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

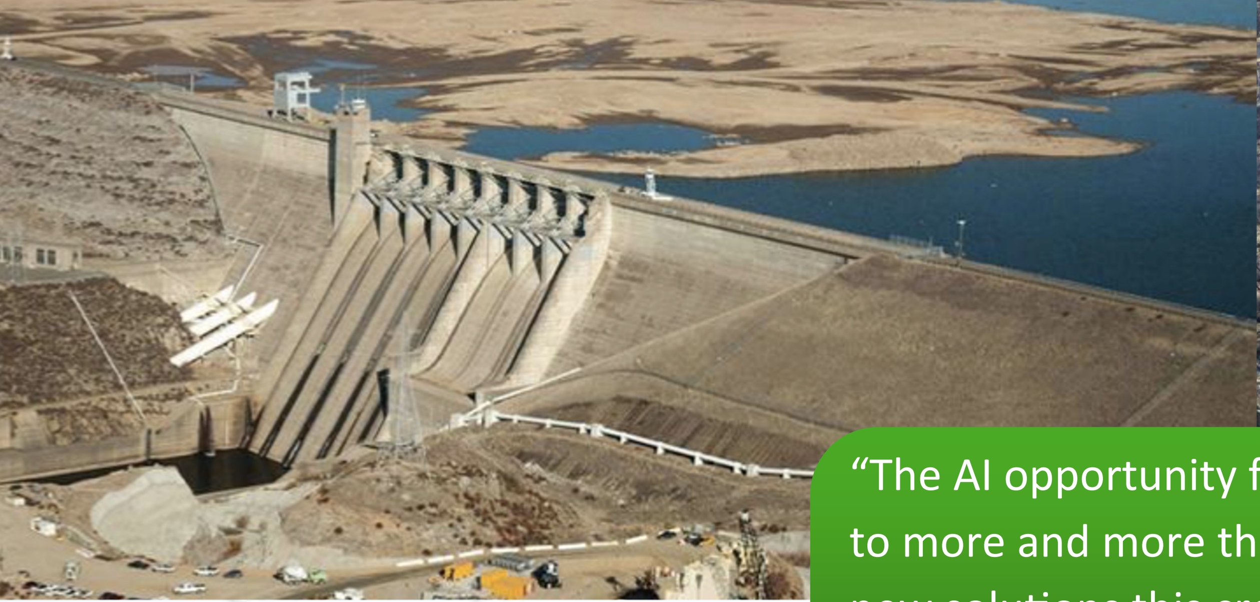
- Snow drought conditions through fall and winter 2021 created dry land-cover
- 80-100 mph winds, combined with ignition, launched an uncontrollable “fire storm”
- Loss of 2 lives. 1000 homes and 20 businesses were destroyed, and more damaged

