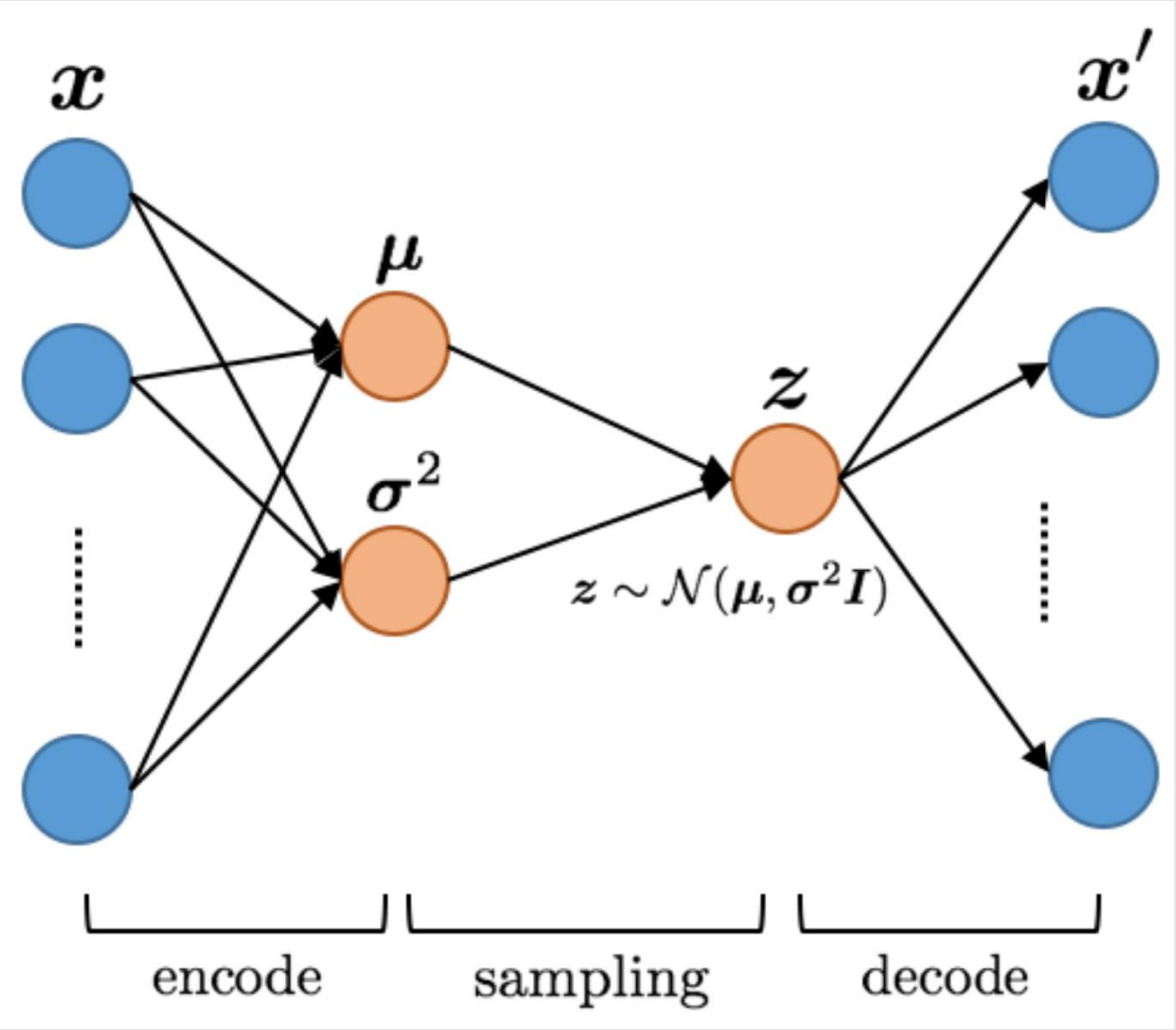
Figure. *Temporal interpolation results on different geospatial datasets.*

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.



Variational Autoencoder (VAE)

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



Learning a Variational Autoencoder (VAE)

- Encoder is a distribution, q, maps data space to latent space
  - Defined by network weight parameters θ: $q_\theta(z|x)$
- Decoder is a distribution, p, maps latent space to data space
  - Defined by network weight parameters ϕ: $p_\phi(x|z)$
- Fit the network weights θ and ϕ, by minimizing:

$$E_{q_\theta(z|x)}[\log p_\phi(x|z)] - \beta KL(q_{\theta(z|x)}||p(z))$$

Reconstruction error on training data

Regularization term: Penalizes distance from a simple Prior

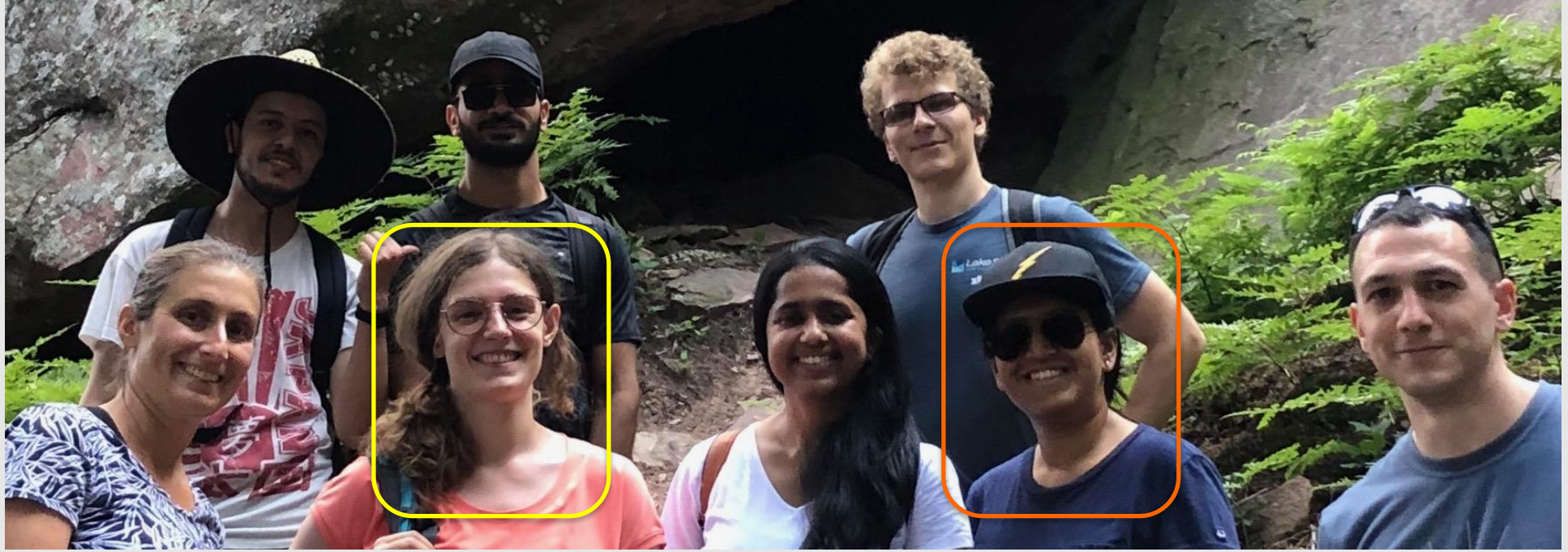
- Optimized using variant of Expectation Maximization (EM), using a variational lower bound

Example: VAE for semi-supervised anomaly detection: application to avalanche detection

Avalanche Detection



Img src : https://www.nytimes.com/2017/01/19/world/europe/italy-avalanche.html



Detecting Avalanche Deposits using VAE on Sentinel-1 Imagery

Saumya Sinha*, Sophie Giffard-Roisin*, Fatima Karbou, Michael Deschatres, Anna Karas, Nicolas Eckert, Cécile Coléou, Claire Monteleoni

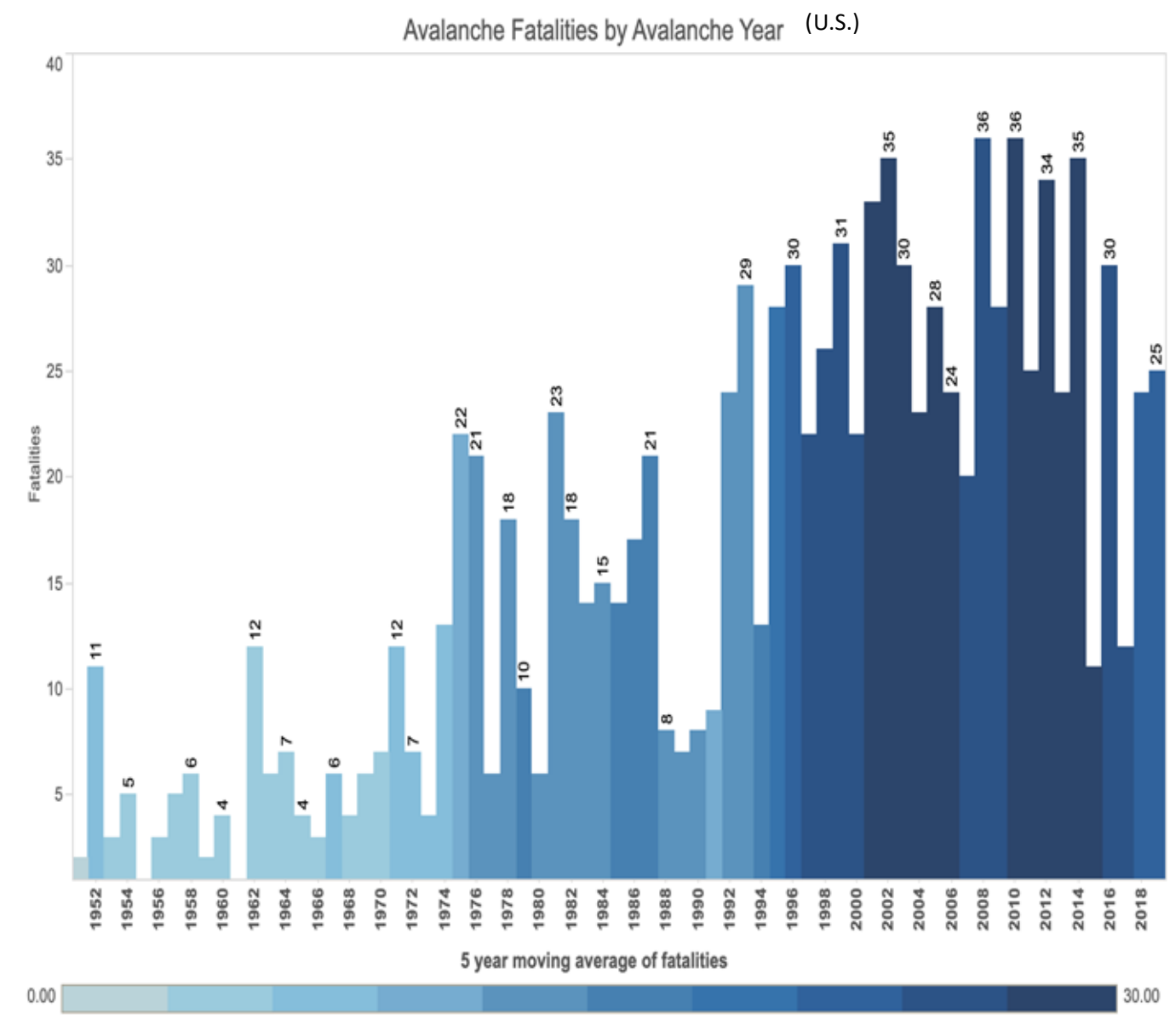
Variational autoencoder anomaly-detection of avalanche deposits in satellite SAR imagery. *Proc. 10th International Conference on Climate Informatics (CI) 2020*

Boulder

*Co-first-authorship

METEO FRANCE

Avalanche Fatalities by Avalanche Year



Img src : https://avalanche.state.co.us/accidents/statistics-and-reporting/

Climate change impacts



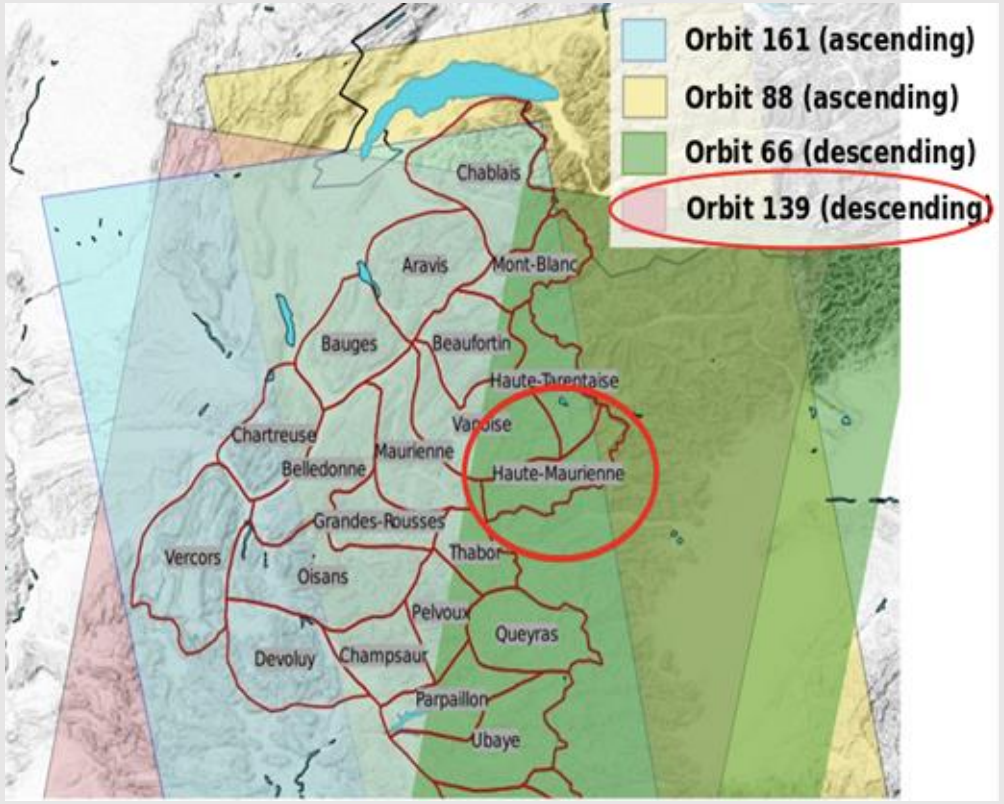
Img src: https://www.independent.co.uk/news/world/europe/mont-blanc-deaths-french-alps-avalanche-skiing-a9227451.html

Challenges for Machine Learning

- Severe class imbalance
  - Avalanches are rare events
- Ground-truth labeled data difficult to obtain
  - Terrain accessibility
  - Weather conditions
  - Danger of avalanches

Avalanche detection problem

- **Limited** in-situ ground-truth measurements
  - Météo-France



- **Unlabeled** SAR imagery
  - Monitoring French Alps in 2017-2018
  - Sentinel-1A and 1B satellites
  - 4 features:
    - Backscatter coefficients at present and previous time
    - Topological features: Slope & Angle

METEO FRANCE

Our Approach

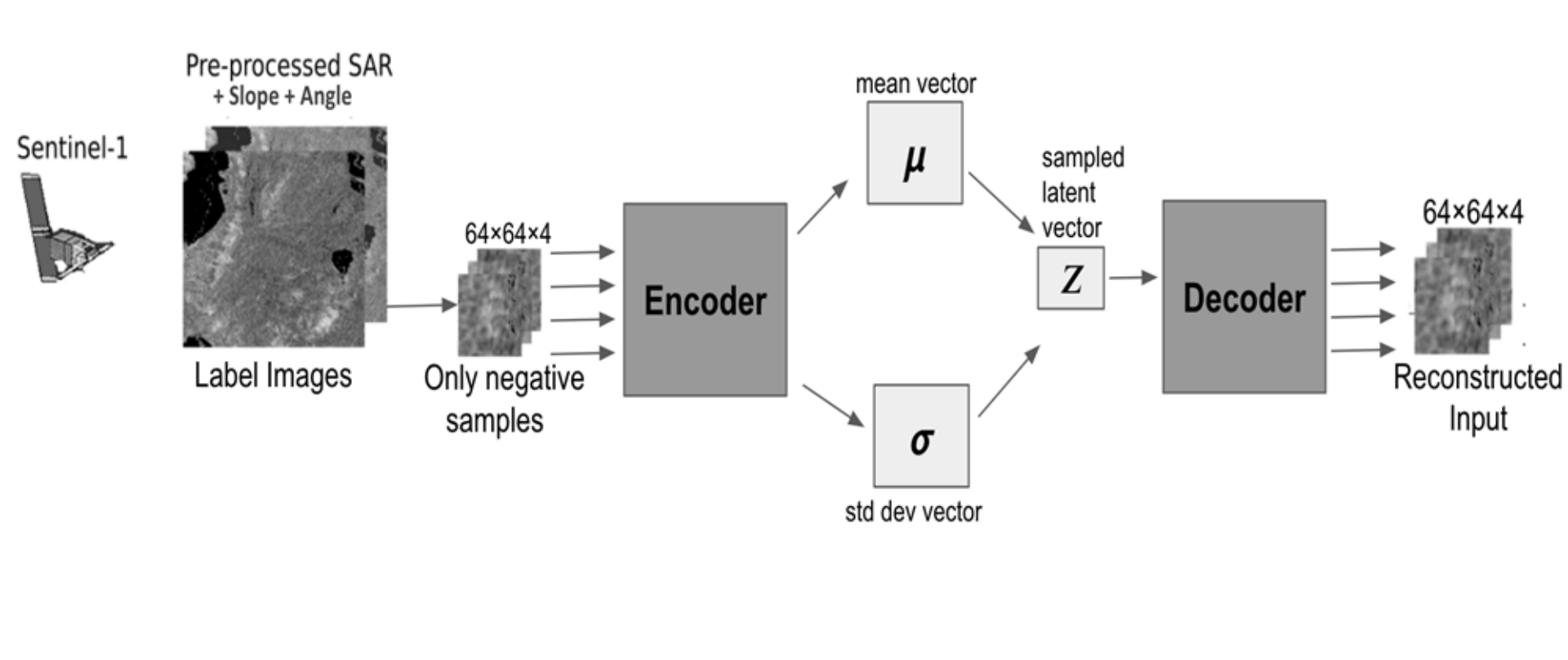
① Treat an avalanche as a rare event, or an anomaly

② Train a variational autoencoder (VAE) on the negative examples

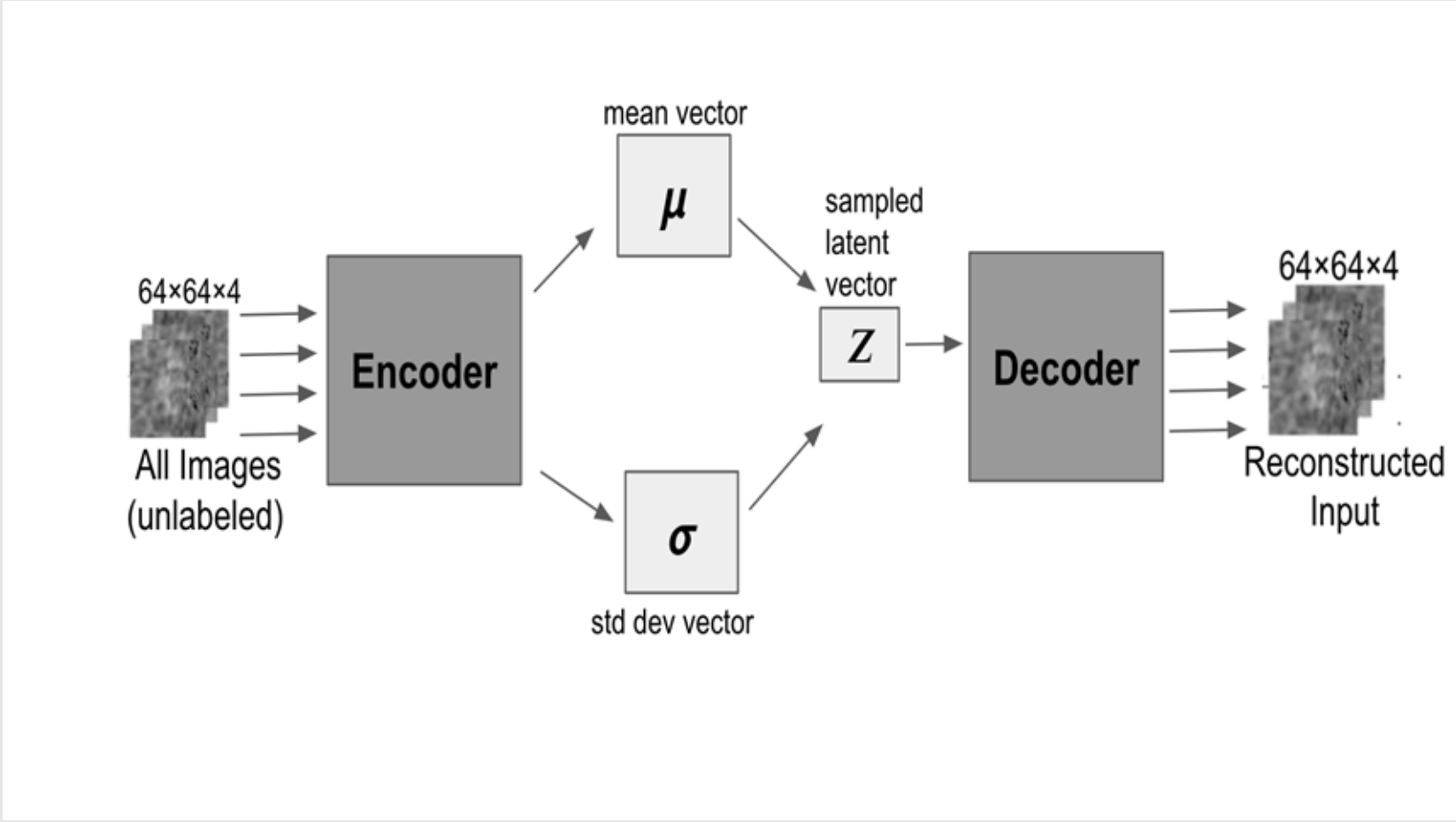
③ Threshold the VAE's reconstruction error to classify a new image

