Machine Learning for Climate Change and Environmental Sustainability

Claire Monteleoni
Choose France Chair in AI
INRIA Paris

University of Colorado Boulder
Inria

December 2021: Boulder County, Colorado

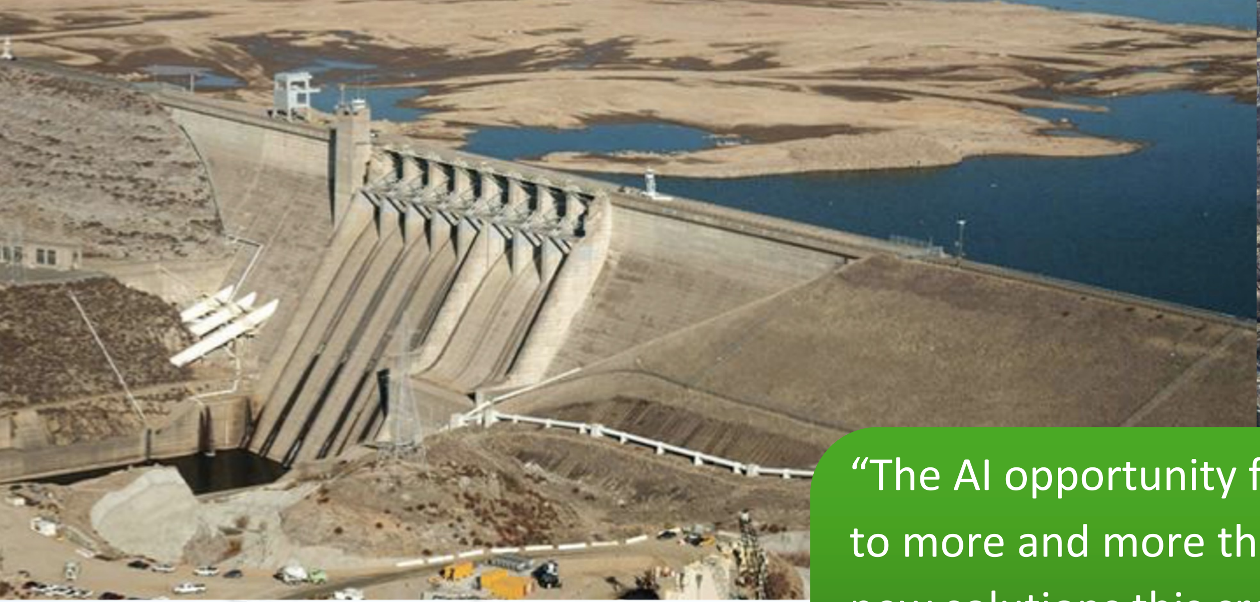
- Snow drought conditions through fall and winter 2021 created dry land-cover
- 80-100 mph winds, combined with ignition, launched an uncontrollable “fire storm”
- Loss of 2 lives. 1000 homes and 20 businesses were destroyed, and more damaged

January 2018: Montecito, Santa Barbara County

- Thomas Fire destroyed 1063 structures and led to poor air quality
- Intense rainfall as the fire was nearing containment produced a debris flow
- 23 lives and over 130 homes were lost
- Damage to critical transportation and water resource infrastructure




Machine learning can shed light on climate change



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

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

Climate Informatics is based on the vision that Machine learning can shed light on climate change

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

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

2011 Launch 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 Launch Climate Informatics Hackathon, Paris and Boulder

2018 World Economic Forum recognizes Climate Informatics as key priority

2019 Climate Informatics Conference held at ENS, Paris

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

2022 11th Conference on Climate Informatics and 8th Hackathon, NOAA, Asheville, NC

2023 12th Conference on Climate Informatics and 9th Hackathon, April 19-21, Cambridge, UK

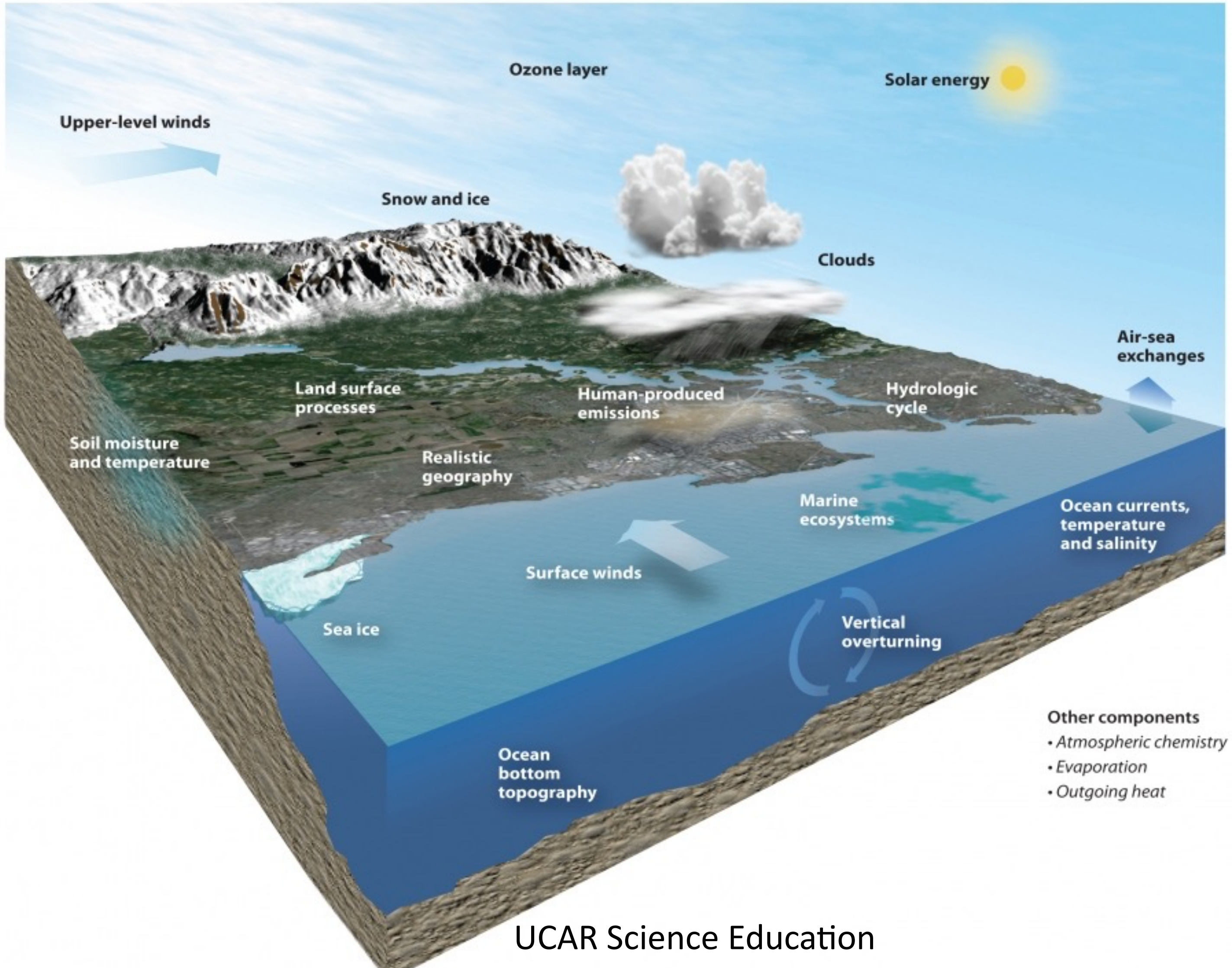
Machine Learning for Climate Change and Environmental Sustainability

- Machine Learning for Understanding and Predicting Climate Change
- Machine Learning for Extreme Weather and Cascading Hazards
- Machine Learning for the Green Transition

Our Climate Informatics research also addresses open problems in Machine Learning

- ❑ Online learning with spatiotemporal non-stationarity
- ❑ Prediction at multiple timescales simultaneously
- ❑ Anomaly detection with limited supervision
- ❑ Tracking highly-deformable patterns

Machine Learning for Understanding and Predicting Climate Change



UCAR Science Education

Online learning from non-stationary spatiotemporal data to adaptively combine climate model ensemble forecasts

[Multiple papers 2009-2020, e.g., AAAI 2012, ALT 2020]

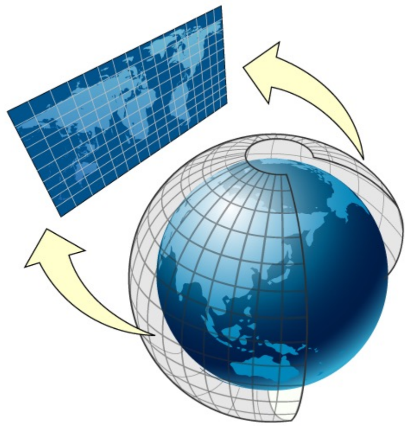
Causal information hubs in Pacific ENSO region

[Saha et al. Climate Informatics 2019]

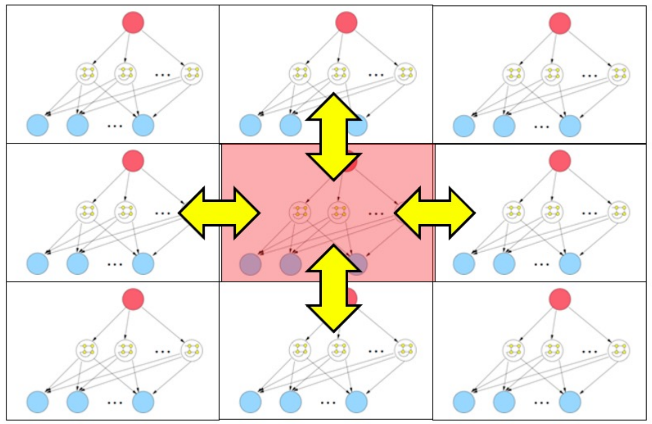
NASA project to attribute and forecast sea-level rise using climate models and satellite altimetry

[Sinha et al., AGU 2022] with NCAR

Online learning with spatiotemporal non-stationarity



Learning when the target concept can vary over time, and multiple other dimensions (e.g., latitude, longitude)



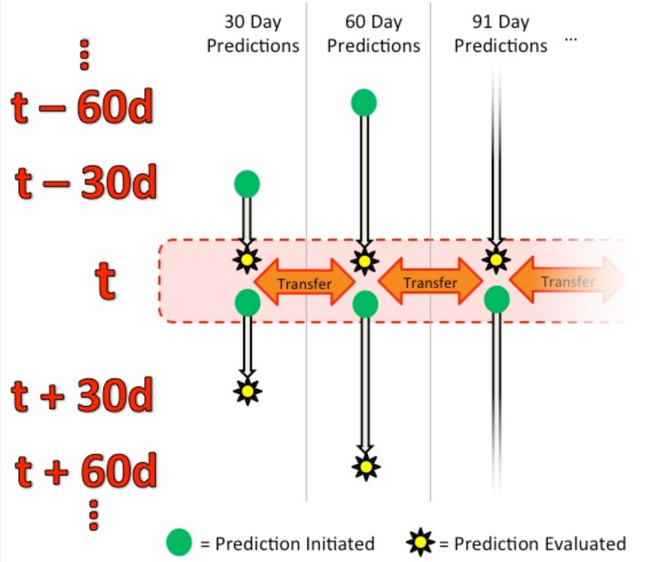
We can exploit local structure in space and time

We can learn the level of non-stationarity in time and space

[McQuade and Monteleoni, AAAI 2012] extended [Monteleoni & Jaakkola, NeurIPS 2003; Monteleoni et al. SAM 2011] to multiple dimensions