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 Climate Science

  Understanding and Predicting Climate Change

- Machine Learning for Climate Adaptation

  Extreme Weather and Cascading Hazards

- Machine Learning for Climate Mitigation

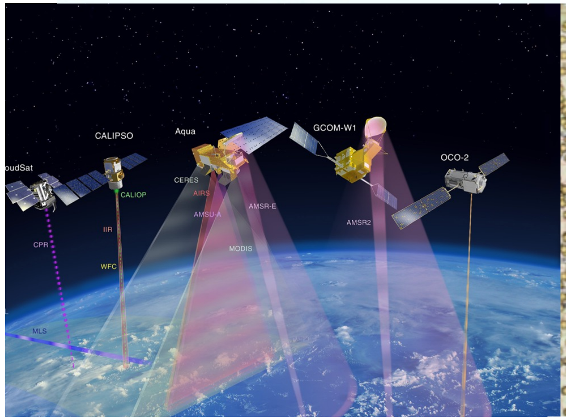
  Accelerating the Green Transition

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

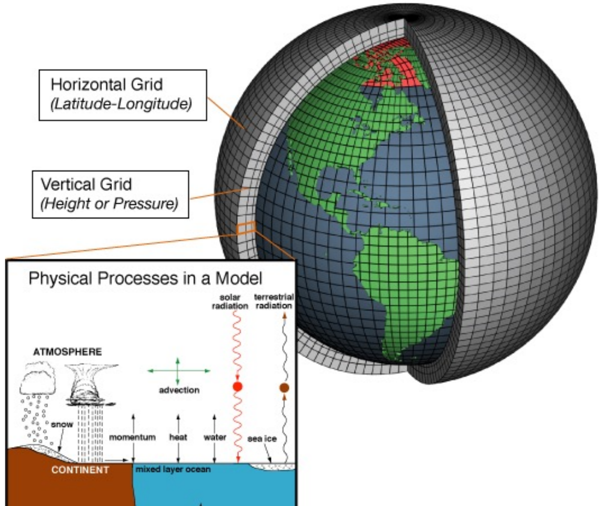
Climate data types

- Past: Historical data
  - Limited amounts
  - Very heterogeneous

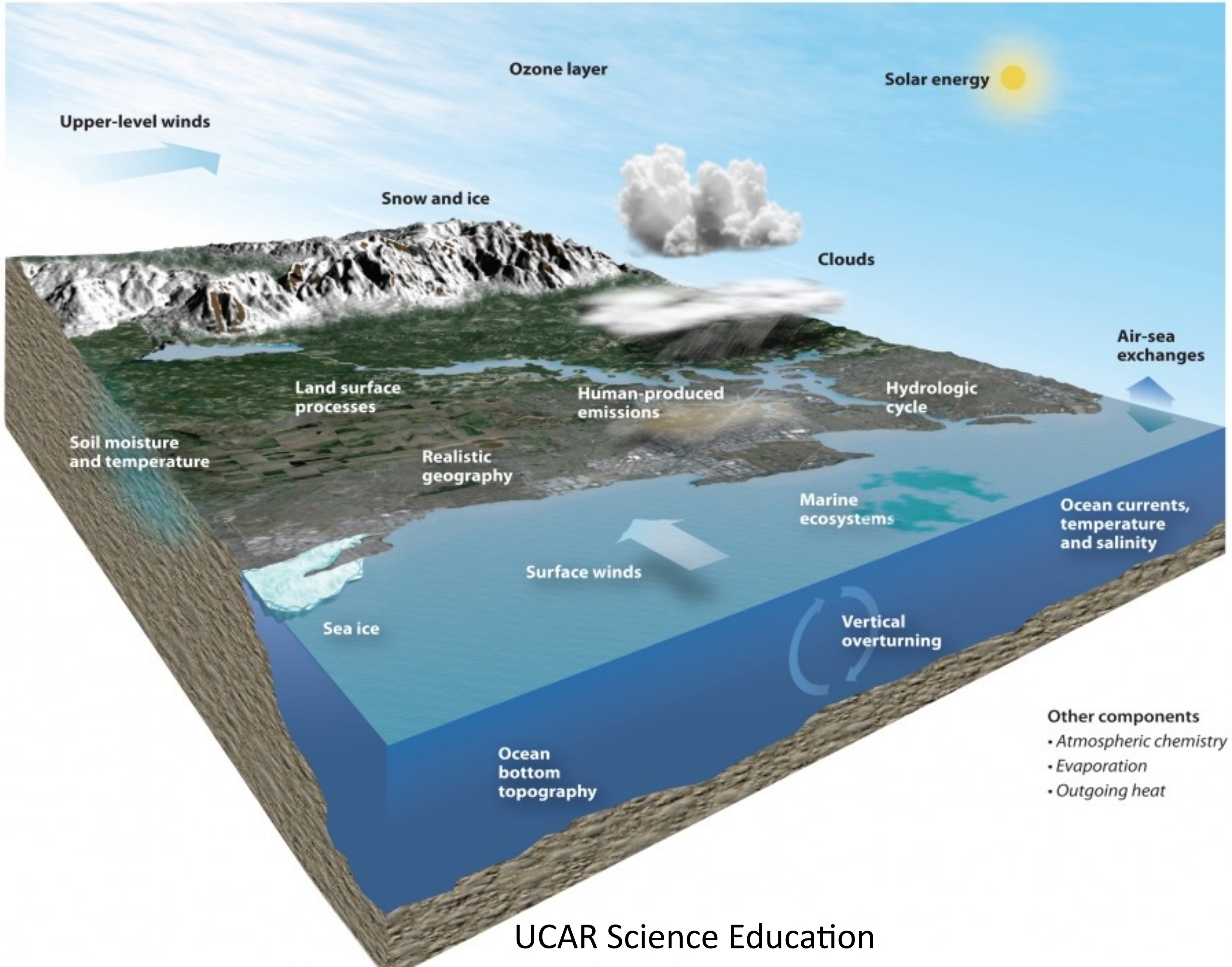


- Present: Observation data
  - Large quantities recently
  - High-dimensional
  - Can be unlabeled, sparse

- Past, Present, Future: Climate model simulations
  - Massive, high-dimensional
  - Encodes scientific domain knowledge, physics
  - Some information lost in discretizations
  - Future predictions cannot be validated



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]

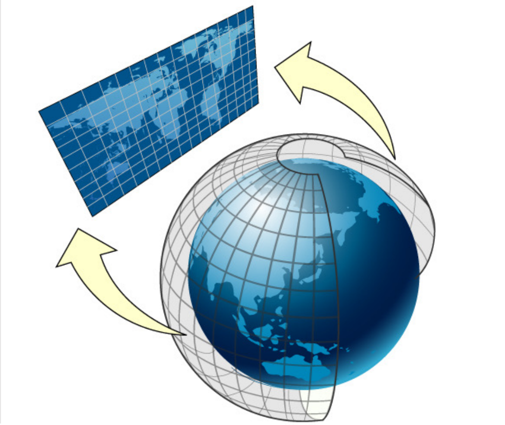
Causal information hubs in Pacific ENSO region

[Saha et al., Climate Informatics 2019]

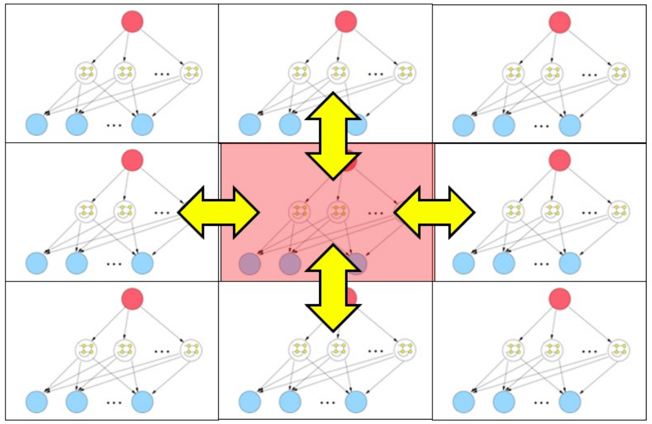
NASA / NCAR project to attribute and forecast sea-level rise using climate models and satellite altimetry

[Sinha et al., AGU 2022, ICLR 2023 workshop]

Online learning with spatiotemporal non-stationarity



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



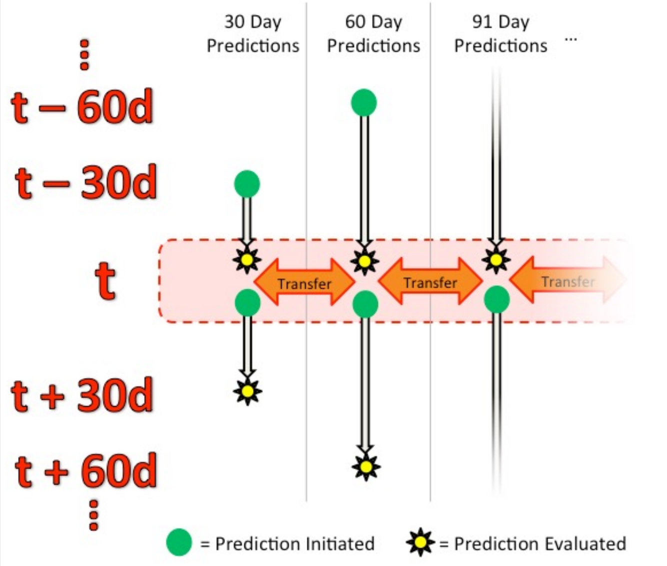
We can exploit local structure in space and time

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

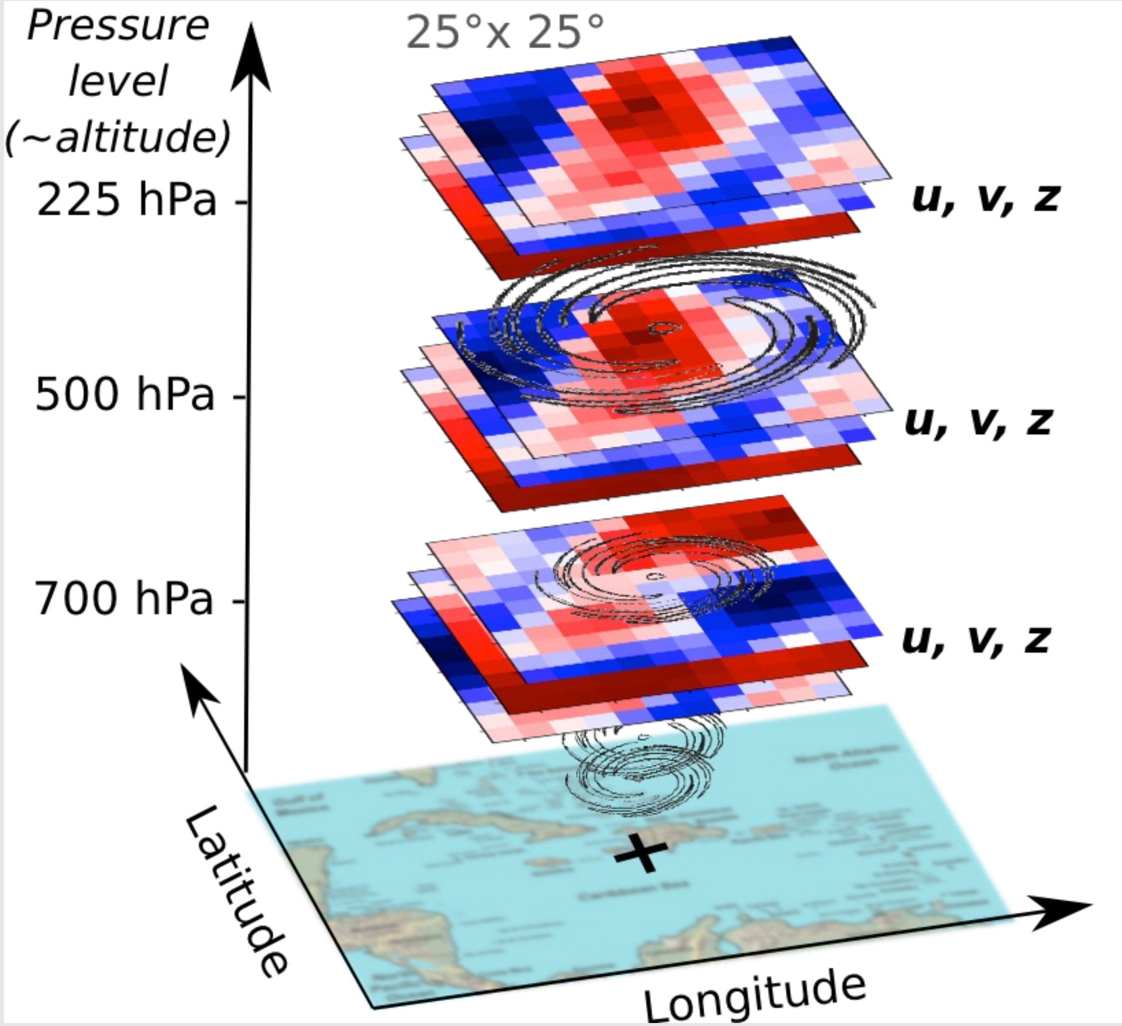
New "regret" framework: [Cesa-Bianchi, Cesari, & Monteleoni, ALT 2020]



