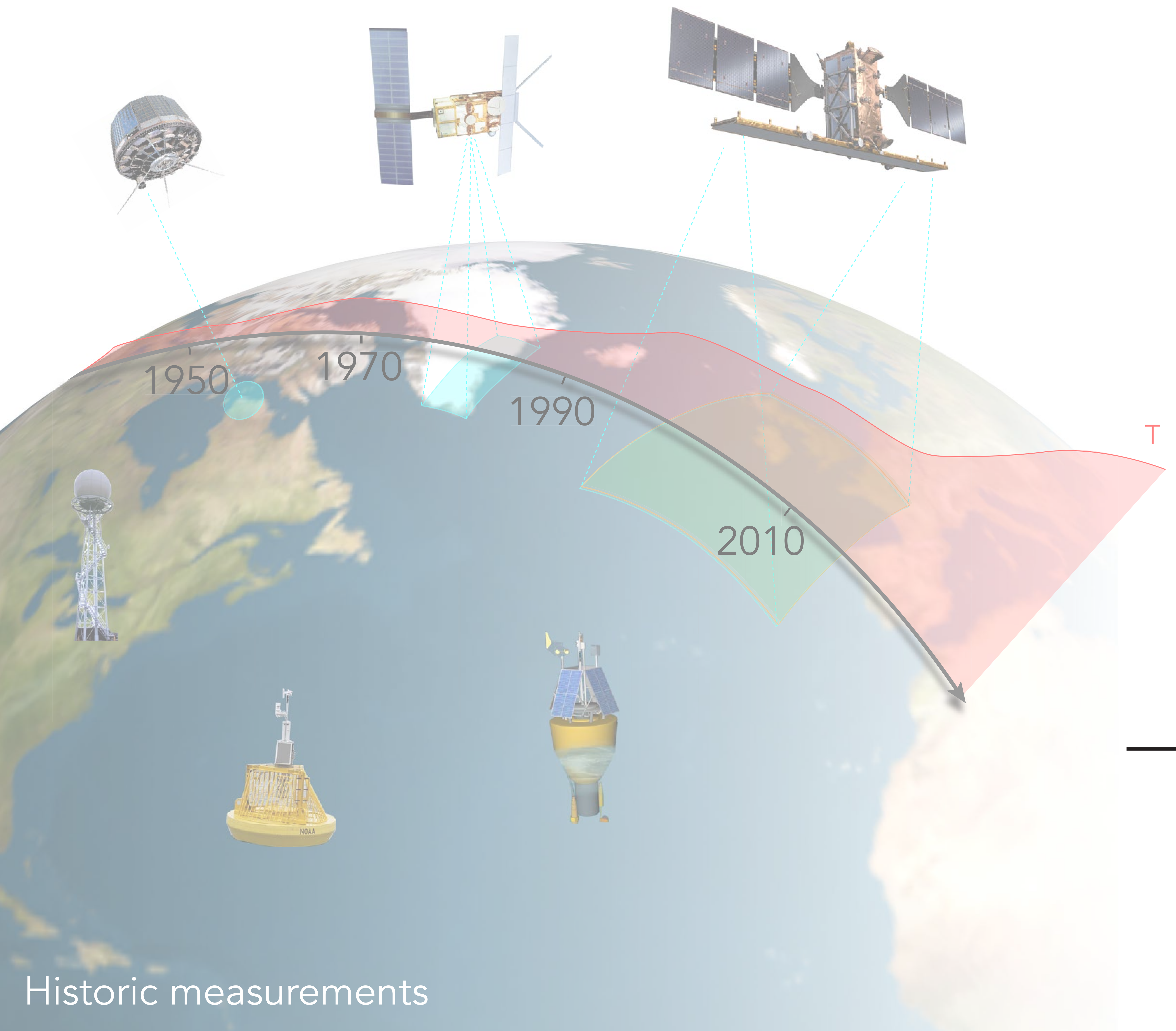


AtmoRep

Representation learning on atmospheric data to address climate change

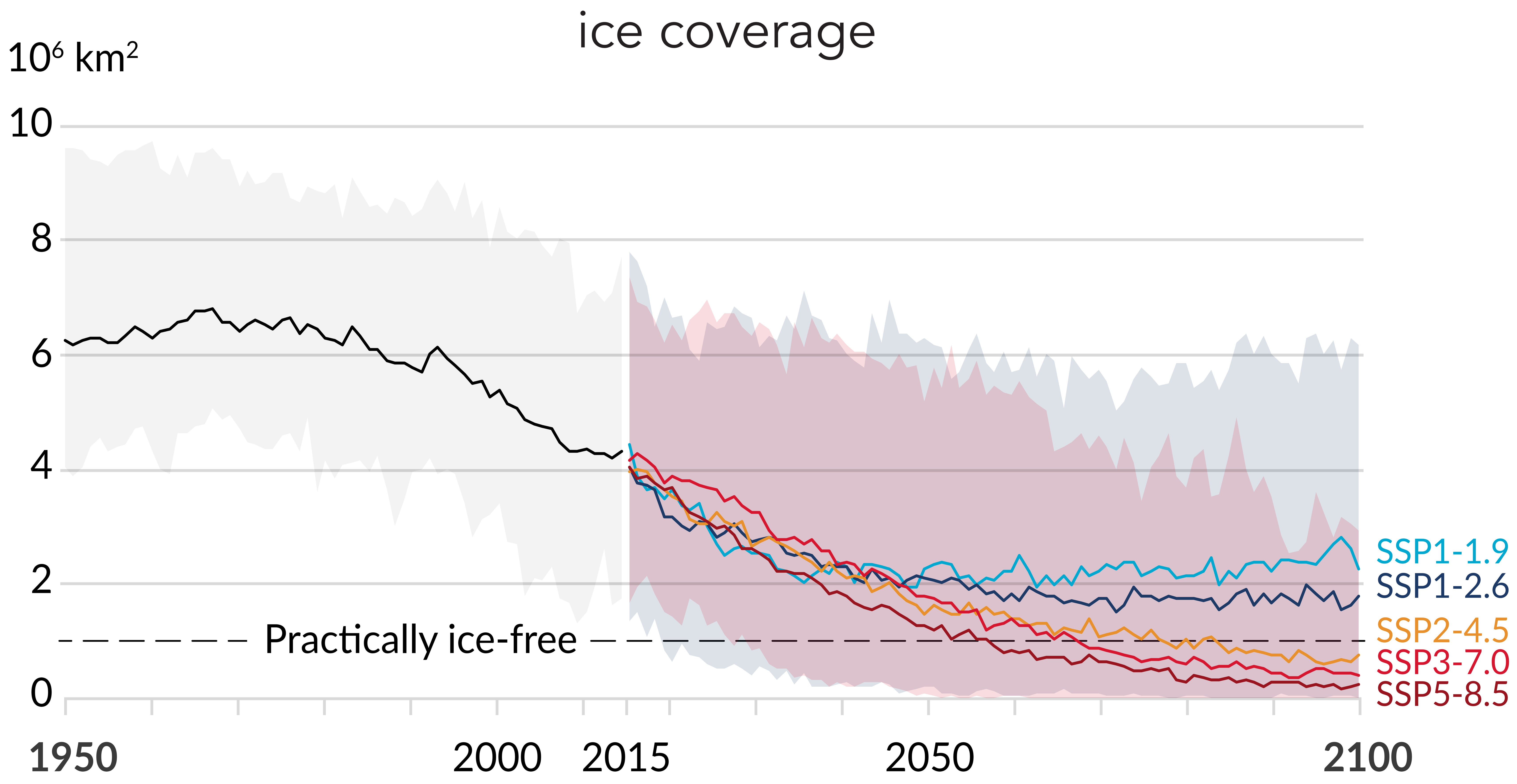
Ilaria Luise, Maike Sonnewald, Martin Schultz, Annesh Subramanian, Christan Lessig

Motivation

- Climate projections have very significant uncertainties

Motivation

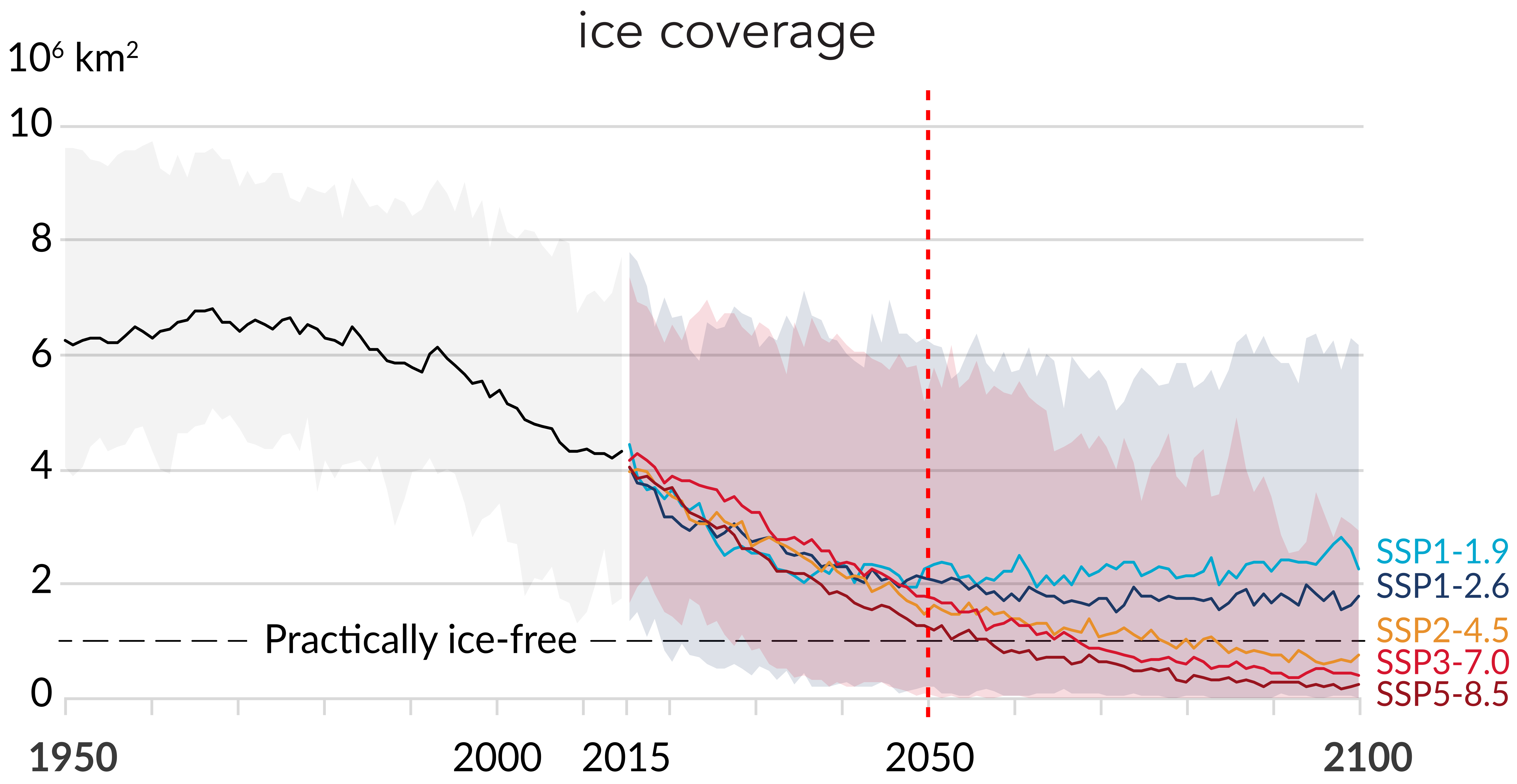
- Climate projections have very significant uncertainties



IPCC, 2021: Climate Change 2021: The Physical Science Basis. Masson-Delmotte, V., et al. Cambridge University Press. In Press.