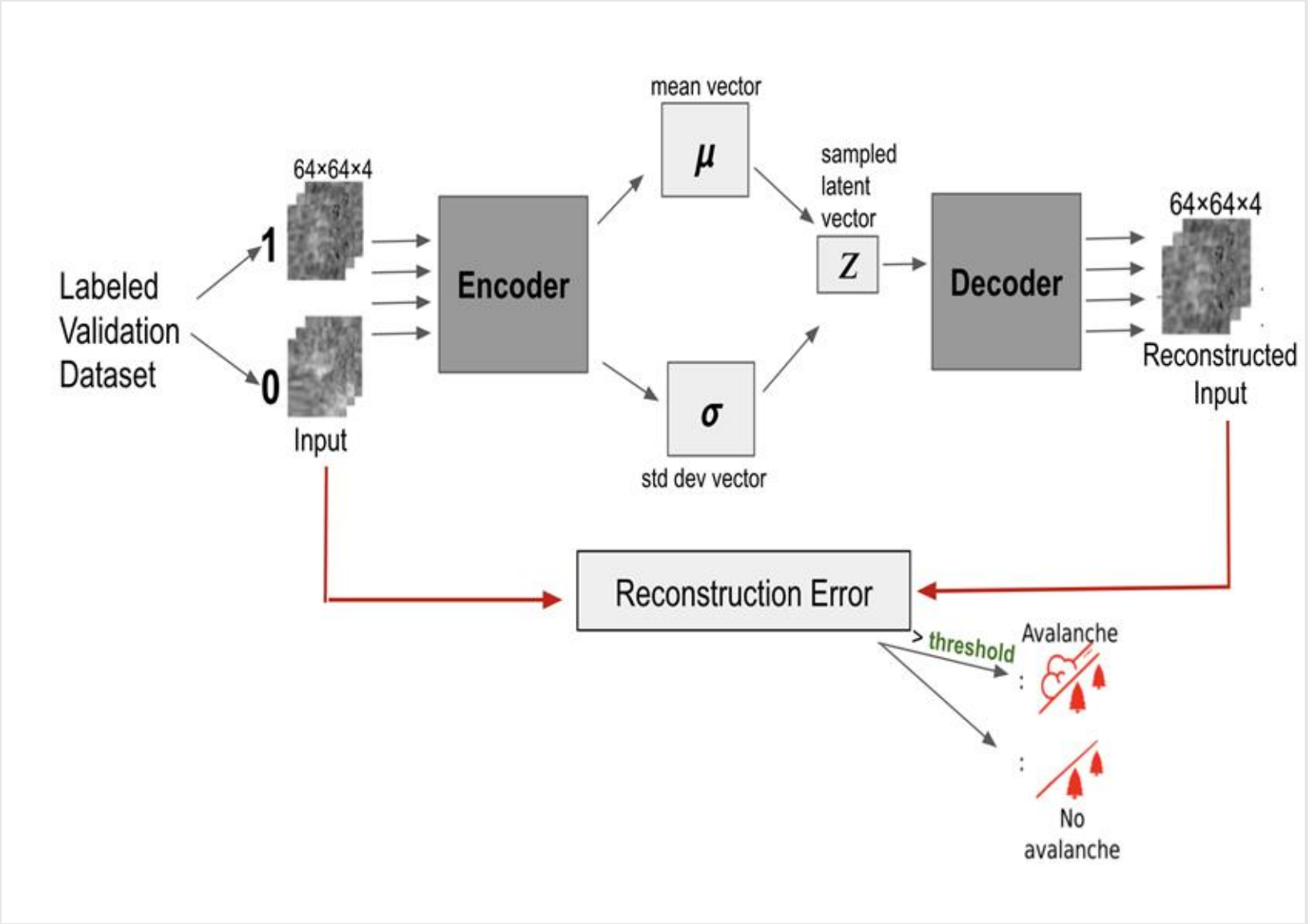
- When labeled data is scarce, the VAE can instead be trained **without** supervision!

VAE trained on negative examples

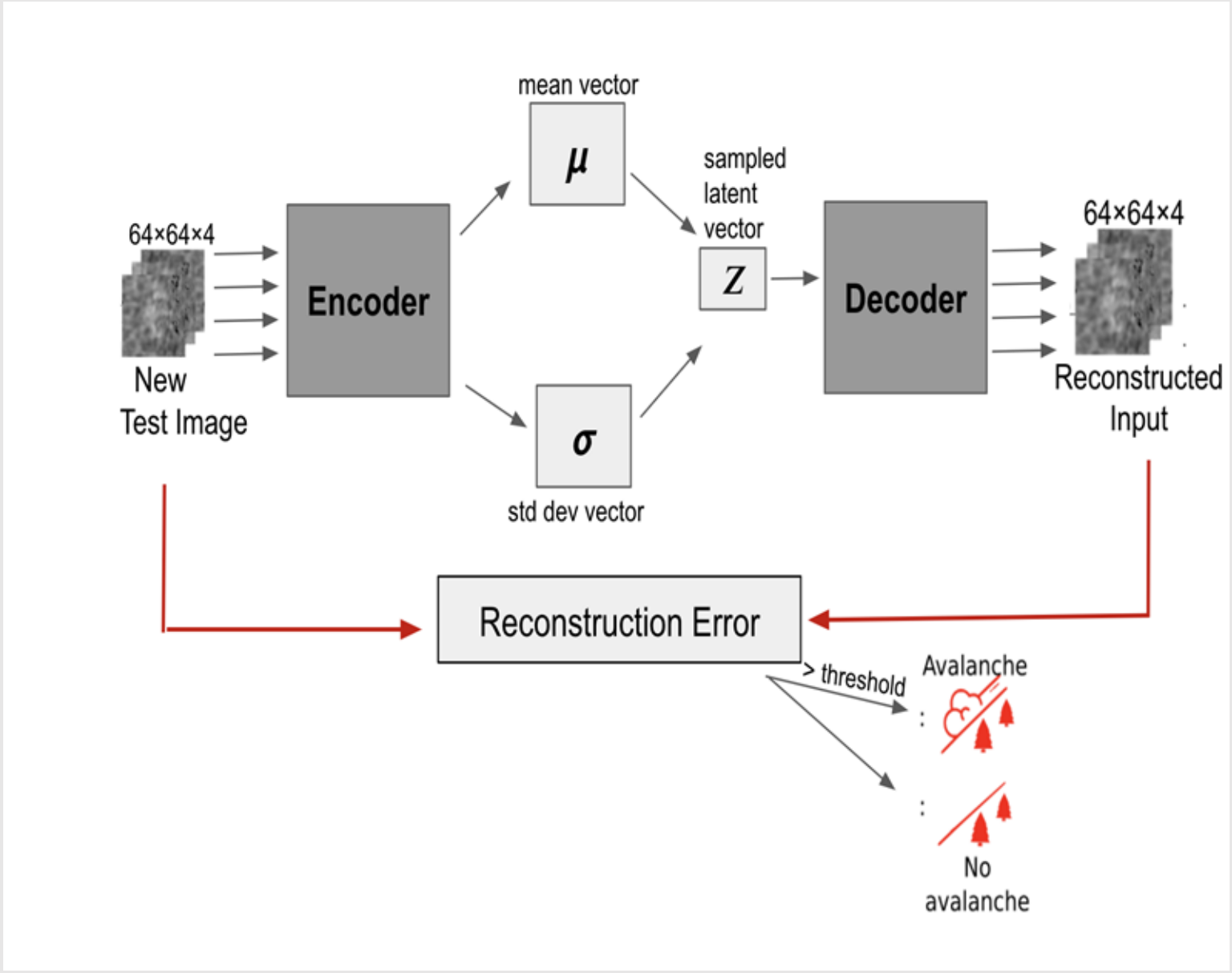


VAE trained on ALL examples



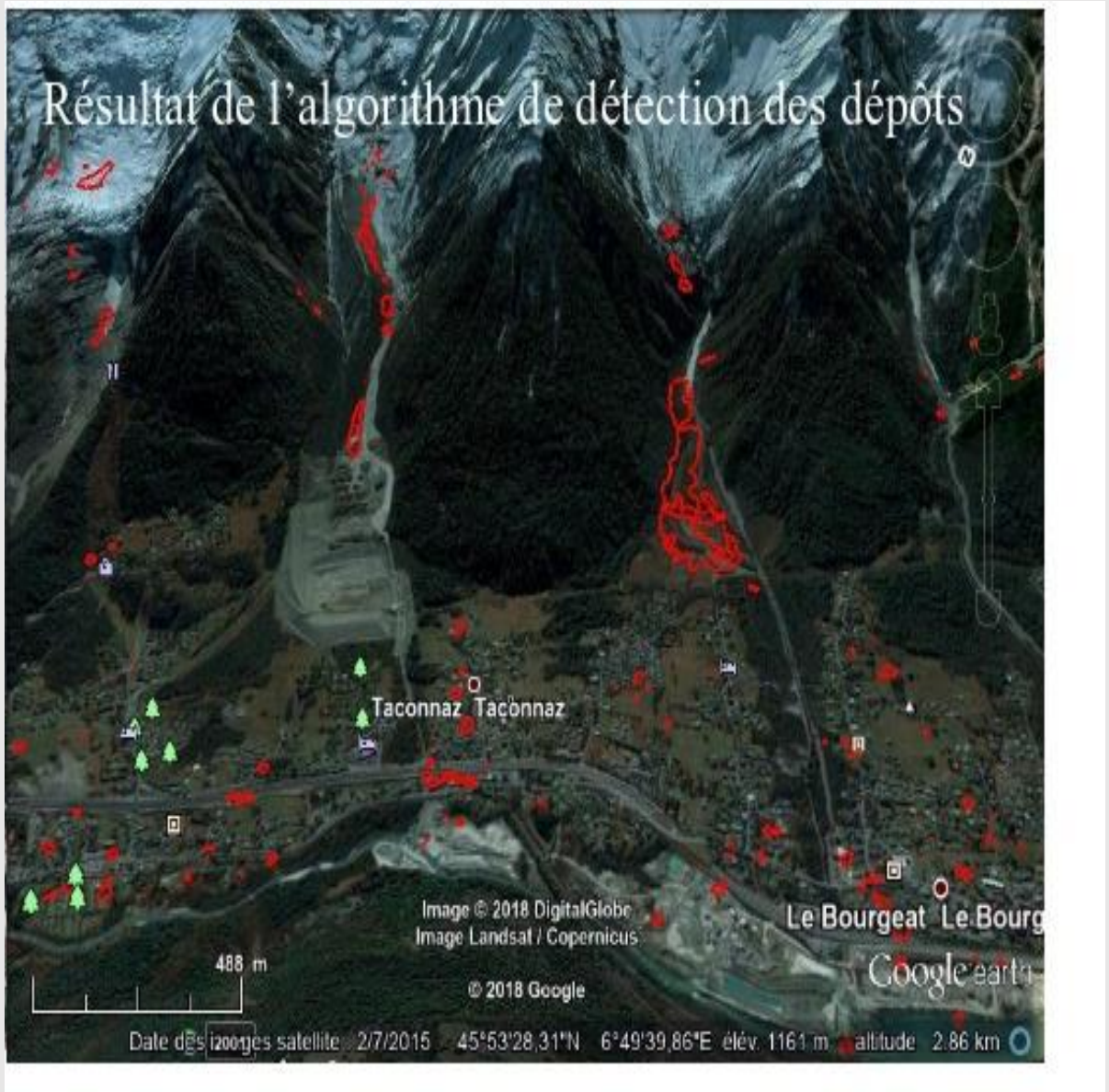


Tuning the threshold for Anomaly Detection



Anomaly Detection Pipeline

Baseline method: Thresholding



[Karbou et al., International Snow Science Workshop 2018 & EGU 2018]

Evaluation

One of the most susceptible mountain chains (out of the 18 in “All Alps”)

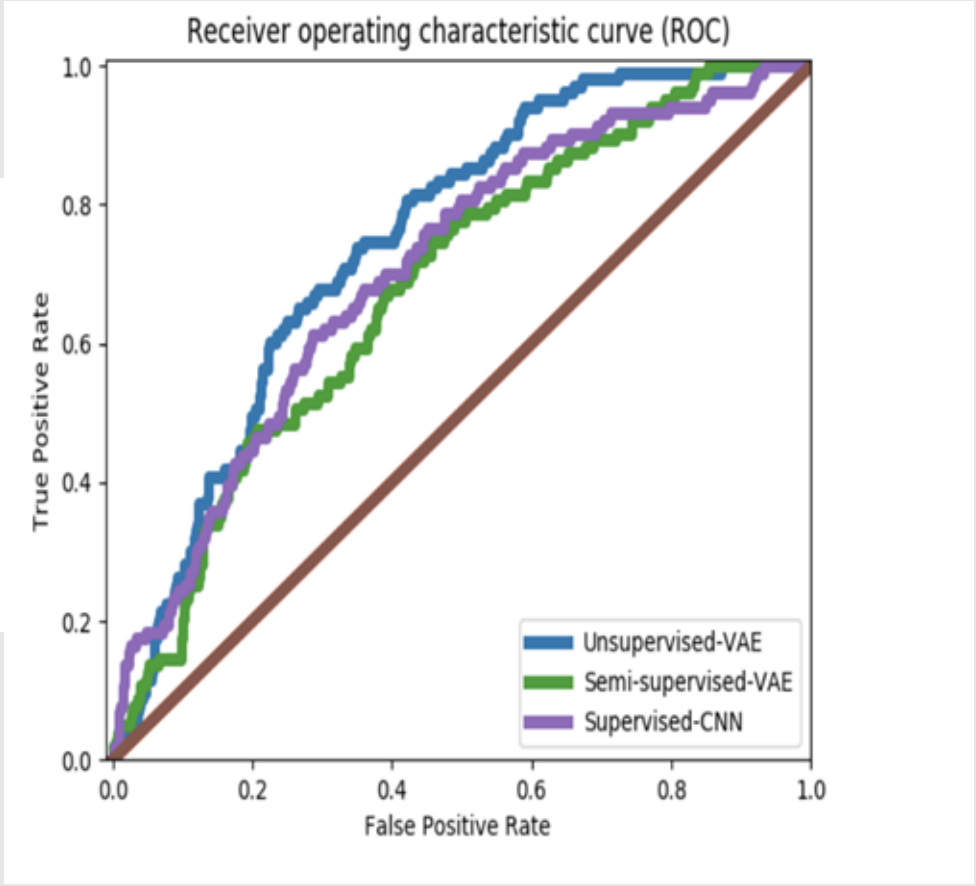
	All Alps		Haute Maurienne	
	Balanced Accuracy	F1-score	Balanced Accuracy	F1-score
Baseline	0.58	0.05	0.58	0.12
Supervised - CNN	0.53	0.10	0.53	0.12
Semi-supervised - VAE	0.59	0.11	0.6	0.23
Unsupervised - VAE	0.69	0.14	0.68	0.26

- Held-out test set: 6,498 labeled examples
- Baseline method from avalanche-detection literature: Thresholding [Karbou et al., ISSW 2018]
- Supervised-learning benchmark method: Convolutional Neural Network (CNN) trained on artificially balanced dataset [Sinha et al., Climate Informatics 2019]

Evaluation

Method	AUC ROC
Supervised - CNN	70.7
Semi-supervised - VAE	68.3
Unsupervised - VAE	75

ROC Curves for Haute Maurienne region



Summary

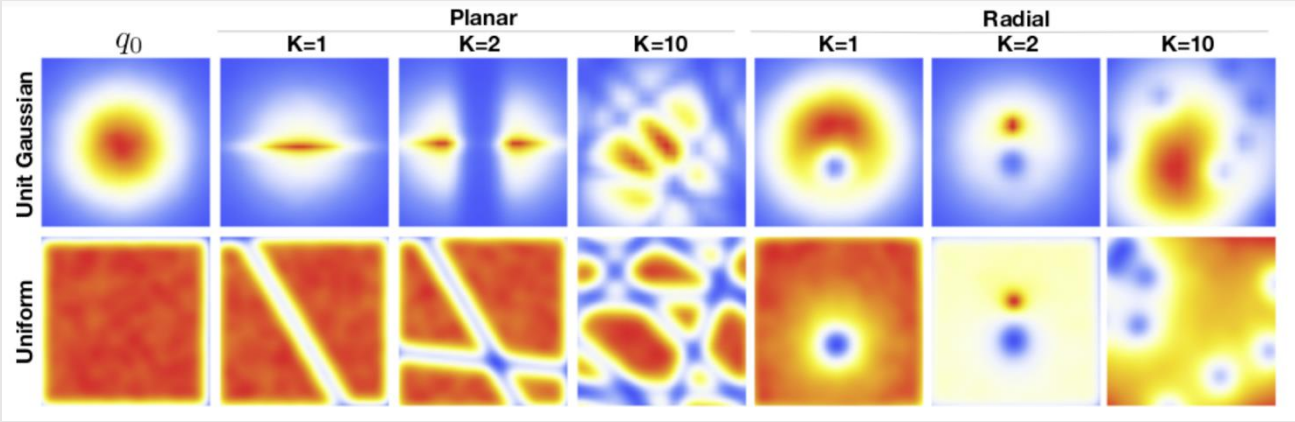
- Provided a semi-supervised approach to detecting rare events when labeled data is limited
  - This semi-supervised pipeline relies heavily on unsupervised learning
- Can be viewed as a form of virtual sensor

Part III: AI Methods: Generative AI

Generative AI: Distinction

- You are probably used to the term "generative AI" referring to large commercial AI chat-bots and agents such as ChatGPT, Gemini, or Claude.
  - For these systems, generative AI models have been trained on video and text scraped from the internet
  - But this is **not the ONLY definition** or use of generative AI!
- We can **train generative AI models from scratch** for our problems
  - For climate and weather problems, we will NOT train on text and video but on the relevant data, e.g. fields of real-values measured at different latitudes, longitudes, times, and perhaps height in atmosphere or depth in ocean, including:
    - Temperature
    - Solar radiation
    - Humidity
- Today we'll learn about the types of generative models that you can then train for your own problems

Normalizing Flows

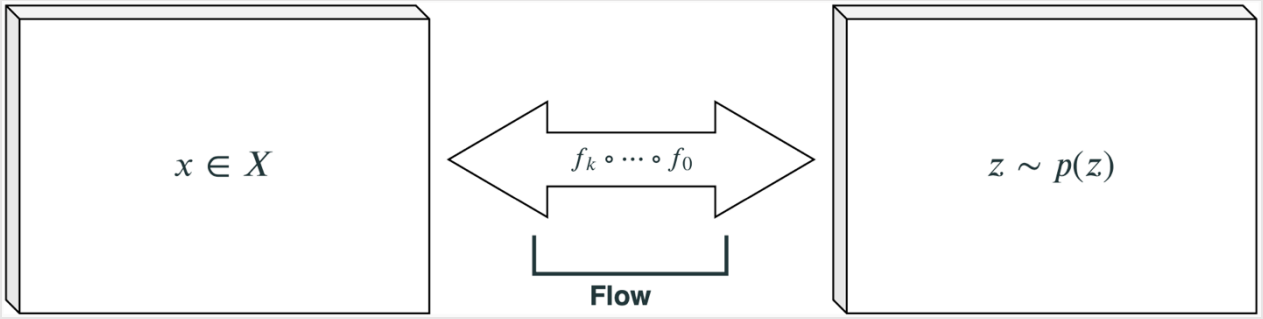


[Rezende & Mohamed, ICML 2015]

