Self-Supervised Learning for the Geosciences

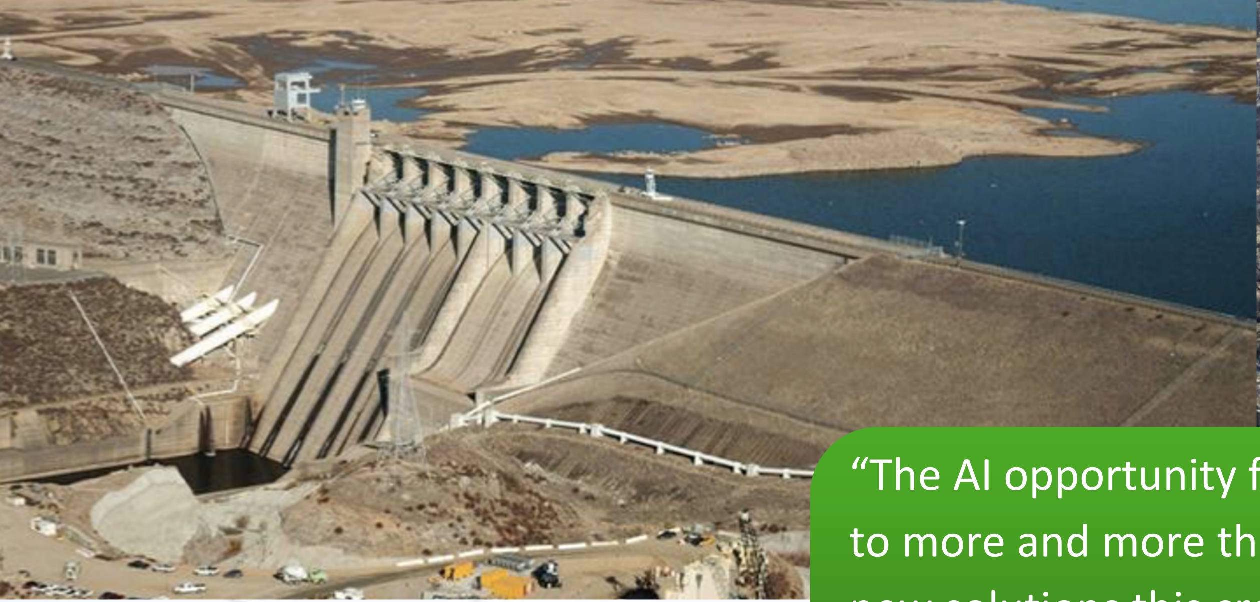
Claire Monteleoni

INRIA Paris

University of Colorado Boulder

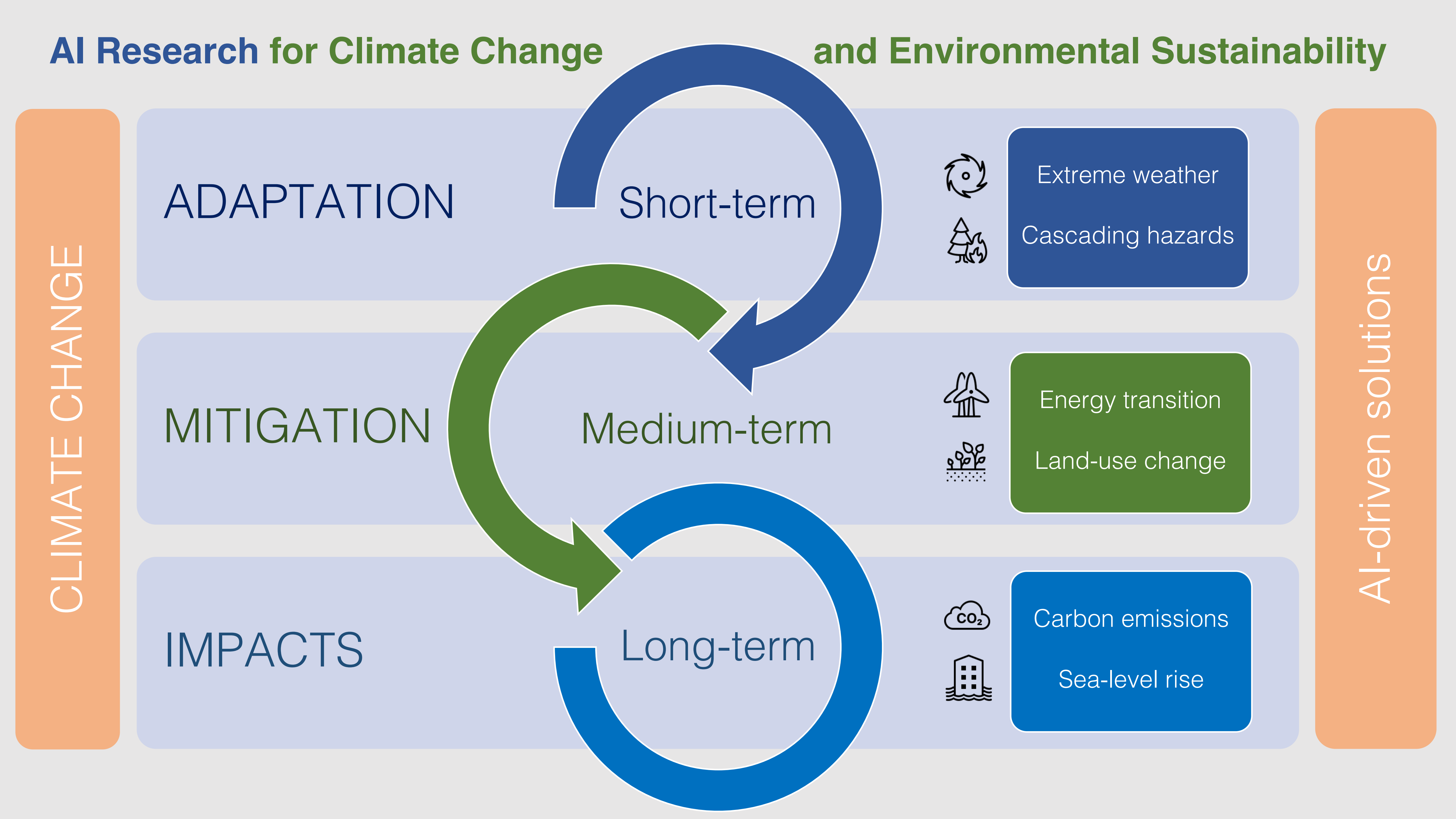
Inria

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



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
CO2
AI-driven solutions

Today: Self-supervised learning for geospatial data

What is self-supervised learning?

Normalizing flows for downscaling geospatial data

A pretext task for temporal downscaling of geospatial data

Outline

What is self-supervised learning?