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)

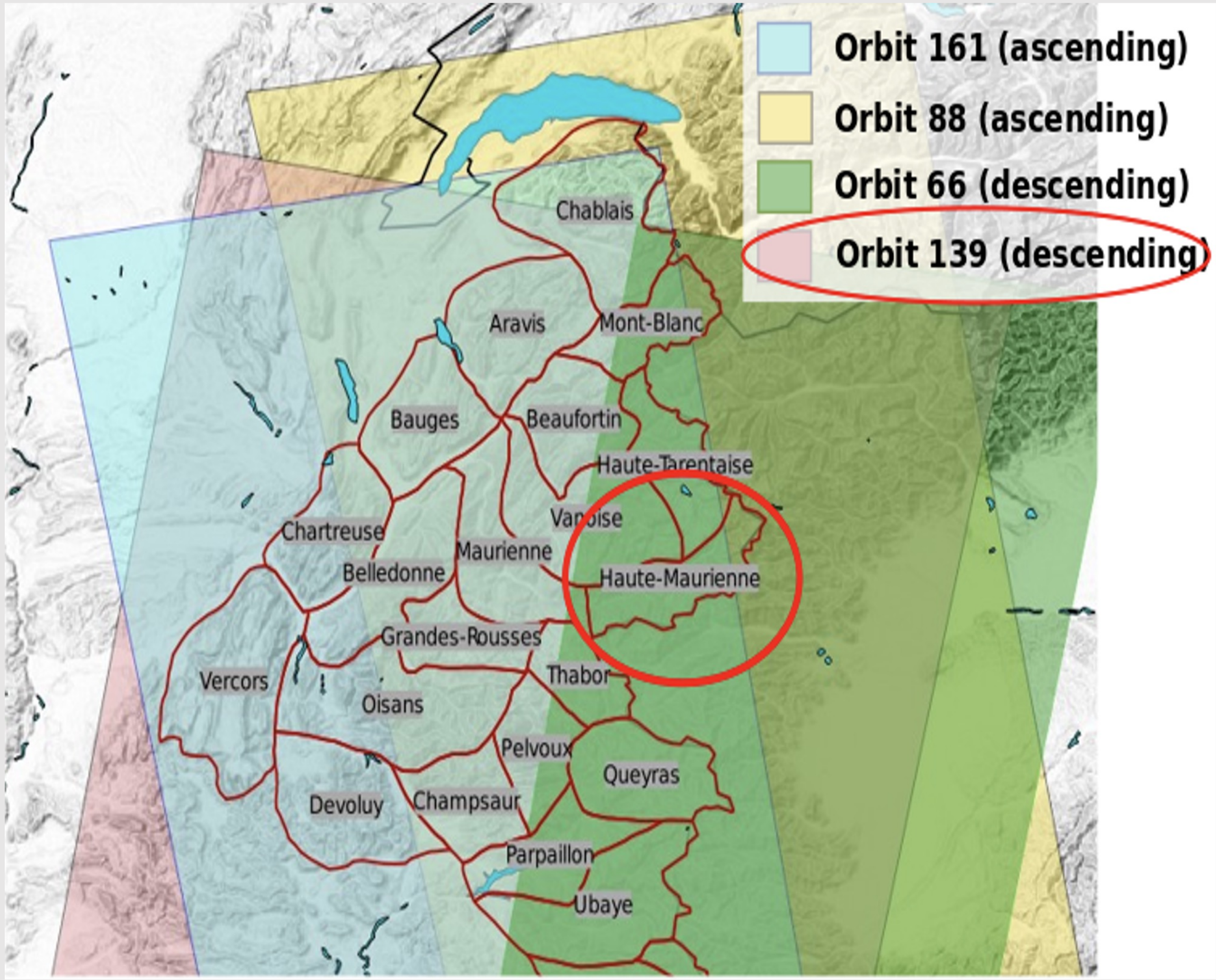
Avalanche detection using CNN; VAE

[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



METEO FRANCE

[Sinha et al., Climate Informatics 2020]

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

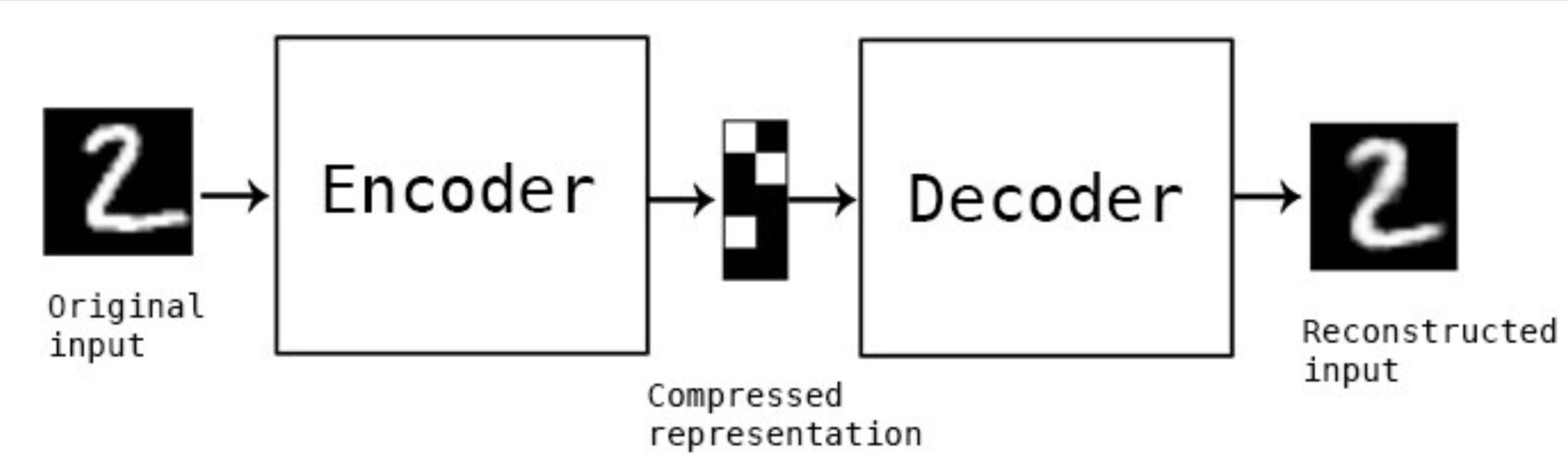
② Train a variational autoencoder (VAE) on the negative examples