This framework for online learning was open in machine learning

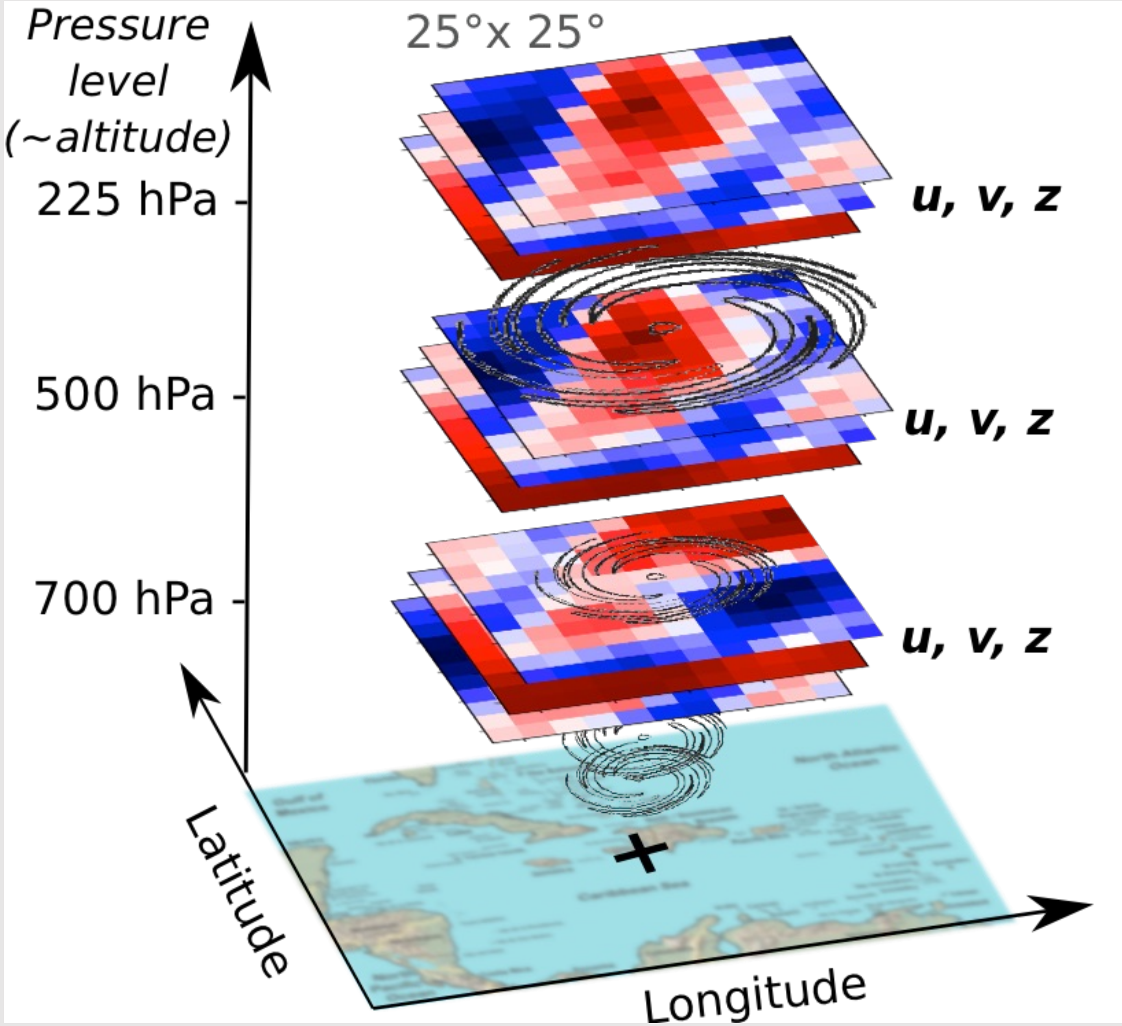
New “regret” framework: [Cesa-Bianchi, Cesari, & Monteleoni, ALT 2020]



Prediction at multiple timescales simultaneously

Applications to both climate science, and financial volatility: [McQuade and Monteleoni, CI 2015; SIGMOD DSMM 2016]

Machine Learning for Extreme Weather and Cascading Hazards



[Giffard-Roisin et al., Frontiers 2020]

Defining and detecting diverse, multivariate extreme events with topic modeling

[Tang & Monteleoni, Climate Informatics 2014; IEEE CISE 2015]

Hurricane track prediction via fused CNNs

[Giffard-Roisin et al., Climate Informatics 2018; Frontiers 2020]

Forecasting Indian Summer Monsoon precipitation extremes

[Saha et al. Climate Informatics 2019; 2020] with India Meteorological Department (IMD)

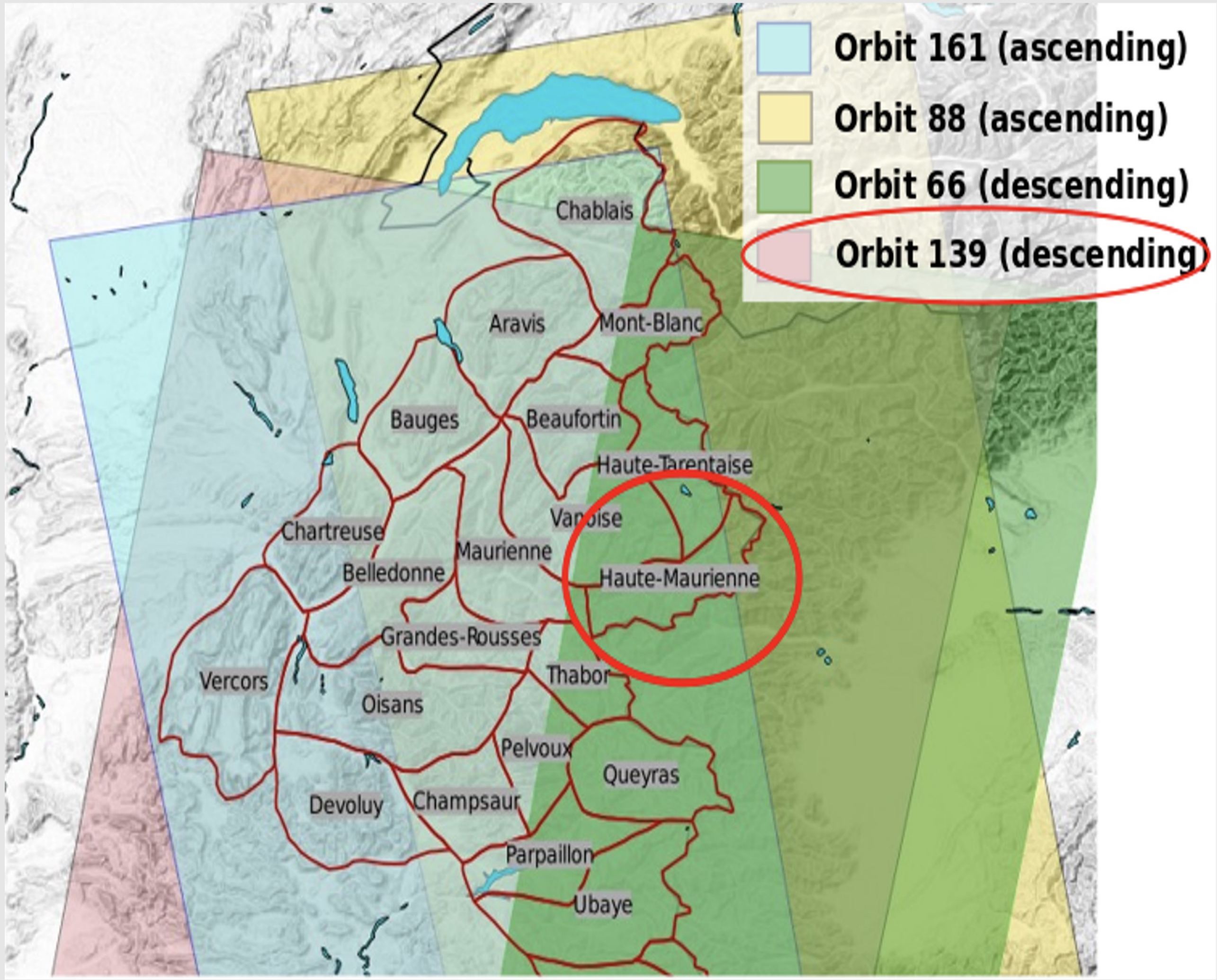
Avalanche detection using CNN; VAE

[Sinha et al., Climate Informatics 2019; 2020] with Météo-France

Avalanche detection

- **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



[Sinha et al., Climate Informatics 2020]

METEO FRANCE

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

Approach

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

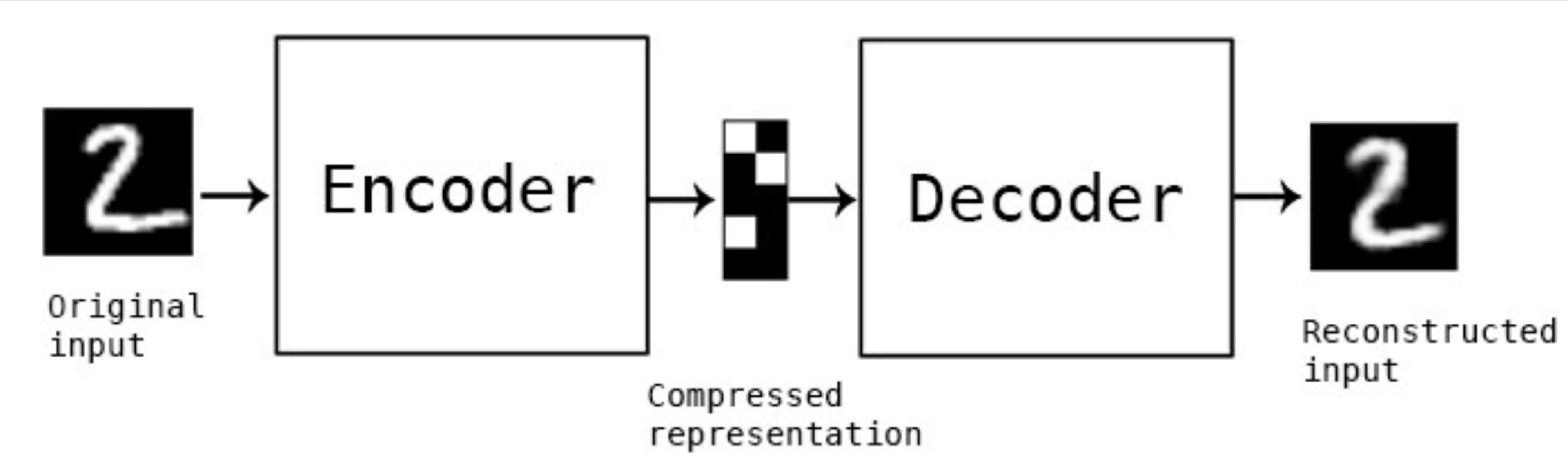
② Train a variational autoencoder (VAE) on the negative examples

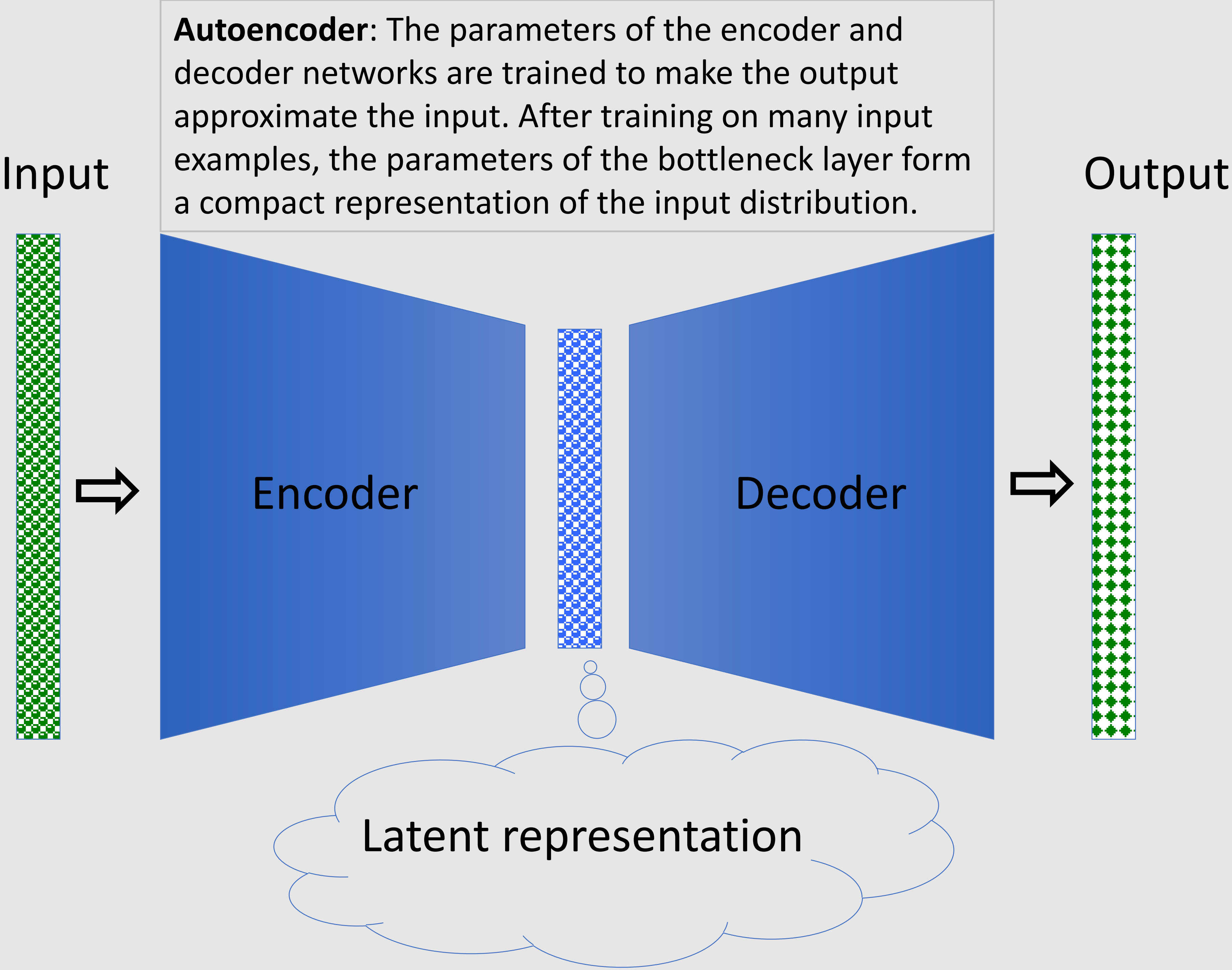
③ Threshold the VAE's reconstruction error to classify a new image

- Our idea: when labeled data is scarce, the VAE can instead be trained **without** supervision!

What is an Auto-encoder?