Normalizing flows for downscaling geospatial data

A pretext task for temporal downscaling of geospatial data

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:

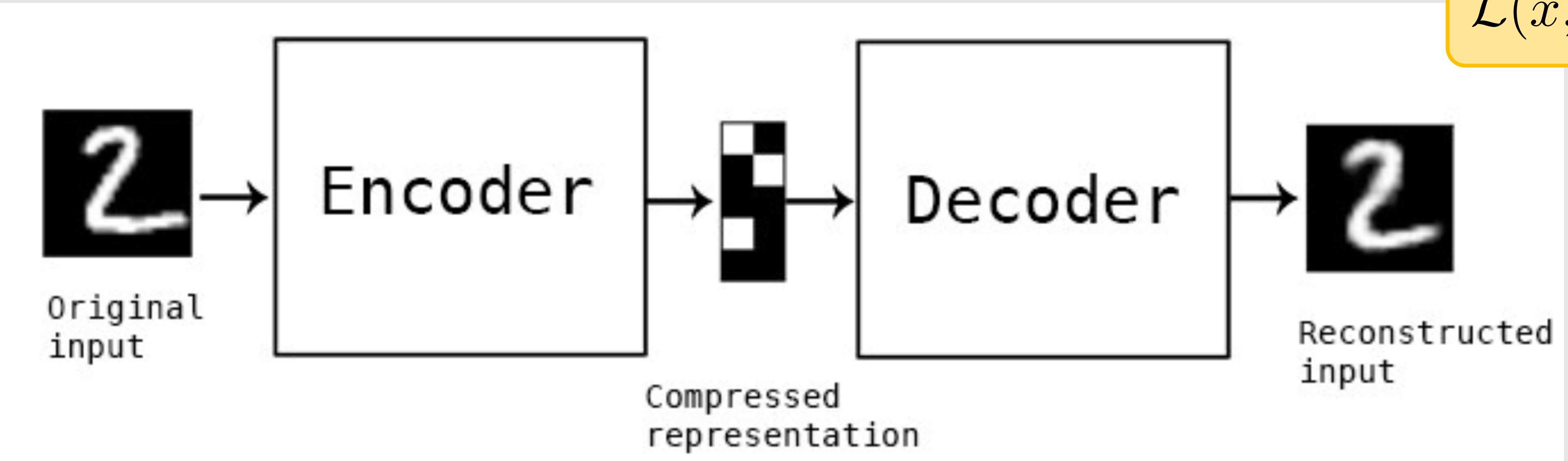
- Design a supervised learning task using only the available data.
- Train a model on this task such that,
- the learned features (or the learned posterior over a feature space) will be useful for another (down-stream) task.