Motivation

- Climate projections have very significant uncertainties



IPCC, 2021: Climate Change 2021: The Physical Science Basis. Masson-Delmotte, V., et al. Cambridge University Press. In Press.

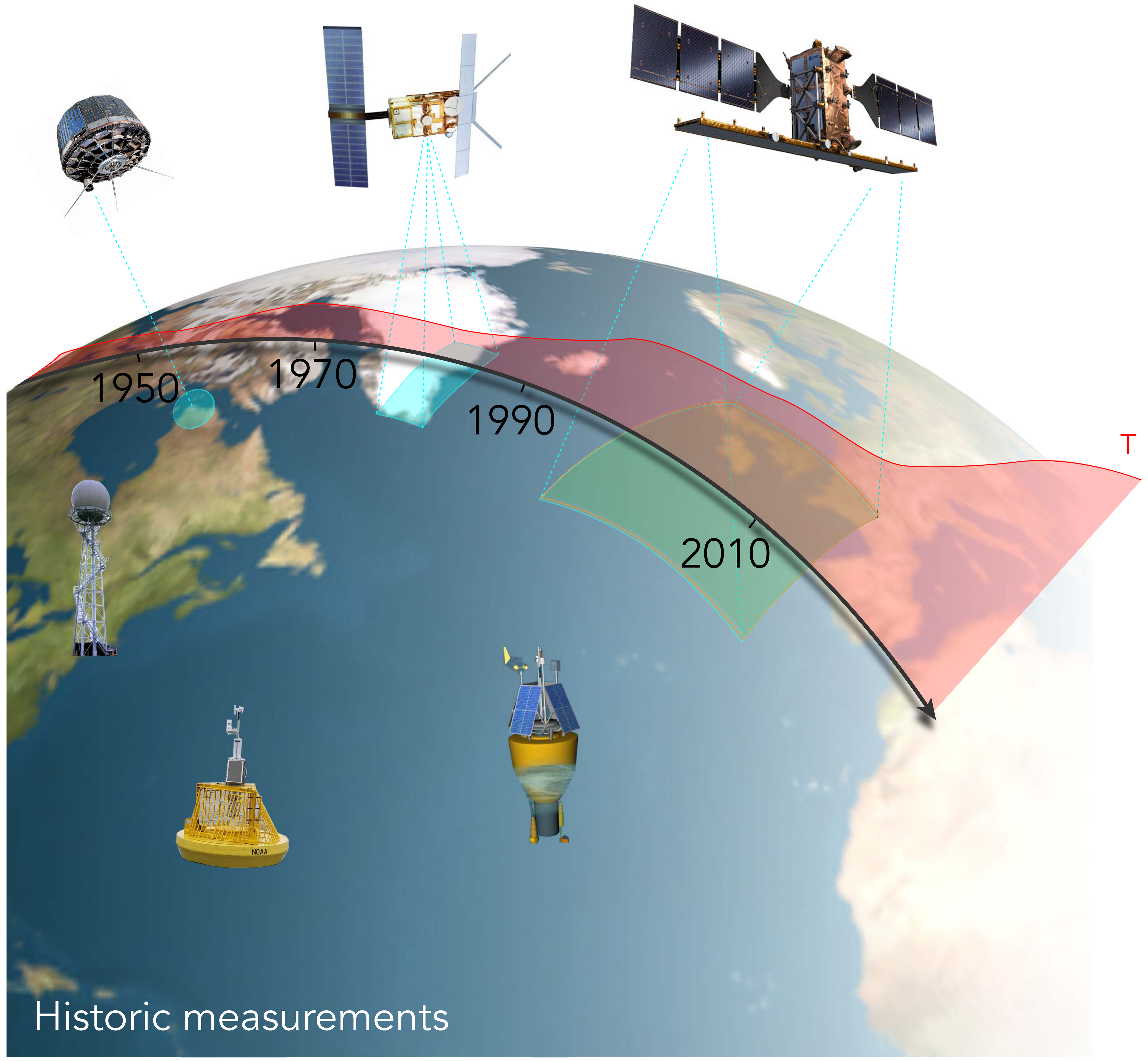
Motivation

- Climate projections have very significant uncertainties
- Only incomplete description of physical processes in the atmosphere in current simulations