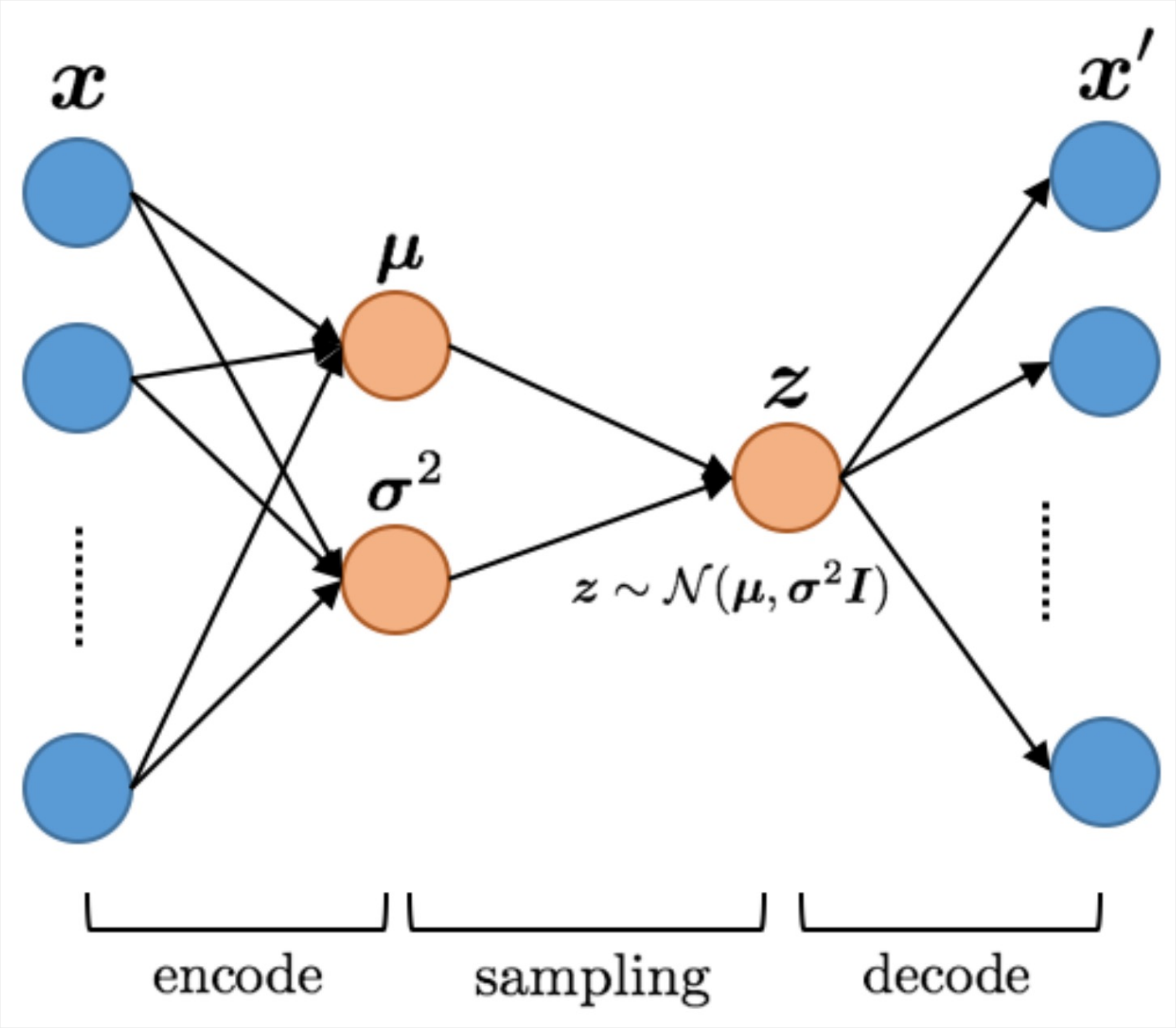
- Train a neural network in an unsupervised way
  - Use the unlabeled data both as input, and to evaluate the output
- After training, the bottleneck layer will be a compact representation of the input distribution



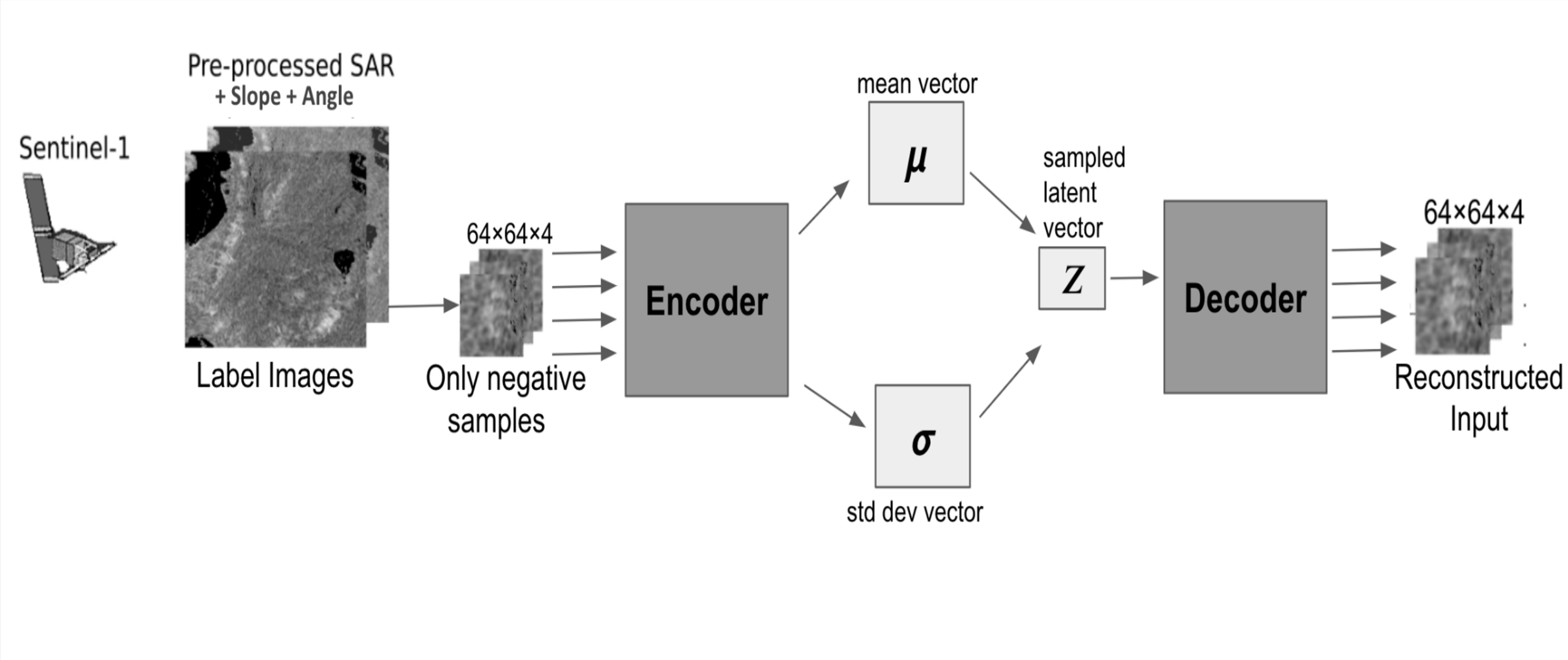

Autoencoder: The parameters of the encoder and decoder networks are trained to make the output approximate the input. After training on many input examples, the parameters of the bottleneck layer form a compact representation of the input distribution.
Input
Output
Encoder
Decoder
Latent representation

Variational Autoencoder (VAE)

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

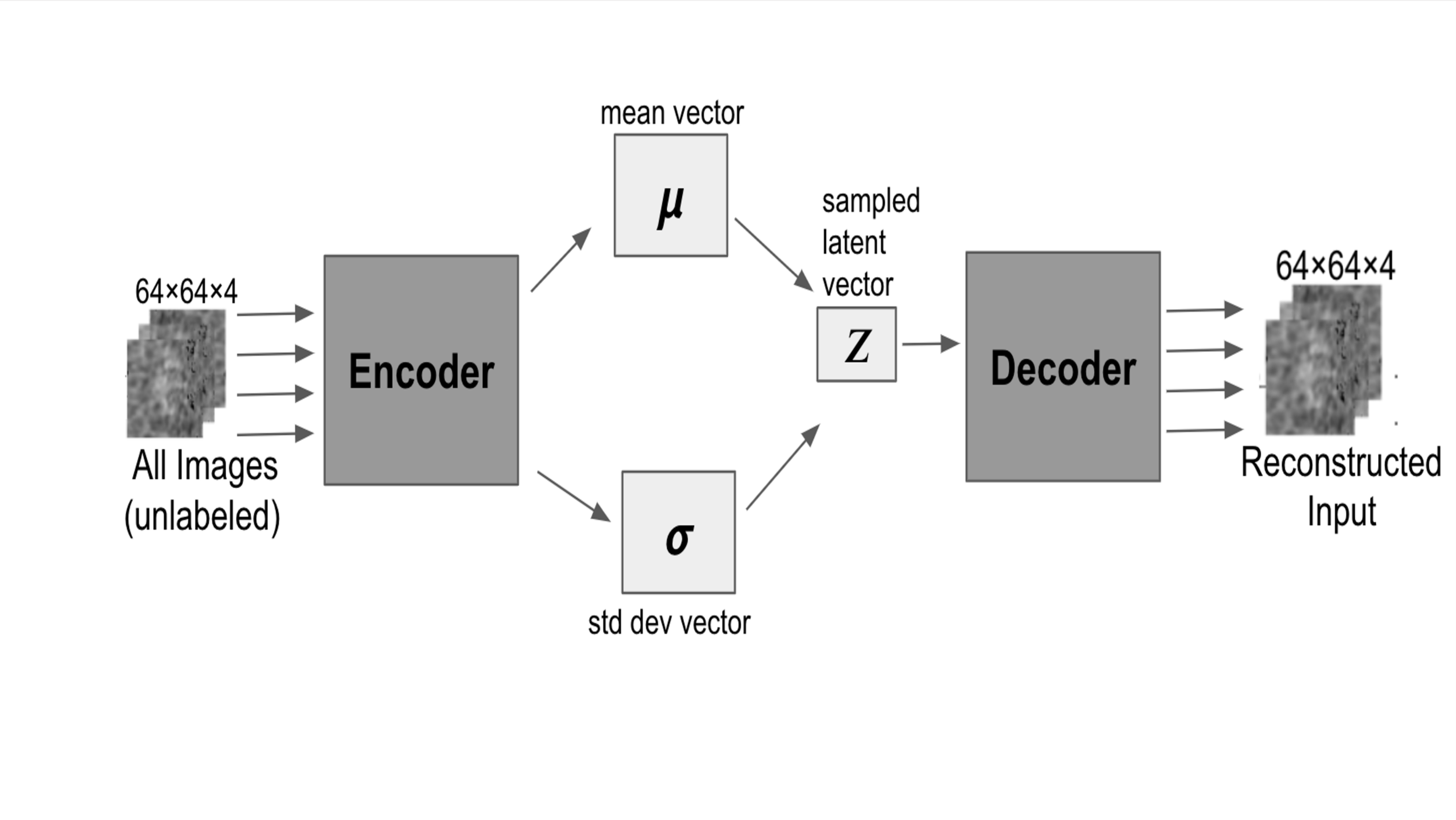


VAE for anomaly detection is typically trained on **negative** examples only



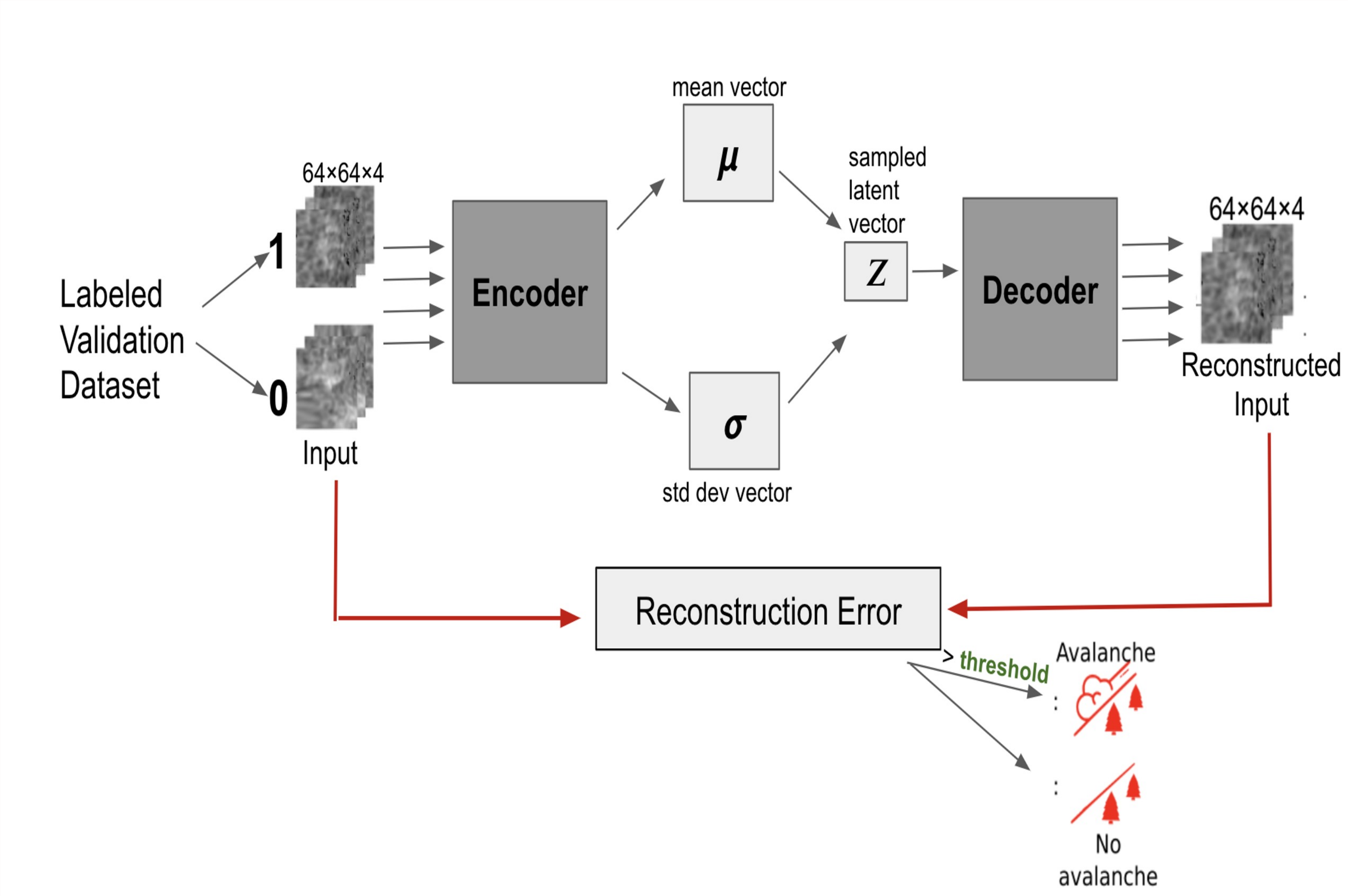
[Sinha et al., Climate Informatics 2020]