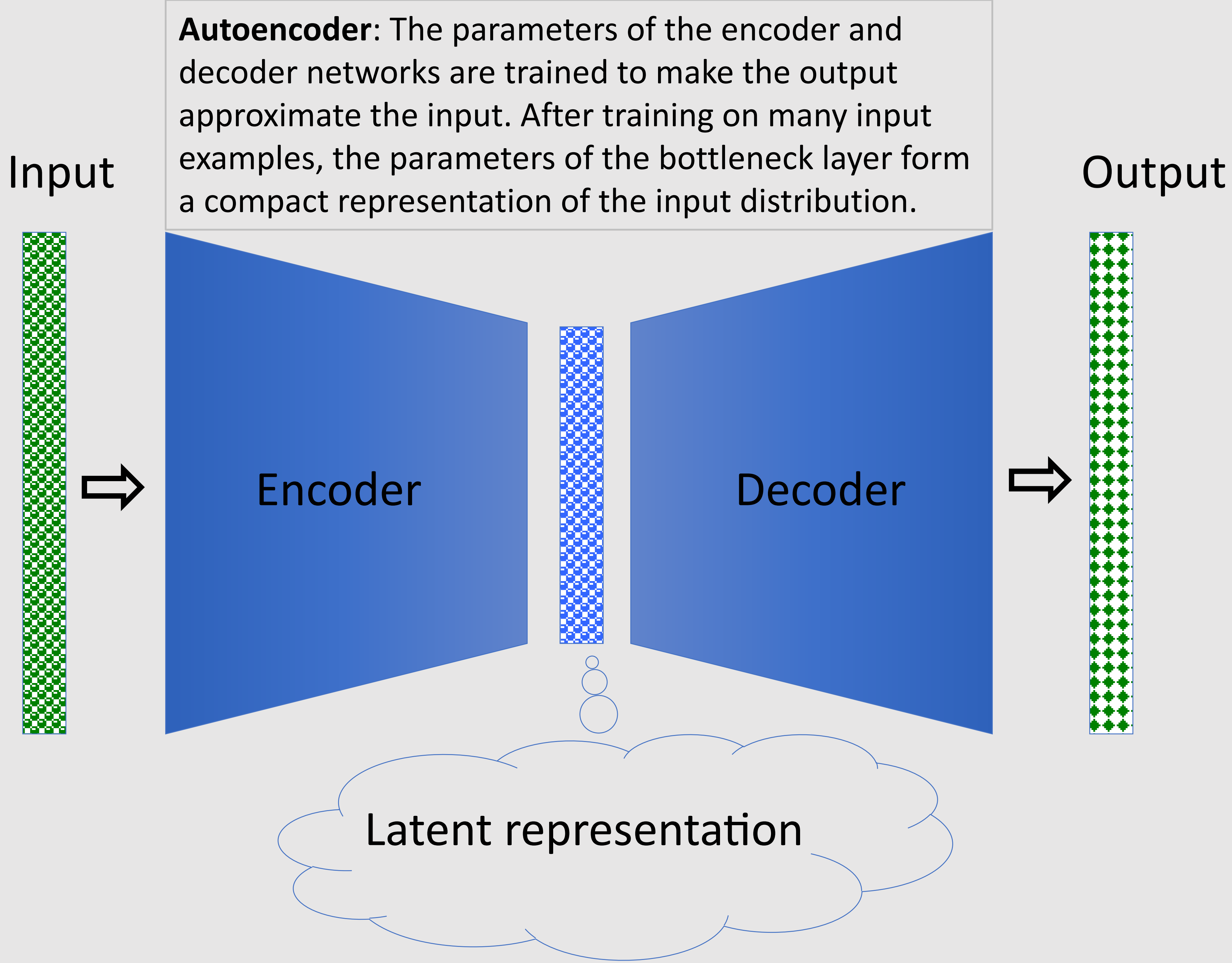
Pretext Task: Example

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

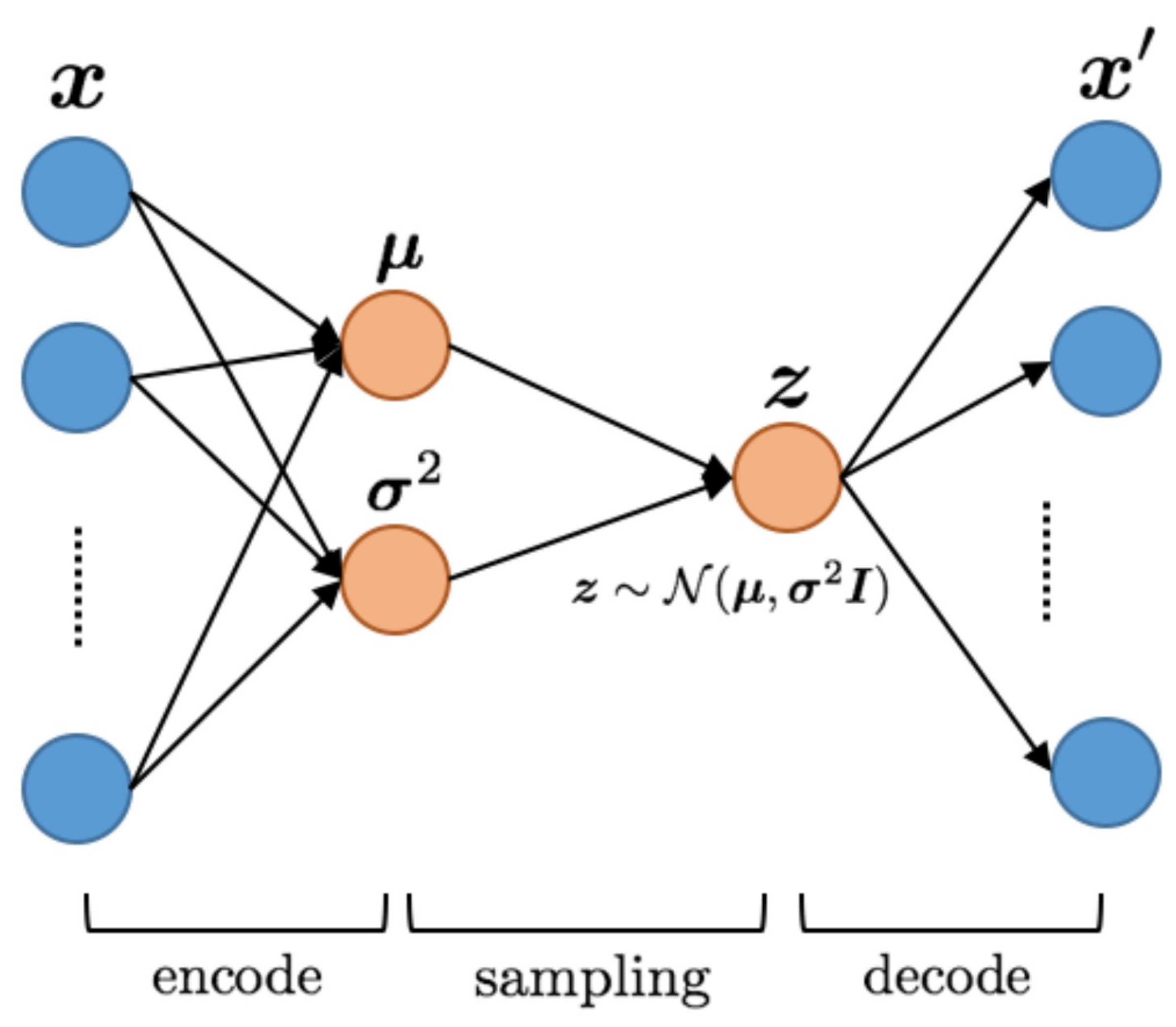
$\mathcal{L}(\hat{x}, x)$



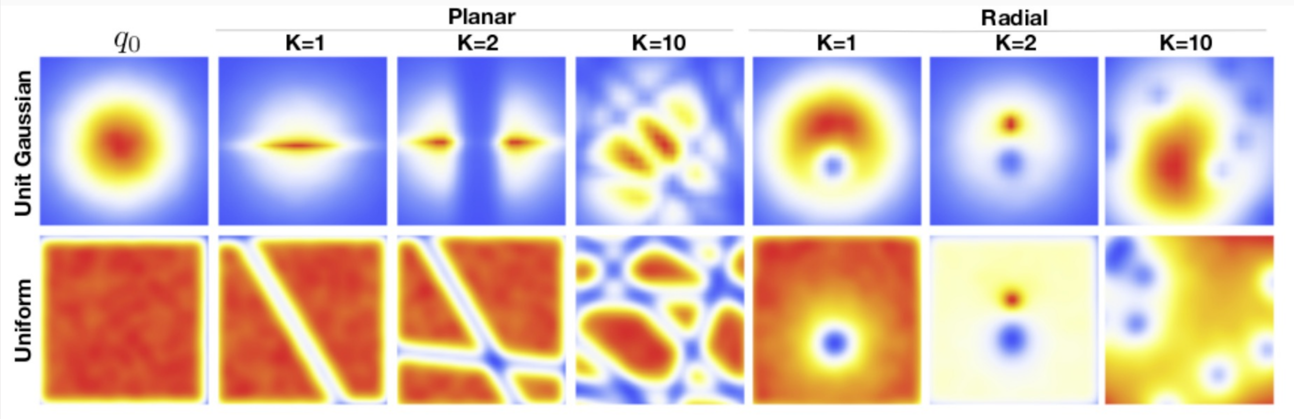

Autoencoder: The parameters of the encoder and decoder networks are trained to make the output approximate the input. After training on many input examples, the parameters of the bottleneck layer form a compact representation of the input distribution.
Input
Output
Encoder
Decoder
Latent representation

Variational Autoencoder (VAE)

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



Normalizing Flows

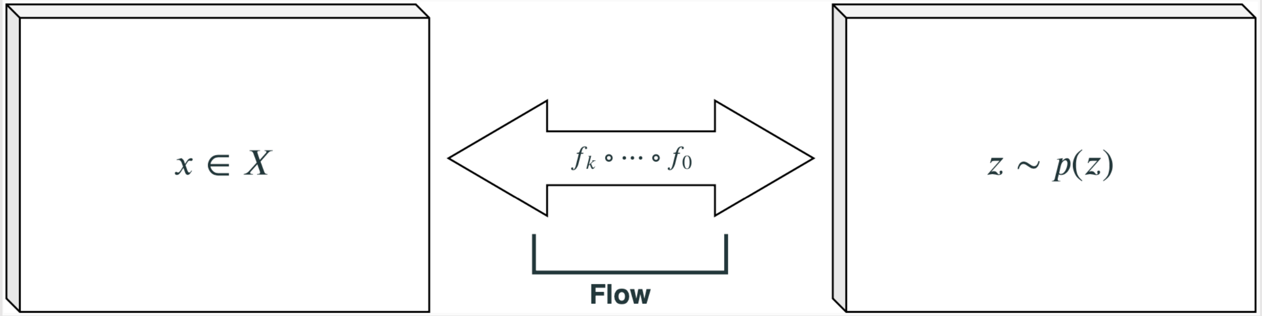


[Rezende & Mohamed, ICML 2015]

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?

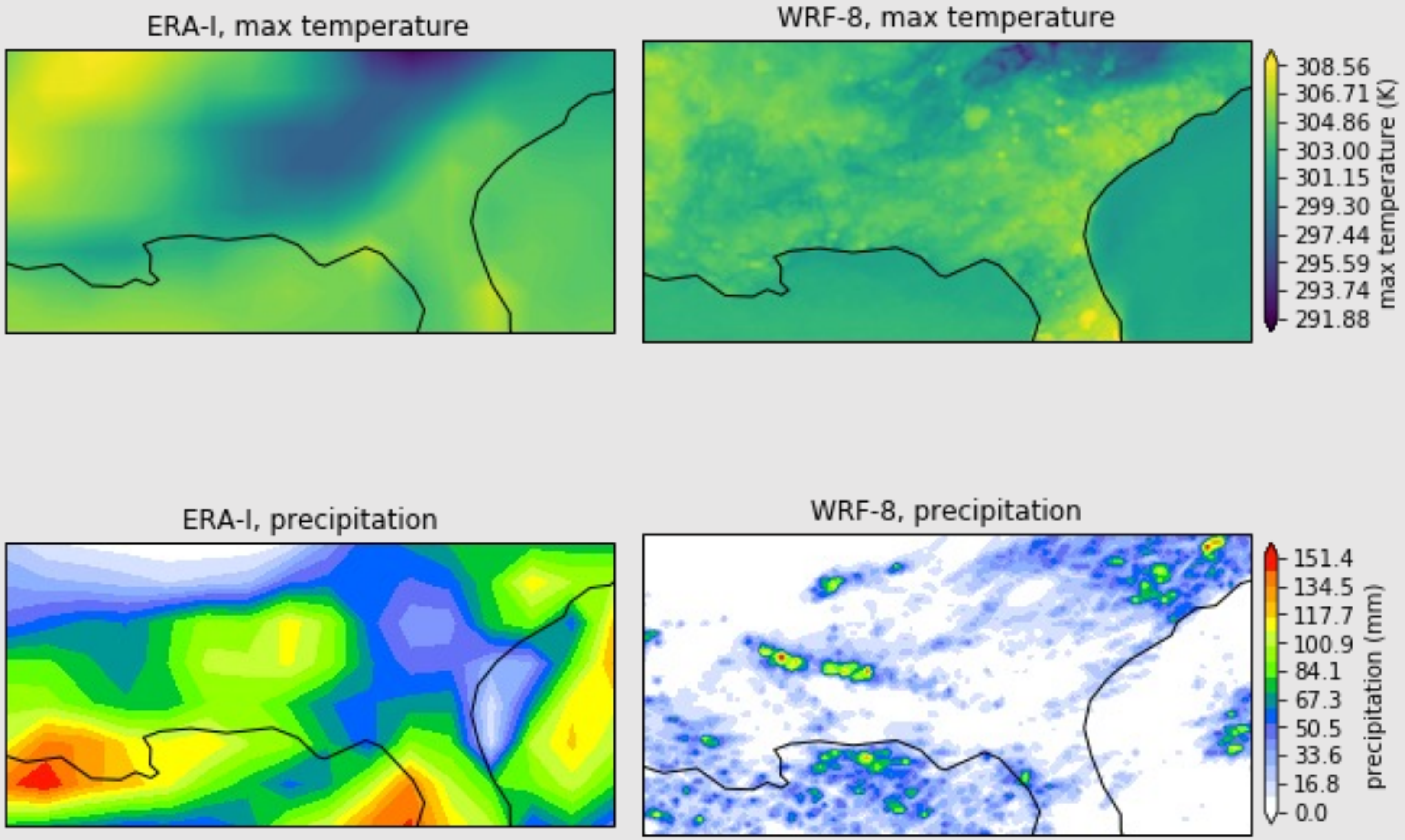
Normalizing flows for downscaling geospatial data