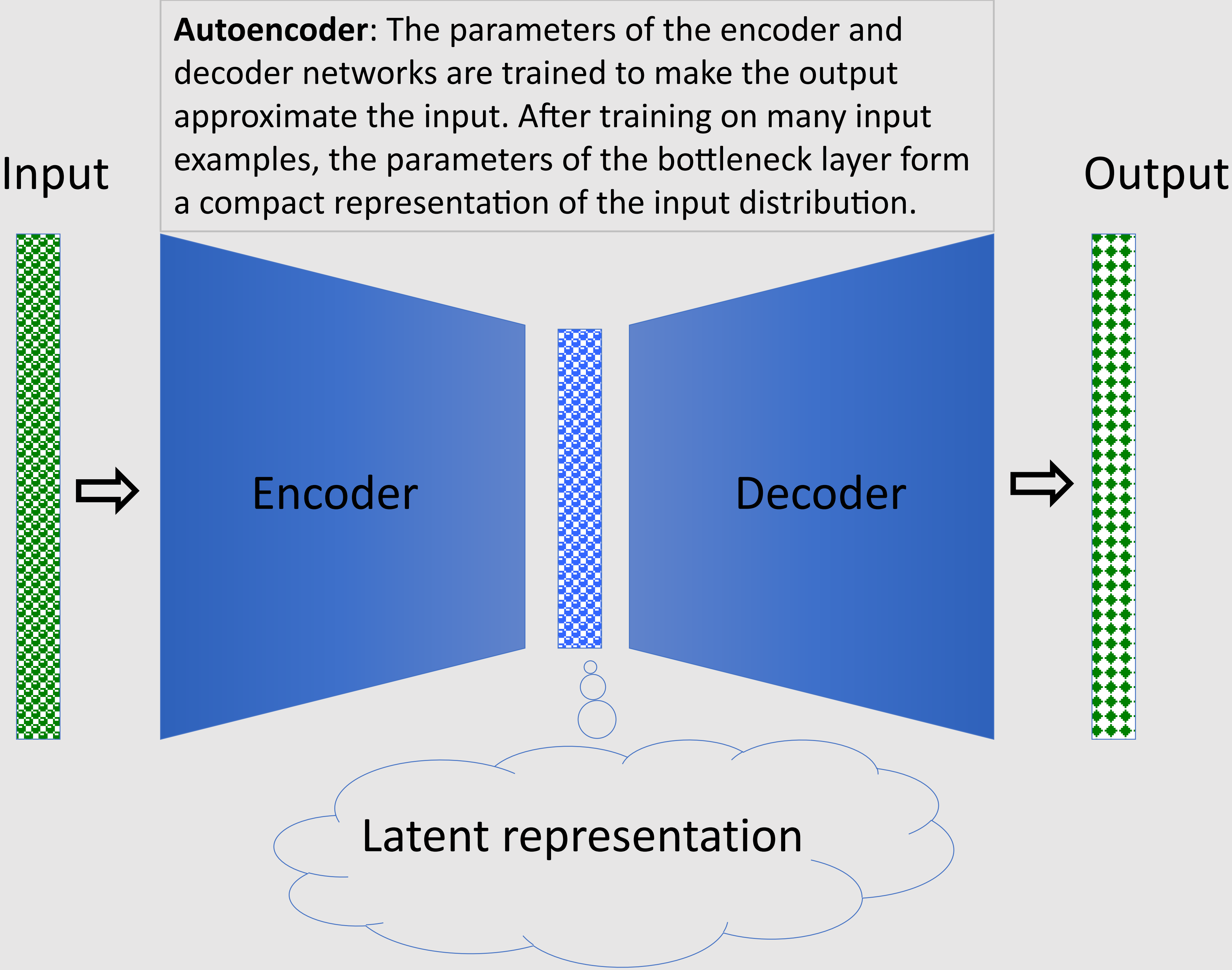
③ Threshold the VAE's reconstruction error to classify a new image

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

What is an Auto-encoder?

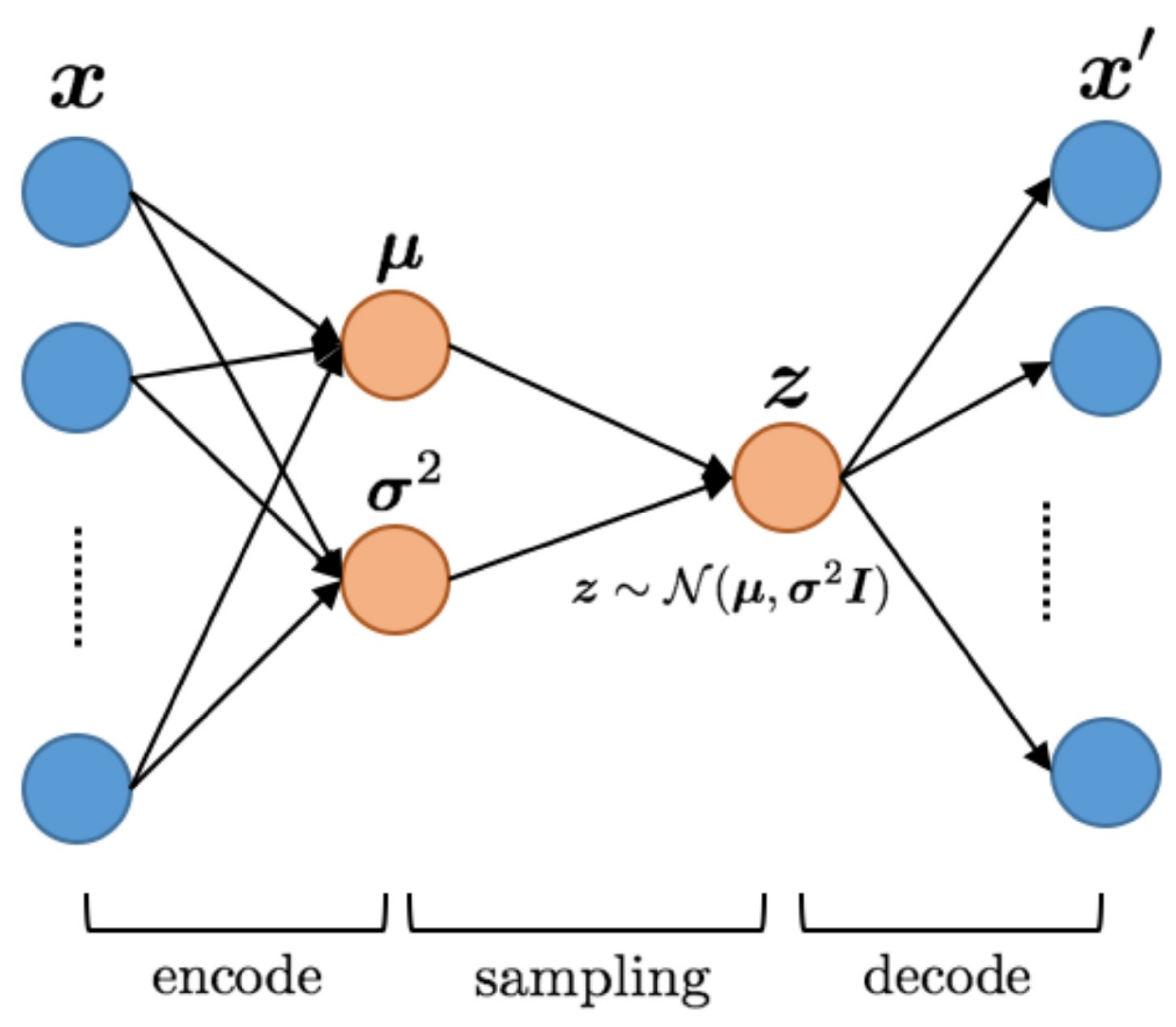
- Train a neural network in an unsupervised way
  - Use the unlabeled data both as input, and to evaluate the output
- After training, the bottleneck layer will be a compact representation of the input distribution