Our approach: Train a VAE on **unlabeled** examples



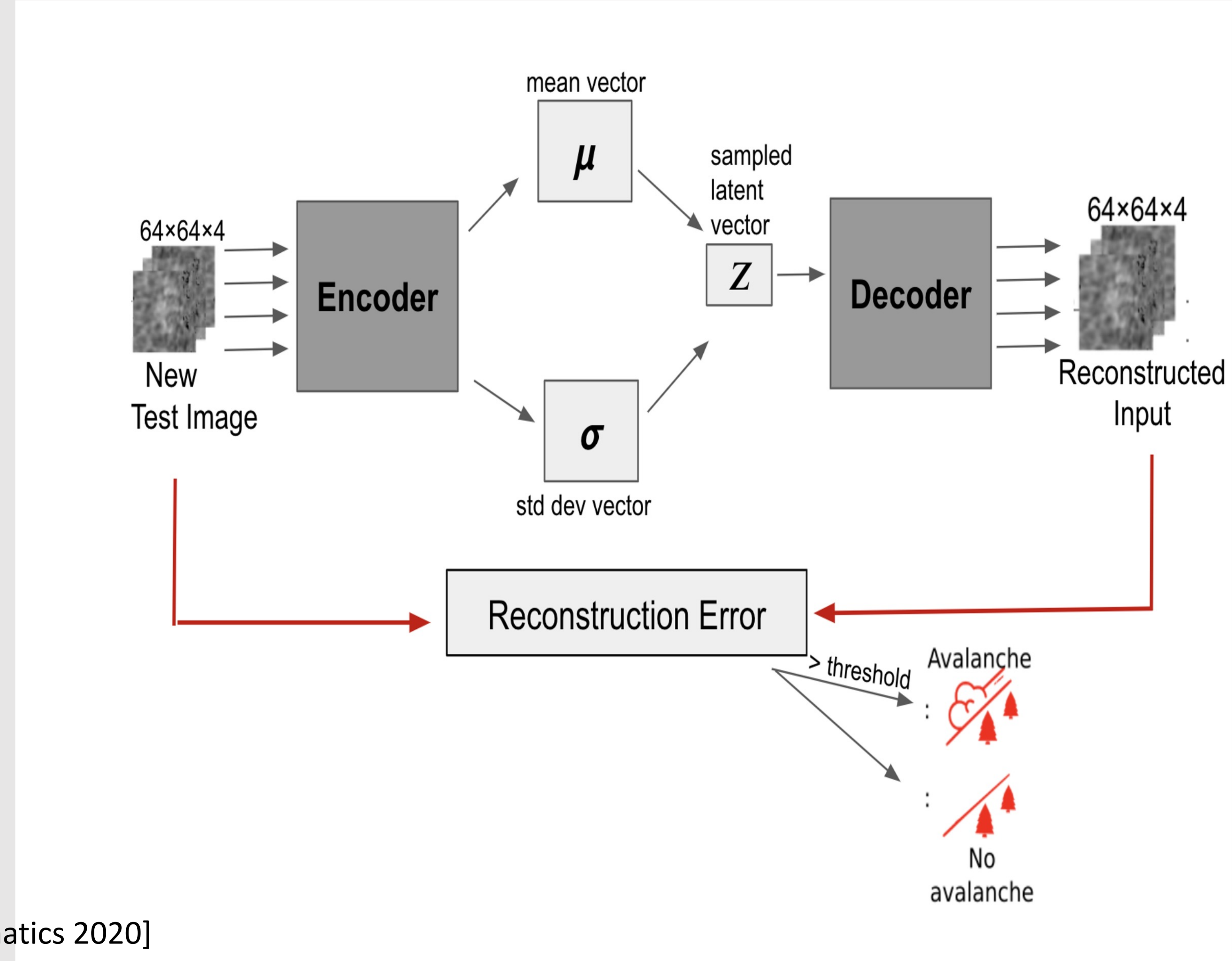
[Sinha et al., Climate Informatics 2020]

Tuning the hyperparameter for avalanche detection



[Sinha et al., Climate Informatics 2020]

Avalanche detection on a test image



[Sinha et al., Climate Informatics 2020]

Evaluation

One of the most avalanche-prone mountain chains in the Alps data set

	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]

[Sinha et al., Climate Informatics 2020]

ML contribution

- Provided a semi-supervised approach to detecting rare events when labeled data is limited
  - Key idea: lean heavily on unsupervised learning and use labeled data ONLY for hyperparameter tuning

- Can be viewed as a form of virtual sensor

[Sinha et al., Climate Informatics 2020]

ML for the Green Transition

Week-ahead solar irradiance forecasting via deep sequence learning

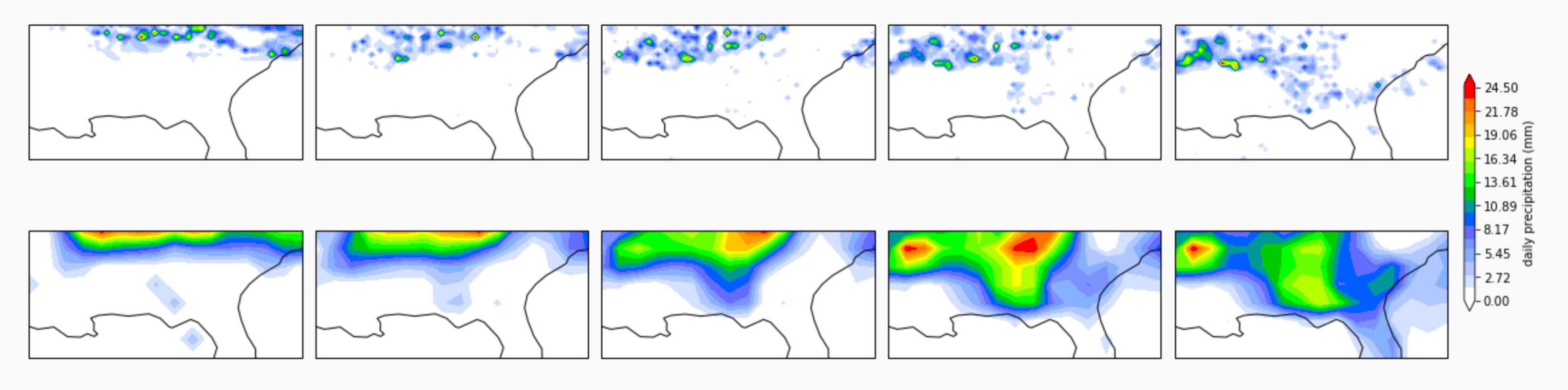
[Sinha et al., CI 2022] with NREL

ML to downscale climate model data for renewable energy planning in U.S. and India

Climate Change AI / Future Earth project with NREL, IIT-Roorkee

[Harilal et al., NeurIPS workshop 2022]

ClimAlign: Unsupervised, generative downscaling

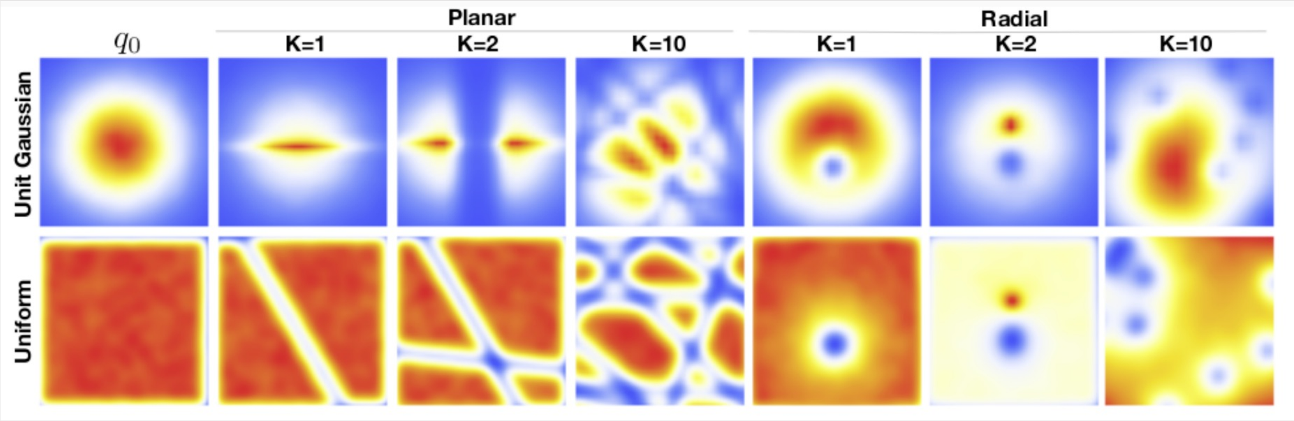


General downscaling technique via domain alignment with normalizing flows
[AlignFlow: Grover et al., AAAI 2020][Glow: Kingma & Dhariwal, NeurIPS 2018]

- **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, et al., Climate Informatics 2020]

Normalizing Flows

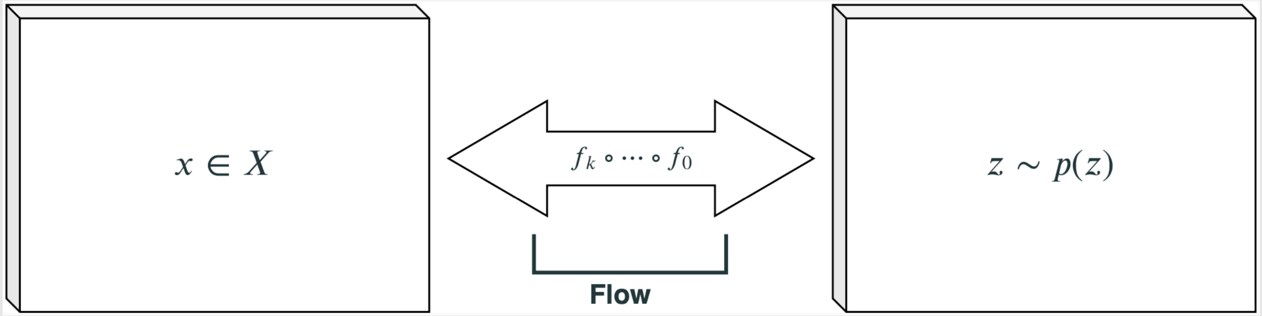


[Rezende & Mohamed, ICML 2015]

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:

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



Our Climate Informatics research also addresses open problems in Machine Learning

- ❑ Online learning with spatiotemporal non-stationarity
- ❑ Prediction at multiple timescales simultaneously
- ❑ Anomaly detection with limited supervision
- ❑ Tracking highly-deformable patterns

Summary and Outlook

Data limitations

- Limited labeled data: unsupervised learning, dimensionality reduction
- Class imbalance: e.g., extreme events are rare by definition!
- Data is limited along the time dimension. Can we substitute data diversity and granularity over space?