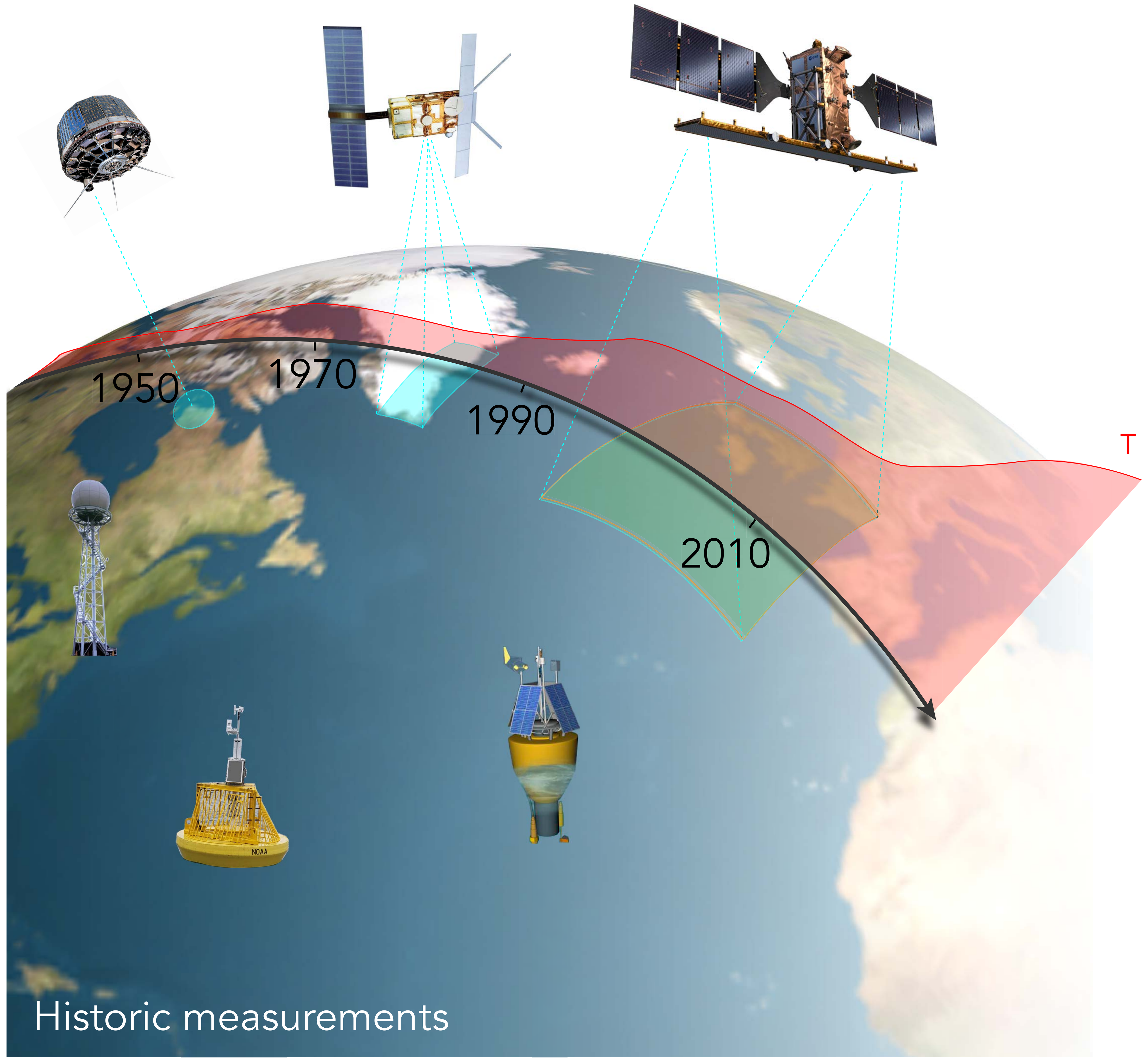
Motivation

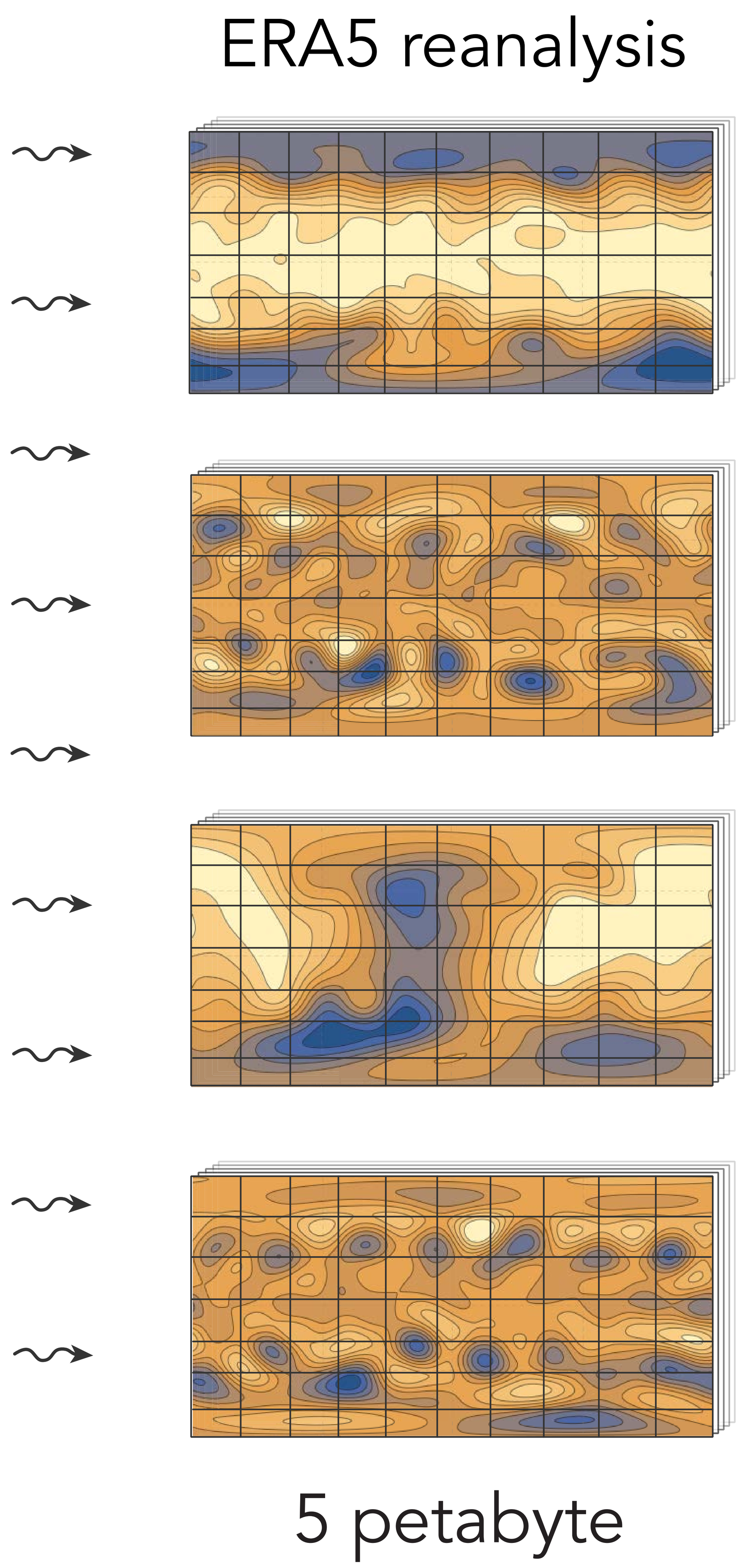
- Climate projections have very significant uncertainties
- Only incomplete description of physical processes in the atmosphere in current simulations
  - No (effective) models for cloud formation, interaction with biosphere, ...
  - Very large number of *interacting* scales (1 m to 10^7 m)
  - ...