A pretext task for temporal downscaling of geospatial data

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 $y_i \in Y$,

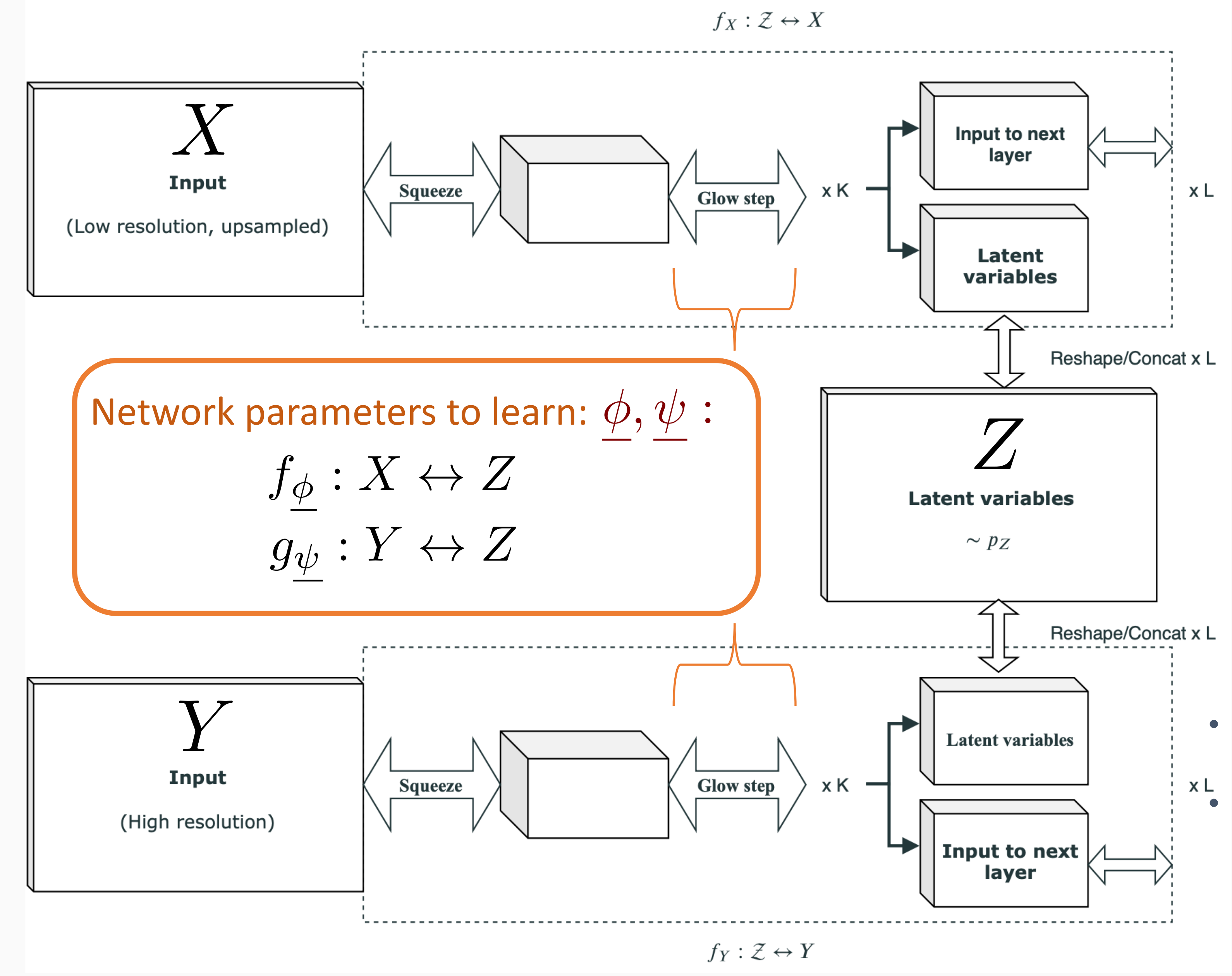
$$f(x_i) \sim P_Y \qquad\qquad f^{-1}(y_i) \sim P_X$$

- 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,

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

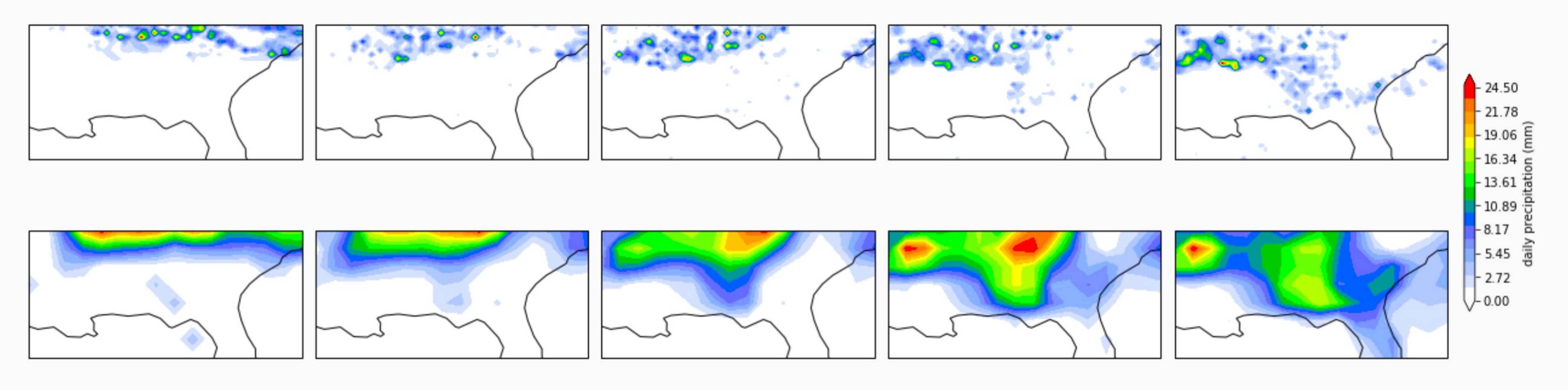
  - Model $P(x|z), P(y|z)$ using AlignFlow [Grover et al. 2020]
    - Starting with a simple prior on P_Z, learn normalizing flows
    - No pairing between x and y examples needed!

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]

Outline

What is self-supervised learning?