Scale resolution challenges

- Downscaling spatiotemporal data fields
- Climate model parameterization problems

Non-stationarity

- Climate *change* means we cannot assume i.i.d. data!
- ML models need to adapt over time, and space

Interpretability

- Evaluation of generative models is an active research area of core ML

Long-term Inspirations

Cascading Hazards

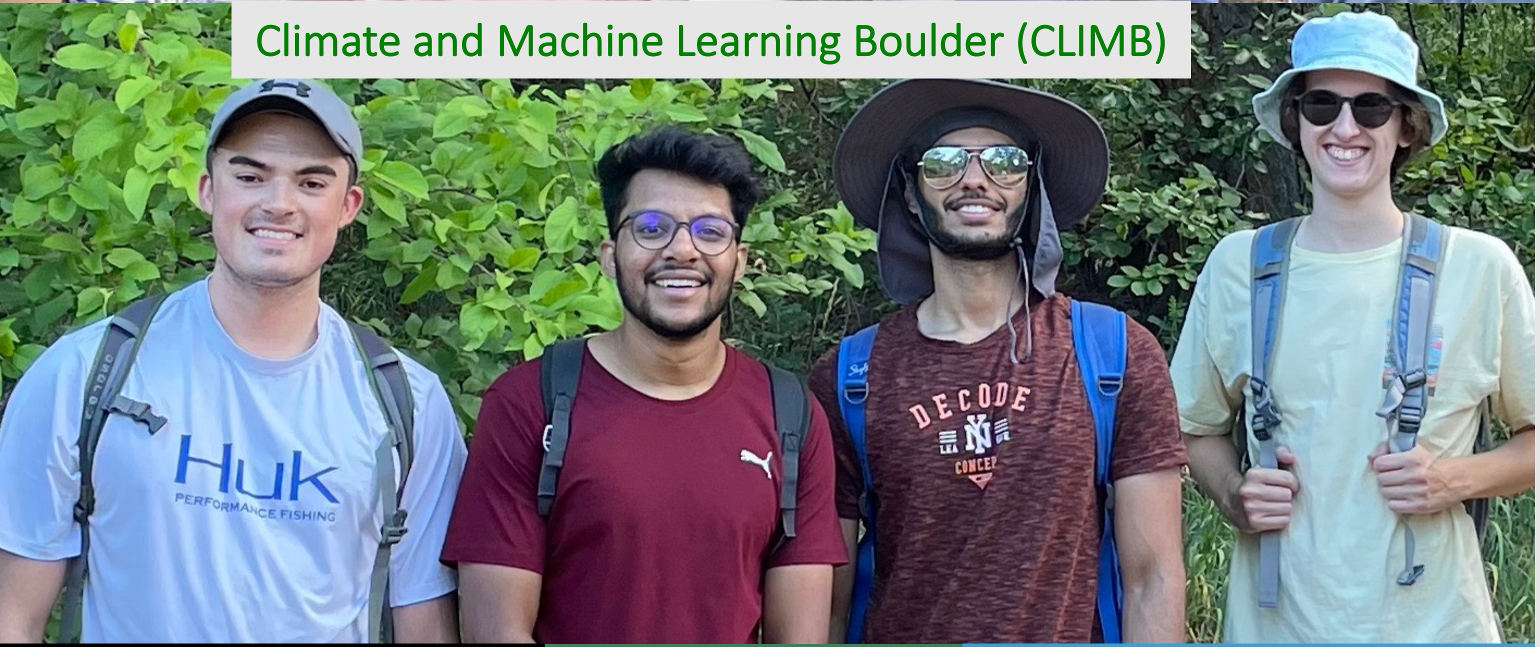
- Goal: move beyond individual weather extremes, to how they couple
- With massive wildfires in France and the U.S., there is extreme urgency!

Climate Justice

- Our research should always help increase climate equity
- Ultimately, we should strive for approaches to help UNDO the legacy of climate IN-justice

Thank you!

Climate and Machine Learning Boulder (CLIMB)



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*
Sophie Giffard-Roisin, *IRD Grenoble*
Brian Groenke, *Alfred Wegener Institute, Potsdam*
Tommi Jaakkola, *MIT*
Anna Karas, *Météo-France & CNRS*
Fatima Karbou, *Météo-France & CNRS*
Balázs Kégl, *Huawei Research & CNRS*
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, *University of Colorado Boulder*
Cheng Tang, *Amazon*

Boulder
Climate Change AI
futurearth
Research. Innovation. Sustainability.
METEO FRANCE
NASA

Inria

PhD, Postdoc and Researcher positions available at INRIA Paris!

ARCHES: AI research for climate change & environmental sustainability

Choose France™

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

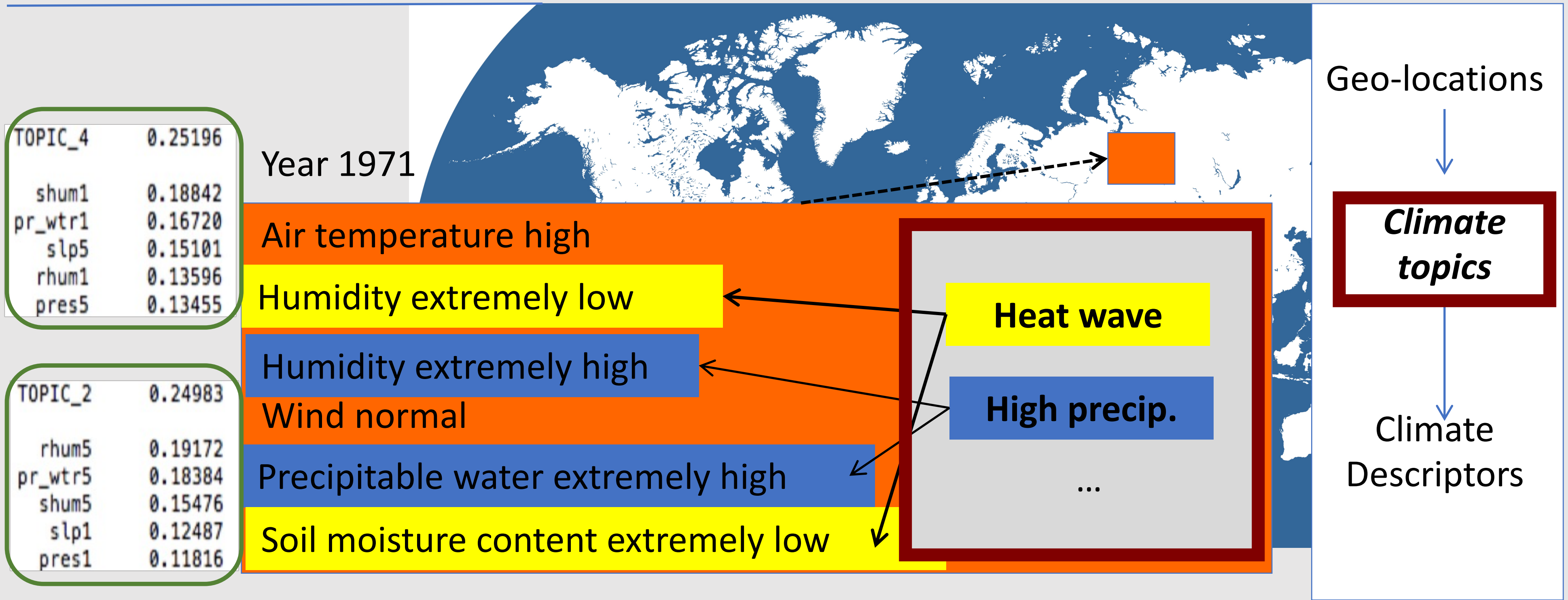
Bonus slides

Unsupervised learning to define/detect multivariate extremes

[Tang & M, Climate Informatics 2014; IEEE CISE 2015]

Extend probabilistic topic modeling (Latent Dirichlet Allocation [Blei et al., 2001]) from NLP

- Multiple variables (complex, multivariate extreme events)
- Ability to detect multiple types of events
- Multiple degrees of severity
- Uses all data, not just extreme values



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

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

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 $y_i \in Y$,

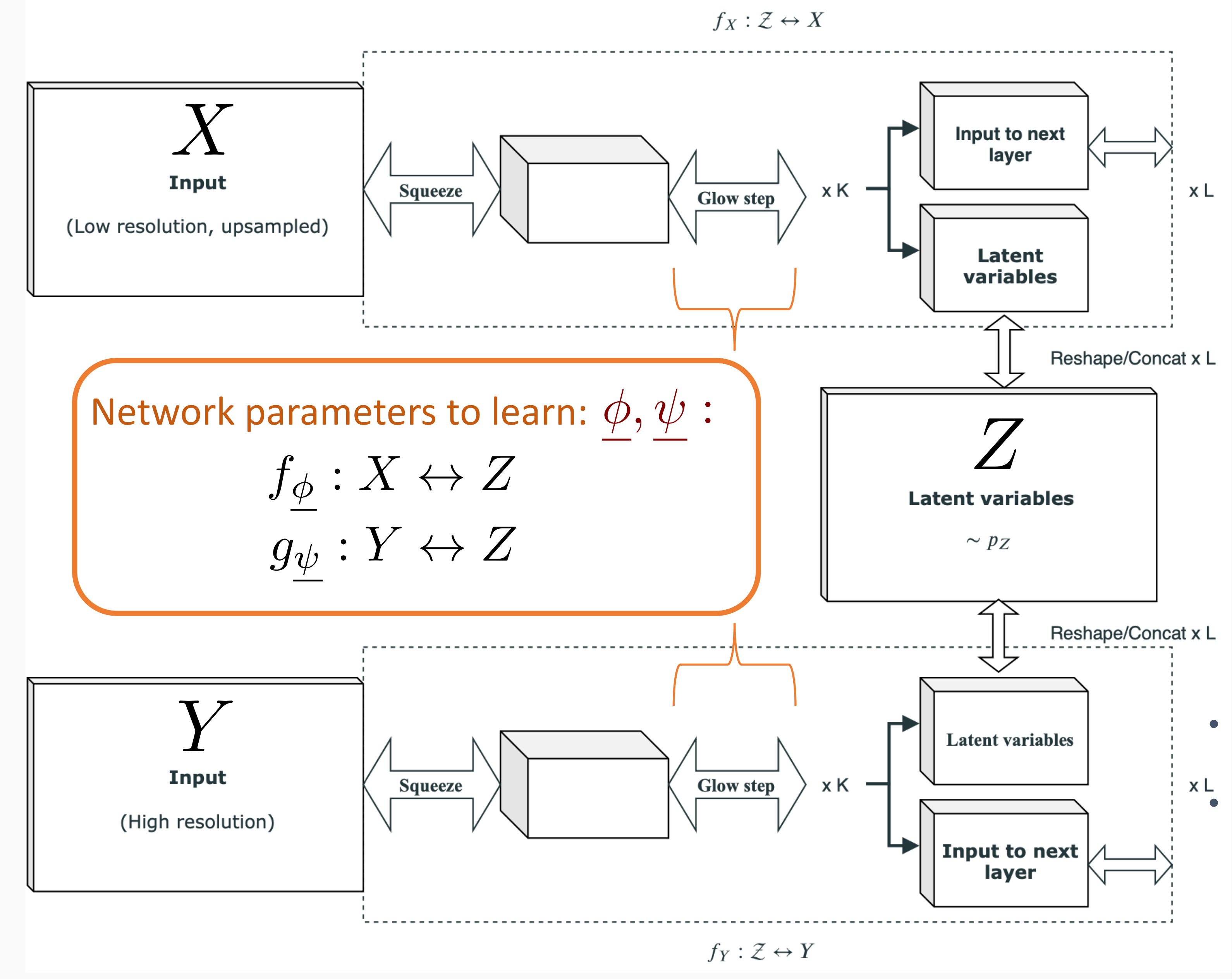
$$f(x_i) \sim P_Y \qquad f^{-1}(y_i) \sim P_X$$

- Downscaling as domain alignment
  - 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