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:

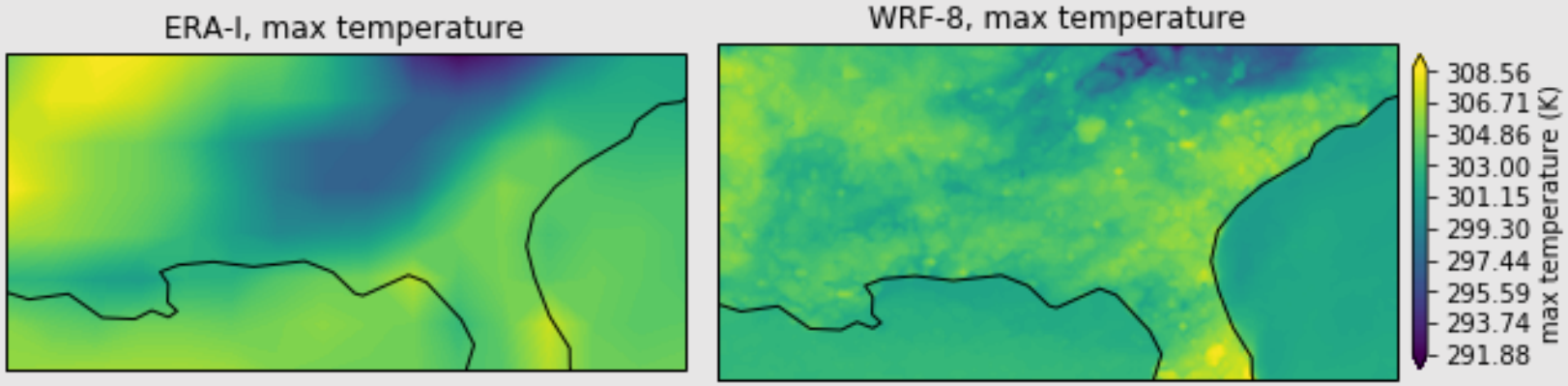
$$z_k = f_k \circ f_{k-1} \circ \cdots \circ f_1(z_0)$$

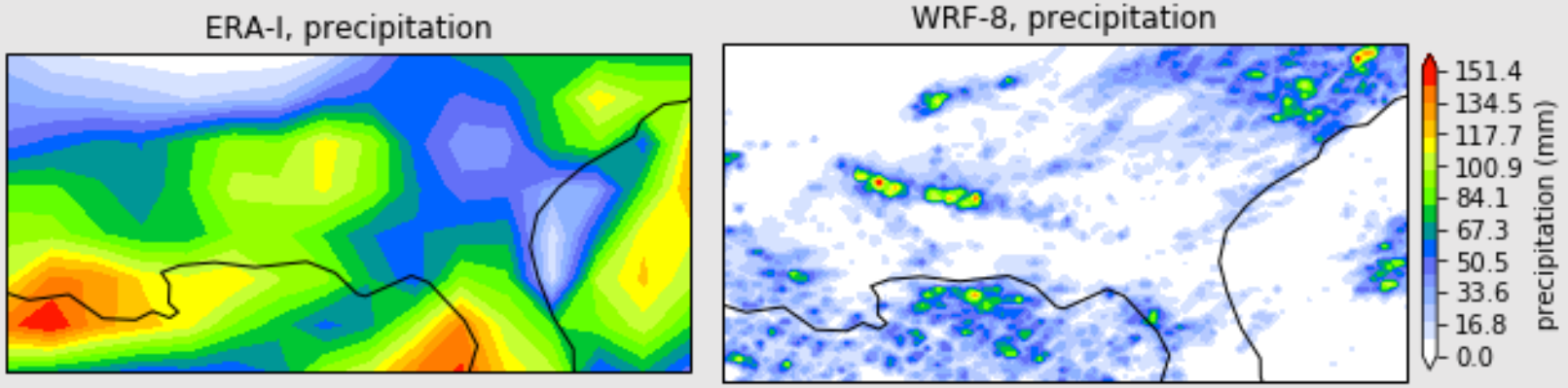


Normalizing Flows: Application to Spatial Downscaling

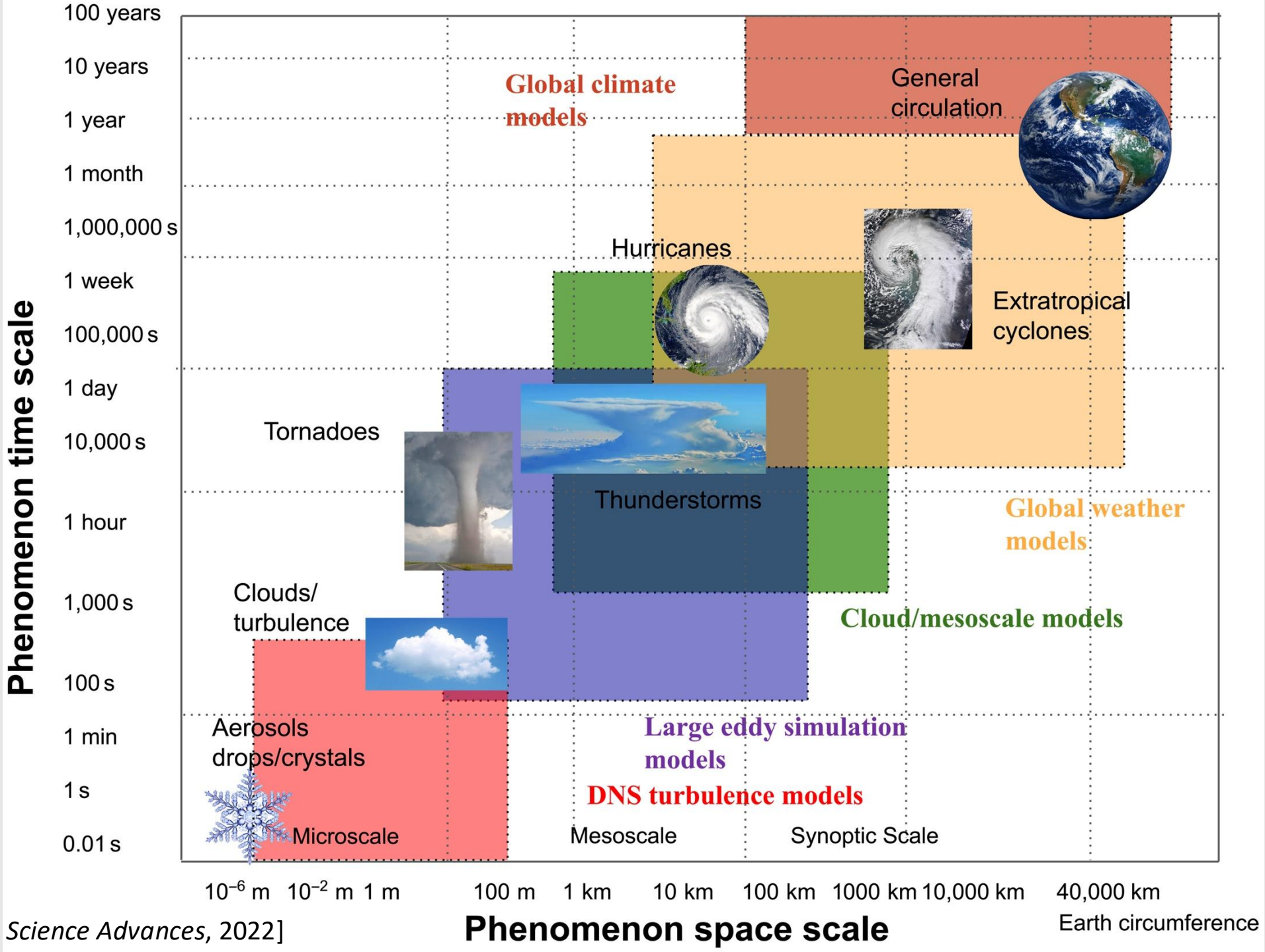
[Groenke, Madaus, & Monteleoni, Climate Informatics 2020]

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









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

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 $y_i \in Y$,

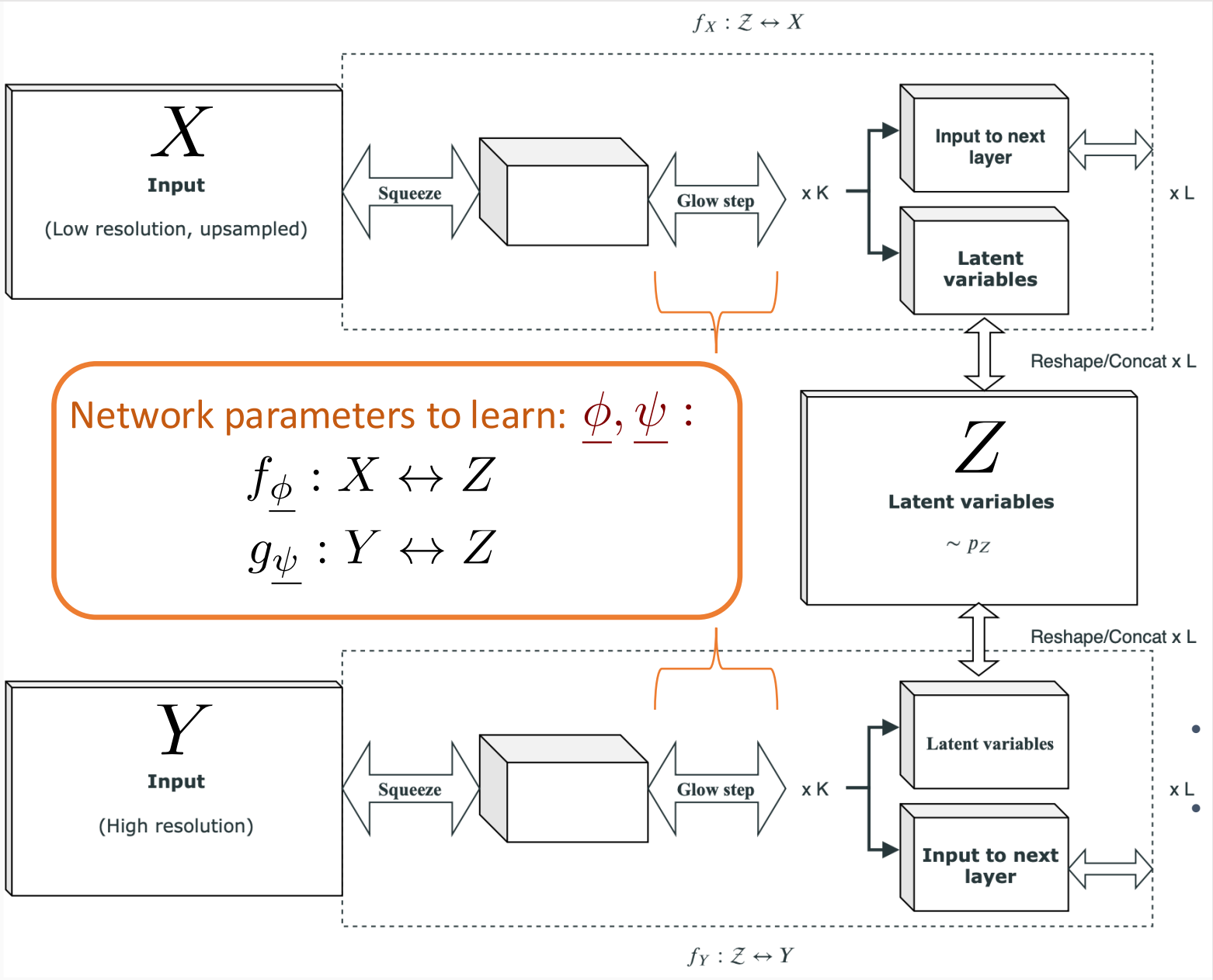
$$f(x_i) \sim P_Y \qquad\qquad f^{-1}(y_i) \sim P_X$$

- 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]

Comparison with supervised benchmarks

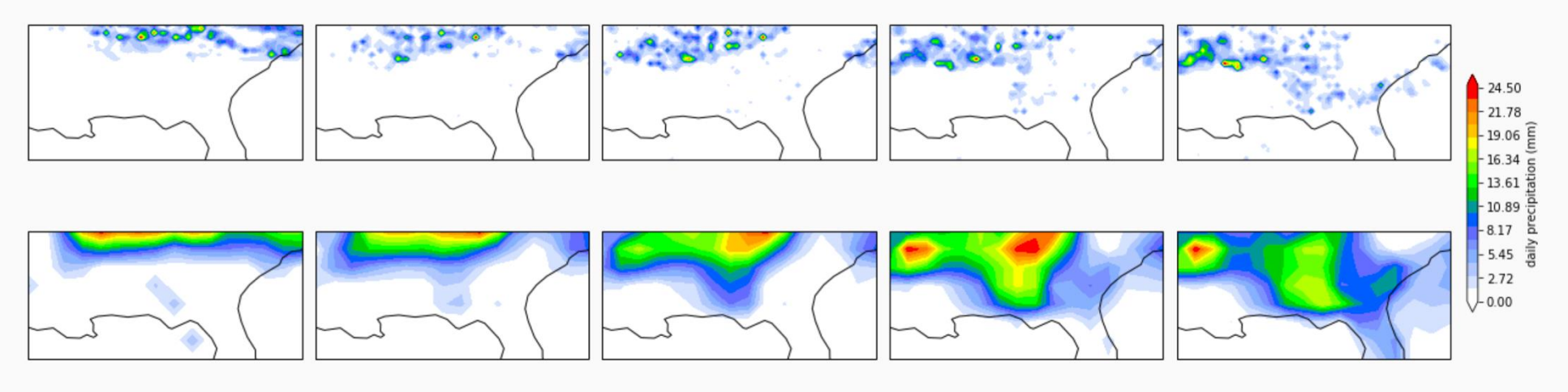
Daily Max Temperature

Region	Method	RMSE	Bias	Corr
SE-US	BCSD	1.51 ± 0.15	-0.02 ± 0.21	**0.93** ± 0.05
	BMD-CNN	**1.30** ± 0.12	0.03 ± 0.13	0.90 ± 0.05
	ClimAlign (ours)	1.56 ± 0.13	**-0.005** ± 0.22	0.87 ± 0.06
P-NW	BCSD	1.54 ± 0.23	**0.01** ± 0.10	**0.95** ± 0.03
	BMD-CNN	**1.25** ± 0.14	-0.06 ± 0.05	0.93 ± 0.02
	ClimAlign (ours)	1.58 ± 0.18	0.03 ± 0.15	0.89 ± 0.04

Daily Precipitation

Region	Method	RMSE	Bias	Corr
SE-US	BCSD	27.32 ± 5.0	0.95 ± 1.4	0.39 ± 0.07
	BMD-CNN	**14.11** ± 2.18	-0.23 ± 0.47	**0.50** ± 0.10
	ClimAlign (ours)	18.40 ± 2.64	**0.08** ± 0.86	0.42 ± 0.07
P-NW	BCSD	8.90 ± 2.30	0.41 ± 0.26	0.61 ± 0.06
	BMD-CNN	**5.77** ± 0.72	**-0.18** ± 0.61	**0.70** ± 0.03
	ClimAlign (ours)	7.33 ± 0.69	0.54 ± 0.54	0.67 ± 0.03

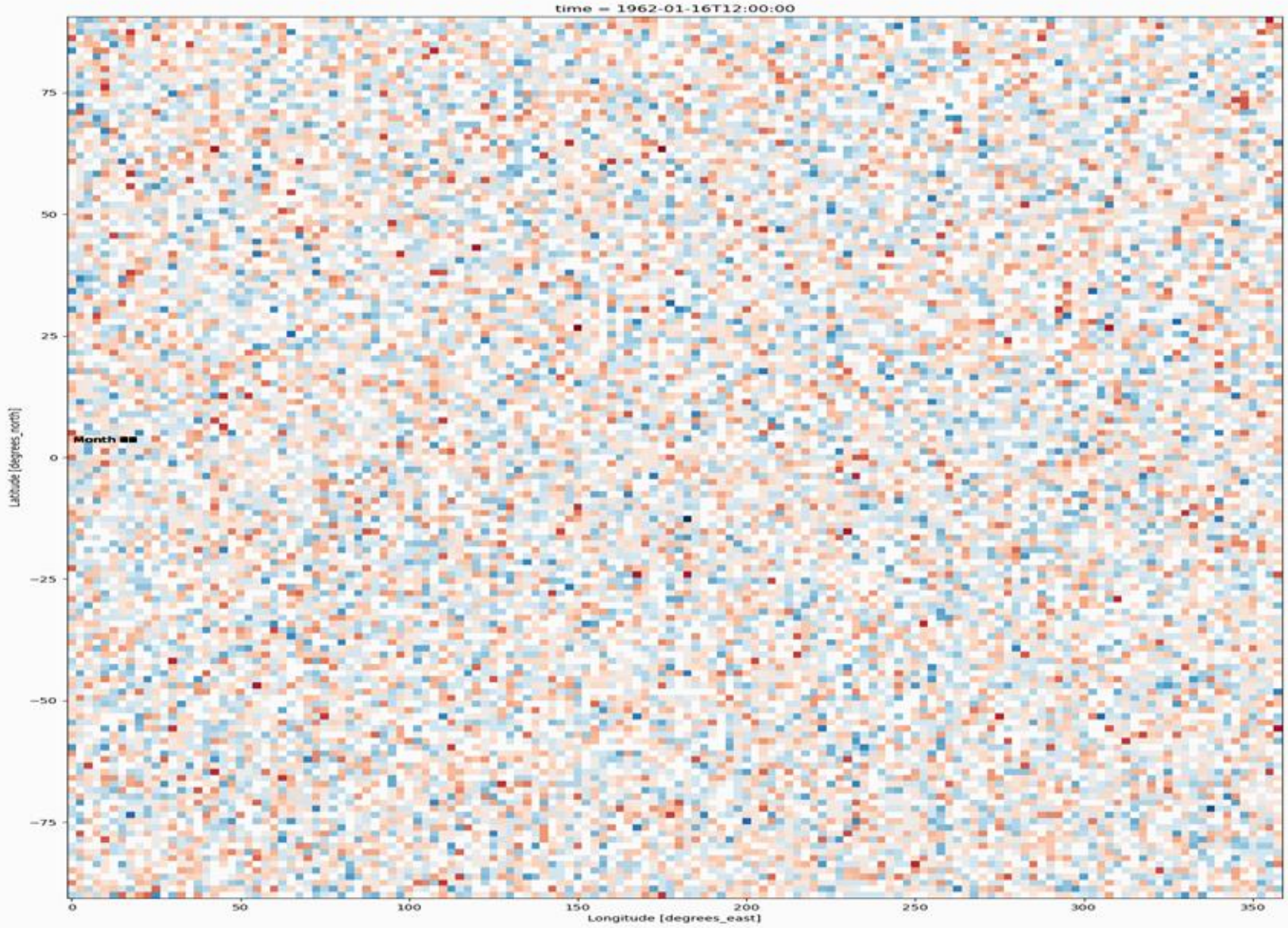
ClimAlign: Unsupervised, generative downscaling

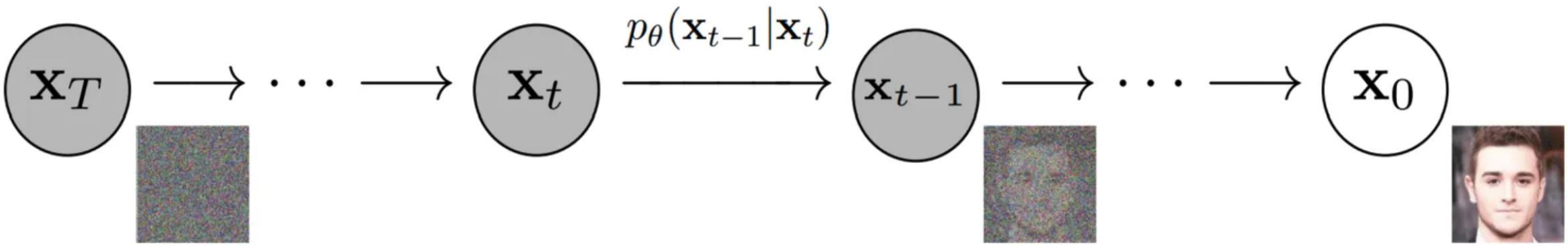


General downscaling technique via domain alignment with normalizing flows
[AlignFlow: Grover et al., AAAI 2020][Glow: Kingma & Dhariwal, NeurIPS 2018]

- **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

Flow-based generative models





Diffusion training: Learn to generate realistic images from random noise

Training the diffusion model

1. Create a pretext task from unlabeled training data: For each training data point, x_0, create a sequence of synthetic images, by adding Gaussian noise to the previous image in the sequence, until you reach an image indistinguishable from Gaussian noise.