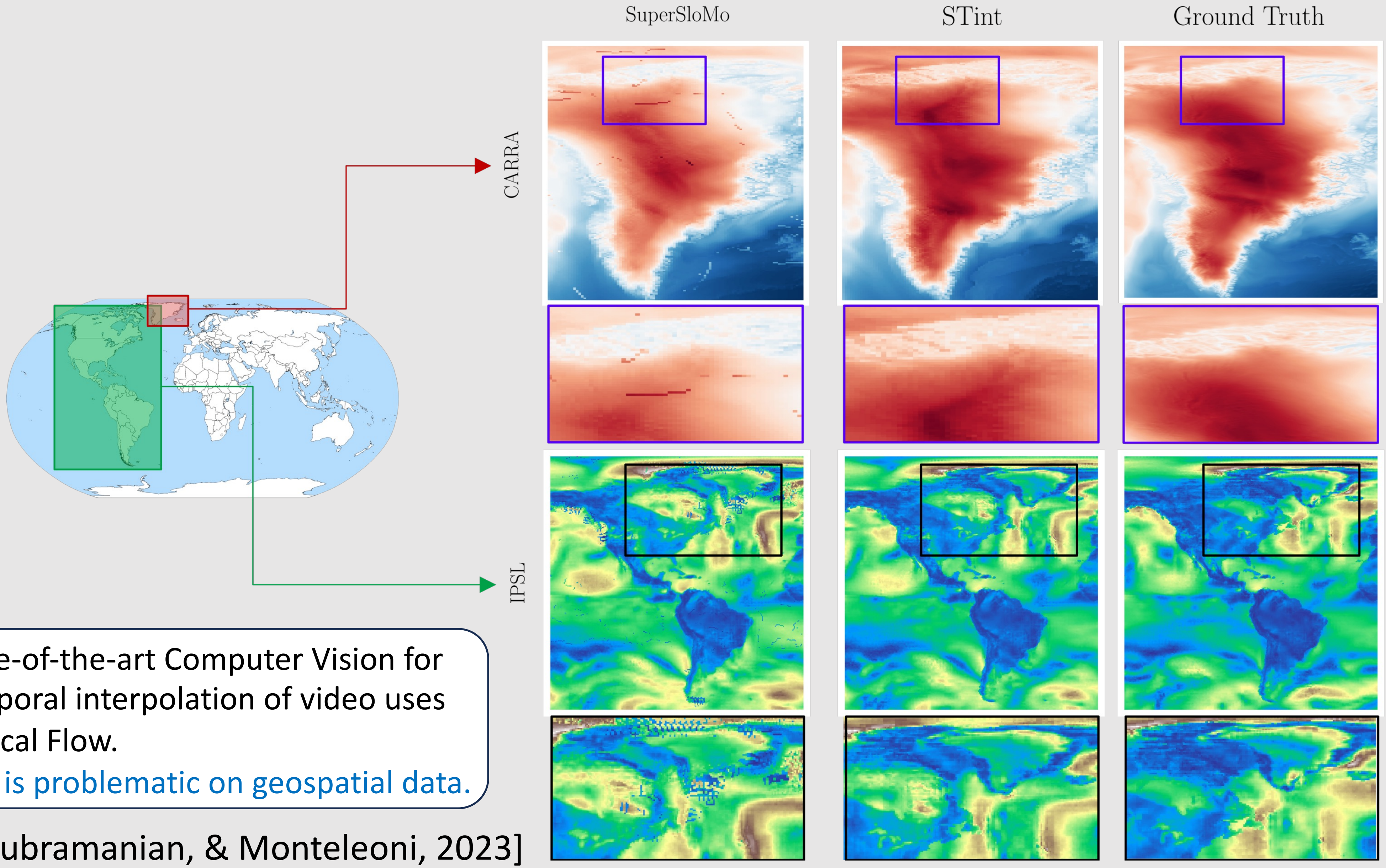
Normalizing flows for downscaling geospatial data

A pretext task for temporal downscaling of geospatial data

The problem

- 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



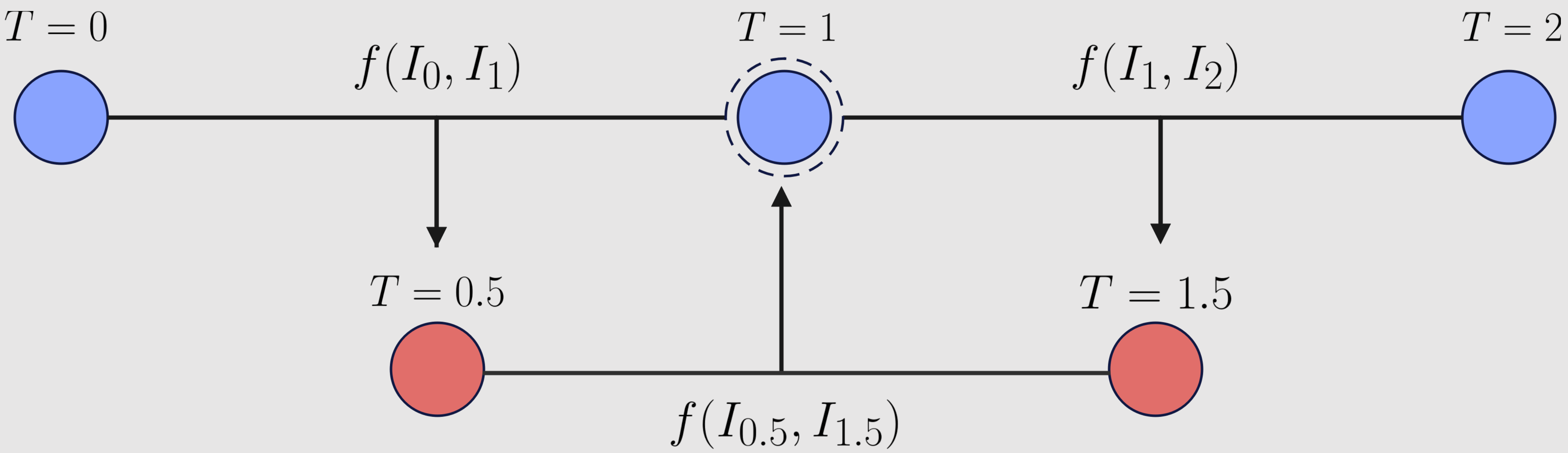
State-of-the-art Computer Vision for temporal interpolation of video uses Optical Flow.
This is problematic on geospatial data.

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

A pretext task for temporal downscaling

STINT: Self-supervised Temporal Interpolation for Geospatial Data

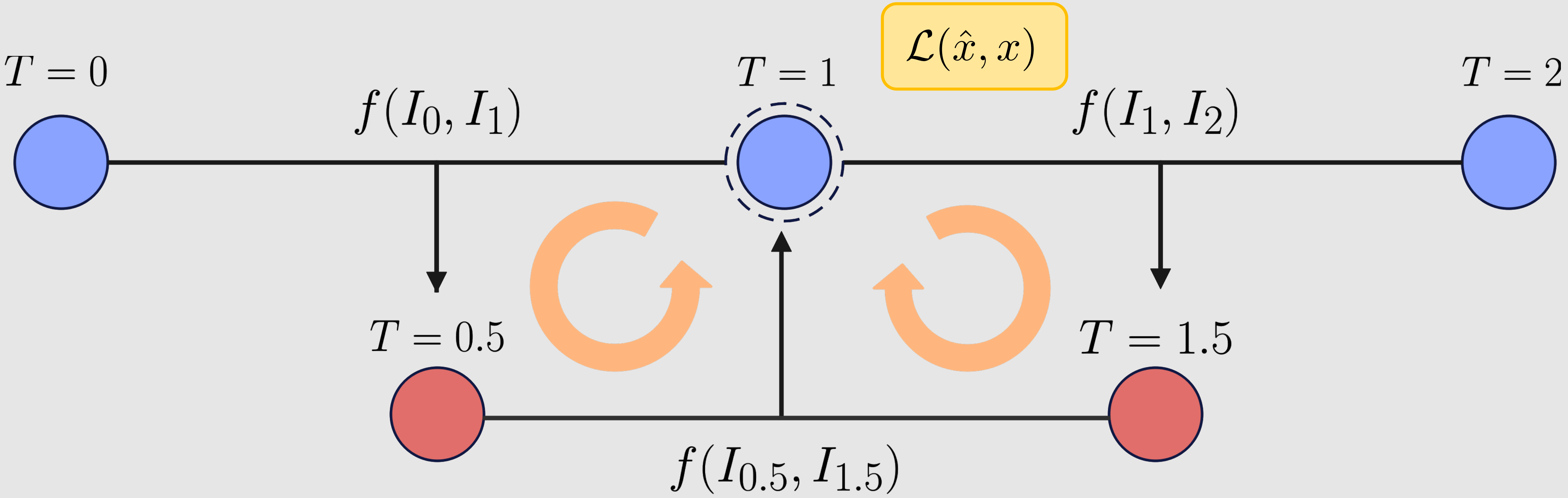
[Harilal, Hodge, Subramanian, & Monteleoni, 2023]