Overview

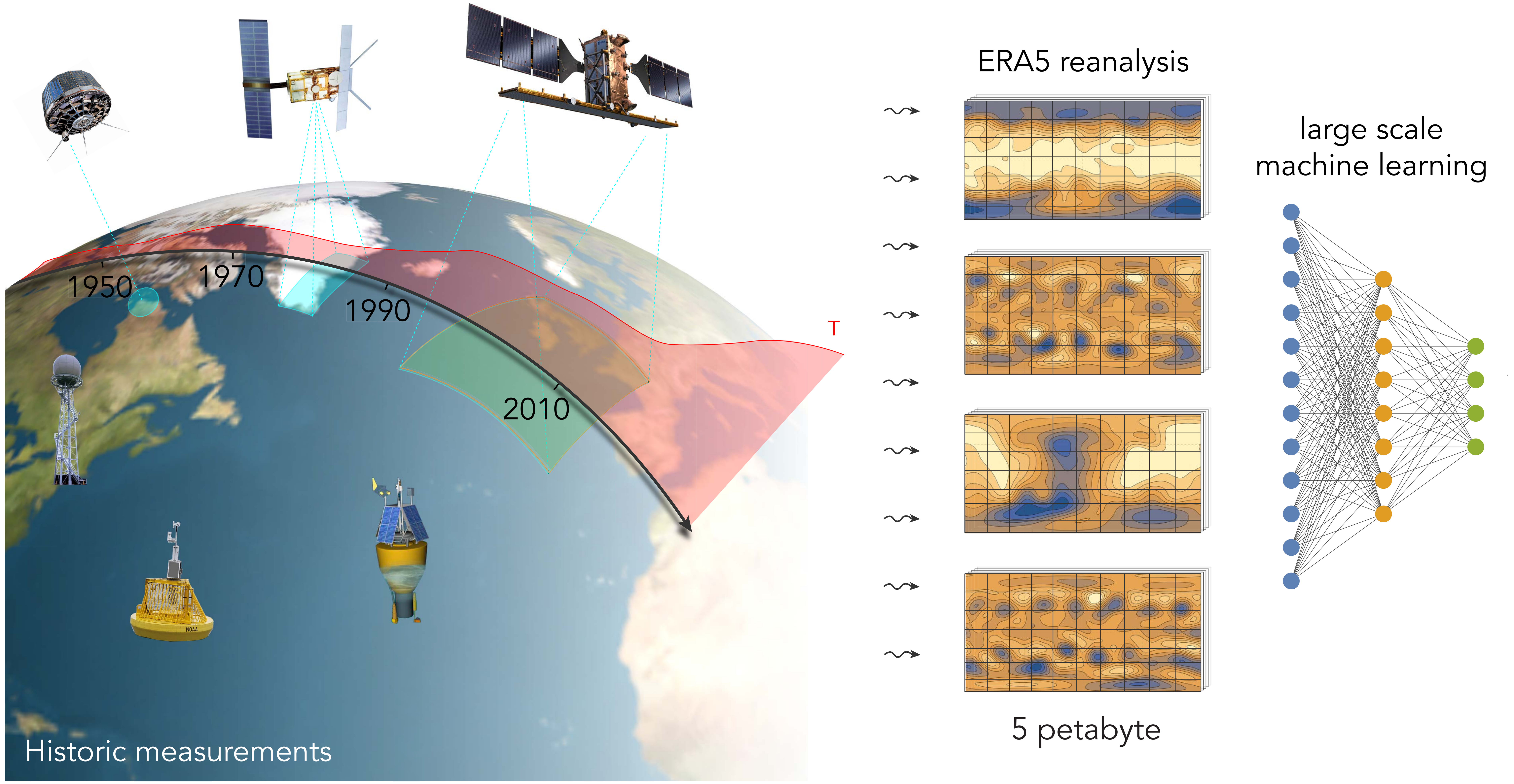


Overview

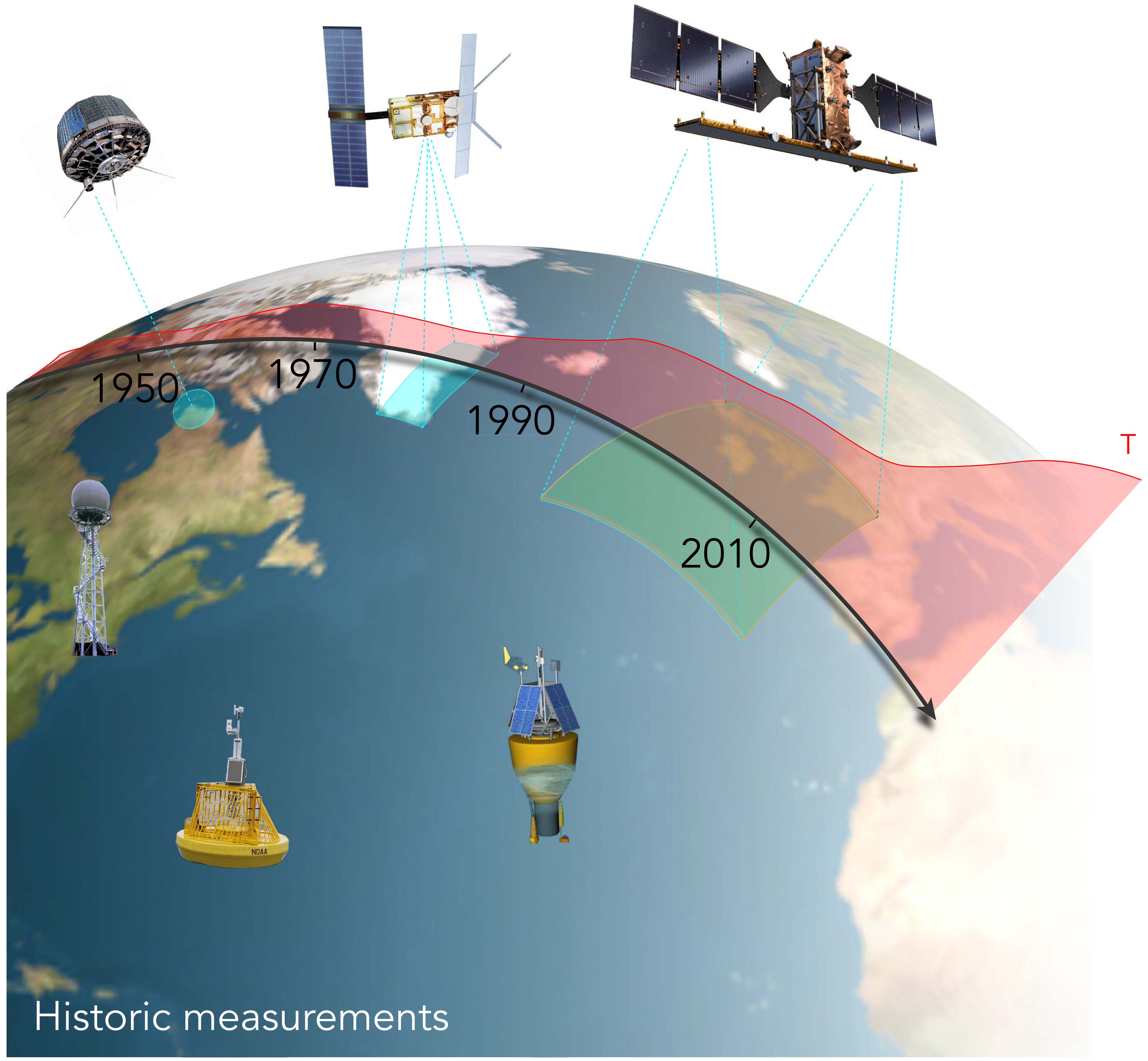

1950
1970
1990
2010
T
Historic measurements

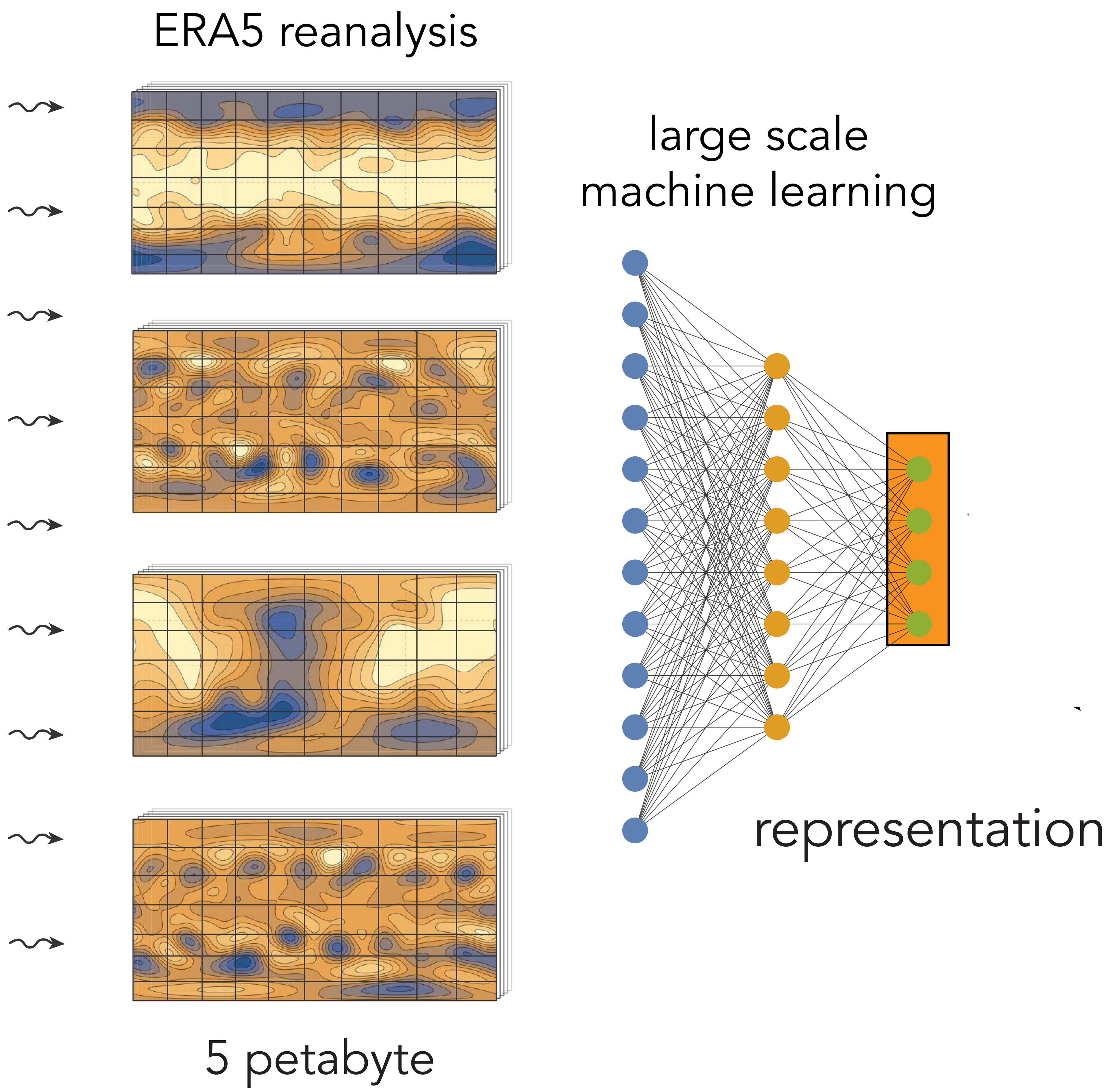

ERA5 reanalysis
5 petabyte

Overview


1950
1970
1990
2010
T
Historic measurements
ERA5 reanalysis
5 petabyte
large scale
machine learning

Overview


1950
1970
1990
2010
T
Historic measurements


ERA5 reanalysis
5 petabyte
large scale
machine learning
representation

Overview


1950
1970
1990
2010
T
Historic measurements
ERA5 reanalysis
5 petabyte
large scale
machine learning
representation
address climate change
Downscaling
Classification
Predictability
Missing data

Overview

1950
1970
1990
2010
T
Historic measurements
ERA5 reanalysis
5 petabyte
large scale
machine learning
representation
address climate change
Downscaling
Classification
Predictability
Missing data
scientific insight

Overview

AtmoRep
1950
1970
1990
2010
T
Historic measurements
ERA5 reanalysis
5 petabyte
large scale
machine learning
representation
address climate change
Downscaling
Classification
Predictability
Missing data
scientific insight

Why machine learning?

- Incomplete description of physical processes in the atmosphere
- Large amounts of historical atmospheric observations

=> Representation learning on observations

Why machine learning?

- Incomplete description of physical processes in the atmosphere
- Large amounts of historical atmospheric observations

=> Representation learning on observations

Impact

- Address climate change
  - Super-resolution, classification, predictability, …
  - Machine-learning corrected climate simulations
- Scientific insight
  - Finite space-time description of interaction across scales and phenomena in the atmosphere

AtmoRep@CERN

- CERN expertise of relevance:
  - Handling of large data sets / storage infrastructure
  - Distributed analysis
  - Machine learning
  - Uncertainty / error estimation
  - …

AtmoRep@CERN

- Collaboration with OpenLab (Sofia Vallecorsa) already underway
- Common interests and directions
  - Information extraction from large amounts of data
  - Uncertainty quantification in deep learning
  - Transformer neural networks
  - …

Timeline

AtmoDist: basic prototype

2022

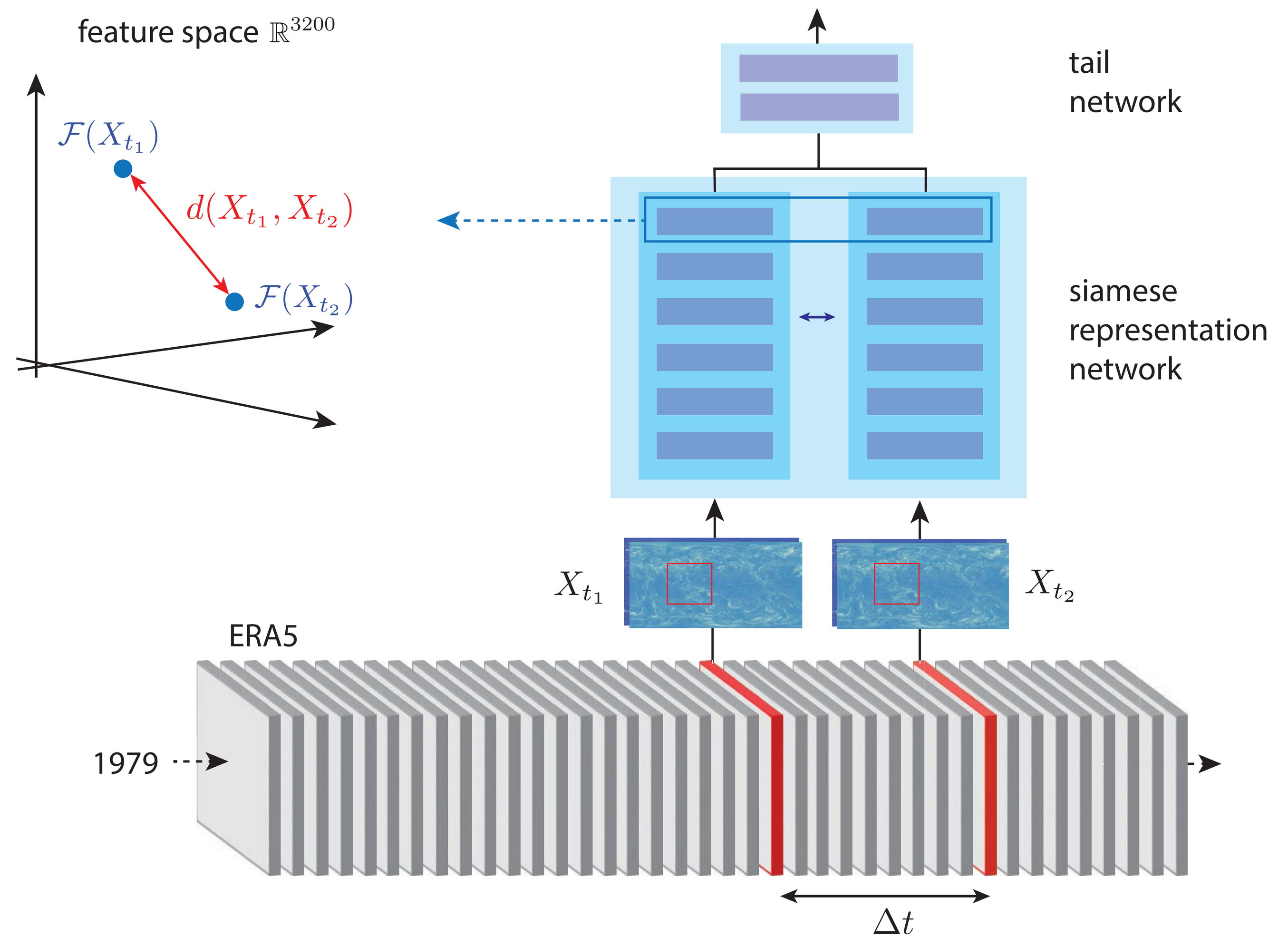
First training and scaling experiments on supercomputer