2. Now do "supervised" learning on pairs (x_{t-1}, x_t) to learn $p_\theta(x_{t-1} \mid x_t)$

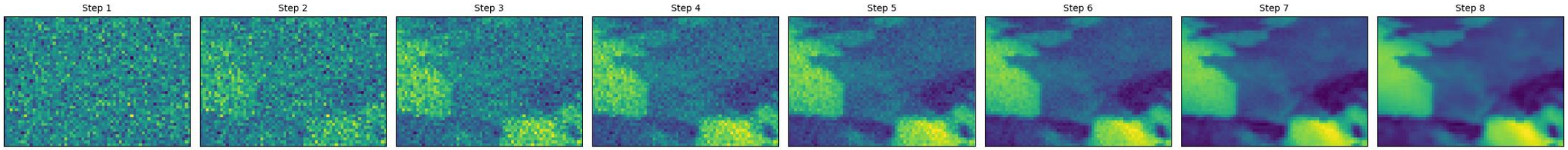
Inference, i.e., how to generate samples similar to the training data

1. Sample x_T randomly (Gaussian).

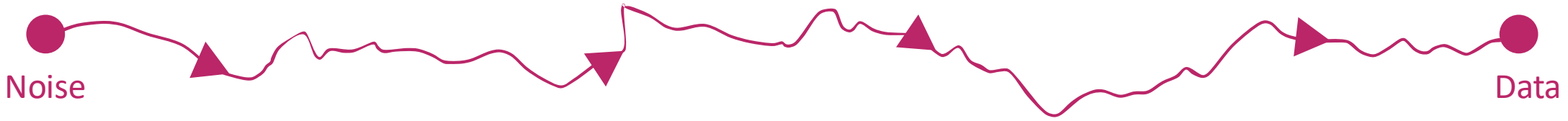
2. Sample T times from p_θ, updating the conditioning to be the previous sample each time

Flow-based generative models

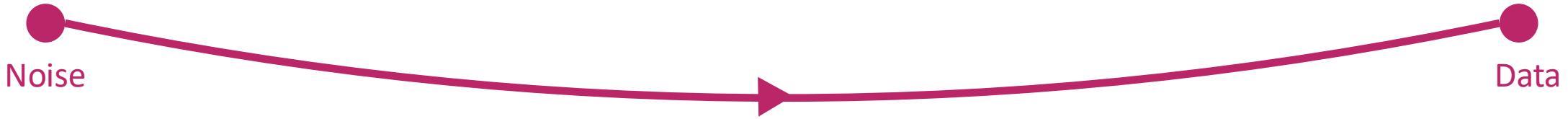
Flow-based generative model: learns a transport (using a neural network) from an easy-to-sample distribution (e.g, Gaussian noise) to the data distribution.



Diffusion: "denoiser" reverses a stochastic noise process, gradually denoising – SDE trajectory



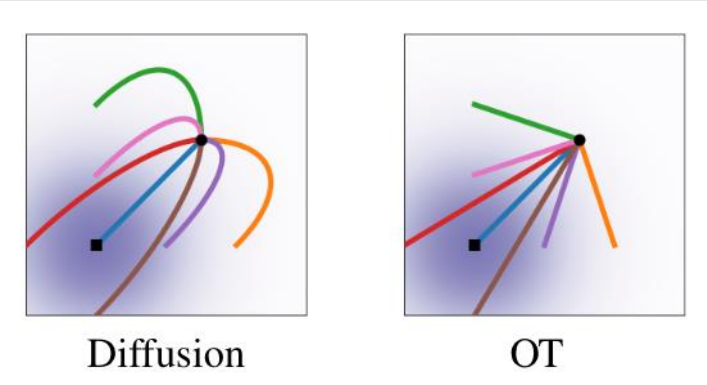
Flow matching [Lipman et al., 2023]: learns a deterministic flow from noise → data – ODE trajectory



Slide credit: Julie Keisler (INRIA)

Flow Matching vs. Diffusion

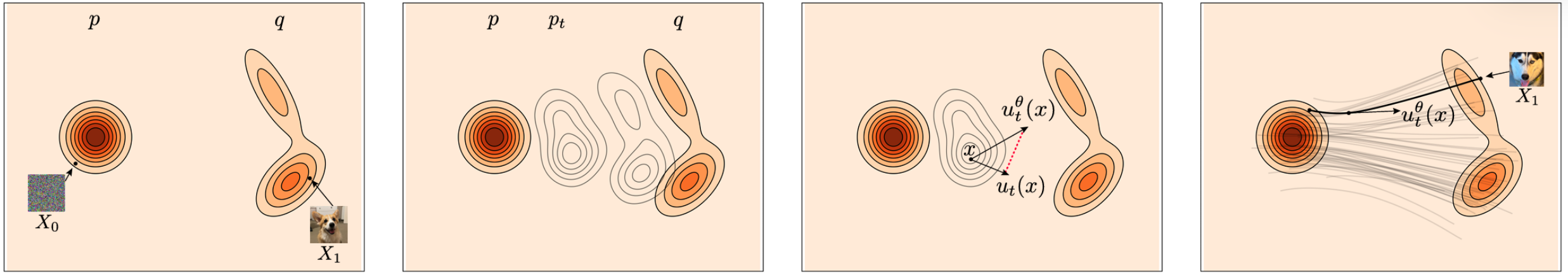
Intuition: Make straighter paths between data samples and noise samples:



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

Generality: Flow matching can connect any two distributions (or stochastic interpolants), unlike diffusion

Flow Matching



(a) Data. **(b)** Path design. **(c)** Training. **(d)** Sampling.

Figure 2 *The Flow Matching blueprint.* (a) The goal is to find a flow mapping samples X_0 from a known source or noise distribution q into samples X_1 from an unknown target or data distribution q. (b) To do so, design a time-continuous probability path $(p_t)_{0\leq t\leq 1}$ interpolating between $p := p_0$ and $q := p_1$. (c) During training, use regression to estimate the velocity field u_t known to generate p_t. (d) To draw a novel target sample $X_1 \sim q$, integrate the estimated velocity field $u_t^\theta(X_t)$ from $t = 0$ to $t = 1$, where $X_0 \sim p$ is a novel source sample.

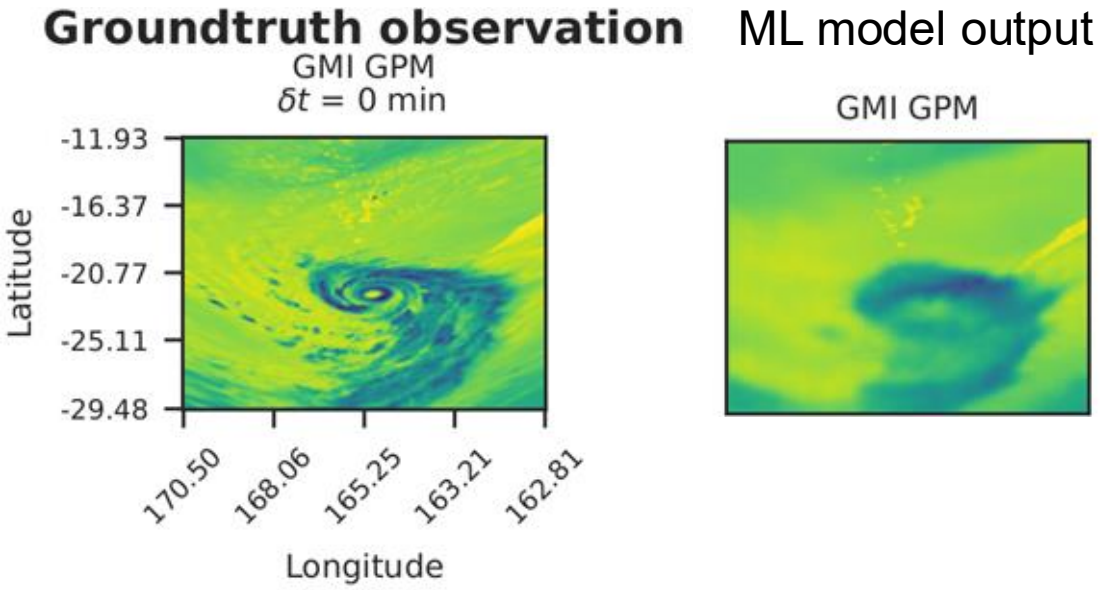
Lipman et al. (2024) Flow Matching Guide and Code

Why use generative AI in climate and weather?

Deterministic ML models often optimize **average** performance

Trained to minimize an average loss, e.g., RMSE

➔ **Blurry, non-physical** model outputs (e.g., images, fields, storm tracks)



ML model output

ML model outputs - sampled after training the same model backbone **generatively**

Why use generative AI in climate and weather?

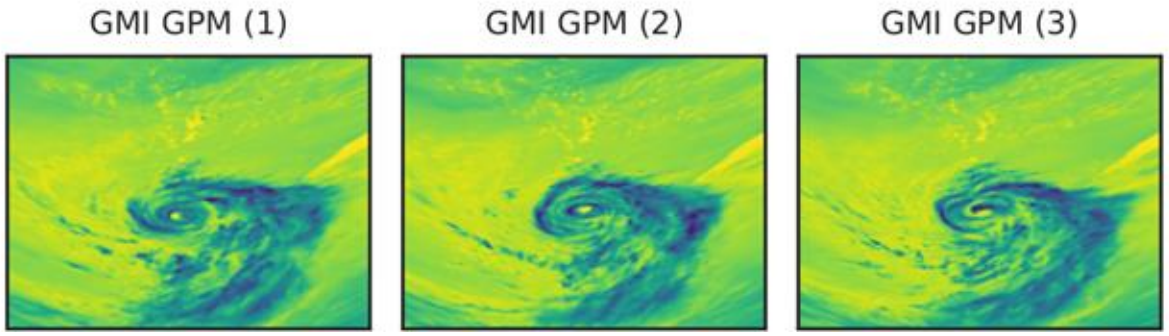
For many problems in climate, weather, geosciences

- Weather forecasting, climate emulation, scenario generation
- Estimation of extremes or rare events, quantiles
- Downscaling to local scales in time or space
- Risk quantification

Generative AI provides benefits

- Improved physicality
  - Improved accuracy of high frequencies
- Probabilistic ensembles
  - Uncertainty quantification
  - Rich, realistic info for decision-makers
- Synthetic data generation

Samples from a **generative model**



Generative AI for:

✓ Probabilistic forecasting

- **ArchesWeatherGen**: Efficient weather forecasting trained at 1.5°
- **ArchesClimate**: Decadal climate emulation, interpolation, extrapolation

❑ Data fusion with probabilistic domain alignment

- **MotifGen**: A generative, multi-source, self-supervised model for learning from geospatial data

❑ Probabilistic domain alignment and downscaling

- **ClimAlign**: Generative unpaired downscaling with domain alignment
- **SerpentFlow**: Generative unpaired domain alignment via shared-structure decomposition (plus application to downscaling)

Challenges of learning from multi-source data

Pléiades Neo

30 cm

Swath: 14 km

Spectral bands:
R, G, B, Deep Blue, Red Edge, NIR, Panchromatic

Revisit capacity:
Daily, anywhere

Daily acquisition capacity:
1,000,000 km²

Pléiades

50 cm

Swath: 20 km

Spectral bands:
R, G, B, NIR, Panchromatic

Revisit capacity:
Daily, anywhere

Daily acquisition capacity:
700,000 km²

Vision-1

87 cm

Swath: 20.8 km

Spectral bands:
R, G, B, NIR, Panchromatic

Revisit capacity:
Daily to 8 days, anywhere depending on latitude and partner satellites

Daily acquisition capacity:
20,000 km²

SPOT

1.5 m

Swath: 60 km

Spectral bands:
R, G, B, NIR, Panchromatic

Revisit capacity:
Up to daily, anywhere

Daily acquisition capacity:
3,000,000 km²

Radar constellation

From 25 cm to 40 m

Swath: 270 km

Imaging modes:
SpotLight, StripMap, ScanSAR

Revisit capacity:
Daily for most latitudes

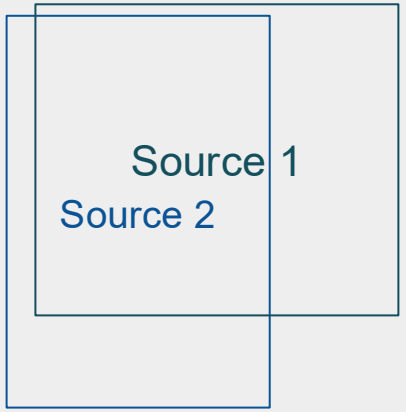
Daily acquisition capacity:
Up to 5,400,000 km²
(Wide ScanSAR Mode)



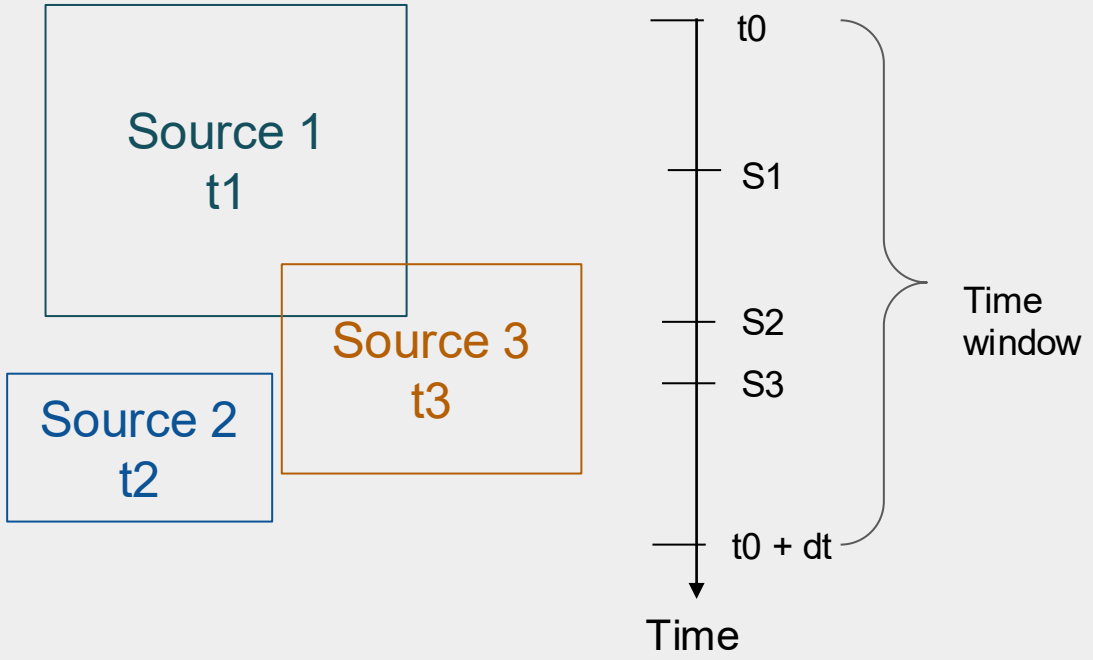
AI for **Fusion** of (Misaligned) Multi-Source Data

[Dauvilliers & Monteleoni, *Environmental Data Science*, 2025]

Traditionally: only train models on collocated data



Our AI model (MOTIF) can learn from **all data sources** even data misaligned in time and space



A multi-source architecture for geospatial data fusion

Multi-modal core-ML	**Geospatial representation learning**
• Embed **values** (e.g., text, images) • Conditioning (“dog or cat”)	• Embed **locations** • Large-scale self-supervised training
DiT model (2022)	SatCLIP (2025), AlphaEarth (2025)

MOTIF: *Multi-source transformer via factorized attention*

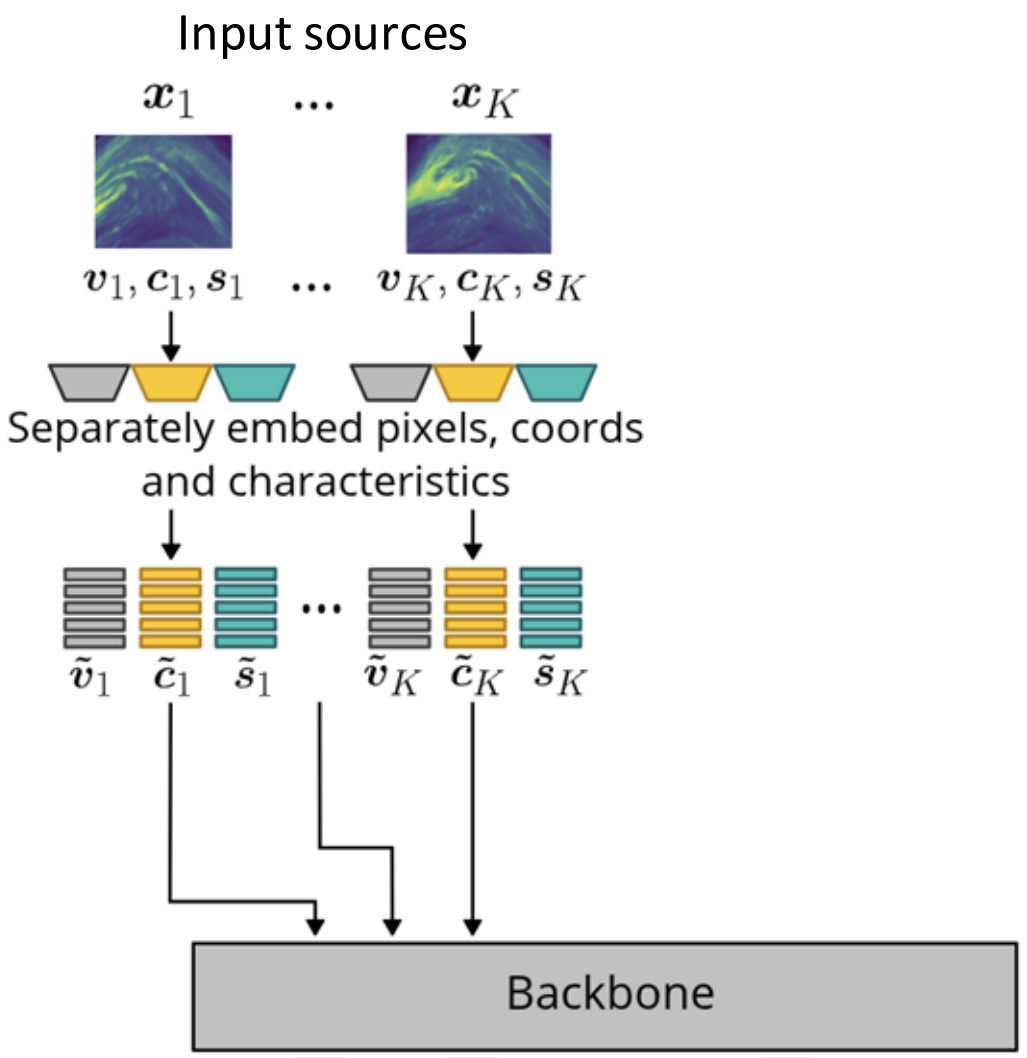
- Separately embed **values** and **coordinates**
- Condition on source characteristics
- Self-supervised pre-training

[Dauvilliers & Monteleoni, Environmental Data Science, 2025]



Training pipeline

Embed separately:
- values,
- coordinates,
- source characteristics



Relative positional embedding

- Usual relative positional embedding in Transformers (Shaw et. al, 2018)

$$V_{out} = \frac{Softmax(QK^T + Q_{pos}K_{pos}^T)}{\sqrt{d}}$$

- Here: spatiotemporal coordinates are used for the positional embedding:

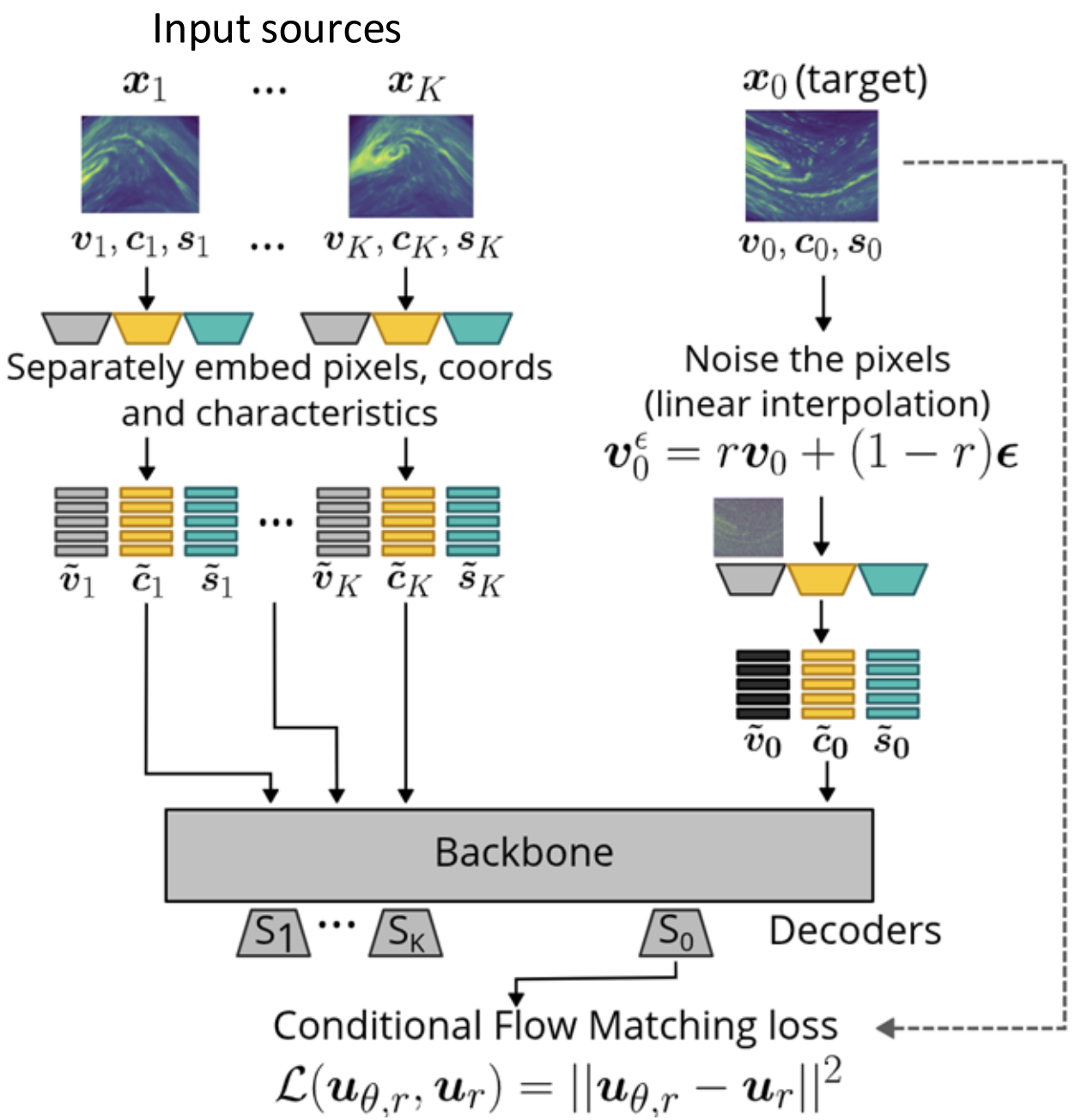
$$\tilde{v}_{out} = \frac{Softmax\left(Q_V K_V^T + Q_C K_C^T\right)}{\sqrt{d}} \tilde{v}_{in}$$

Values (e.g. pixels)

Spatio-temporal coordinates

Training pipeline

Embed separately:
- values,
- coordinates,
- source characteristics



SSL training task: reconstruct a masked source

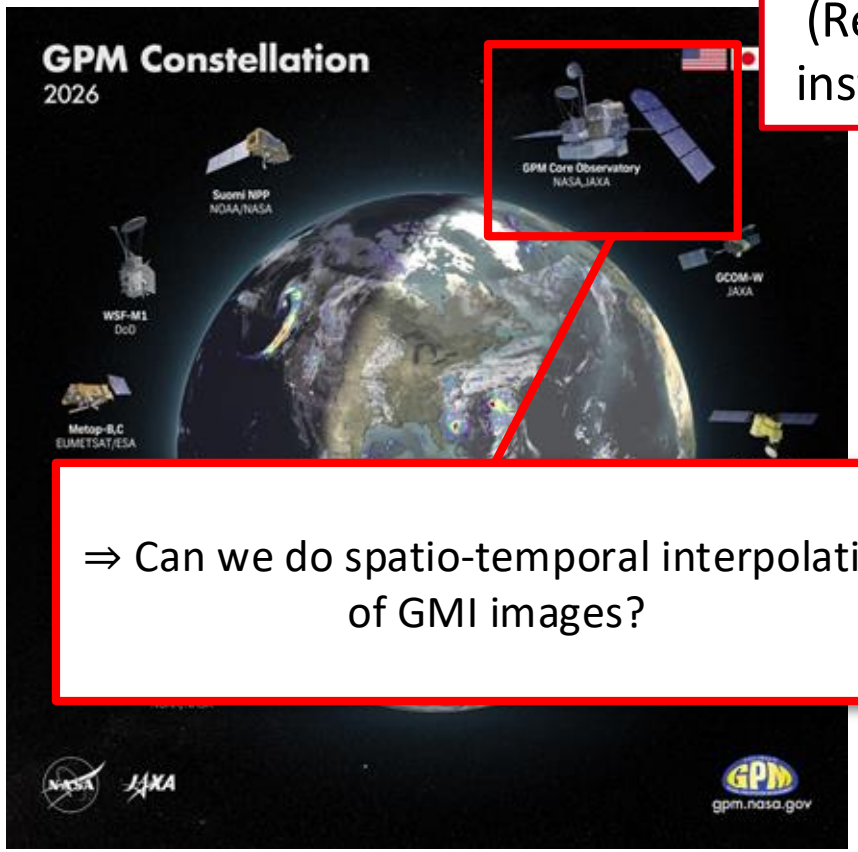
Generative version trained via Flow Matching

[Dauvilliers & Monteleoni, Climate Informatics 2026]

Application to tropical cyclones

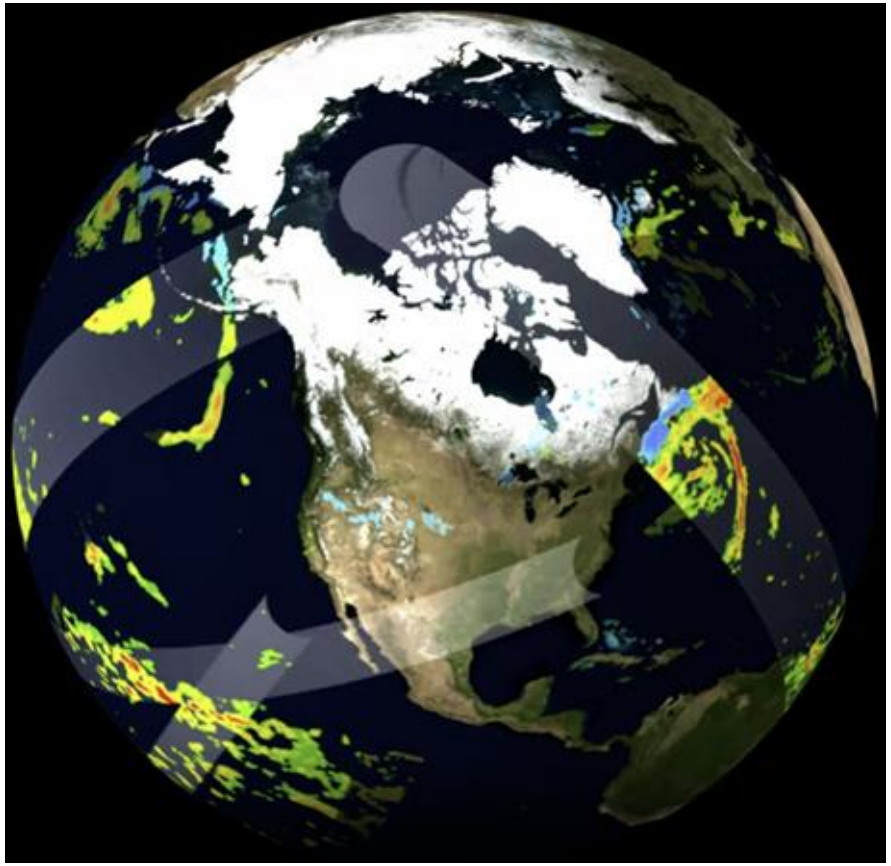
The GPM Constellation

GMI (Reference instrument)



⇒ Can we do spatio-temporal interpolation of GMI images?

The GPM constellation, NASA



GPM Constellation covers the Earth, NASA's Scientific Visualization Studio

156

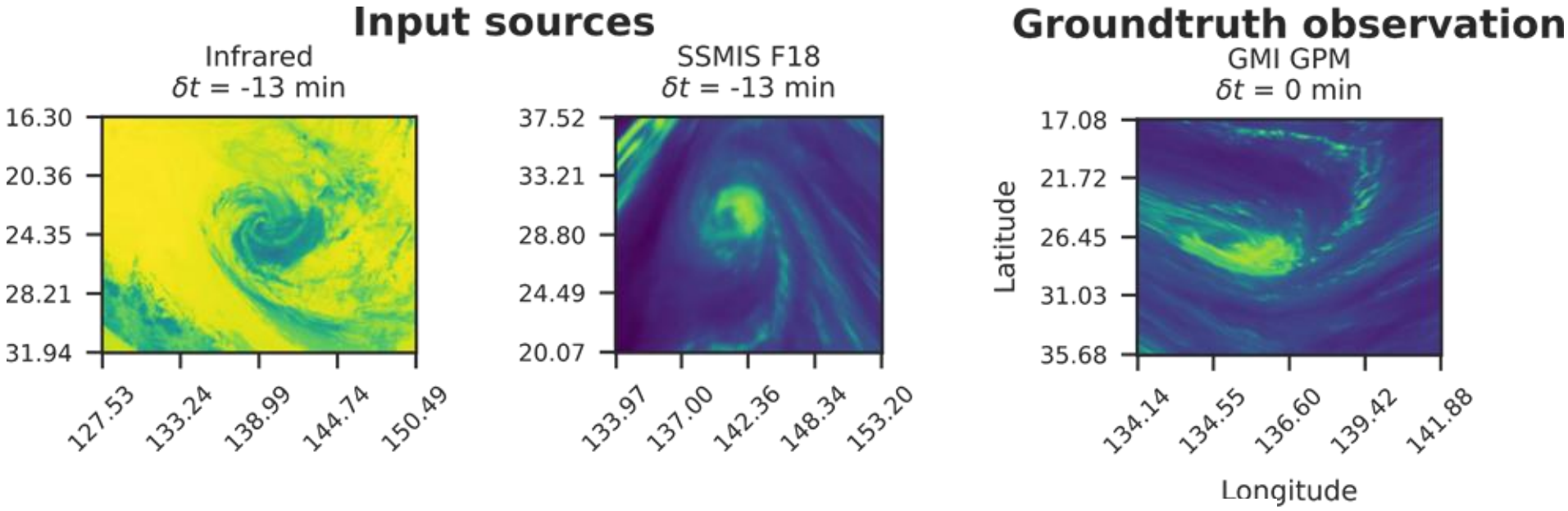
Dataset

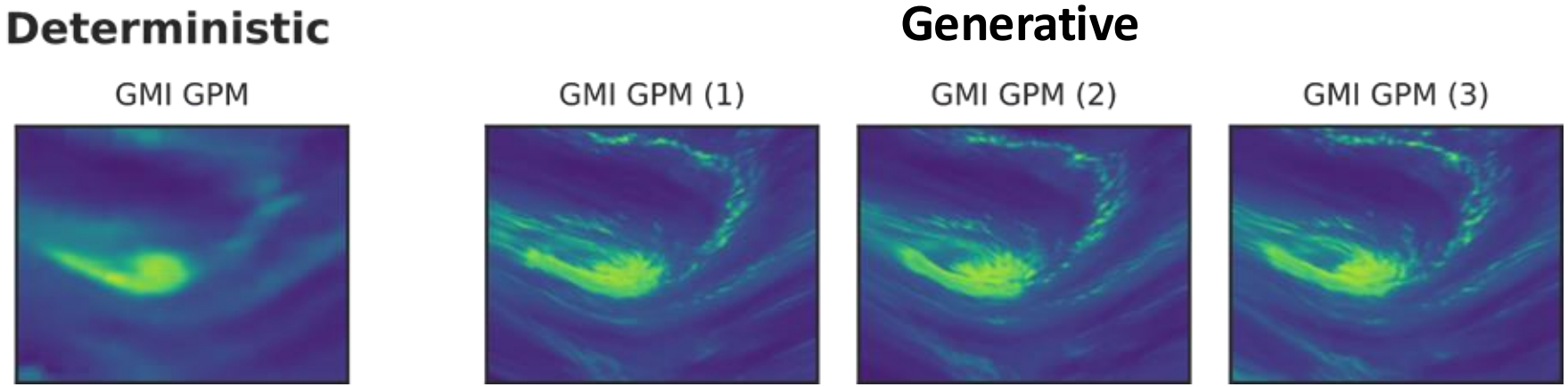
TC-PRIMED (Razin et al. 2023)

- ~3k storms, 1987 to 2024, all basins
- 2005-07-14-21-23 for validation, 2006-08-15-22-24 for testing
- 11 microwave sensors
- Infrared observations