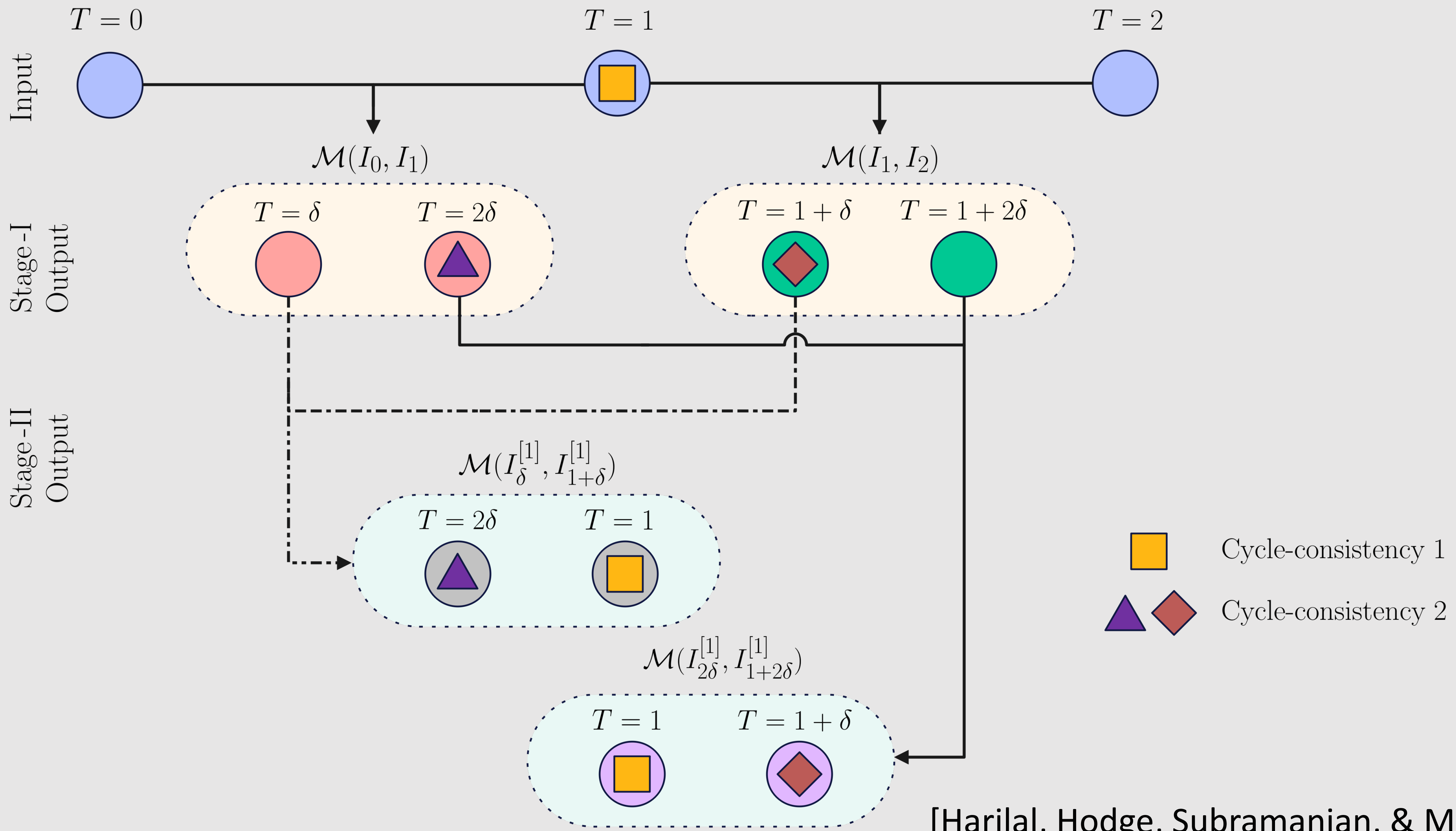
A pretext task for temporal interpolation