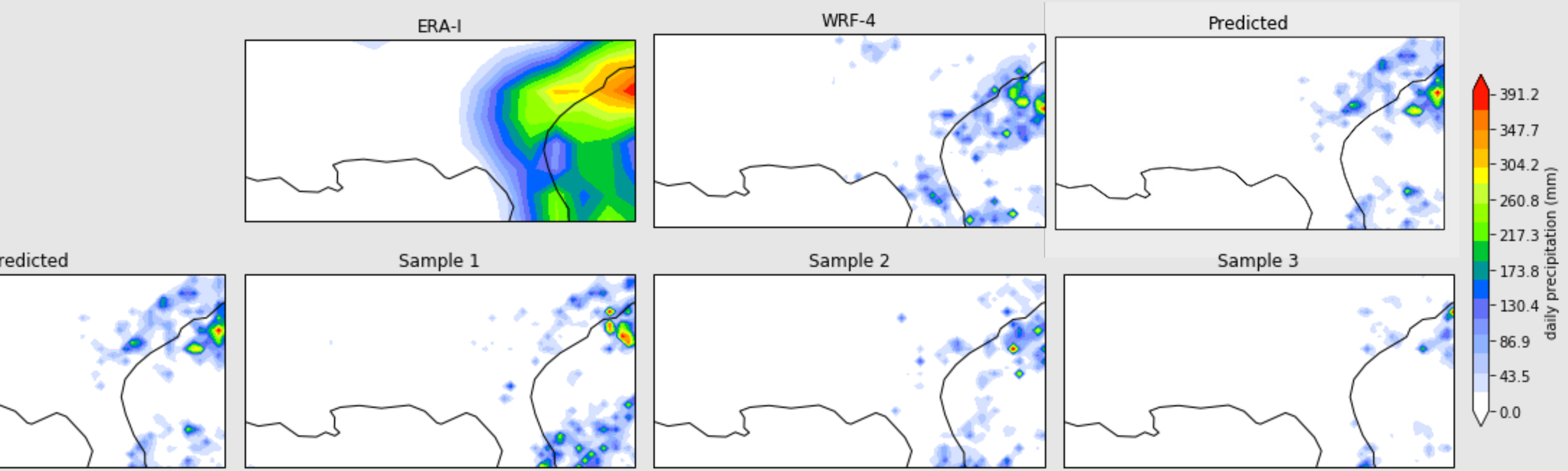
Daily Max Temperature

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

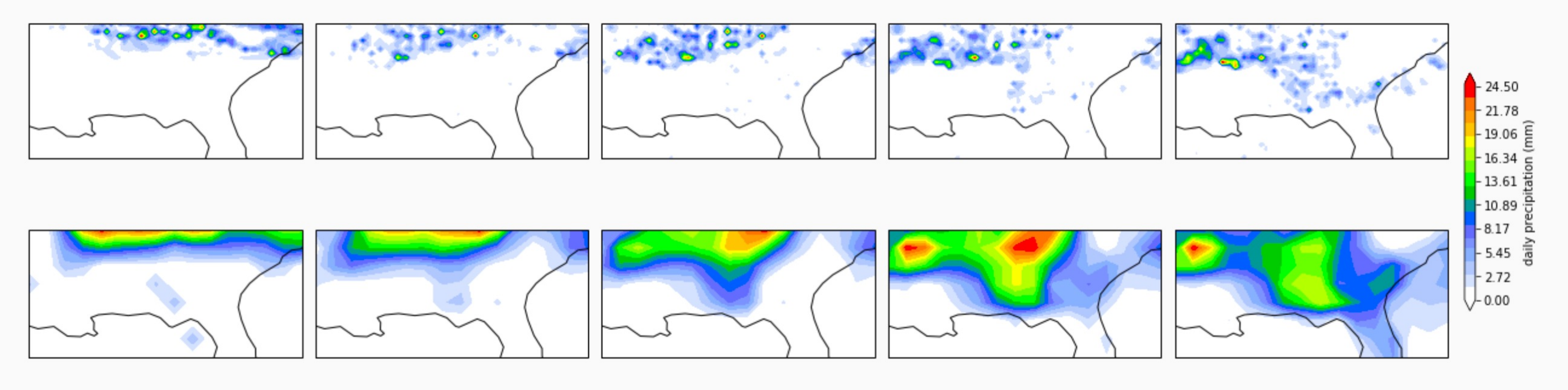
Daily Precipitation

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

Point prediction example



ClimAlign: Unsupervised, generative downscaling

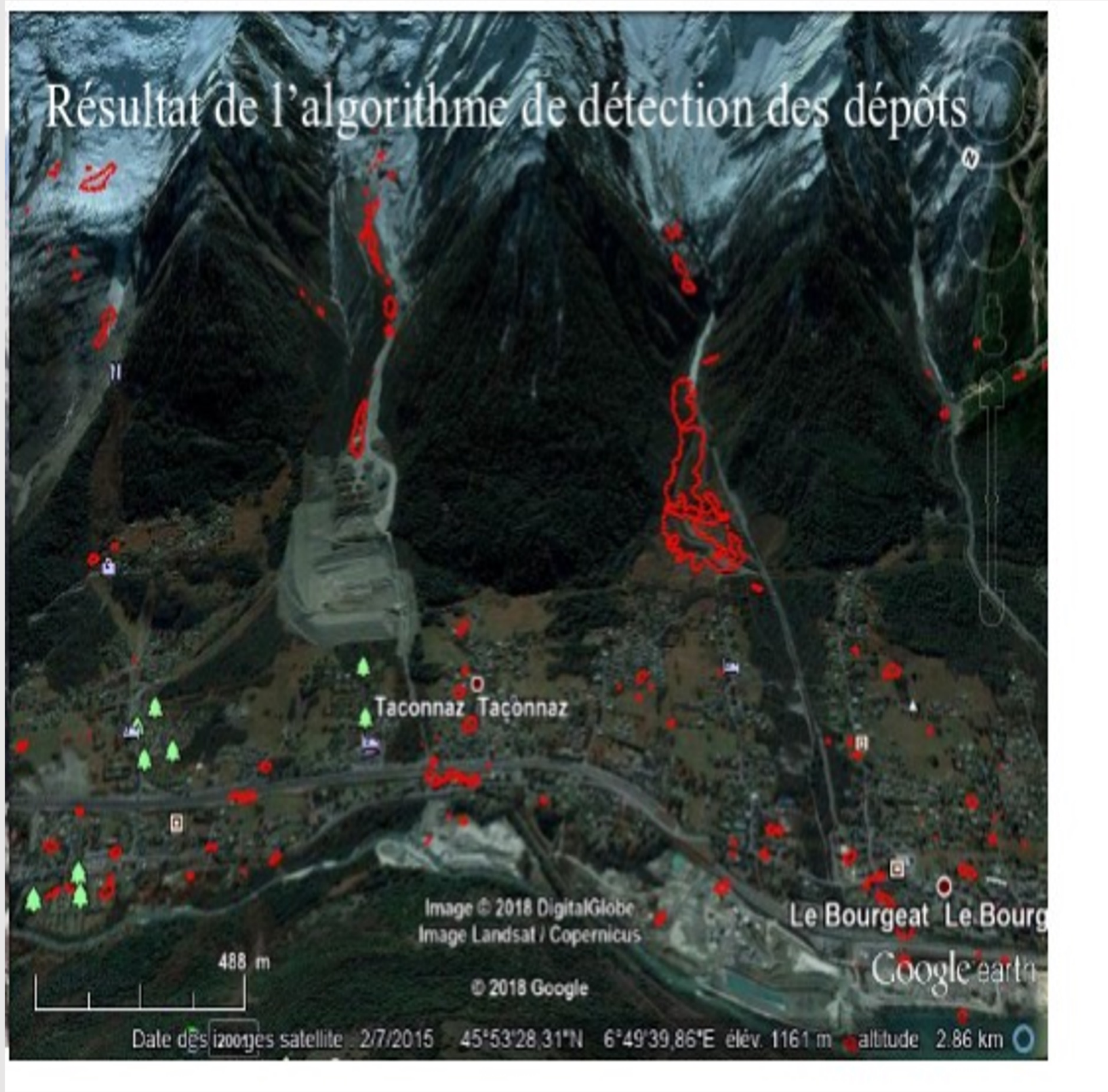


[Groenke et al., CI 2020]