Proof-of-concept for spatio-temporal attention maps

today

Compute time on Jülich supercomputer

2023



Timeline

AtmoDist: basic prototype

2022

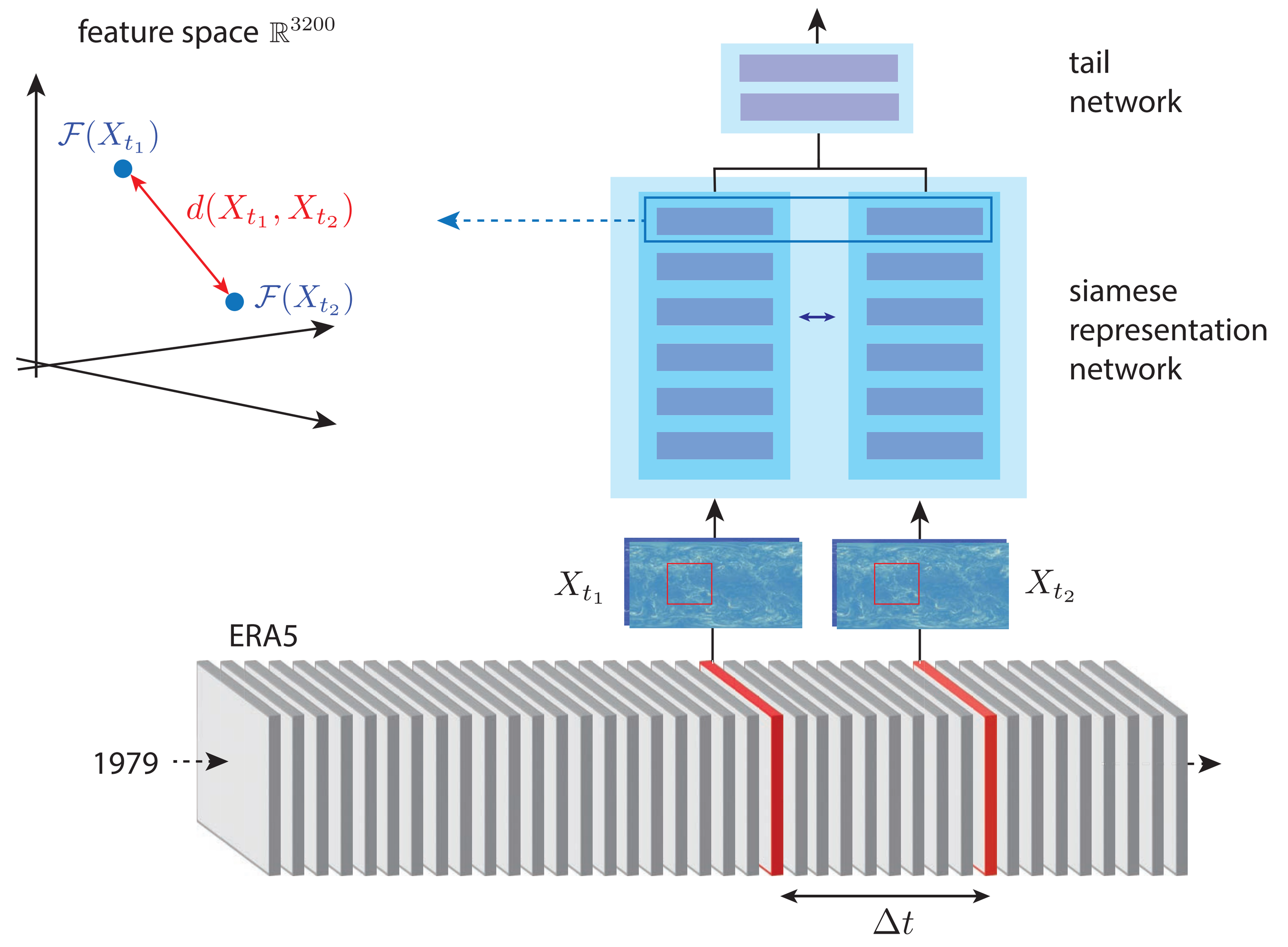
First training and scaling experiments on supercomputer

Proof-of-concept for spatio-temporal attention maps

today — Compute time on Jülich supercomputer

Finalize neural network architecture and training

2023



Timeline

AtmoDist: basic prototype

2022

First training and scaling experiments on supercomputer

Proof-of-concept for spatio-temporal attention maps

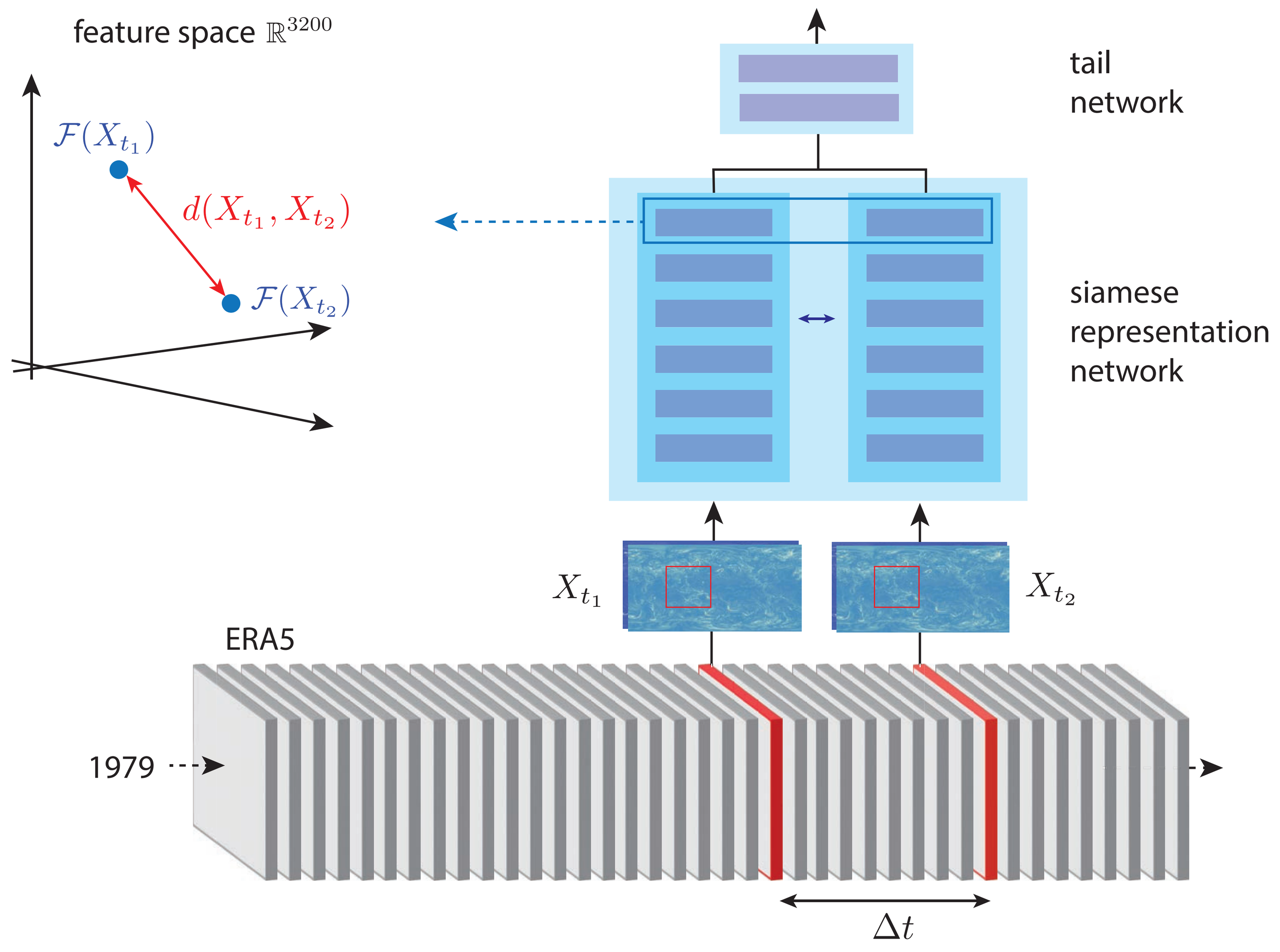
today — Compute time on Jülich supercomputer

Finalize neural network architecture and training

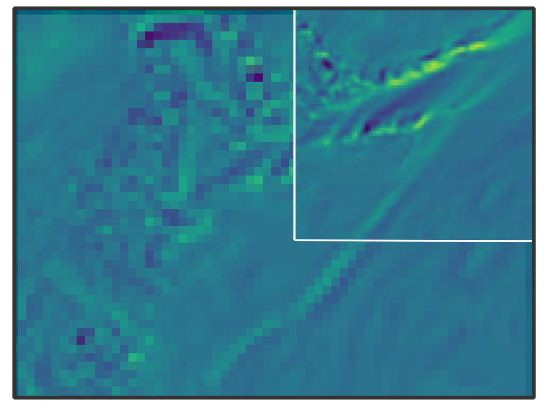
Parallel scaling and training on O(100) TB

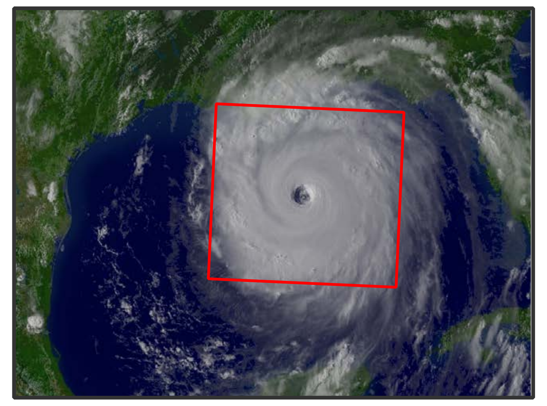
Implementation of demonstration applications

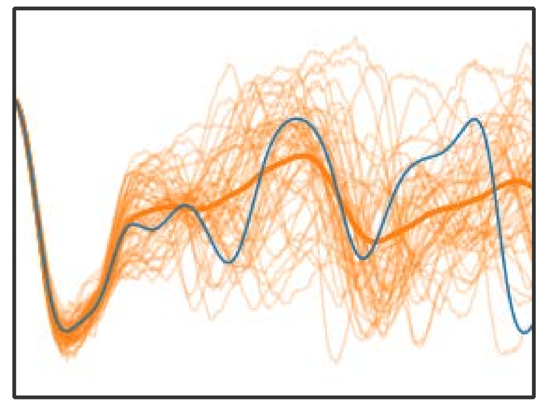
2023

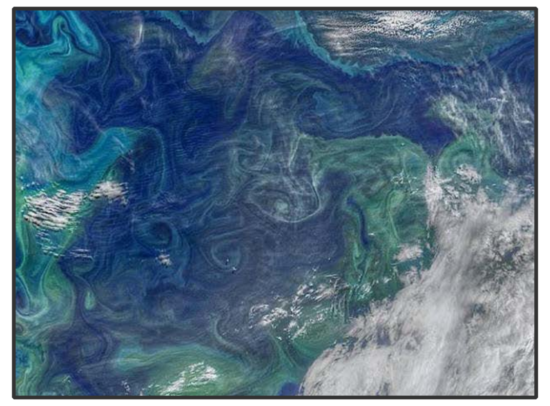


address climate change


Downscaling


Classification


Predictability


Missing data

Timeline

2022

- AtmoDist: basic prototype
- First training and scaling experiments on supercomputer
- Proof-of-concept for spatio-temporal attention maps