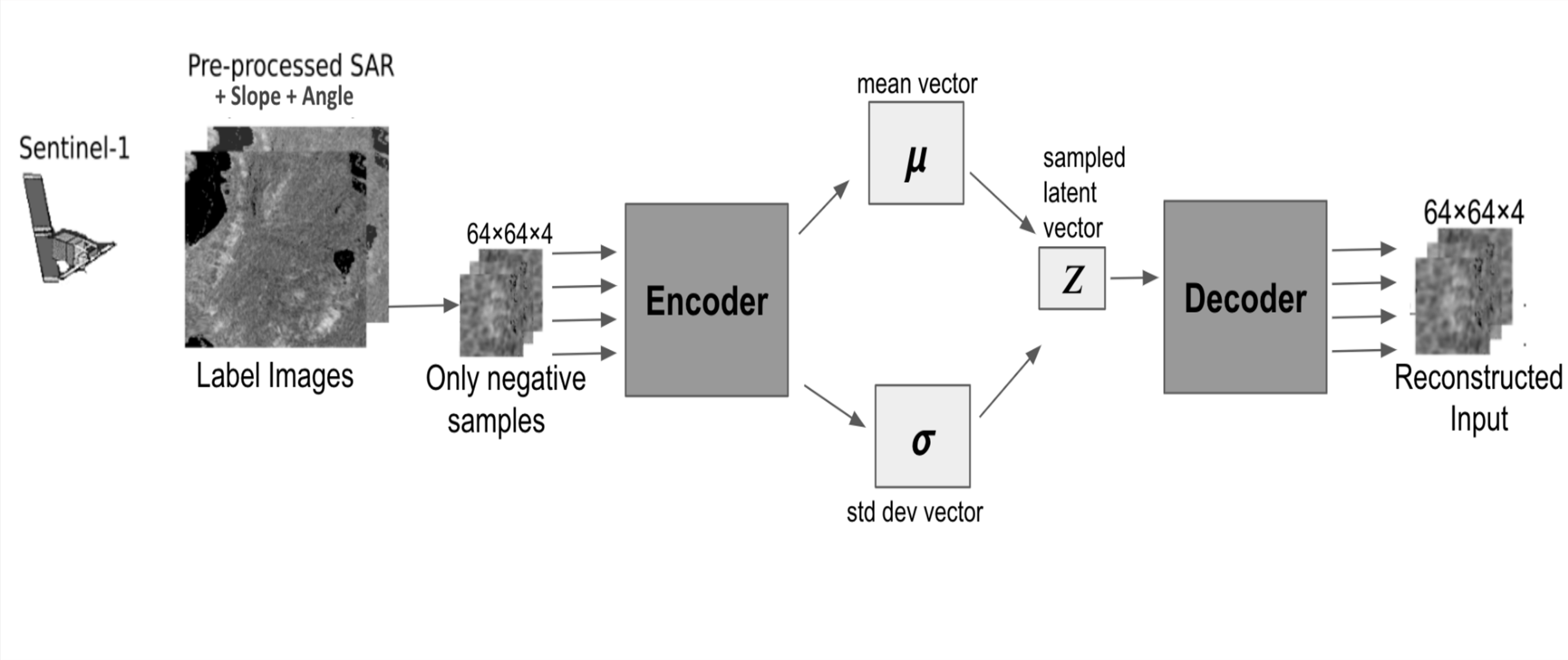

Autoencoder: The parameters of the encoder and decoder networks are trained to make the output approximate the input. After training on many input examples, the parameters of the bottleneck layer form a compact representation of the input distribution.
Input
Output
Encoder
Decoder
Latent representation

Variational Autoencoder (VAE)