Interpolation windows of 6 hours

Results: Captures fine structure better than deterministic model





Training pipelines compared for the generative version

- Supervised training

Non-GMI sensors —Model→ GMI

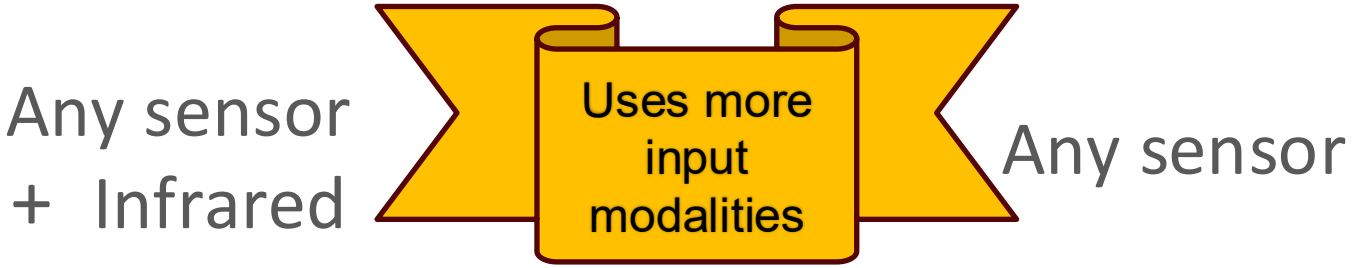
⇒ ~6k samples

- Self-supervised training



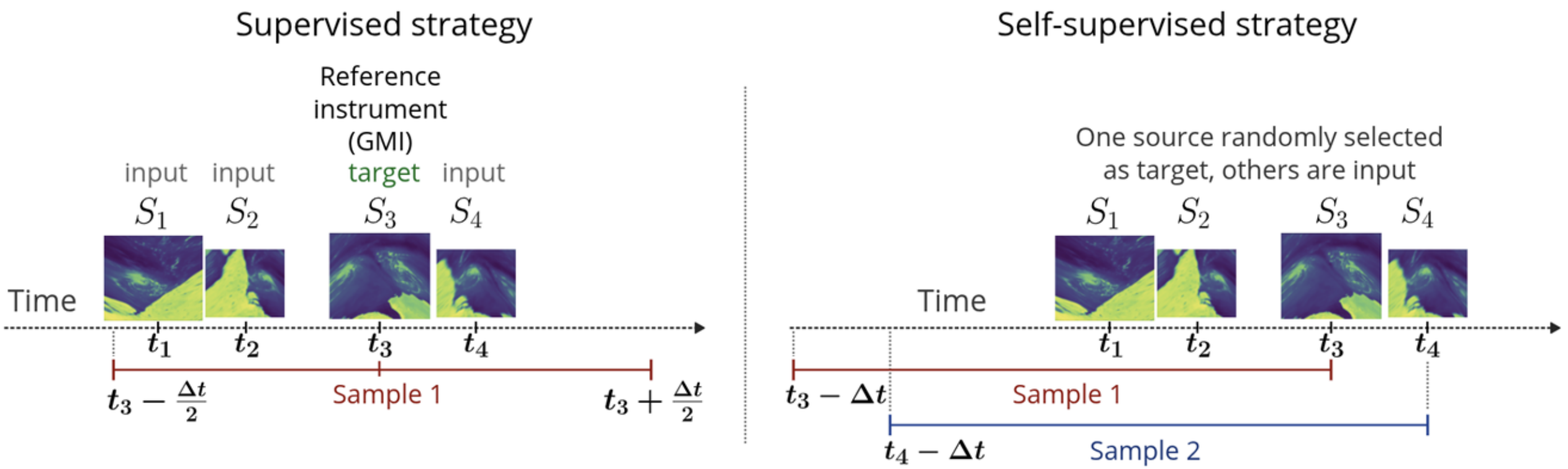
⇒ ~48k samples

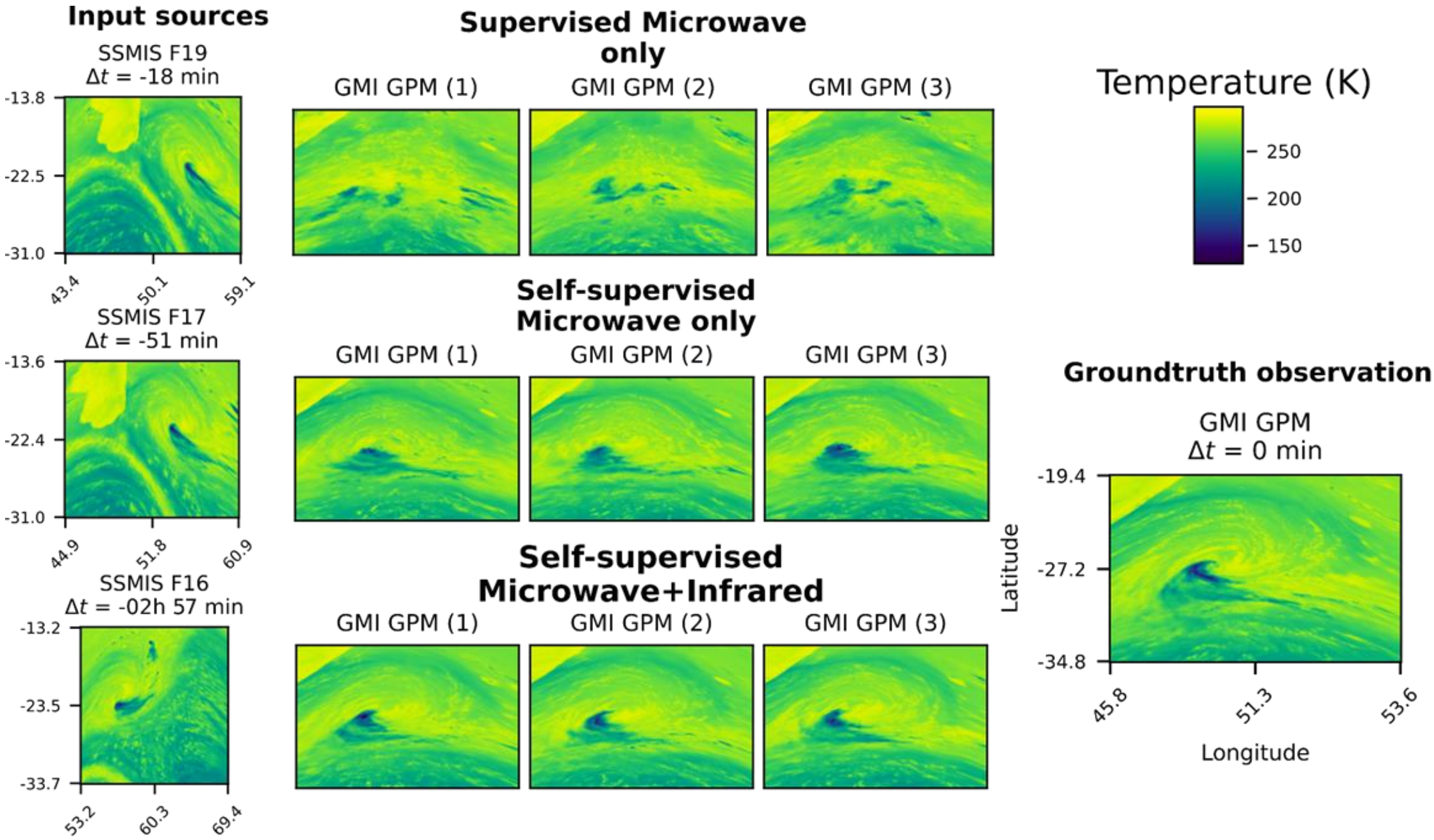
- SSL plus Infrared

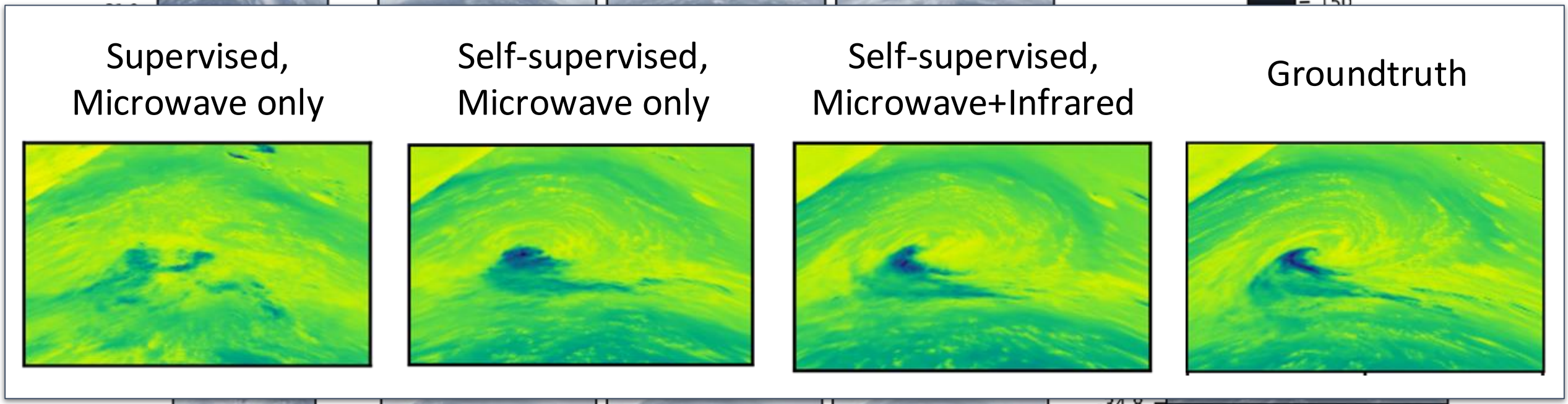


⇒ ~44k samples

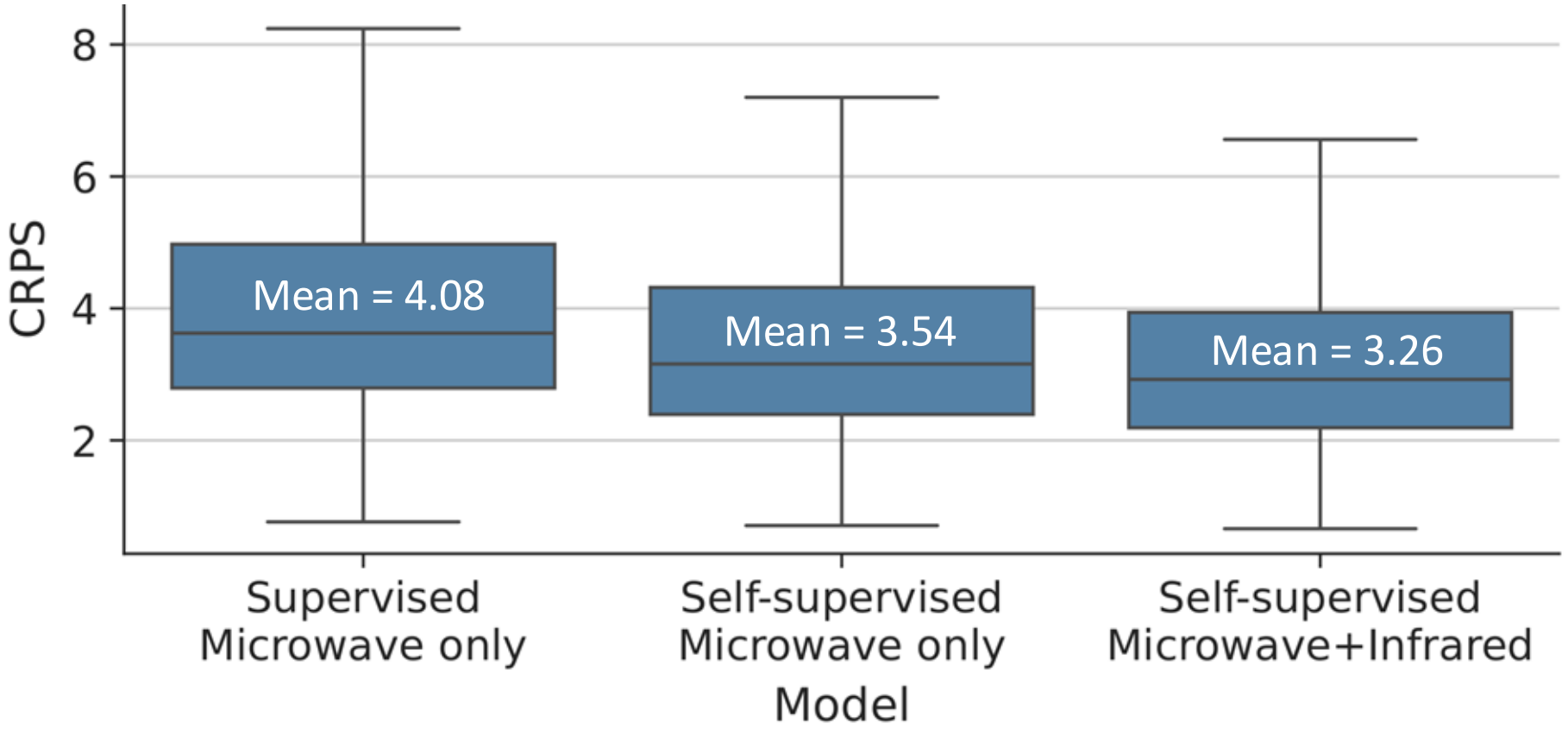
Training tasks




Input sources
SSMIS F19
Δt = -18 min
-13.8
-22.5
-31.0
43.4
50.1
59.1
SSMIS F17
Δt = -51 min
-13.6
-22.4
-31.0
44.9
51.8
60.9
SSMIS F16
Δt = -02h 57 min
-13.2
-23.5
-33.7
53.2
60.3
69.4
Supervised Microwave only
GMI GPM (1)
GMI GPM (2)
GMI GPM (3)
Self-supervised Microwave only
GMI GPM (1)
GMI GPM (2)
GMI GPM (3)
Self-supervised Microwave+Infrared
GMI GPM (1)
GMI GPM (2)
GMI GPM (3)
Temperature (K)
250
200
150
Groundtruth observation
GMI GPM
Δt = 0 min
-19.4
-27.2
-34.8
45.8
51.3
53.6
Latitude
Longitude


Supervised,
Microwave only
Self-supervised,
Microwave only
Self-supervised,
Microwave+Infrared
Groundtruth

Generative version: ablations



Results

Model type	Training strategy	Input sources	CRPS	RMSE	MAE	MAPE (%)
Deterministic	Self-supervised	MW+IR	-	9.32 [9.19, 9.45]	5.15 [5.07, 5.22]	2.26 [2.22, 2.30]
Flow matching	Supervised Baseline	MW only	4.08 [4.02, 4.14]	10.92 [10.77, 11.07]	6.29 [6.19, 6.38]	2.79 [2.73, 2.84]
Flow matching	Self-supervised	MW only	3.54 [3.49, 3.60]	9.72 [9.58, 9.85]	5.31 [5.23, 5.39]	2.34 [2.30, 2.39]
Flow matching	Self-supervised	MW+IR	**3.26** [3.21, 3.31]	**9.01** [8.89, 9.13]	**4.76** [4.69, 4.83]	**2.09** [2.05, 2.13]

Brackets indicate 95% confidence intervals

Conditional Ranked Probability Score (CRPS)

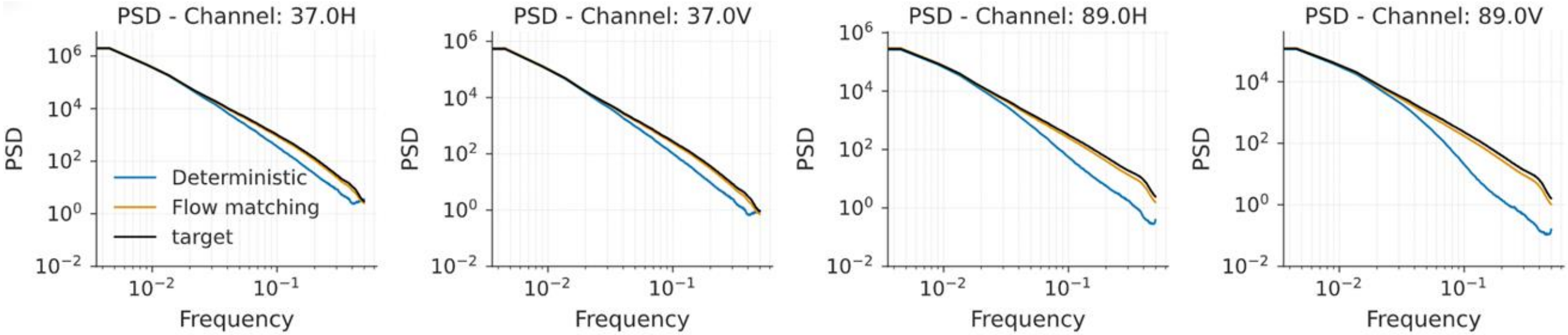
$$\mathrm{CRPS}(F, x) = -\int_{-\infty}^{\infty} (F(y) - \mathbb{1}\{y \geq x\})^2 \, \mathrm{d}y$$

- A different formulation avoids computation of this integral
- CRPS can be directly compared to the MAE of a deterministic forecast
- Energy Score is the multi-variate extension of CRPS

Further reading on these and other evaluation measures for probabilistic forecasts:

Raftery & Gneiting (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. JASA

Evaluation against deterministic model - Spectrum

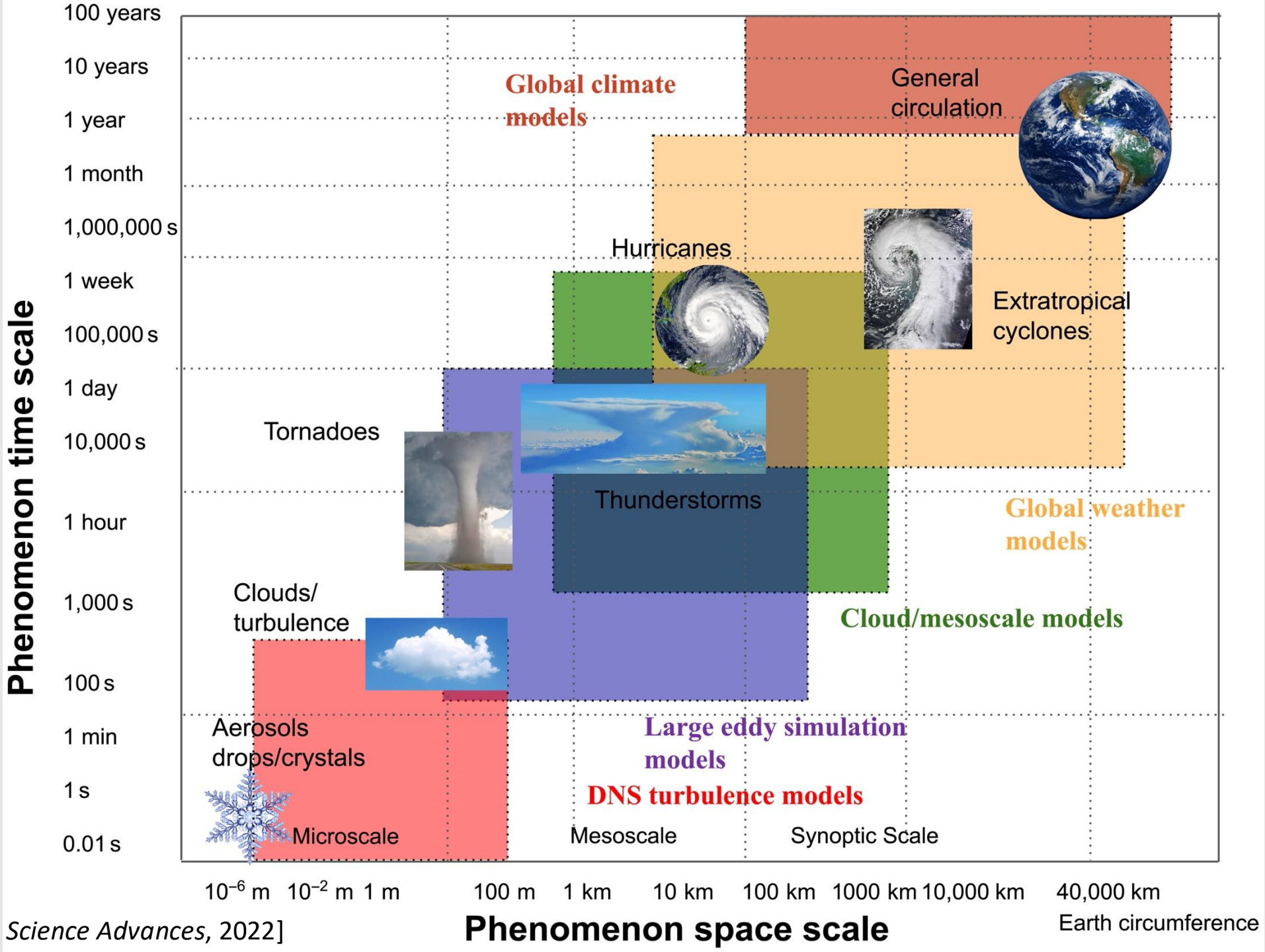


Generative AI for:

✓ Probabilistic forecasting
- **ArchesWeatherGen**: Efficient weather forecasting trained at 1.5°
- **ArchesClimate**: Decadal climate emulation, interpolation, extrapolation

✓ Data fusion with probabilistic domain alignment
- **MotifGen**: A generative, multi-source, self-supervised model for learning from geospatial data

❑ Probabilistic domain alignment and downscaling
- **ClimAlign**: Generative unpaired downscaling with domain alignment
- **SerpentFlow**: Generative unpaired domain alignment via shared-structure decomposition (plus application to downscaling)



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



Unsupervised Downscaling of Climate Models via Domain Alignment: *Application to Wind Fields.*

Julie Keisler, Boutheina Oueslati,
Anastase Charantonis, Yannig Goude,
Claire Monteleoni

Proceedings of Climate Informatics 2026

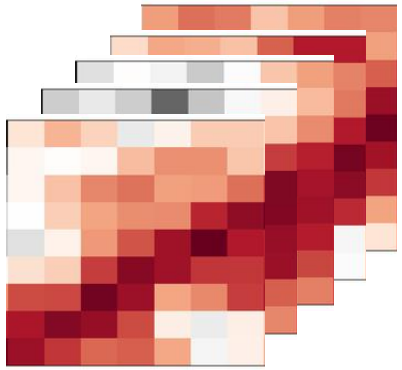
Climate model downscaling

Coarse spatial resolution: GCMs operate at ~100+ km resolution, missing critical local-scale processes

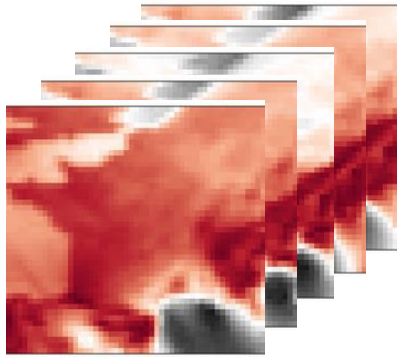
Limited representation: Fine-scale features like topographic effects and land-sea contrasts are poorly captured

Systematic biases: GCM outputs deviate significantly from observed climate patterns

Local impact studies compromised: Direct GCM outputs inadequate to assess future wind patterns for energy planning

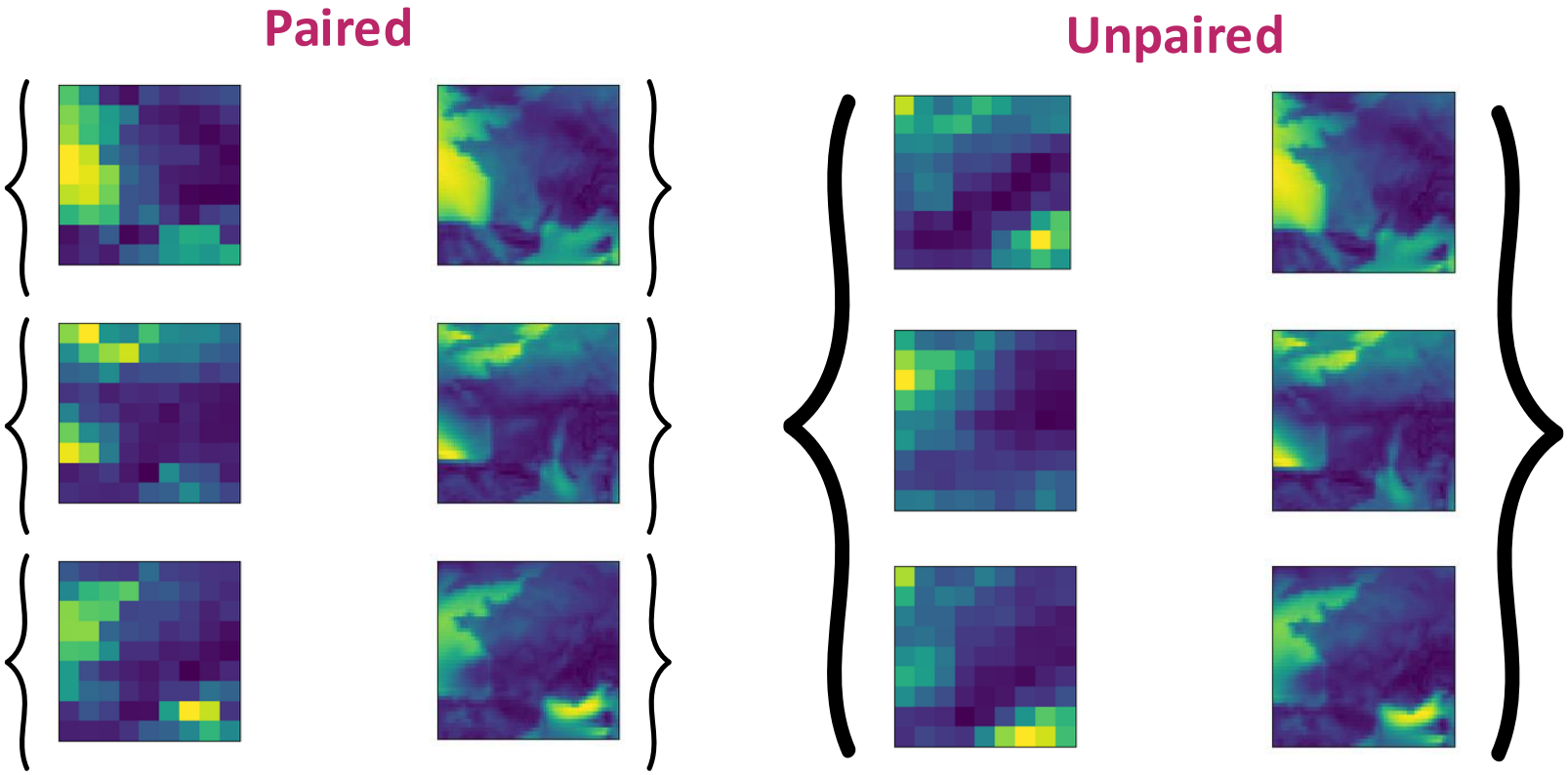


GCM ACCESS-ESM1 over France over 5 days



ERA5 over France over 5 days

Downscaling as unpaired domain alignment



- GCM: **physically plausible trajectories**
- Reanalyses: actual **historical weather** sequence → **GCM and reanalysis are unpaired**
- Downscaling = **distribution alignment**: map GCM outputs onto the reanalysis distribution

Limitations of Traditional Statistical Methods

Distribution-based methods

- Quantile Mapping, CDF-t (Michelangeli et al. 2009), marginal distributions correction