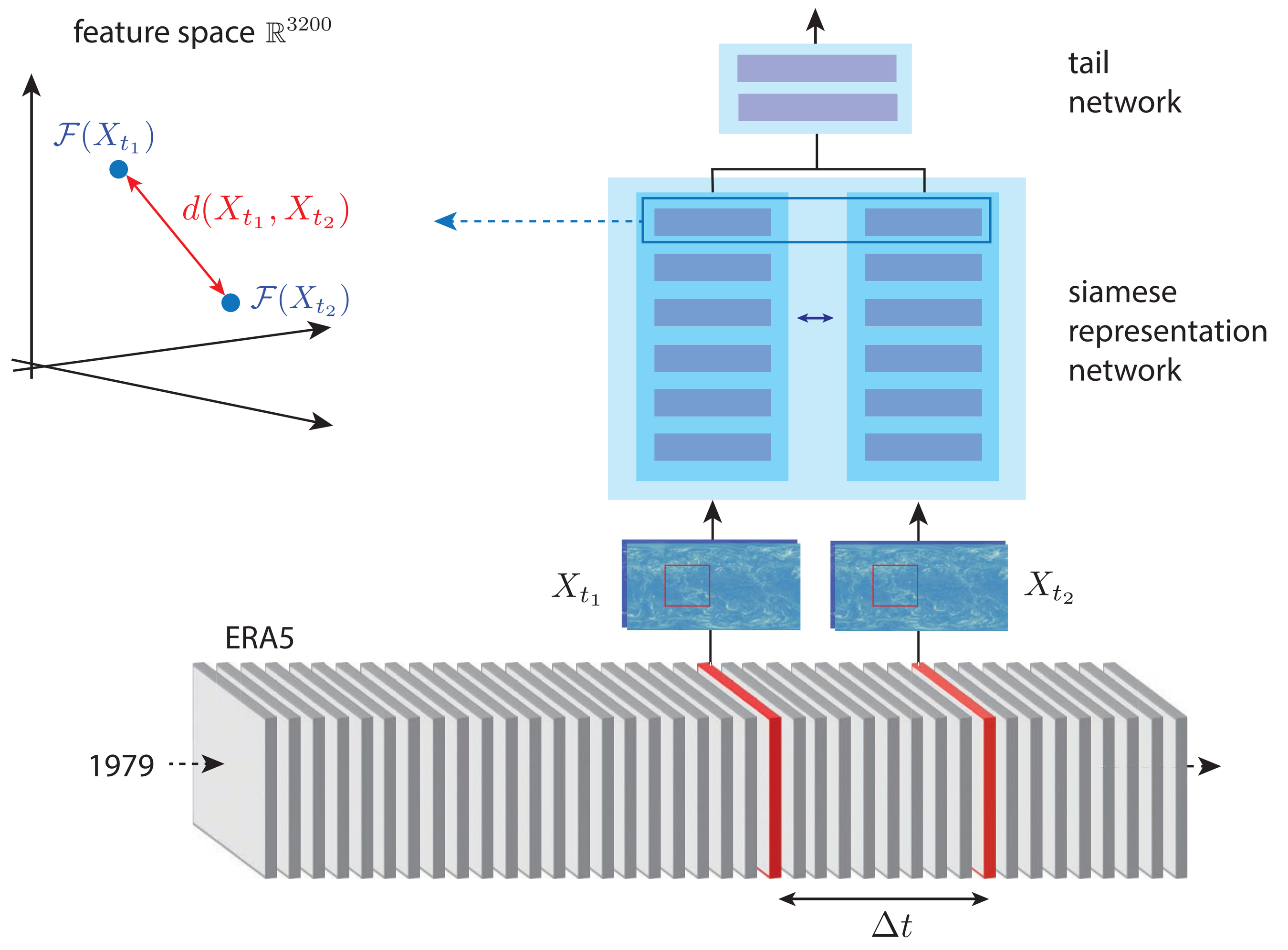
today

- Compute time on Jülich supercomputer

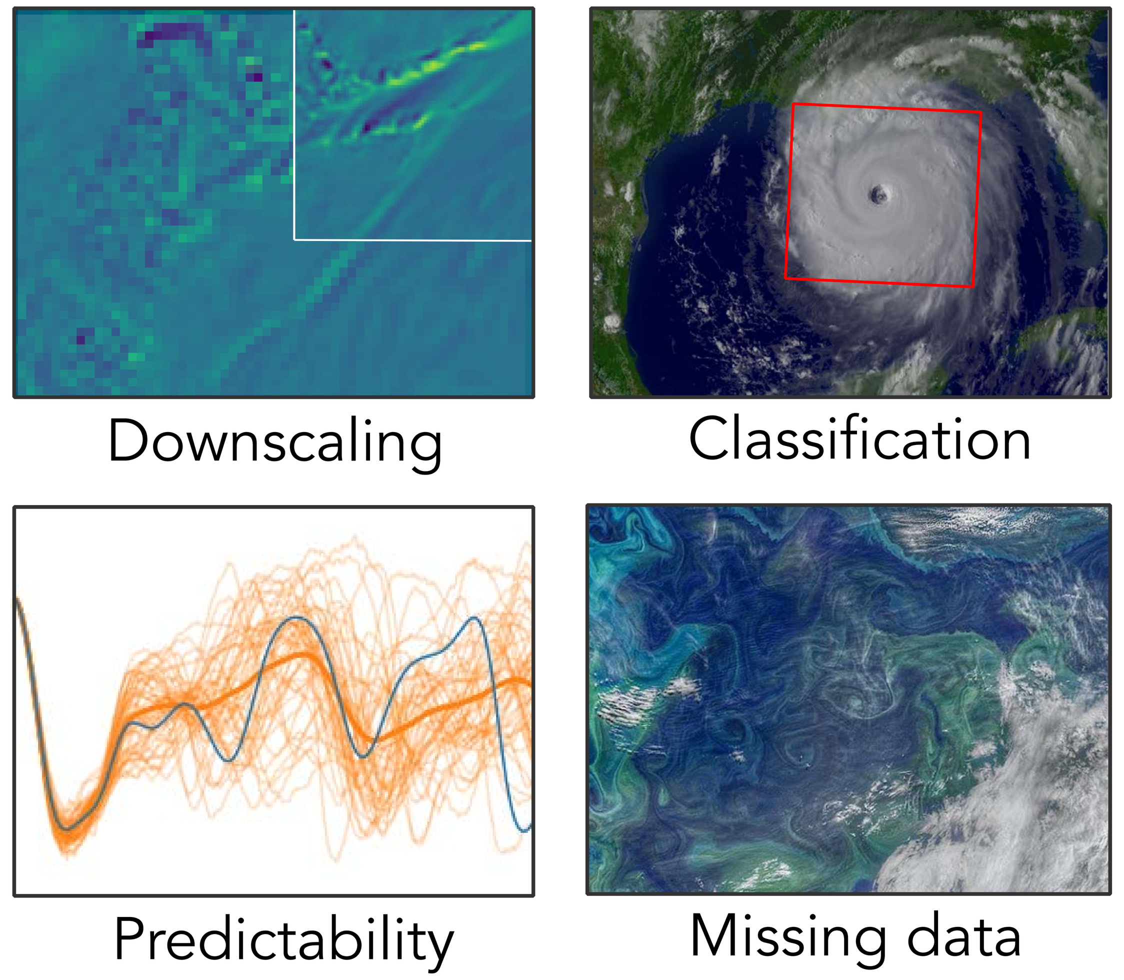
- Finalize neural network architecture and training
- Parallel scaling and training on O(100) TB
- Implementation of demonstration applications

2023

- Finalize applications and publications



address climate change



Timeline

2022

- AtmoDist: basic prototype
- First training and scaling experiments on supercomputer
- Proof-of-concept for spatio-temporal attention maps