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

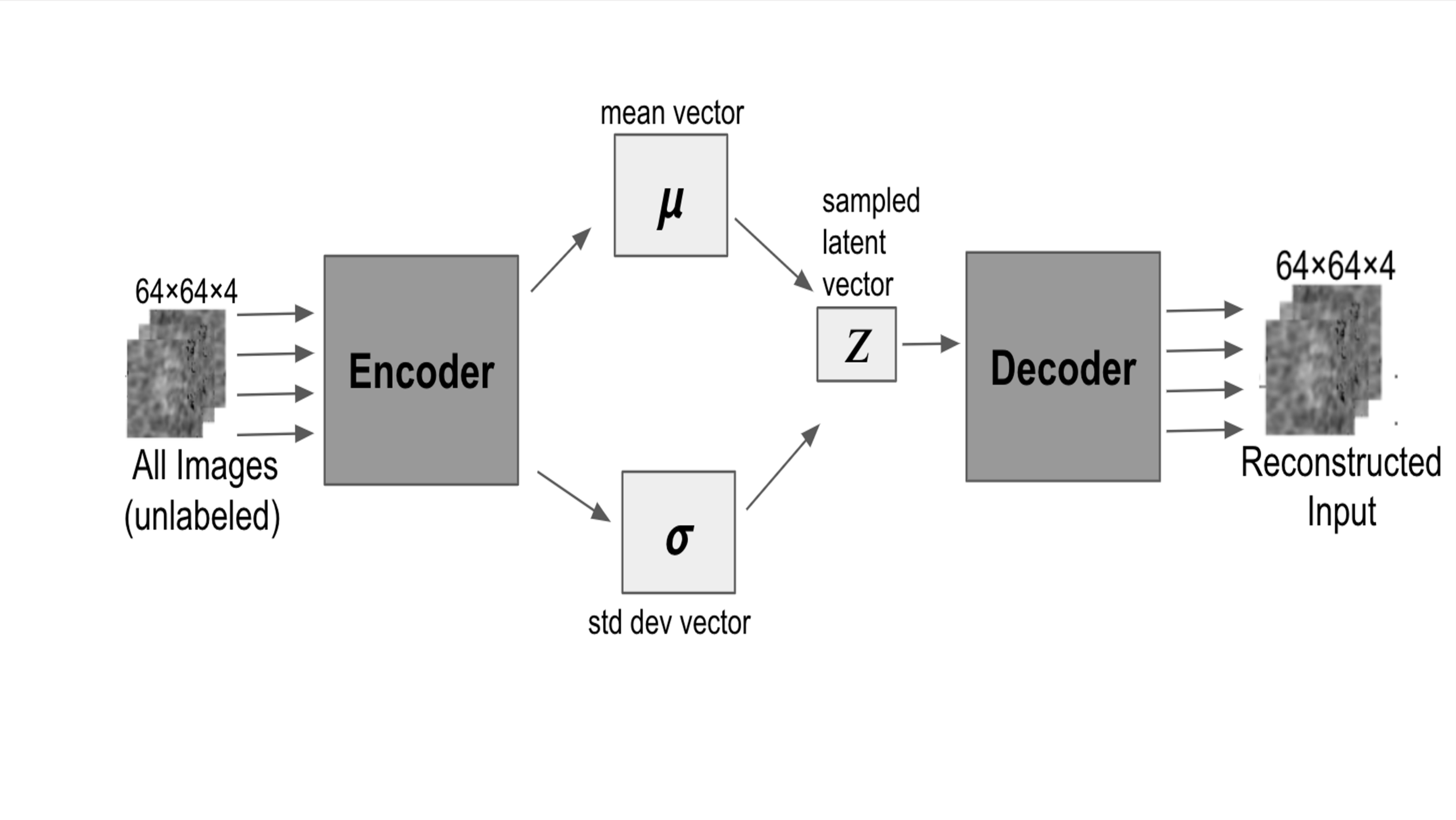


VAE for anomaly detection is typically trained on **negative** examples only



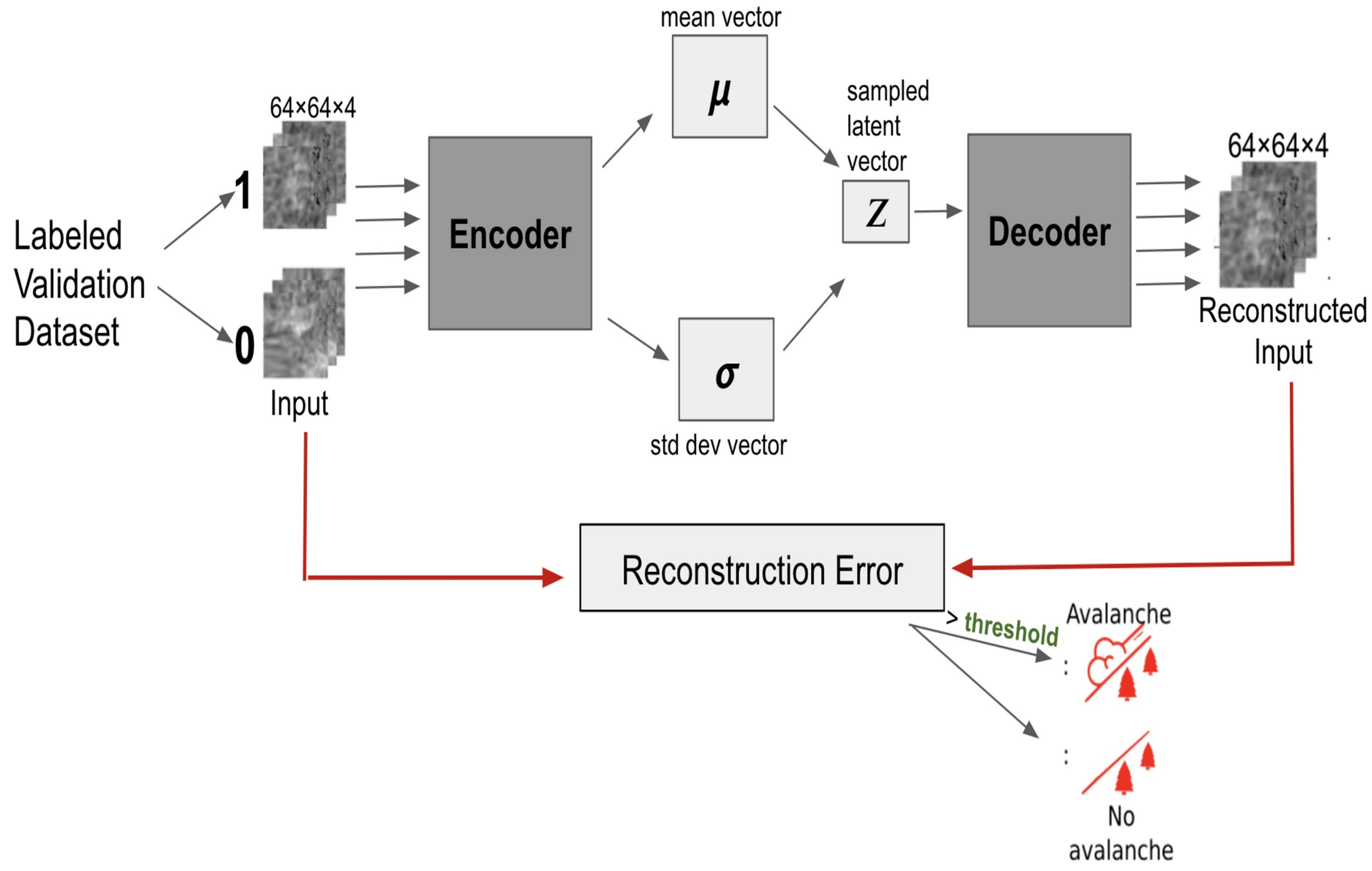
[Sinha et al., Climate Informatics 2020]

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



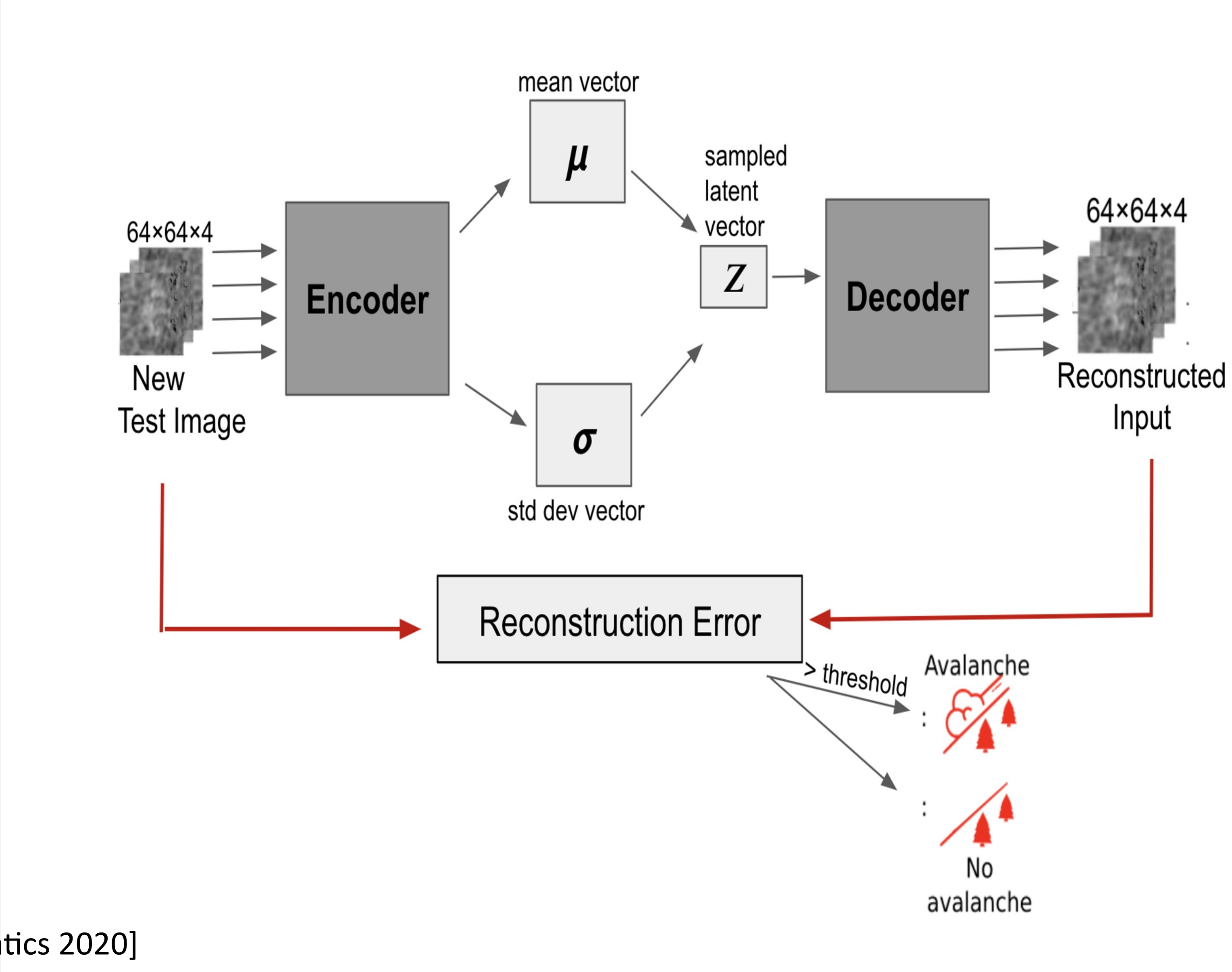
[Sinha et al., Climate Informatics 2020]