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

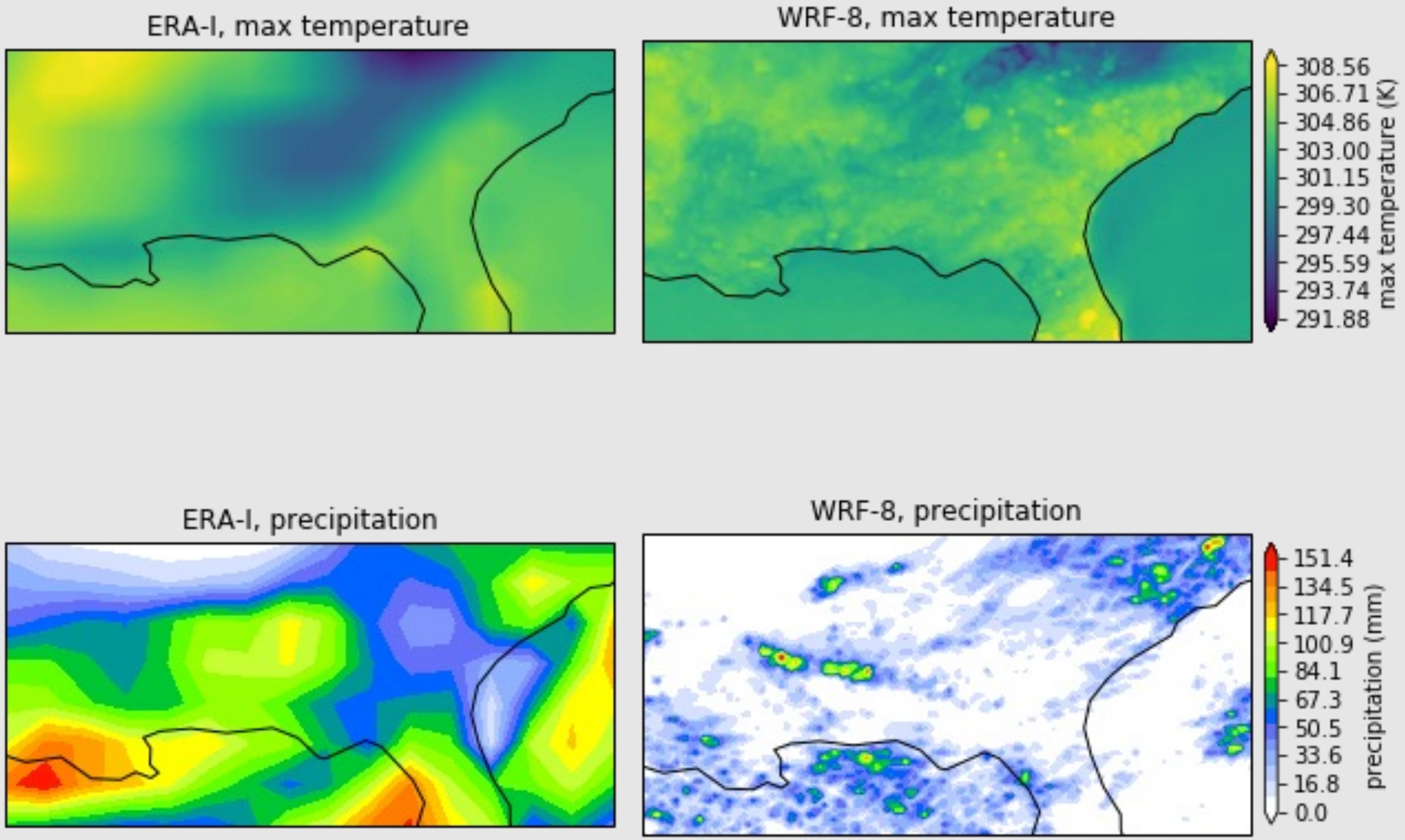
Baseline method: Thresholding



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

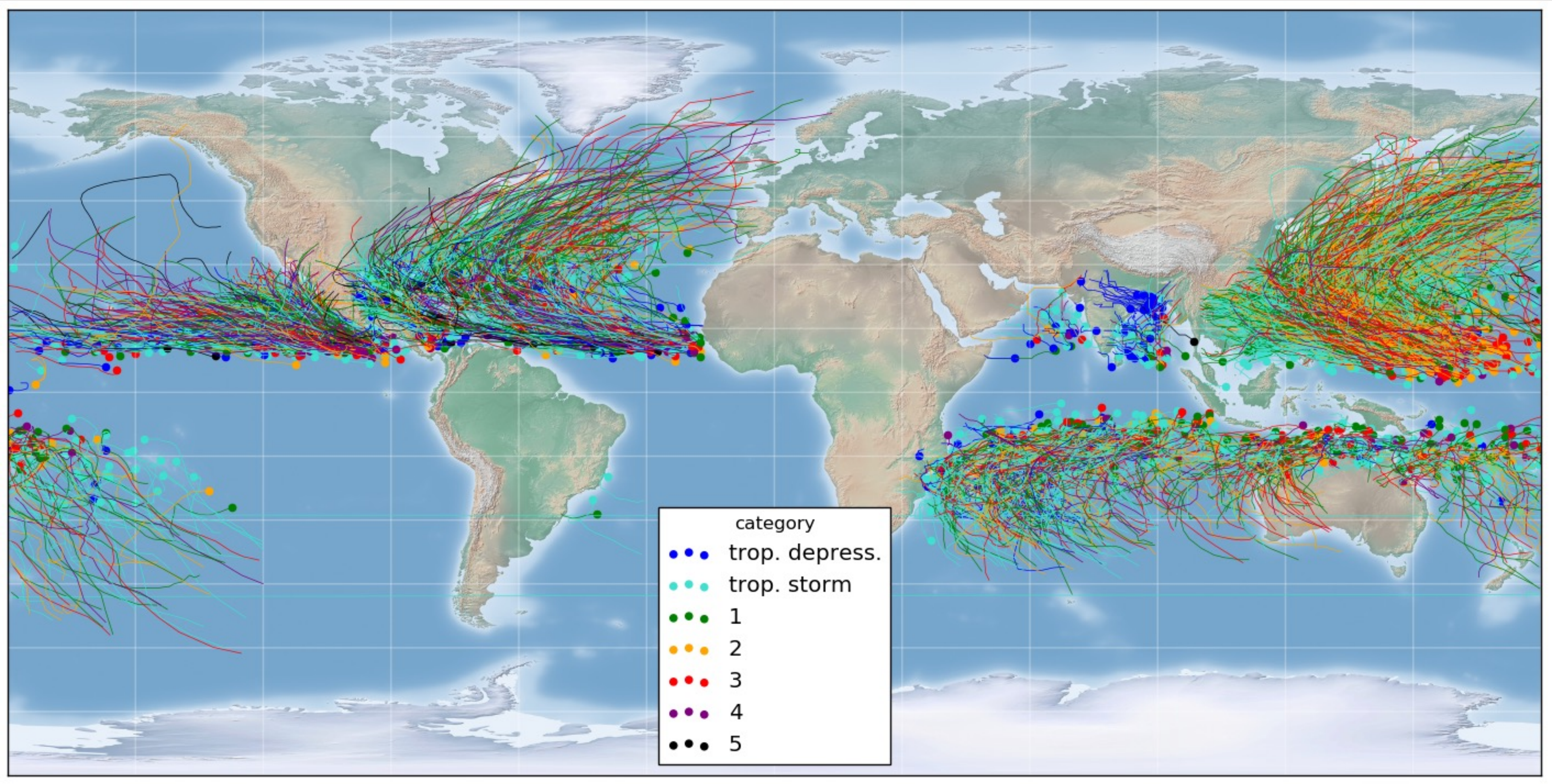
Downscaling: training data

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



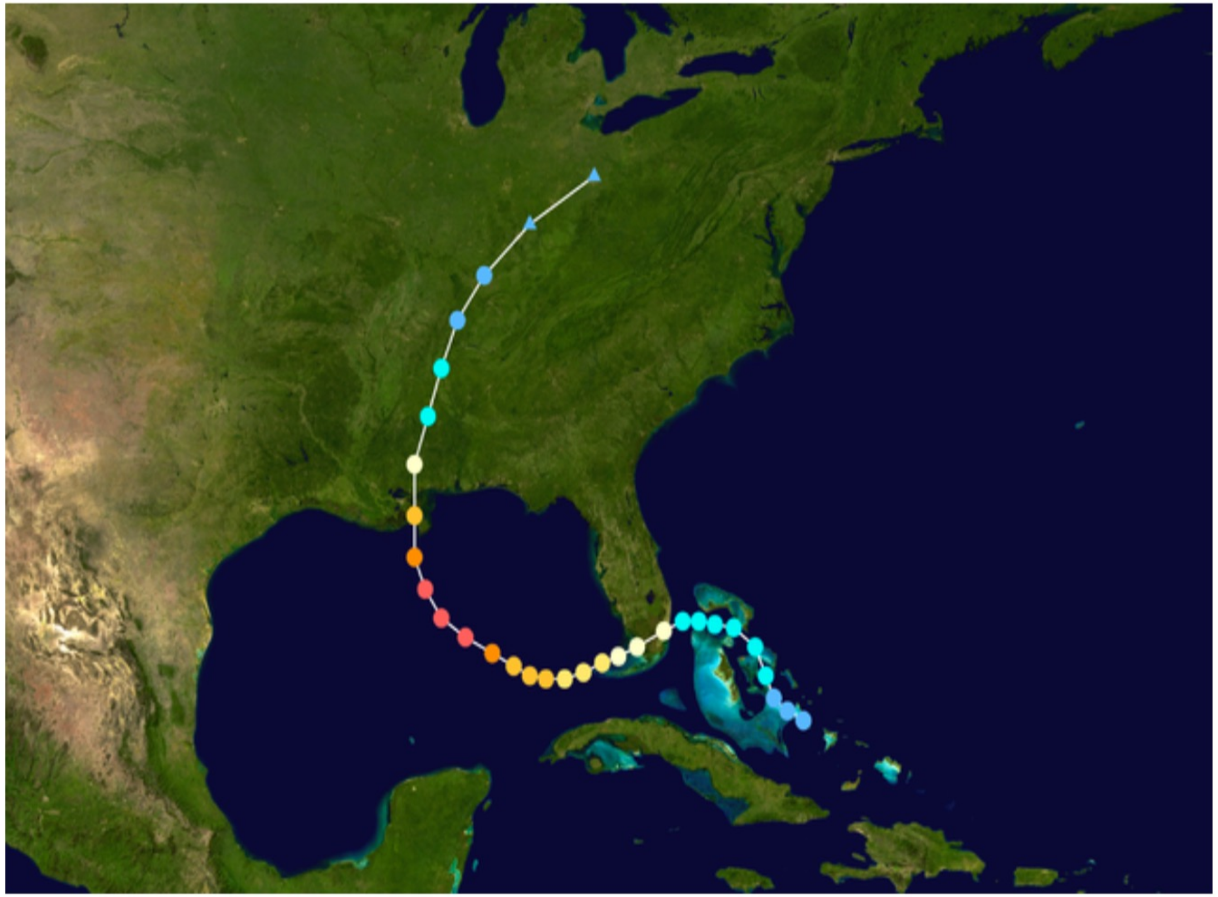
Storm track data

3000 tropical/extra-tropical storm tracks since 1979, measurements every 6 hrs



NOAA IBTrACS database

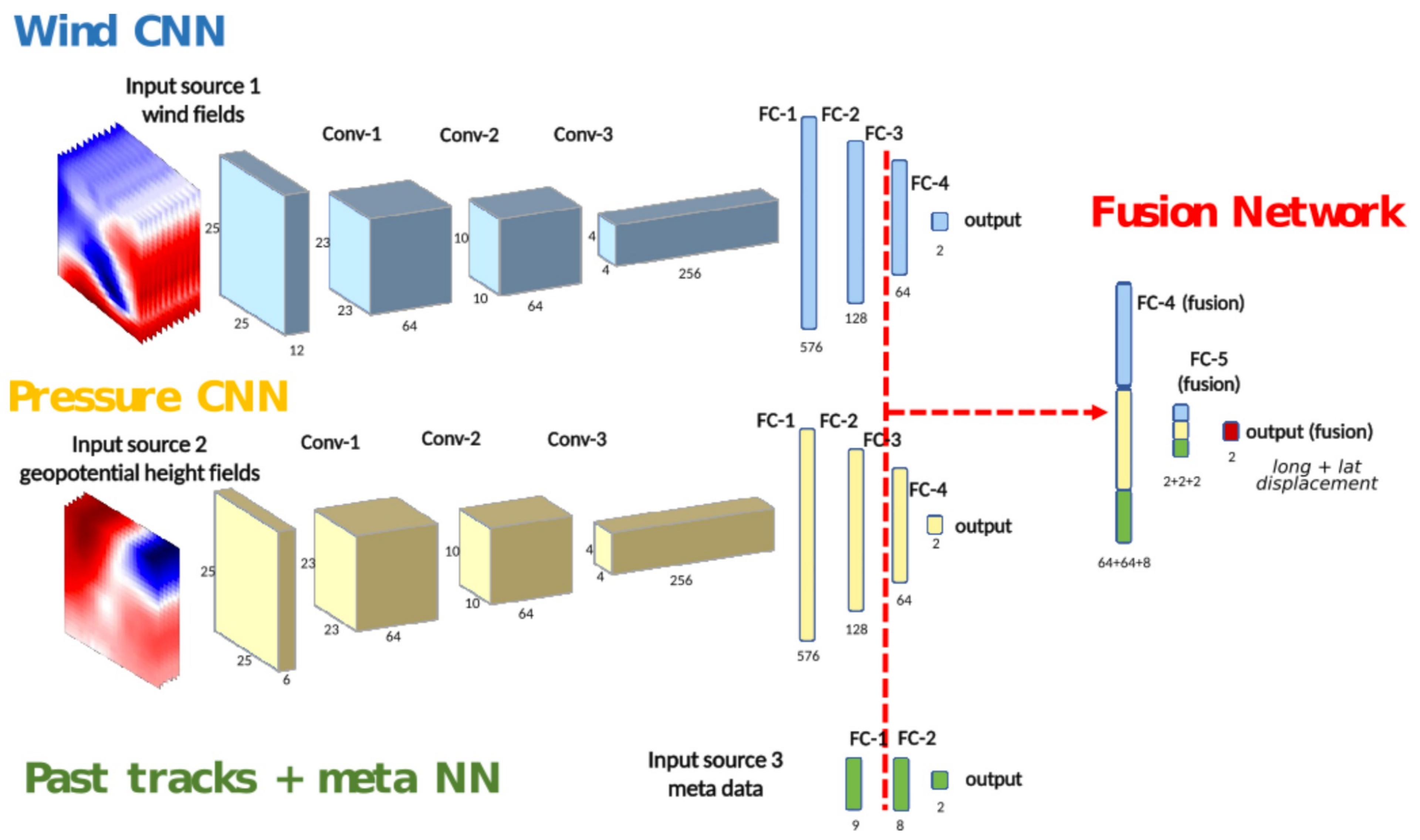
Storm track data



Saffir-Simpson Hurricane Scale		
Category	Wind Speed	
	mph	knots
5	>=156	>=135
4	131-155	114-134
3	111-130	96-113
2	96-110	84-95
1	74-95	65-83
Non-Hurricane Scale		
Tropical Storm	39-73	34-64
Tropical Depression	0-38	0-33

- Hurricane Katrina, 2005. (1 dot every 6 hours).
- **Tracks** and **Intensity** : Two main goals of the forecast

Deep Learning fusion network



[Giffard-Roisin et al.,Frontiers 2020]

Comparison to benchmarks

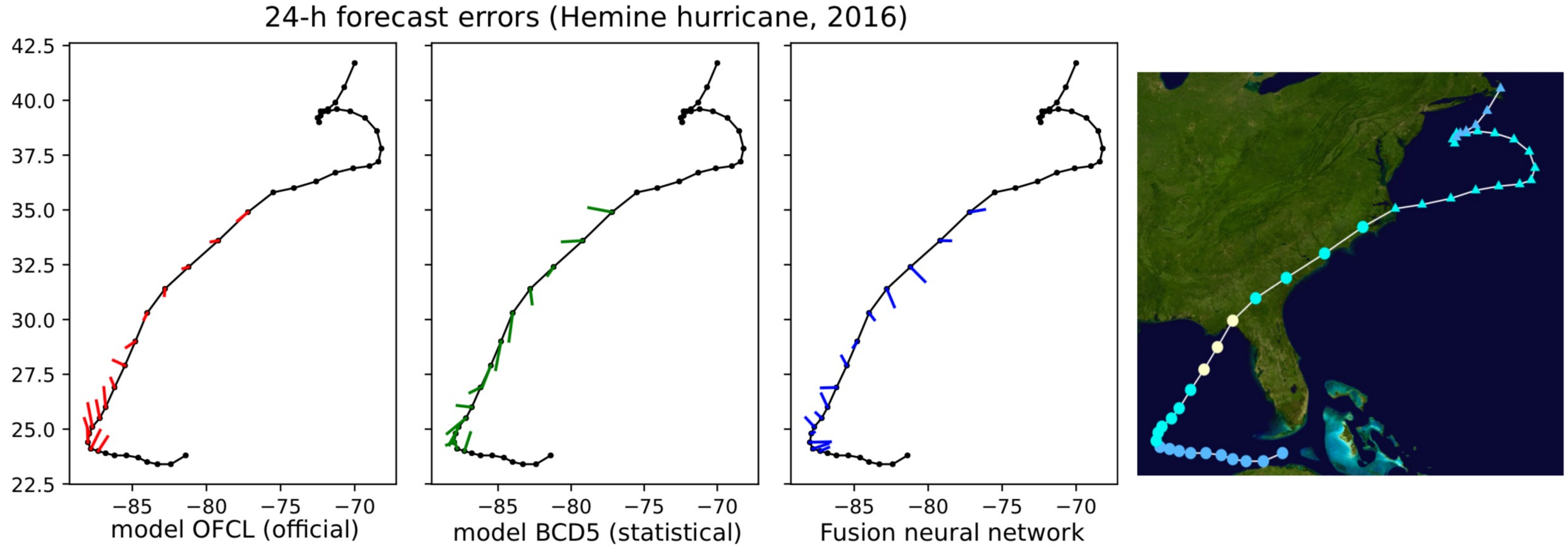
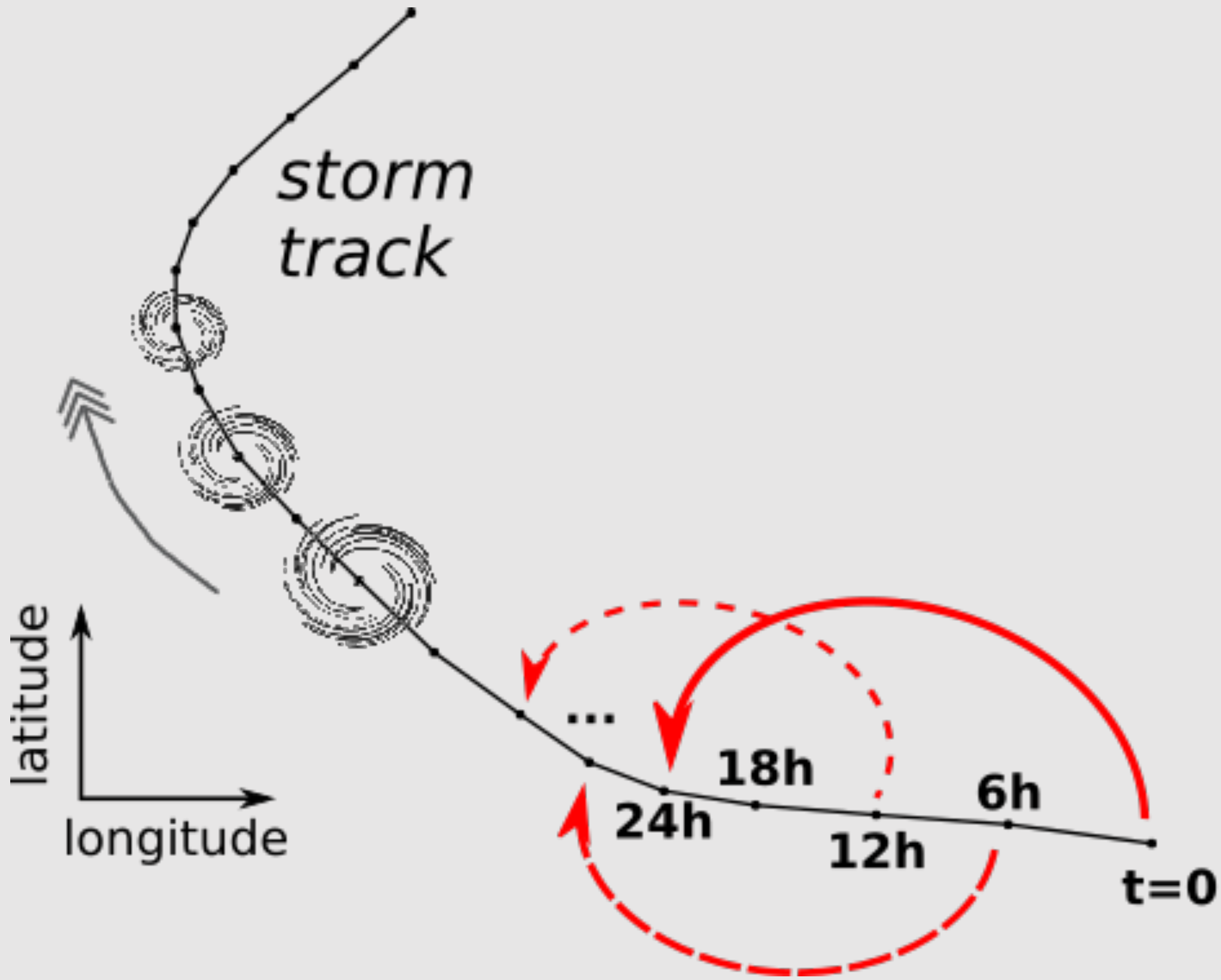


Figure: Hermine hurricane (2016) : 24-h forecast errors (4 time steps ahead). The bars connect each pair of predicted and ground truth location. The larger the length, the larger the error.