STINT: Self-supervised Temporal Interpolation for Geospatial Data

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

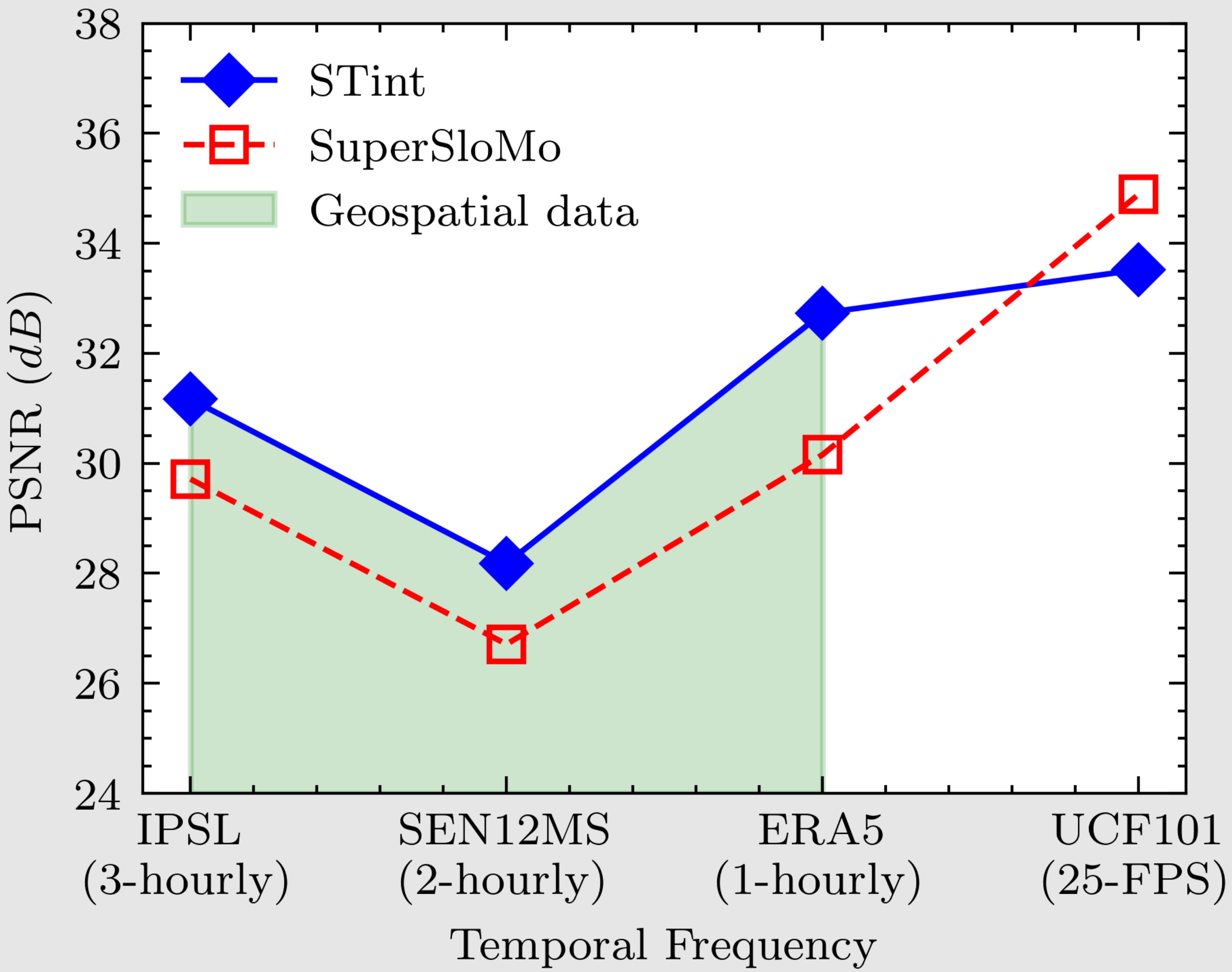


STINT: Self-supervised Temporal Interpolation



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

STINT: Self-supervised Temporal Interpolation

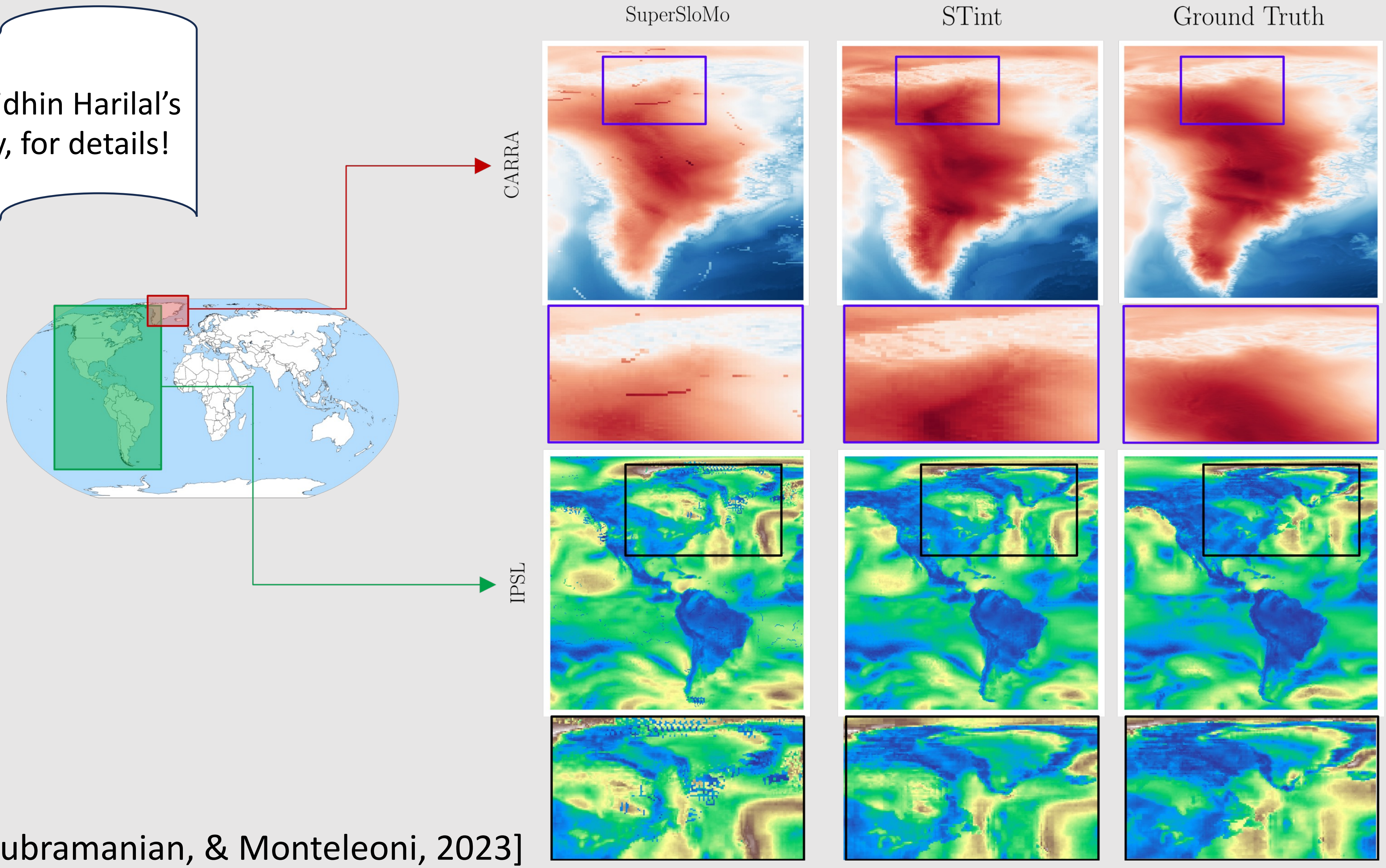


	$\frac{MSE}{Capacity}$ (↓)	PSNR (↑)	SSIM (↑)
ERA5 Solar			
Baseline	0.3086	25.238	0.623
SuperSloMo	0.0907	30.157	0.733
Proposed	**0.0561**	**32.731**	**0.792**
IPSL Wind			
Baseline	0.6206	24.097	0.619
SuperSloMo	0.4150	29.713	0.681
Proposed	**0.2904**	**31.167**	**0.713**
CARRA Temperature			
Baseline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

STINT: Self-supervised Temporal Interpolation

Please see Nidhin Harilal's poster today, for details!