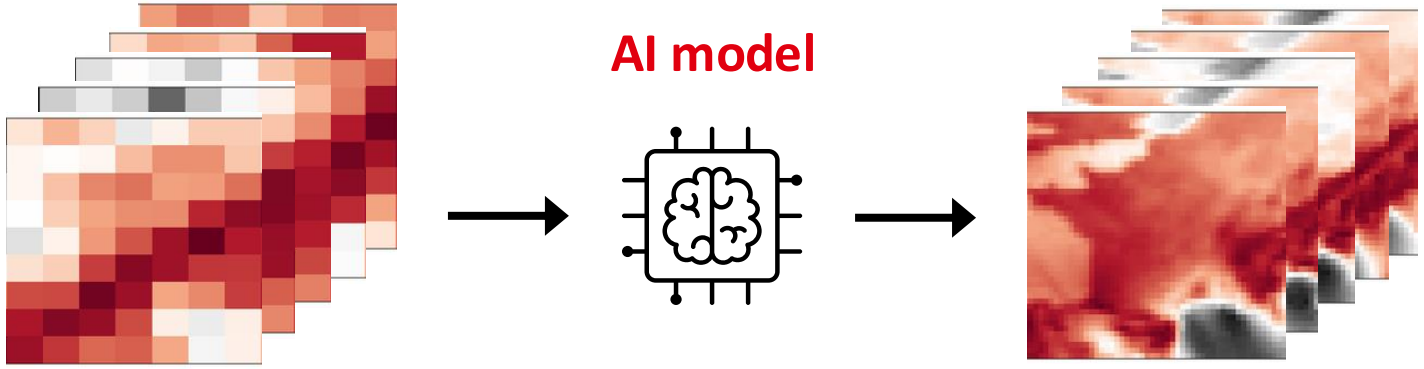
Multivariate methods (optimal transport)

- R2D2 (Mehrotra & Sharma 2020), dOTC (Robin et al. 2019)

Curse of high-dimensionality: traditional methods fail to model all the interactions at stake

⟶ AI / Deep learning based methods

Generative AI for Downscaling



Advantages

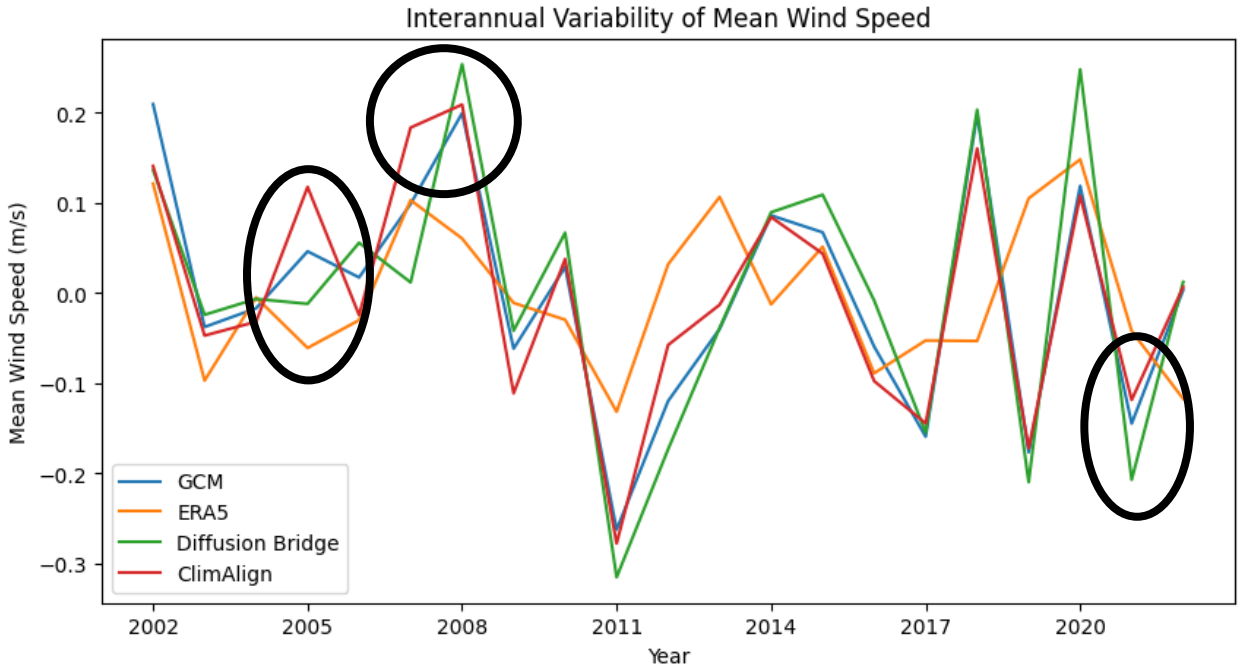
- Learns the **full joint distribution** of high-resolution fields
- Generates **multiple realizations**

Challenges

- Unpaired data ⟶ **No ground truth**
- Implicit methods can drift from the GCM signal

= projections **untrustworthy**

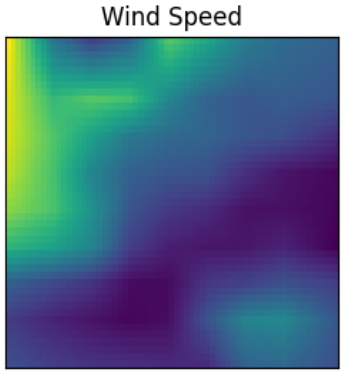
(ClimAlign/Groenke et al. 2020, diffusion bridges/Bischoff & Deck 2024)



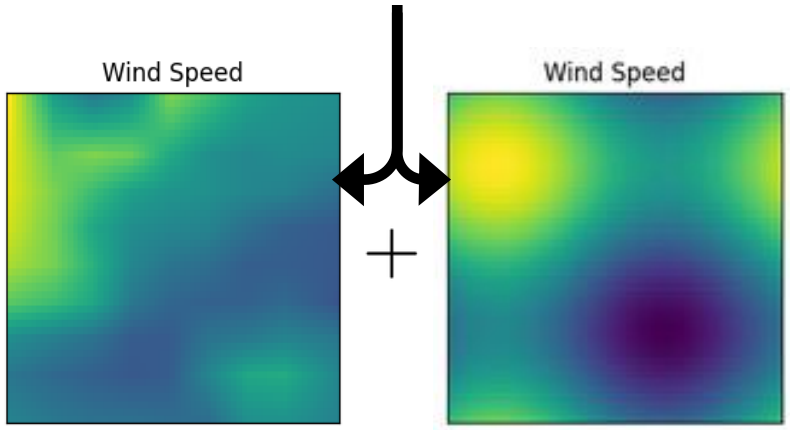
SerpentFlow: introducing pseudo-pairs

SerpentFlow is an AI-based generative method for **unpaired domain alignment** via Shared-Structure Decomposition

GCM average wind speed

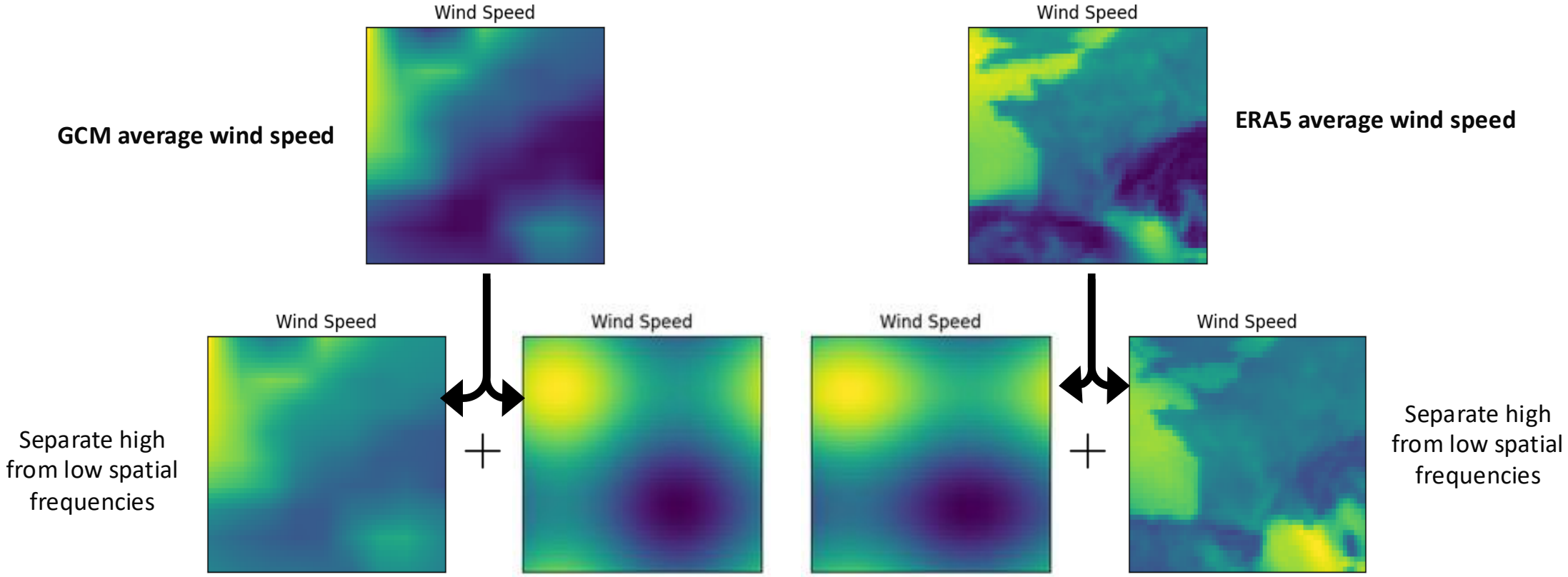


Separate high from low spatial frequencies



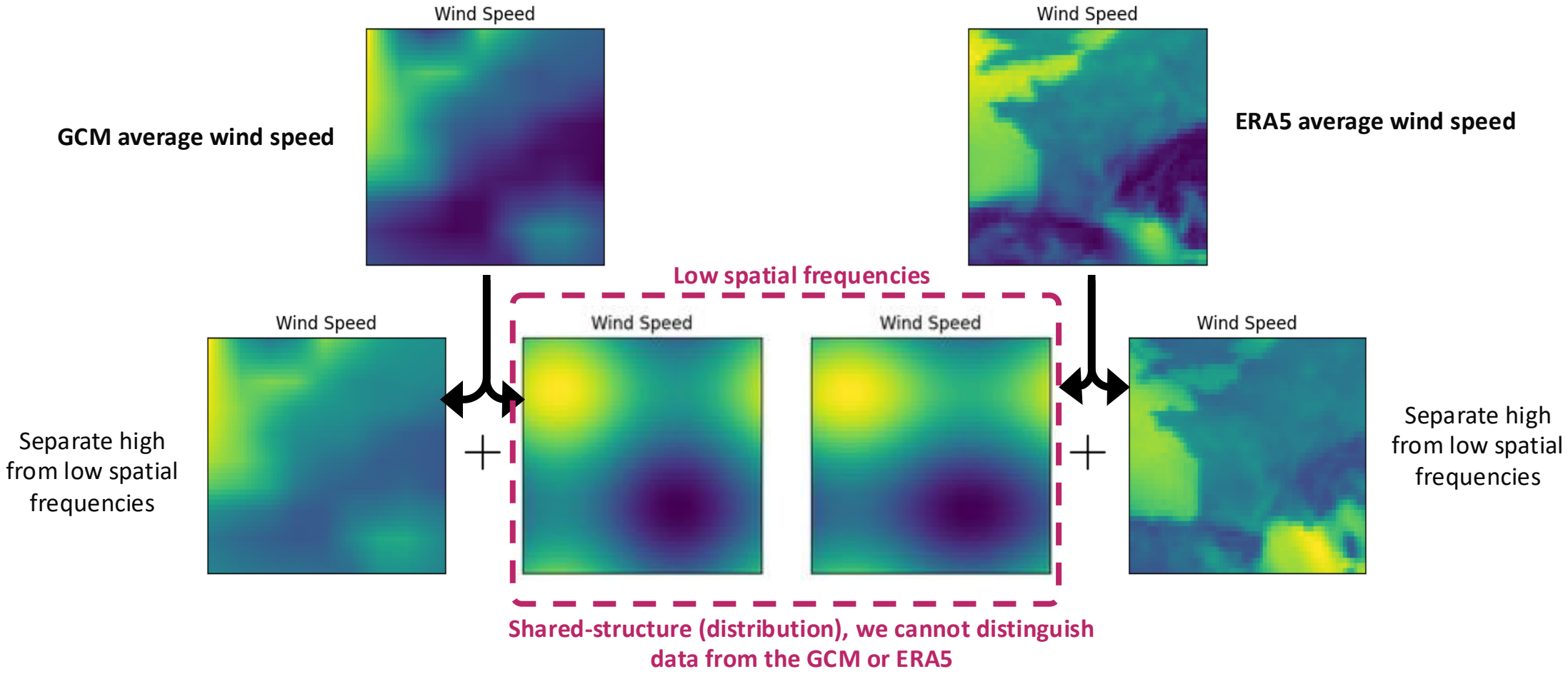
SerpentFlow: introducing pseudo-pairs

SerpentFlow is an AI-based generative method for **unpaired domain alignment** via Shared-Structure Decomposition



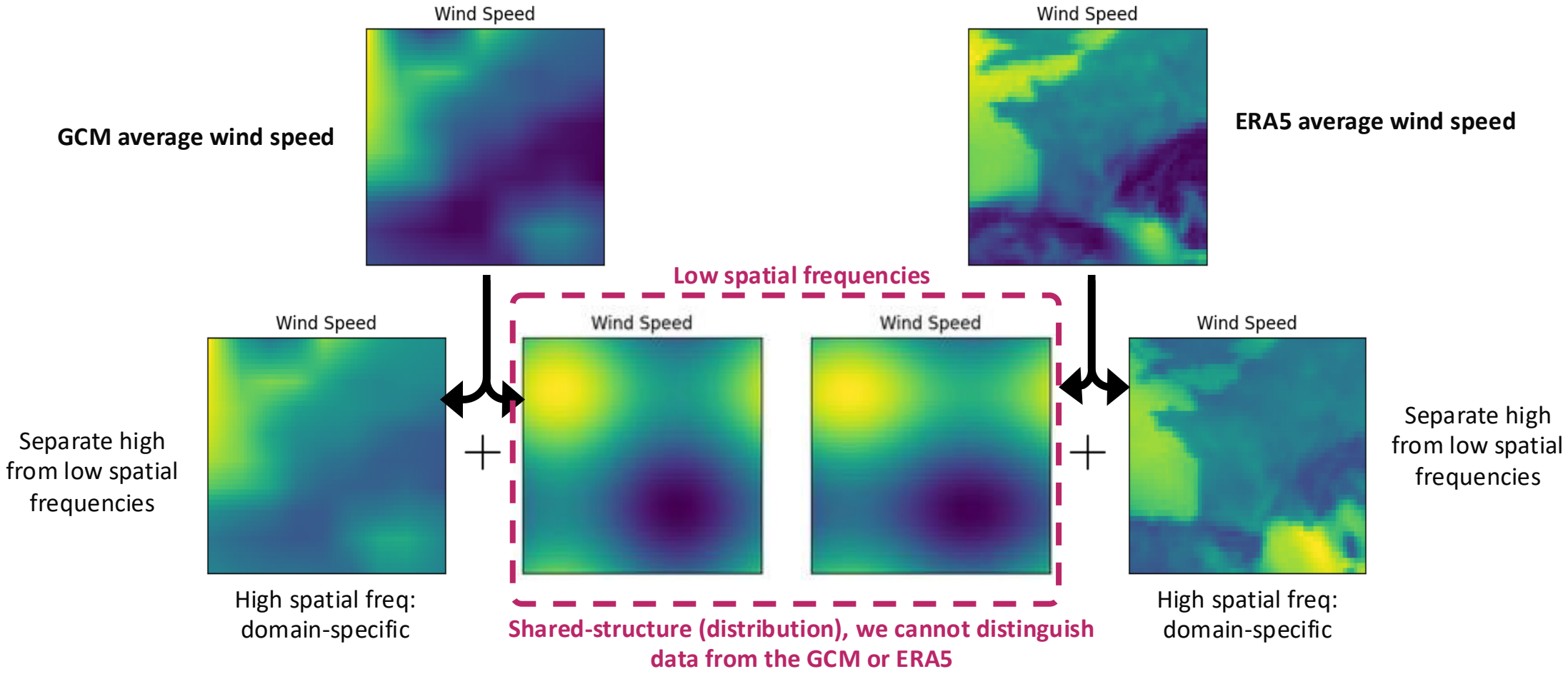
SerpentFlow: introducing pseudo-pairs

SerpentFlow is an AI-based generative method for **unpaired domain alignment** via Shared-Structure Decomposition

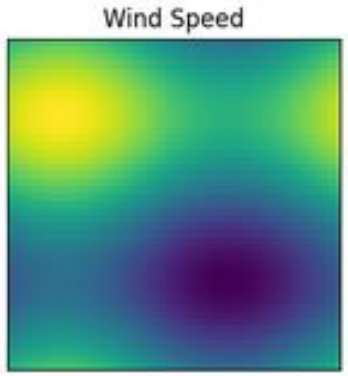


SerpentFlow: introducing pseudo-pairs

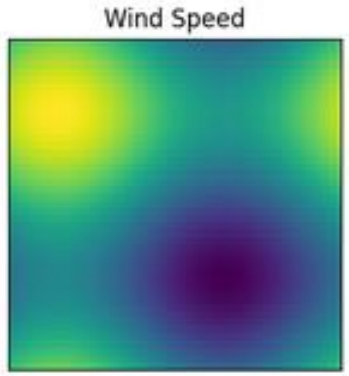
SerpentFlow is an AI-based generative method for **unpaired domain alignment** via Shared-Structure Decomposition



Automatically find the cutoff frequency



Low pass filter(HR) with cutoff frequency w



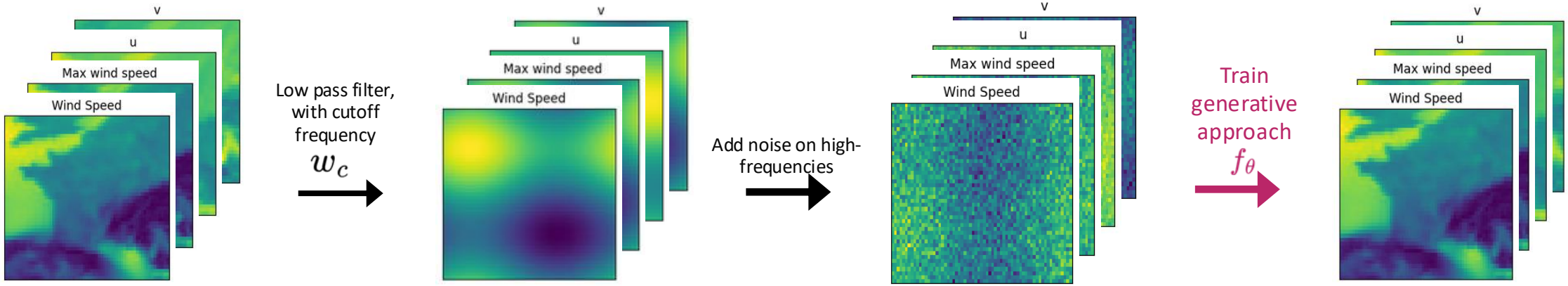
Low pass filter(LR) with cutoff frequency w

→ Classifier → LR or HR data?