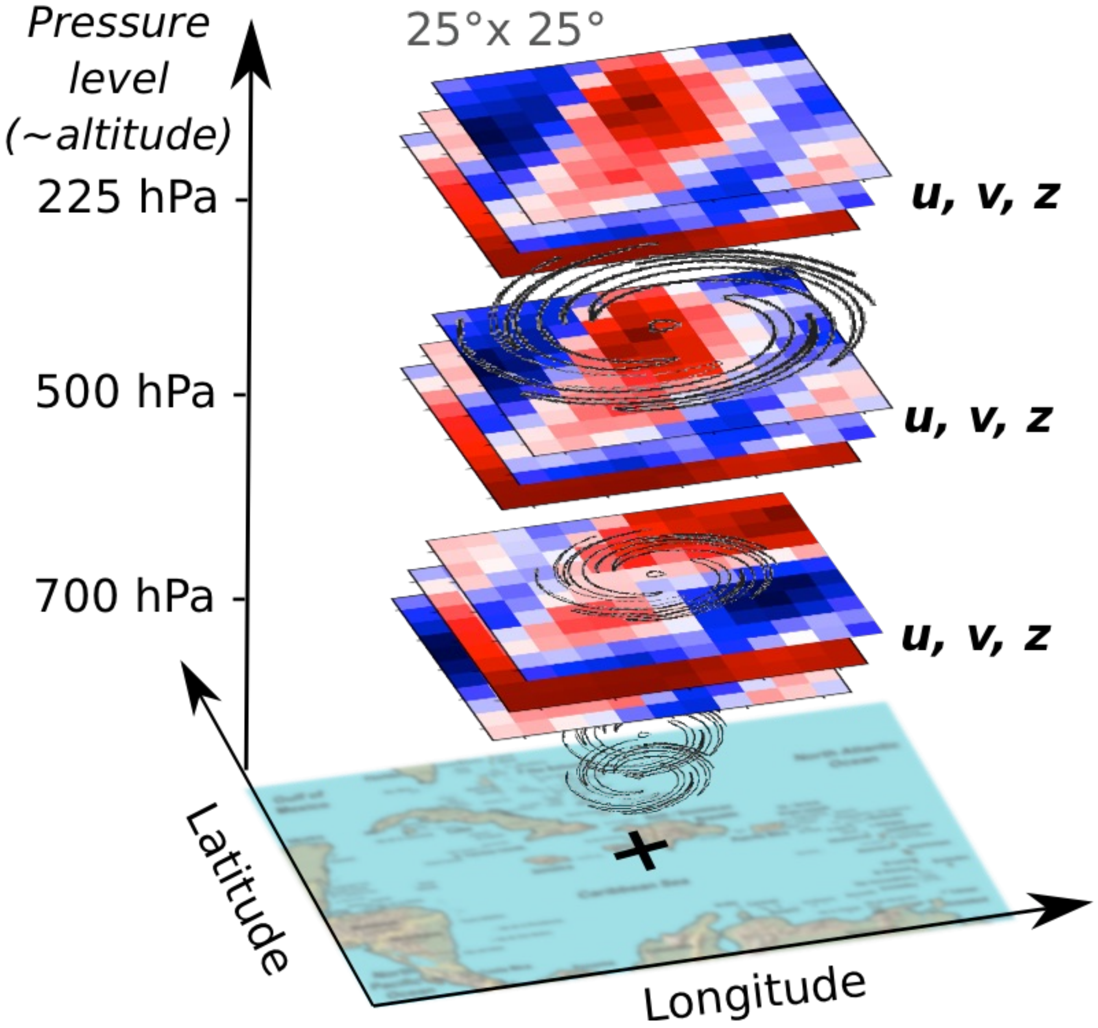
[Giffard-Roisin et al.,Frontiers 2020]

Forecasting task: 24h spatial displacement



Our approach: moving frame-of-reference

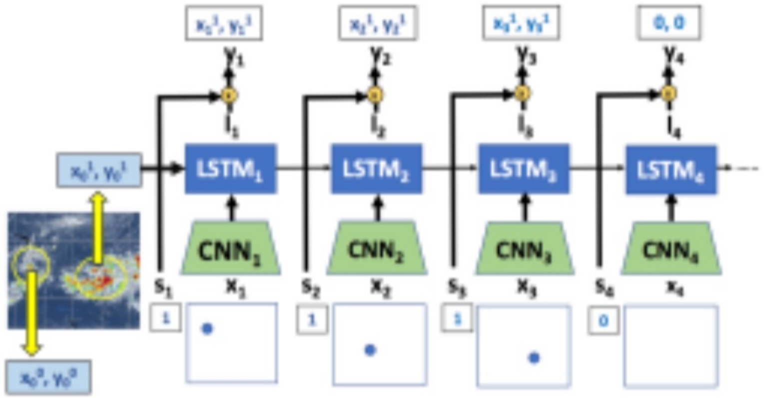
- Estimate future **displacement** as $\vec{u} = (dx, dy)$
- Centered reanalysis data (center = current storm location)



[Giffard-Roisin et al.,Frontiers 2020]

Related work

- Define a region (hurricane basin)
- Estimate future location as (x,y) coordinates
- Training set: storms from the same basin



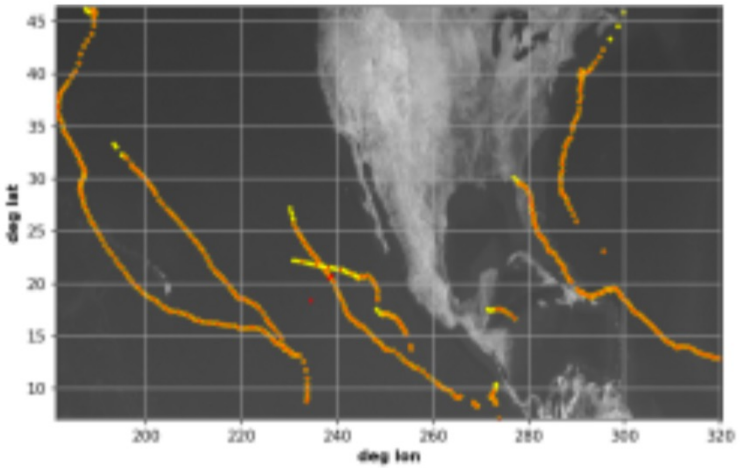


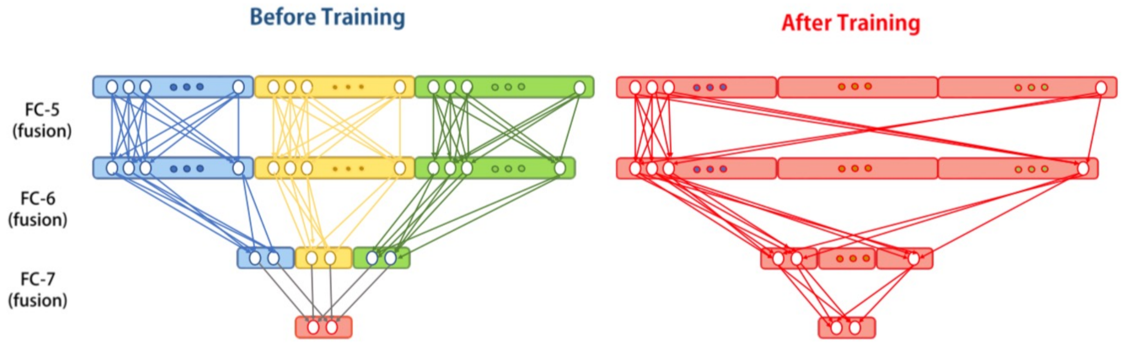
Figure: Mudigonda et. al, DLPS Workshop at NIPS 2017

Data types

- *Wind and pressure fields*: at 3 pressure levels (700 hPa, 500 hPa, and 225 hPa); at times t and $t - 6h$ **(2D+t)**
- *Past displacements*: $\vec{u_{t-6h}}$ and $\vec{u_{t-12h}}$ **(0D+t)**
- *Other hand-crafted features:* **(0D)**:
    - current latitude / longitude
    - windspeed
    - Jday predictor(Gaussian function of "Julian day of storm init - peak day of the hurricane season")
    - current distance to land

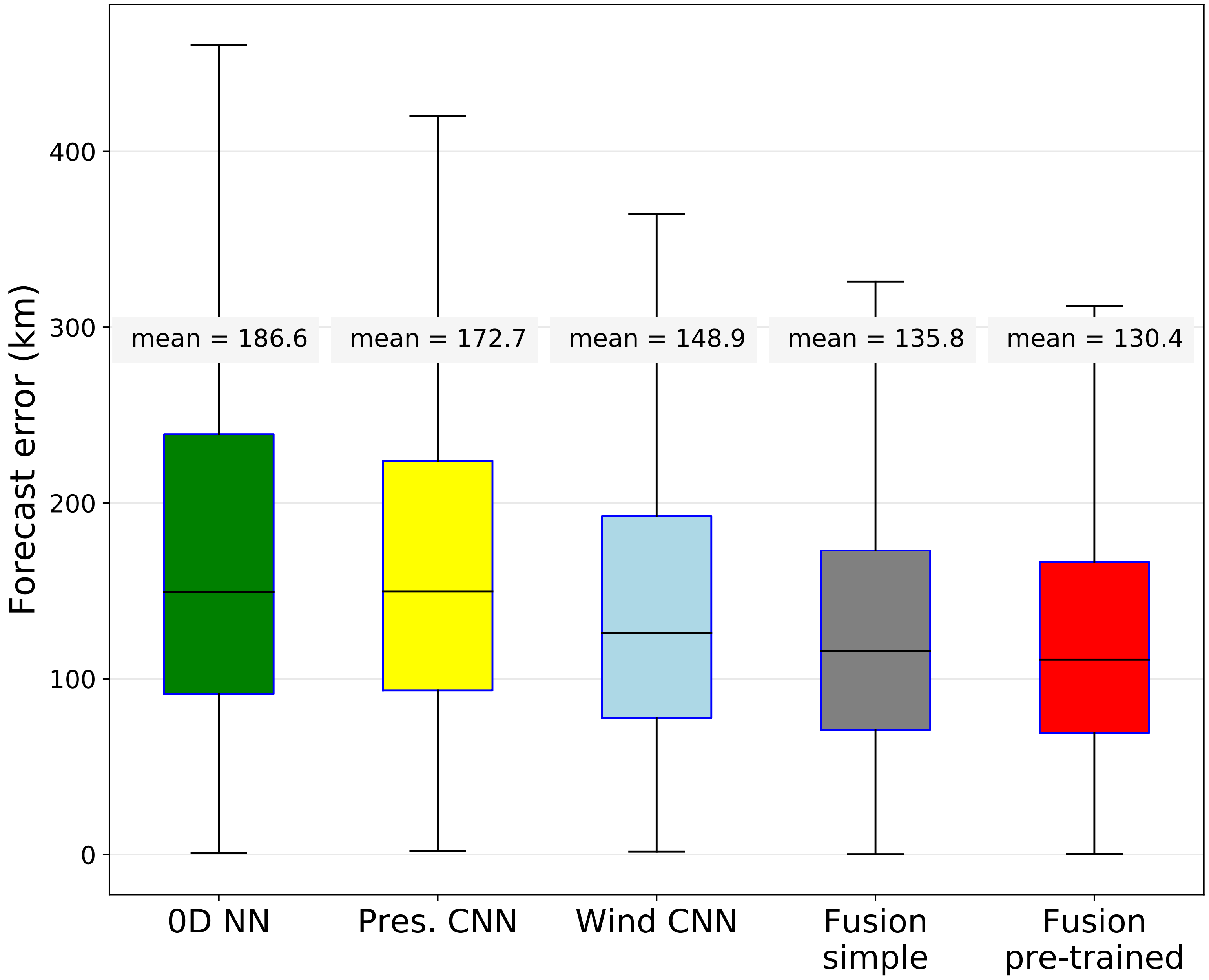
Training the fusion network

- Stage I: Train separate networks
- Stage II: Train the fusion network
  - Zoom in fusion layers:



  - Add connections between different streams in fusion layers
  - Re-train the whole network

Performance of network components



State of the art

- **BCD5** : statistical model, often used to benchmark other storm track forecasting methods
- **OFCL** : National Hurricane Center official forecast (consensus of dynamical models), BUT evolving over years

Model	Atlantic errors (km)		East Pacific errors (km)	
	mean error	std	mean error	std
BCD5	125	90	112	78
Fusion	112	71	88	52

Table: Mean and standard deviation 24h-forecast errors for the Atlantic and Pacific basins on part of the test set (total = 4349 time steps)

- Mean error across all basins, time steps from hurricanes only: 103.9 km
- [Climate Informatics '18]: 6h prediction error, same evaluation: 28.5 km