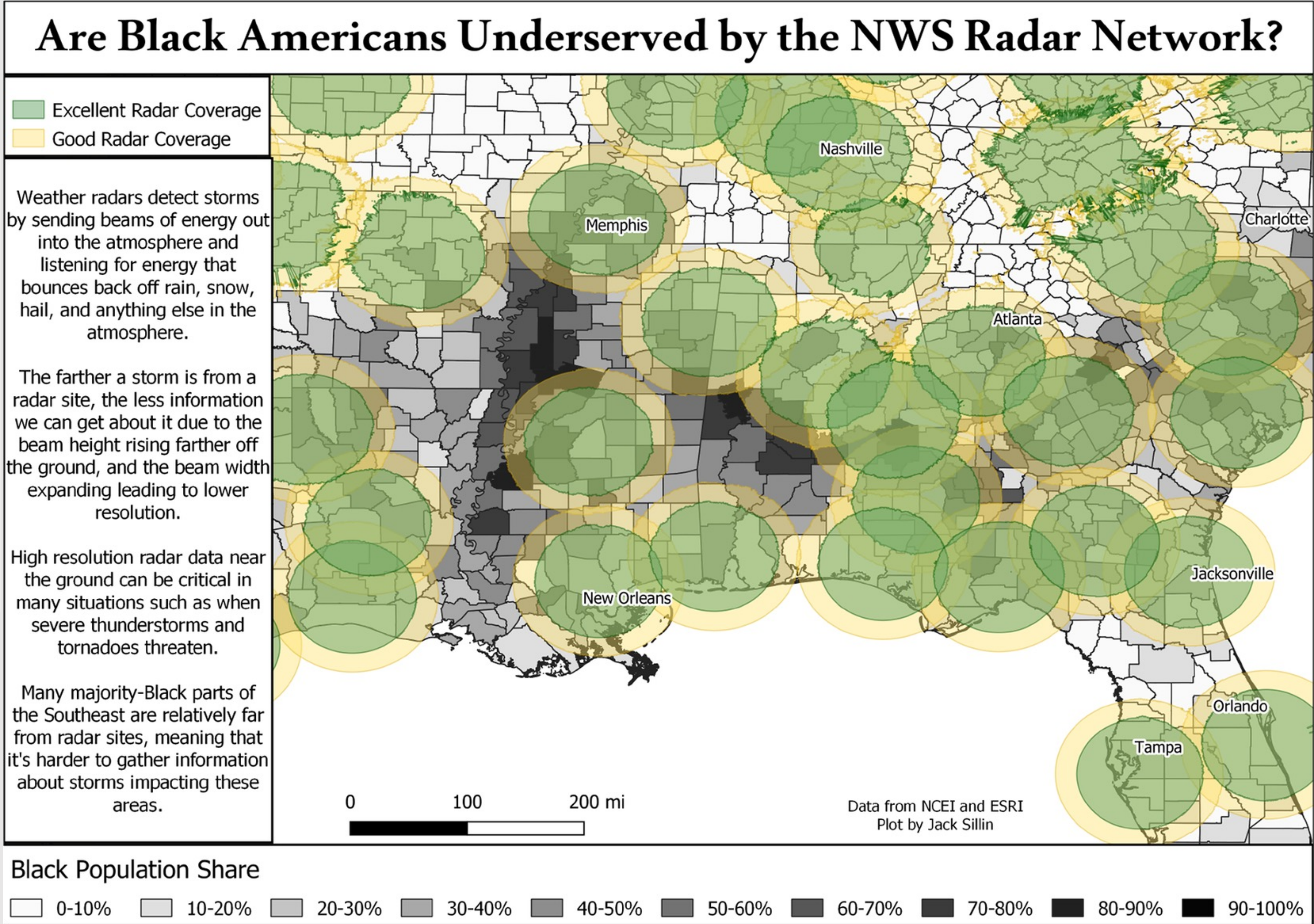


[Harilal, Hodge, Subramanian, & Monteleoni, 2023]

Summary and Outlook

- 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
- A pretext task for temporal downscaling of geospatial data
  - Works best when input 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?
- Implications for data equity in climate and environmental sciences

“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]

Semi/Unsupervised learning: Equity motivation

- 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)

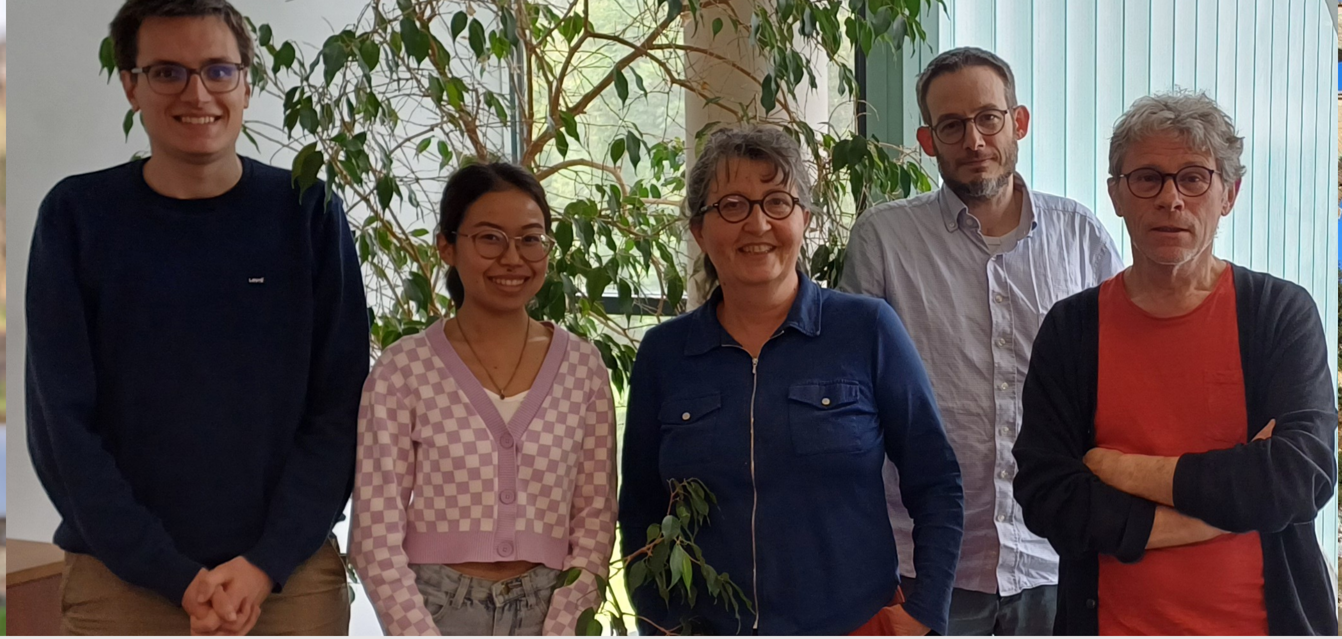
Boulder
NSF
Climate Change AI
futurearth
Research. Innovation. Sustainability.

Thank you!

And many thanks to:

Arindam Banerjee, *University of Illinois Urbana-Champaign*
Nicolò Cesa-Bianchi, *Università degli Studi di Milano*
Tommaso Cesari, *Toulouse School of Economics*
Guillaume Charpiat, *INRIA Saclay*
Cécile Coléou, *Météo-France & CNRS*
Michael Dechartre, *Irstea, Université Grenoble Alpes*
Nicolas Eckert, *Irstea, Université Grenoble Alpes*
Brandon Finley, *University of Lausanne*
Sophie Giffard-Roisin, *IRD Grenoble*
Brian Groenke, *Alfred Wegener Institute, Potsdam*
Nidhin Harilal, *University of Colorado Boulder*
Tommi Jaakkola, *MIT*
Anna Karas, *Météo-France & CNRS*
Fatima Karbou, *Météo-France & CNRS*
Balázs Kégl, *Huawei Research & CNRS*
David Landry, *INRIA Paris*
Luke Madaus, *Jupiter Intelligence*
Scott McQuade, *Amazon*
Ravi S. Nanjundiah, *Indian Institute of Tropical Meteorology*
Moumita Saha, *Philips Research India*
Gavin A. Schmidt, *NASA Senior Advisor on Climate*
Saumya Sinha, *National Renewable Energy Lab*
Cheng Tang, *Amazon*

METEO FRANCE
NASA



Inria

AI Research for Climate Change and Environmental Sustainability (ARCHES)

ENVIRONMENTAL DATA SCIENCE

An interdisciplinary, open access journal dedicated to the potential of artificial intelligence and data science to enhance our understanding of the environment, and to address climate change.

Data and methodological scope: Data Science broadly defined, including:
Machine Learning; Artificial Intelligence; Statistics; Data Mining; Computer Vision; Econometrics

Environmental scope, includes:
Water cycle, atmospheric science (including air quality, climatology, meteorology, atmospheric chemistry & physics, paleoclimatology)
Climate change (including carbon cycle, transportation, energy, and policy)
Sustainability and renewable energy (the interaction between human processes and ecosystems, including resource management, transportation, land use, agriculture and food)
Biosphere (including ecology, hydrology, oceanography, glaciology, soil science)
Societal impacts (including forecasting, mitigation, and adaptation, for environmental extremes and hazards)
Environmental policy and economics

www.cambridge.org/eds

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
2023	12th Conference on Climate Informatics and 9th Hackathon, Cambridge, UK
2024	13th Conference on Climate Informatics, April, Turing Institute, London

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

NSF

CIRES