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

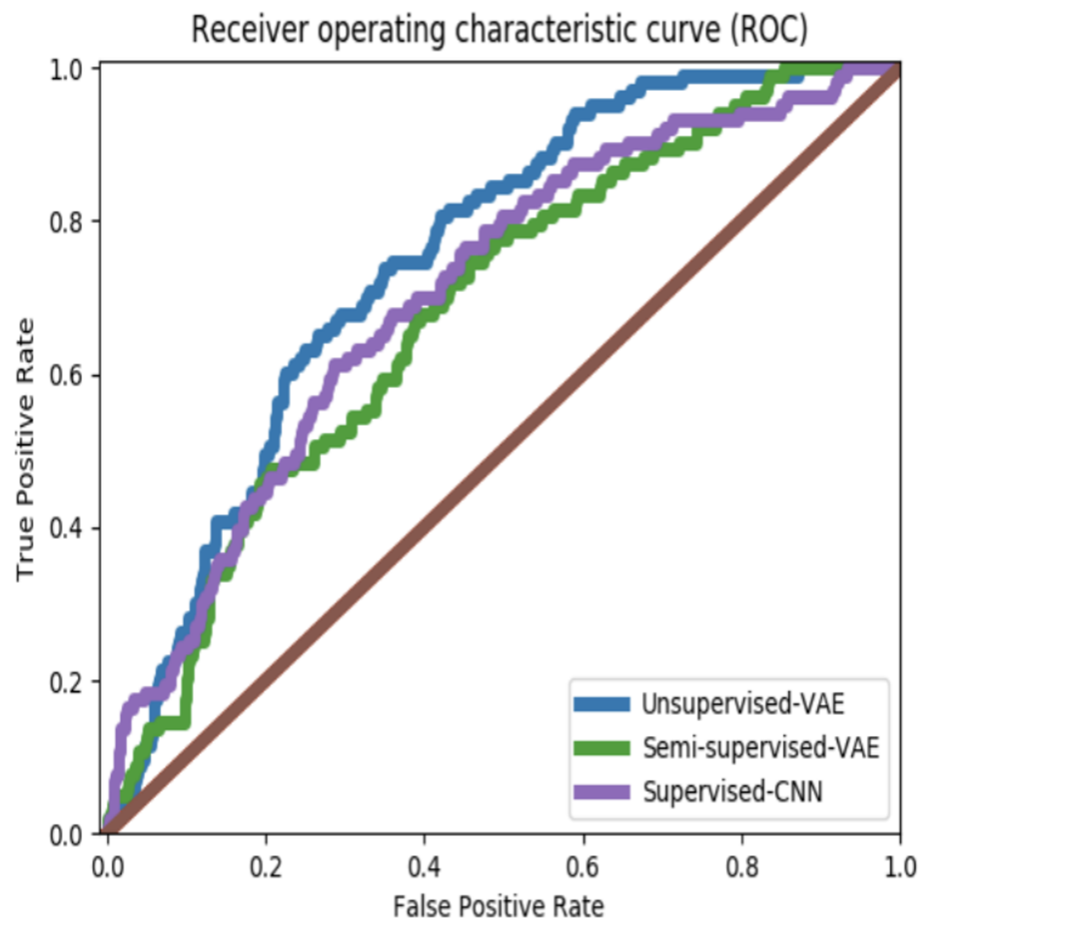
	All Alps		Haute Maurienne	
	Balanced Accuracy	F1-score	Balanced Accuracy	F1-score
Baseline	0.58	0.05	0.58	0.12
Supervised - CNN	0.53	0.10	0.53	0.12
Semi-supervised - VAE	0.59	0.11	0.6	0.23
Unsupervised - VAE	0.69	0.14	0.68	0.26

- 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]