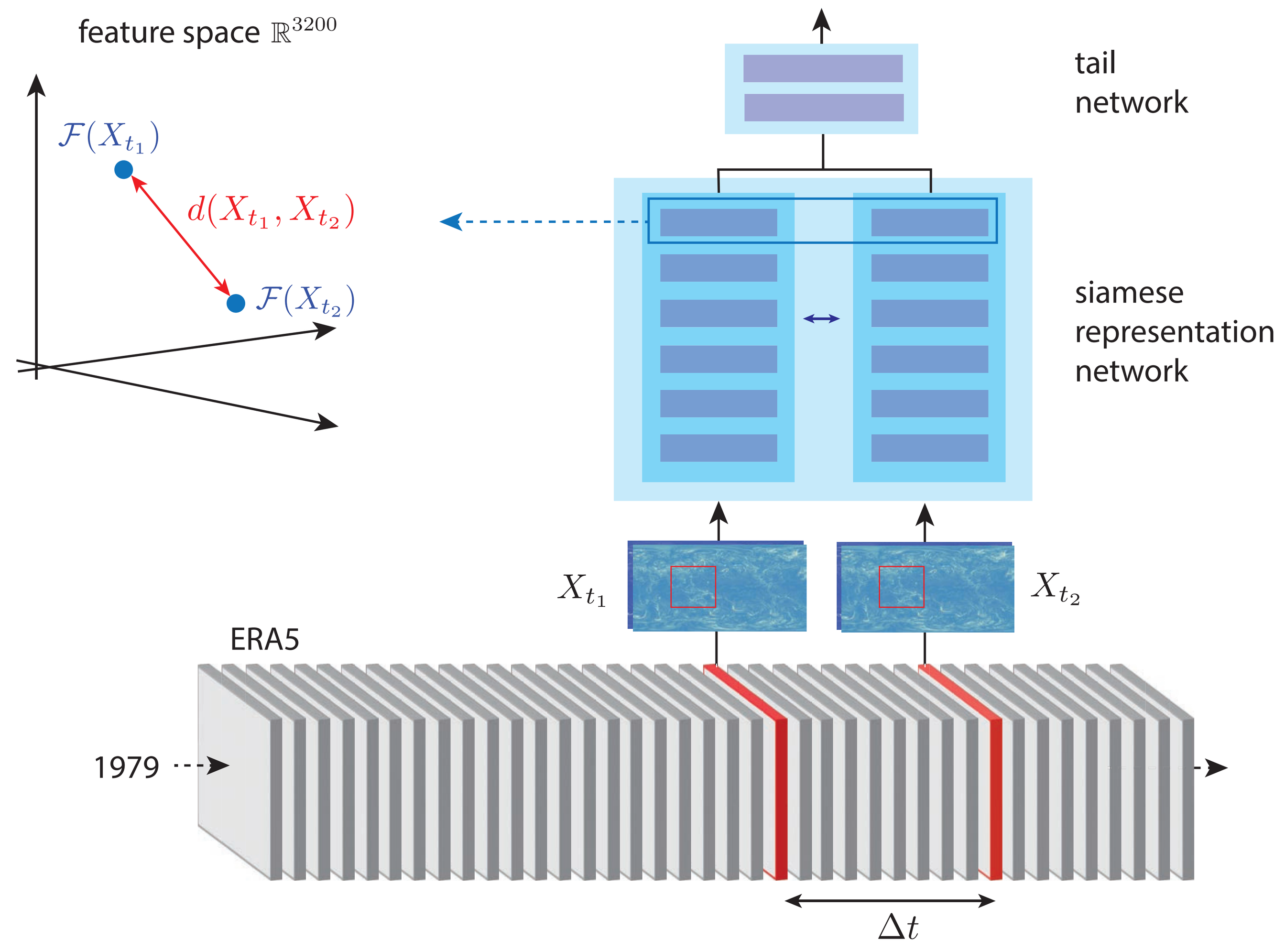
today

- Compute time on Jülich supercomputer

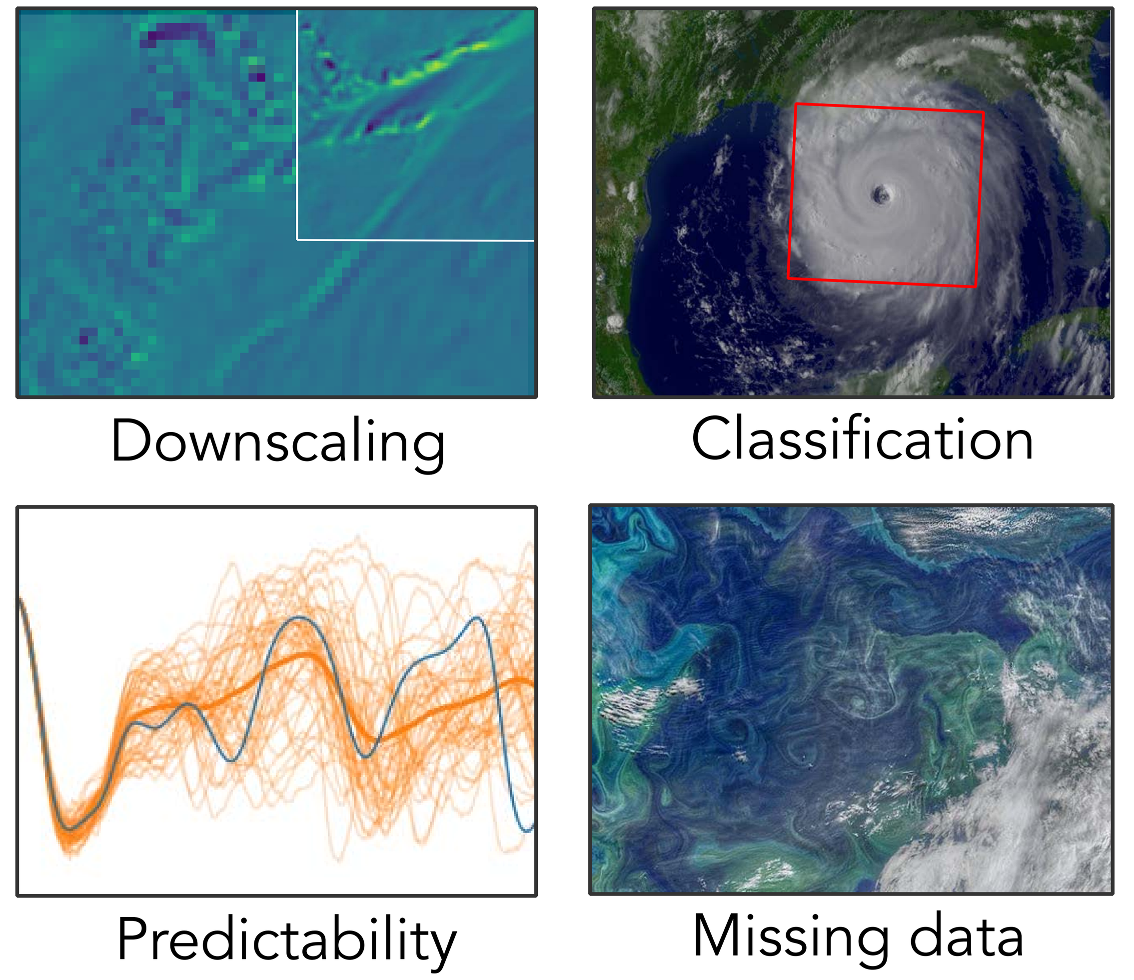
- Finalize neural network architecture and training
- Parallel scaling and training on O(100) TB
- Implementation of demonstration applications

2023

- Finalize applications and publications
- Public availability of pretrained neural network



address climate change



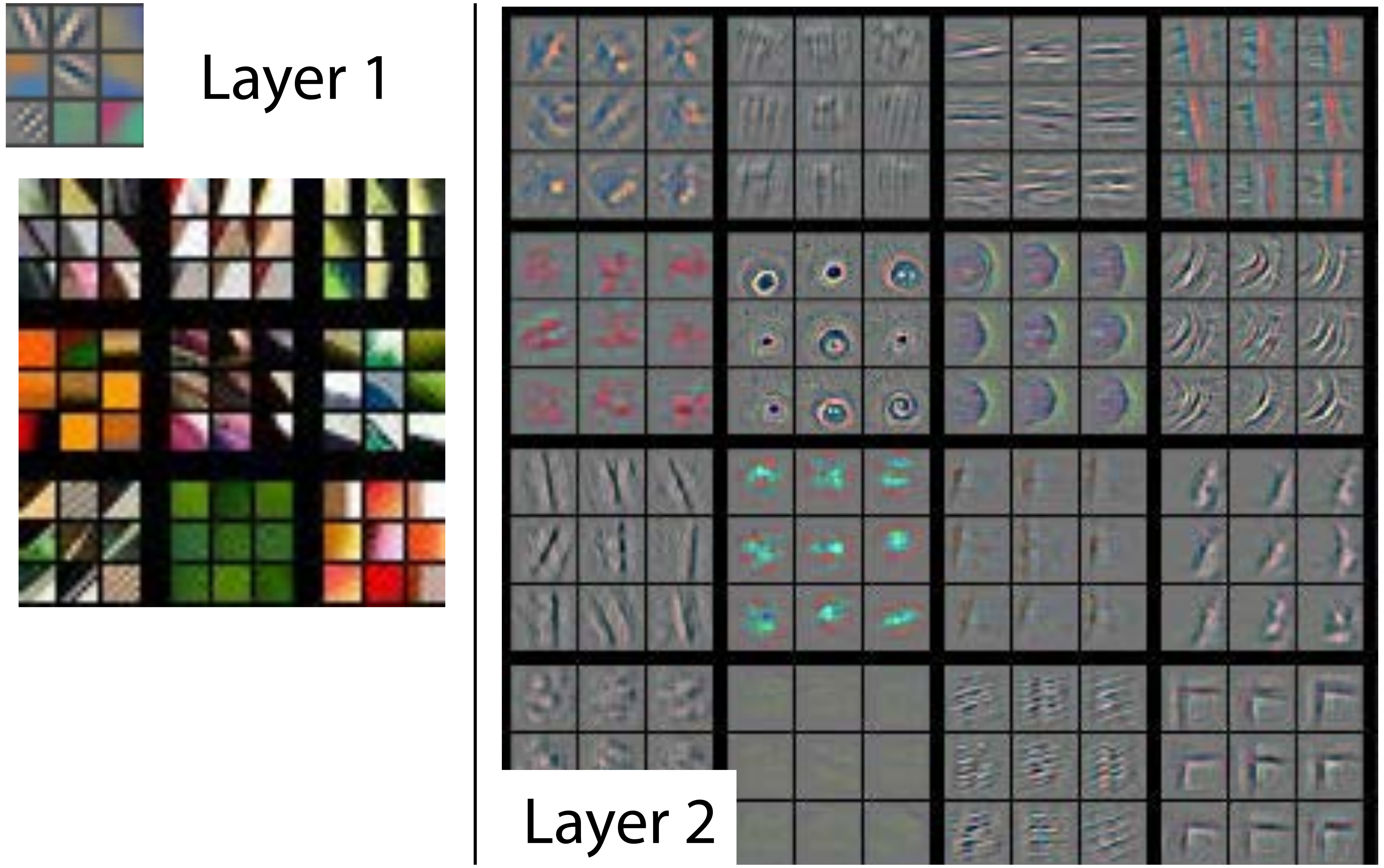
AtmoRep

- Use large amounts of historical observations to improve climate projections and related applications
- Significant potential impact through various applications
- Scientifically interesting and challenging