- Stop reducing cutoff frequency when accuracy <= 60%
- Often found frequency ~= **effective resolution** of the GCM

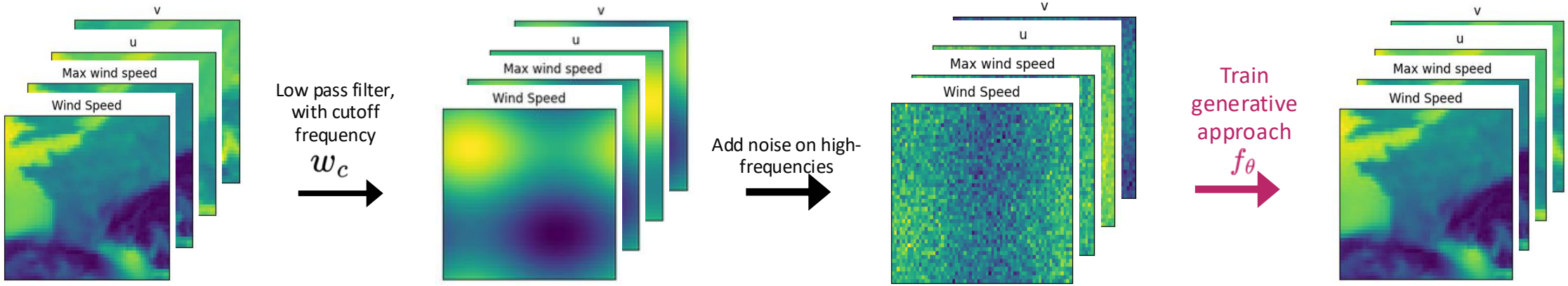
SerpentFlow pipeline

Training

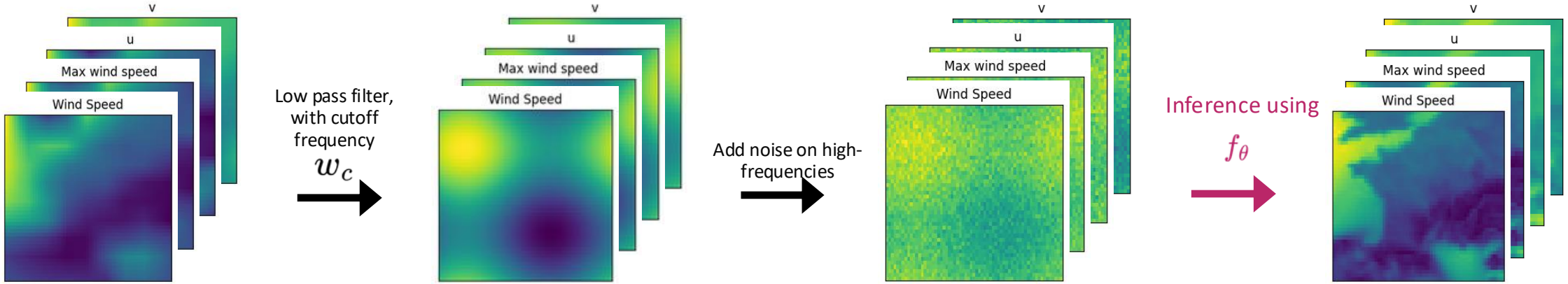


SerpentFlow pipeline

Training

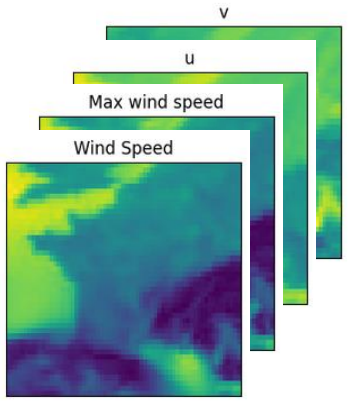


Inference



SerpentFlow pipeline

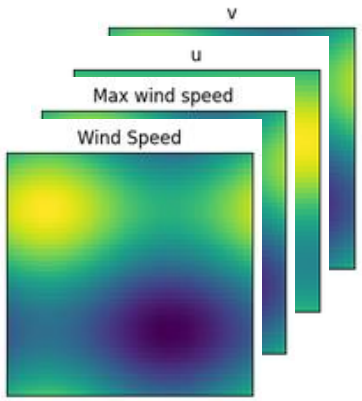
Training



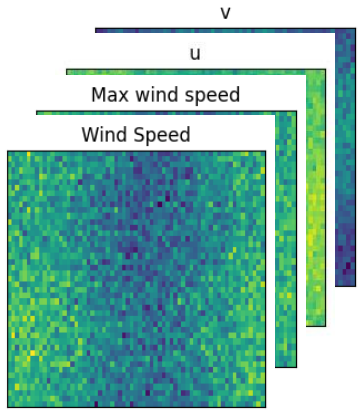
Reference (obs/rea)

Low pass filter, with cutoff frequency w_c

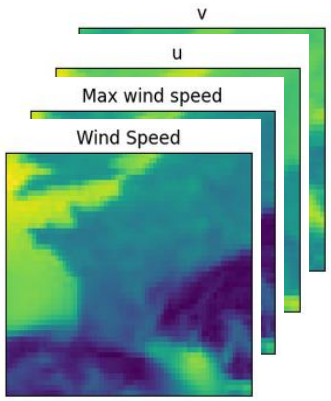
Joint distribution



Add noise on high-frequencies

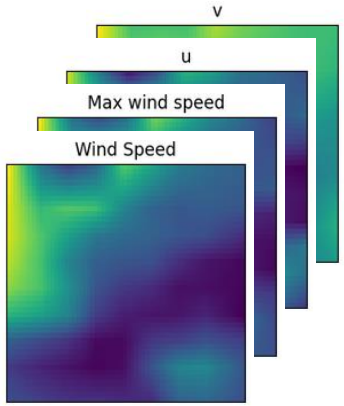


Train generative approach f_θ



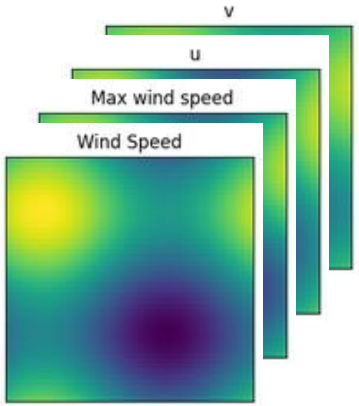
Reference (obs/rea)

Inference

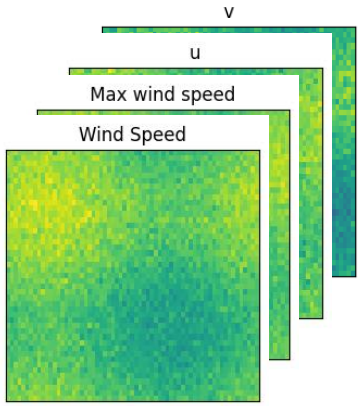


GCM interpolated

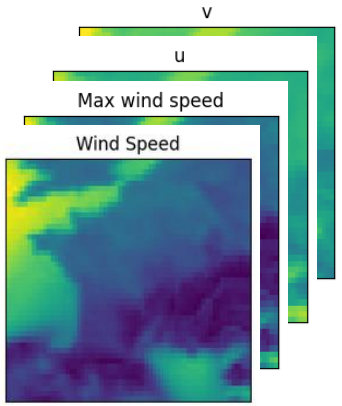
Low pass filter, with cutoff frequency w_c



Add noise on high-frequencies

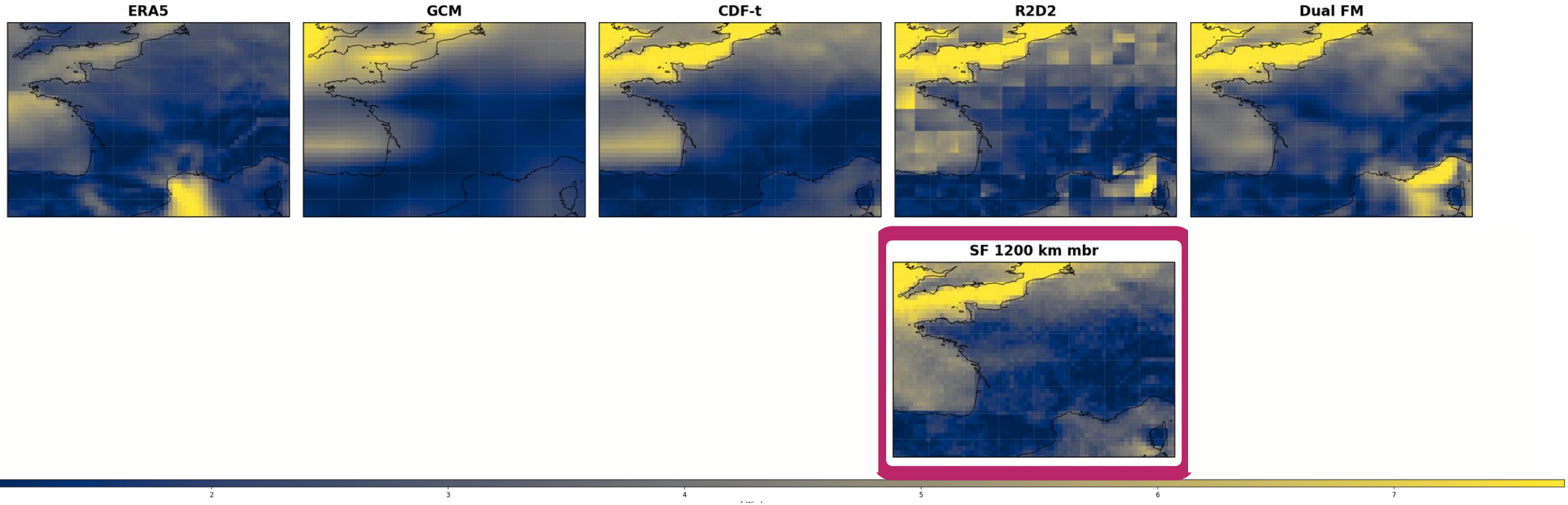


Inference using f_θ



Downscaled GCM

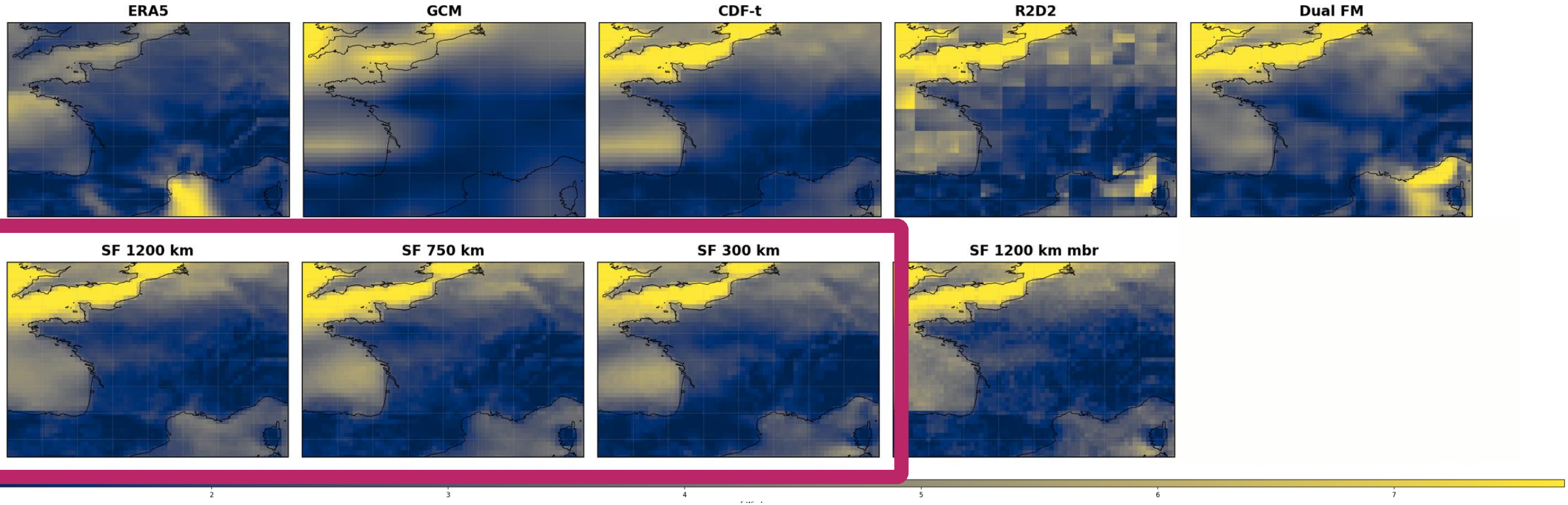
Did we approximate ERA5 distribution?



Key findings

- DL methods outperform CDF-t & R2D2 on spatial structure.
- **SF 1200 km** (1 member) **best approximation** of ERA5 distribution

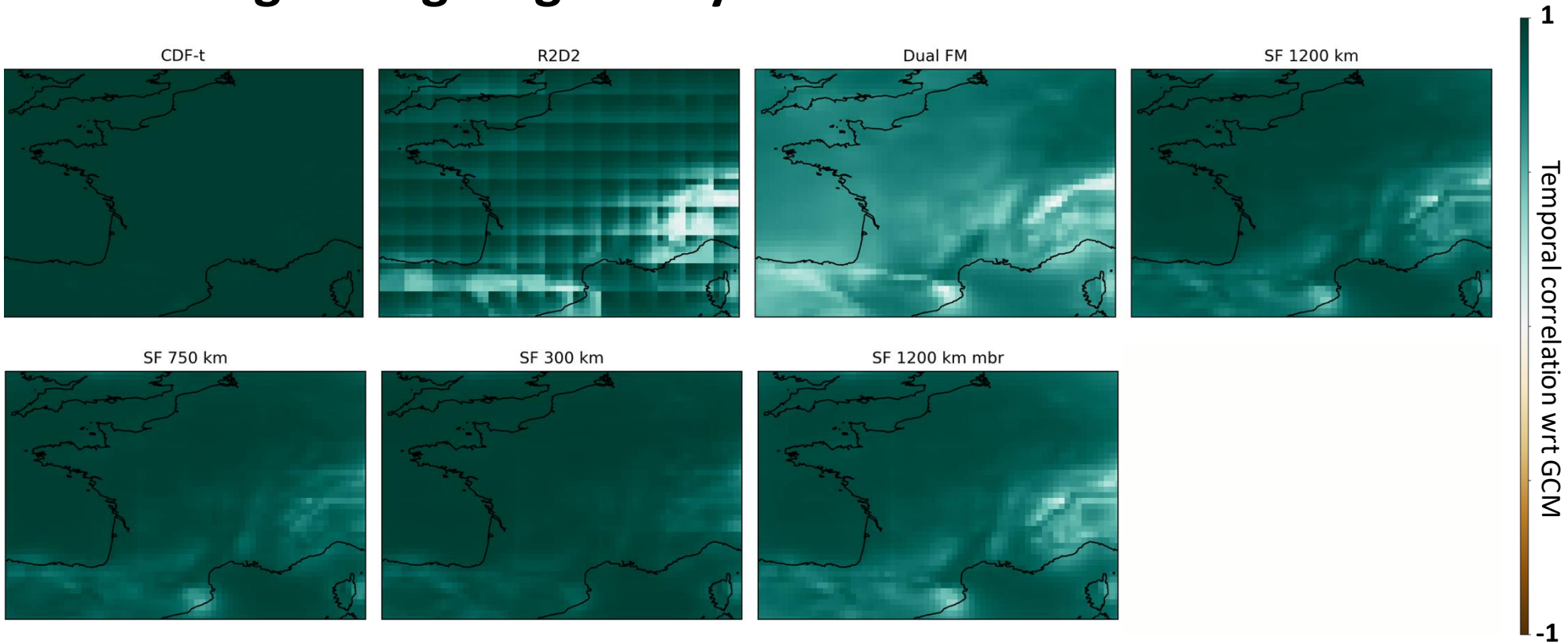
Did we approximate ERA5 distribution?



Key findings

- DL methods outperform CDF-t & R2D2 on spatial structure.
- **SF 1200 km** (1 member) **best approximation** of ERA5 distribution
- **higher cutoff frequency** ⟶ drift from ERA5 distribution.

Are we following the signal given by the GCM?



Key findings

- **CDF-t closest to GCM** dynamics, Dual FM does not follow the GCM track
- SF drift appears in the **mountain** regions
- **Lower cutoff frequency** ⟶ closer to GCM dynamics

Outlook

- Works across **GCMs, resolutions & irregular domains** — ERA5 (25 km) and SAFRAN (8 km)
- Outperforms statistical methods to model the **joint distribution** of the reanalysis data
- Compared to other generative approaches, it stays close to the GCM dynamics
- **Cutoff frequency is interpretable**
  - Links scale separation directly to the GCM's effective resolution
  - Manages **trade-off** between following the GCM track or matching the reanalysis distribution
- **Limits:** near-surface wind only — next: other variables, extremes, temporal downscaling

Generative AI for:

- ✓ Probabilistic forecasting
  - **ArchesWeatherGen**: Efficient weather forecasting trained at 1.5°
  - **ArchesClimate**: Decadal climate emulation, interpolation, extrapolation

- ✓ Data fusion with probabilistic domain alignment
  - **MotifGen**: A generative, multi-source, self-supervised model for learning from geospatial data

- ✓ Probabilistic domain alignment and downscaling
  - **ClimAlign**: Generative unpaired downscaling with domain alignment
  - **SerpentFlow**: Generative unpaired domain alignment via shared-structure decomposition (plus application to downscaling)

Getting started with flow-based generative AI

Flow Matching

Tutorials:

- Peter Holderrieth et al. (2026) “An Introduction to Flow Matching and Diffusion Models” Lecture notes and videos: `diffusion.csail.mit.edu`
- Yaron Lipman et al. (2024) “Flow Matching Guide and Code” arXiv:2412.06264
- Emile Pierret, Valentine Tosel, Julie Delon, Alasdair Newson (2026) “Flow Matching for Applied Mathematicians” hal-05538982v1

Further reading on flow-based methods

Stochastic Interpolants

General flow-based setting to transport from noise to data

- Flow Matching, Rectified Flows are examples
- Generalizes the setting so you can train from ODE and sample from SDE or vice-versa

Neural Operators

Flow-based methods work in operator space, allowing for scale-agnostic generalization

Consistency Models

Single-step version of Flow Matching. Open research area with some theory, practical limitations.

Reverse Markov Chain Learning

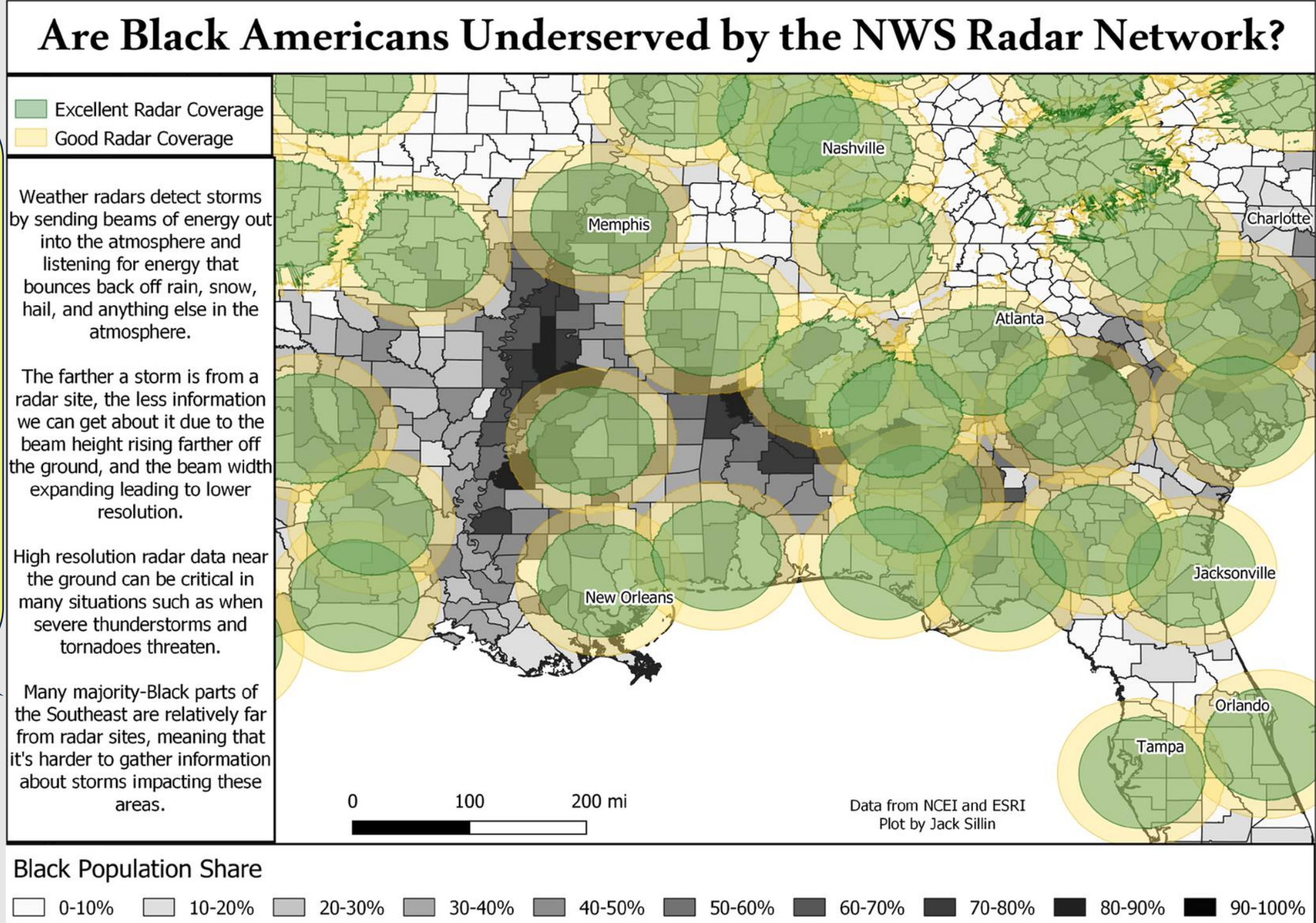
Essentially flow matching from low to high resolution

Xinwei Shen, Nicolai Meinshausen, Tong Zhang (2025) “Reverse Markov Chain Learning: Multi-Step Generative Models for Complex Distributions” arXiv:2502.13747

Bonus section: AI 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
  - Train and validate them in high-data regions
  - Fine-tune them using the limited data in the low-data regions and use them to generate more data.

- 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