Evaluation

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



ROC Analysis for Haute Maurienne region

[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

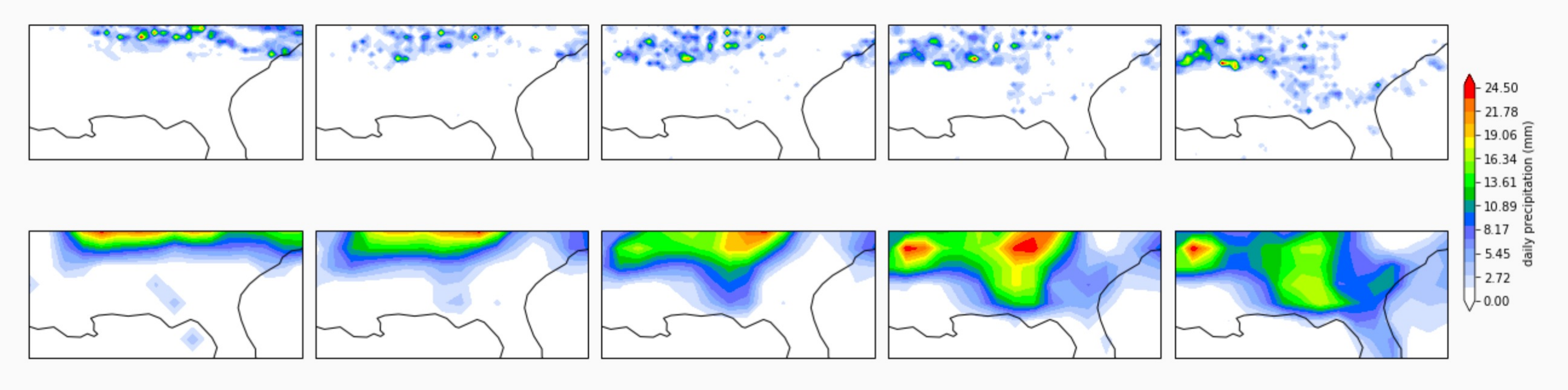
[Sinha et al., CI 2022] with NREL

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

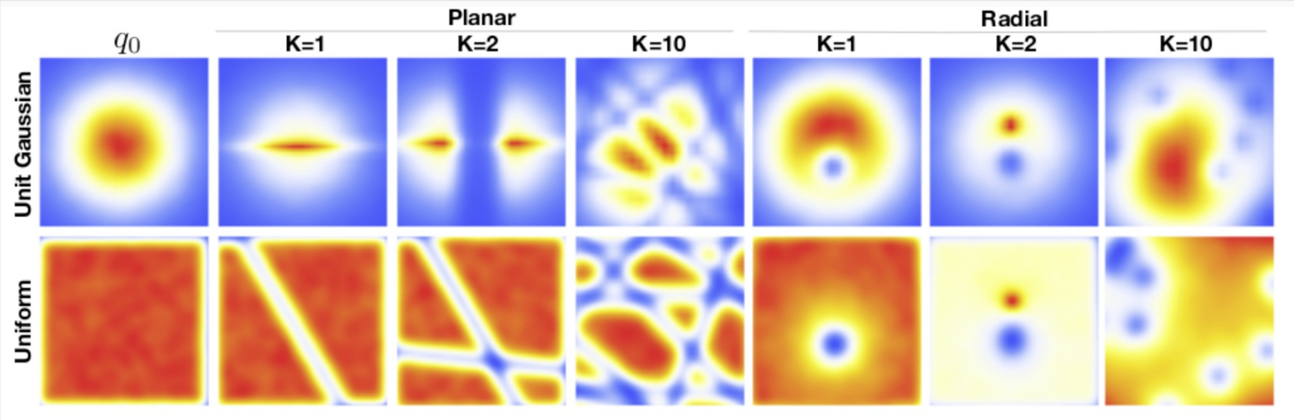


General downscaling technique via domain alignment with normalizing flows
[AlignFlow: Grover et al., AAAI 2020][Glow: Kingma & Dhariwal, NeurIPS 2018]

- **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

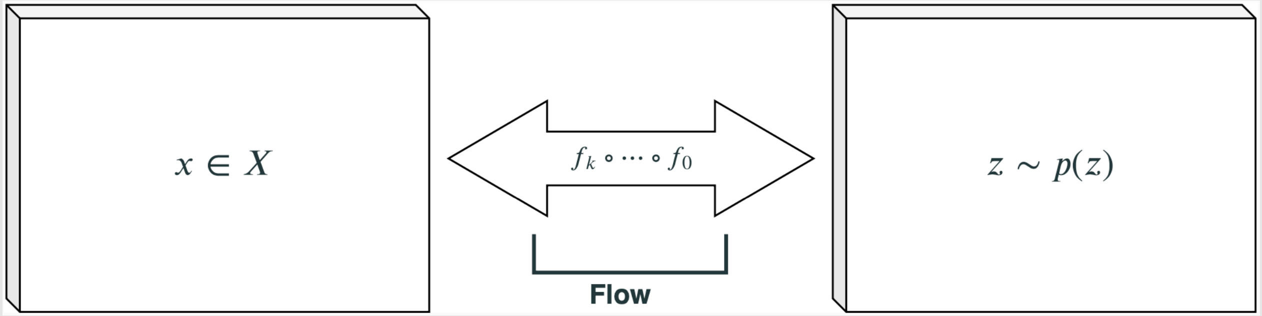


[Rezende & Mohamed, ICML 2015]

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

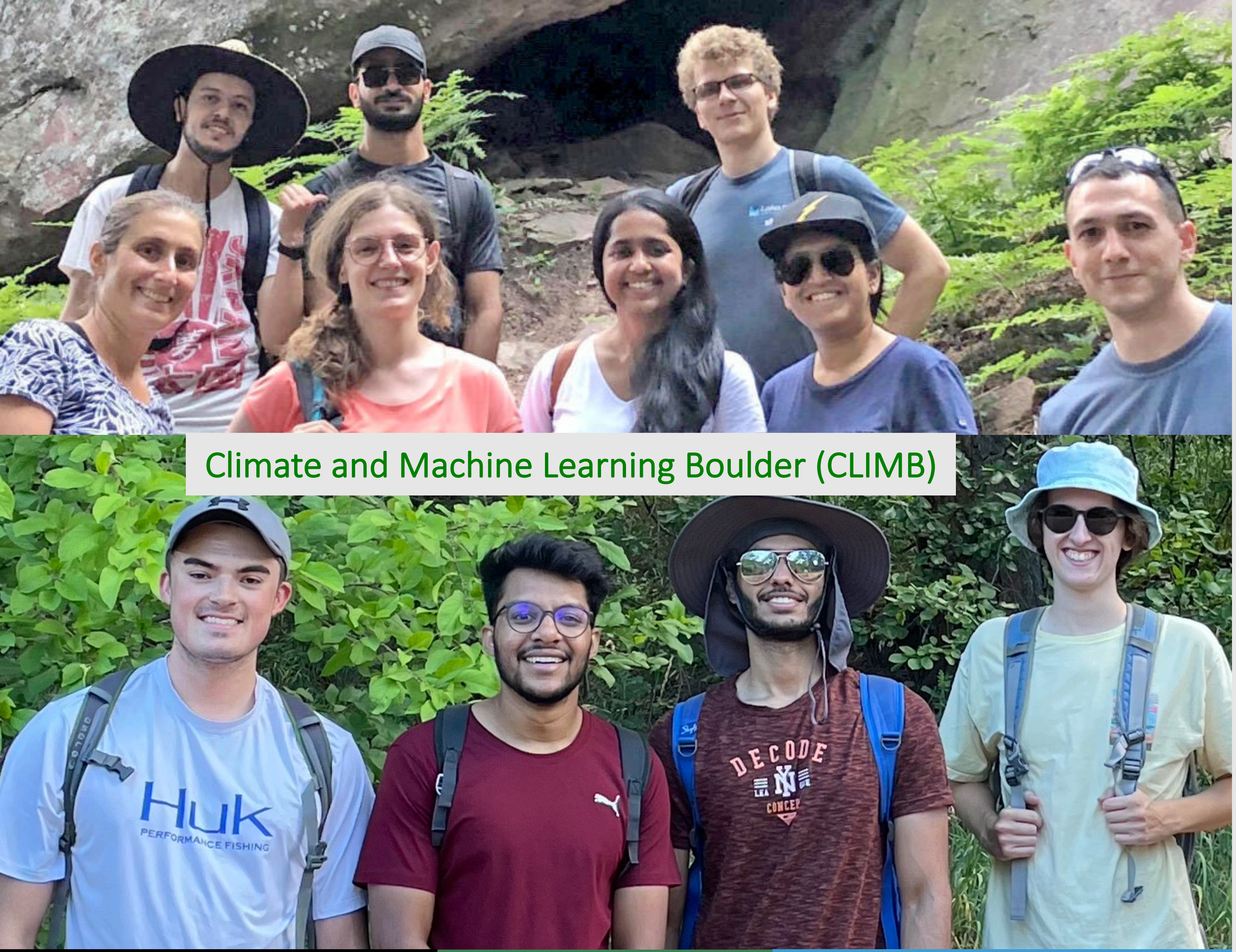
Long-term Inspirations

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



Climate and Machine Learning Boulder (CLIMB)

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*
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 NSF Climate Change AI futurearth Research. Innovation. Sustainability. METEO FRANCE NASA

ARCHES:
AI research for climate change
& environmental sustainability

Join our team at INRIA Paris!

Inria

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