Representation learning

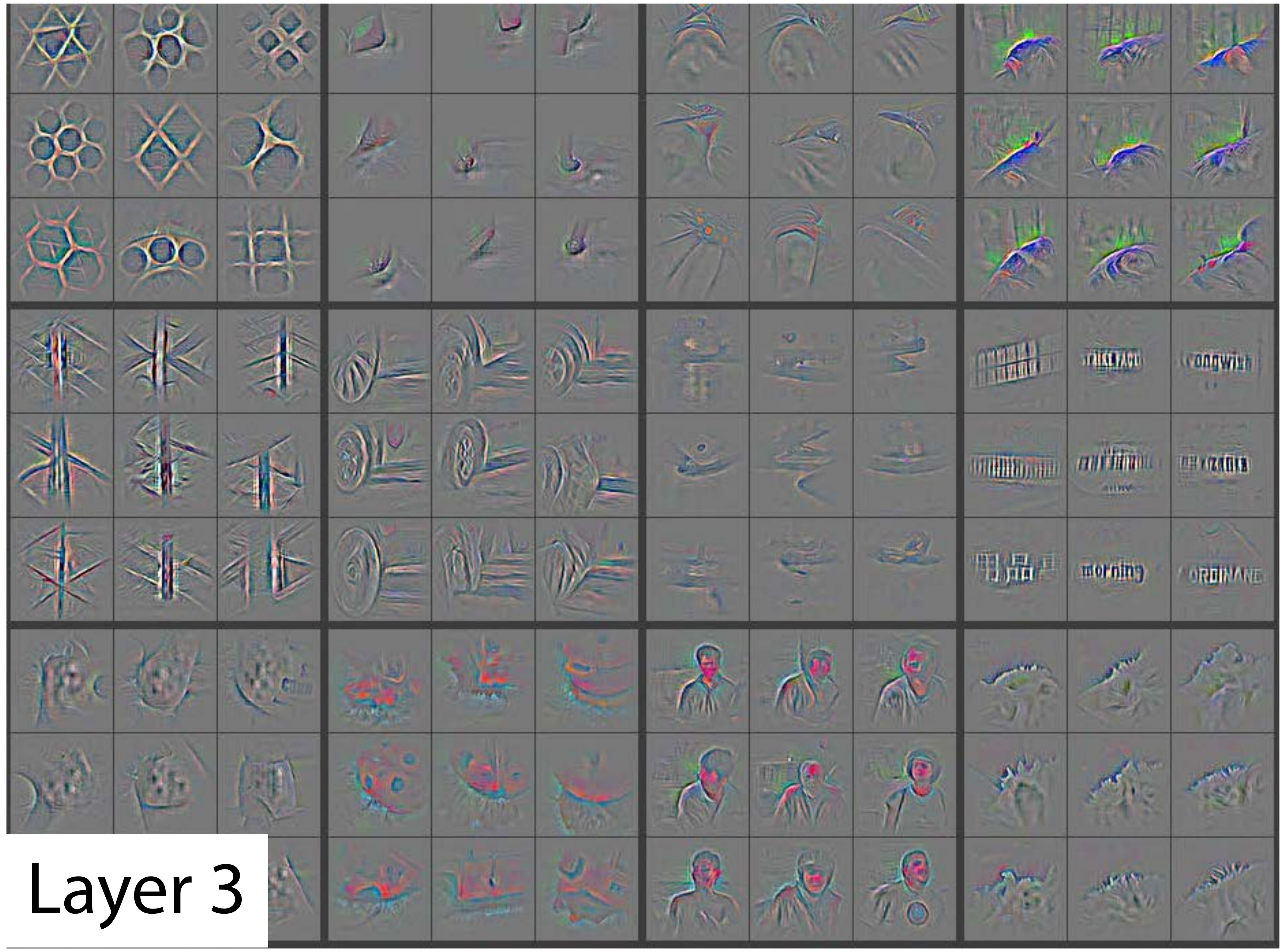
- What is a representation?
  - Layers of neural network are nonlinear maps of input data to intermediate representation (elements in a vector space)
  - For a trained neural network the intermediate layers hence provide transformed data adapted to a domain
  - Often the same intermediate representations are useful for a range of tasks

Representation learning



From M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, Computer Vision – ECCV 2014, pages 818–833, Cham, 2014. Springer International Publishing.

Representation learning

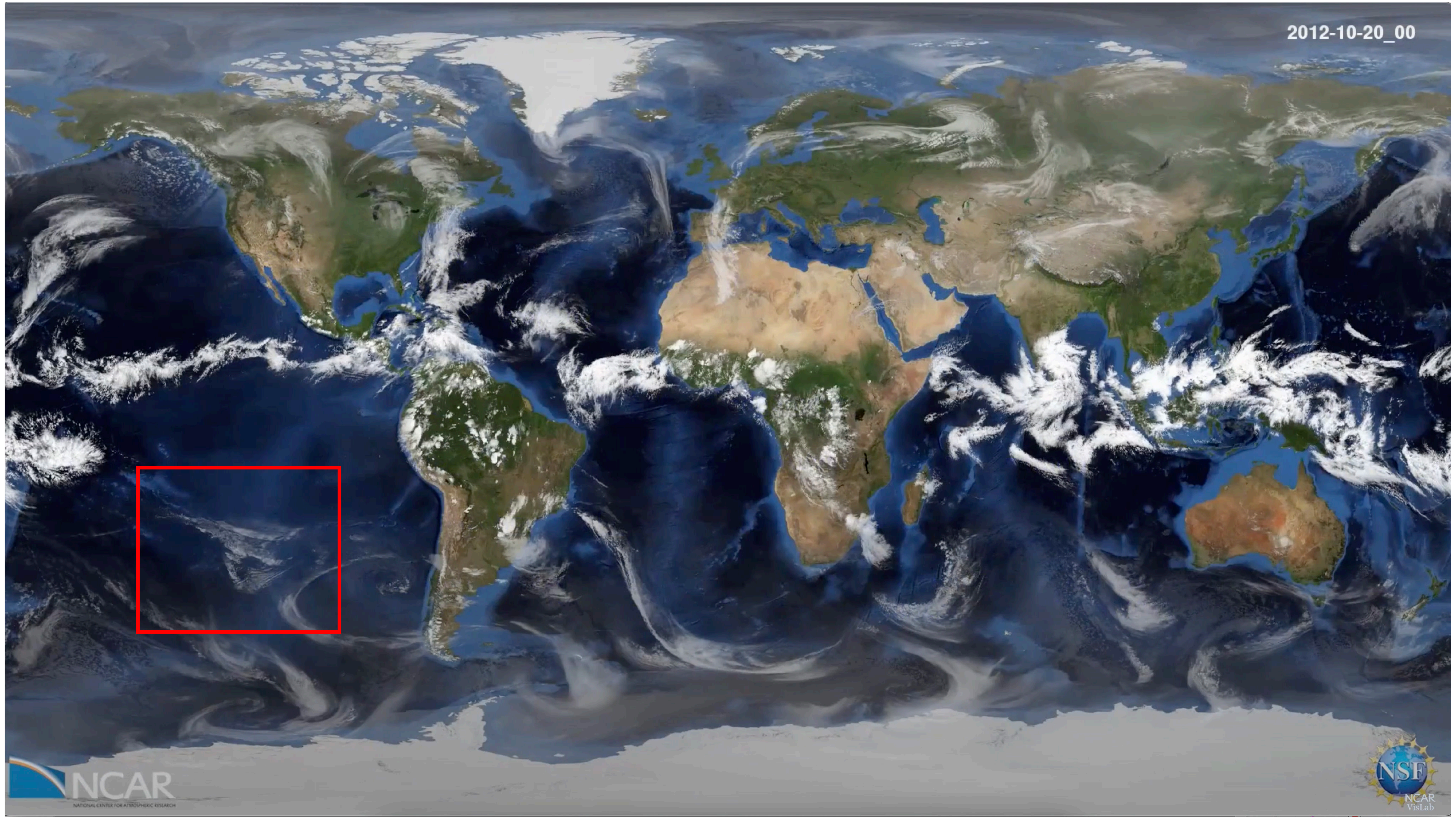


From M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, Computer Vision – ECCV 2014, pages 818–833, Cham, 2014. Springer International Publishing.

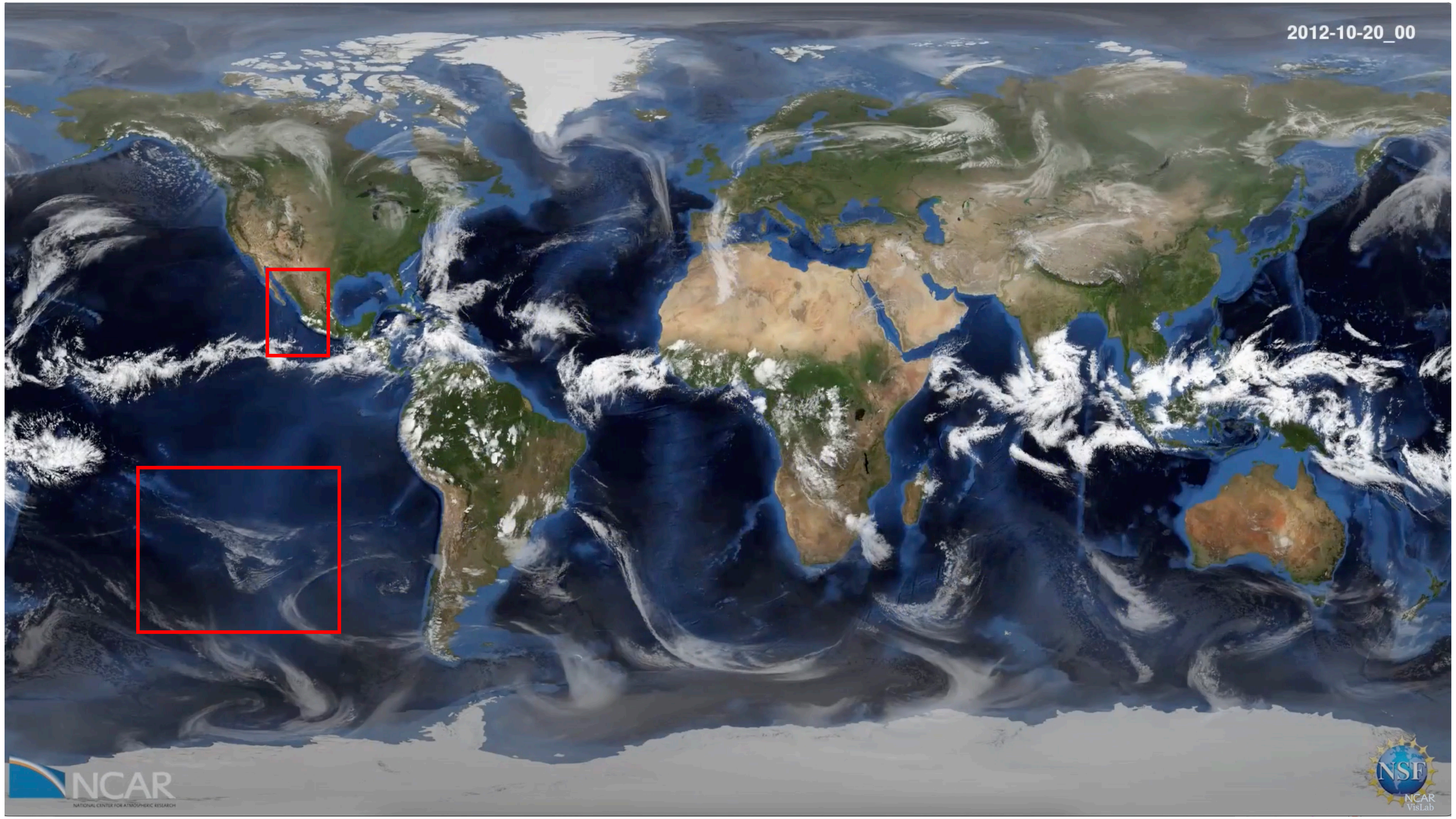
Scientific insight

- Spatio-temporal representation learning:
  - Learn representation that describe finite time spatio-temporal interactions across scales

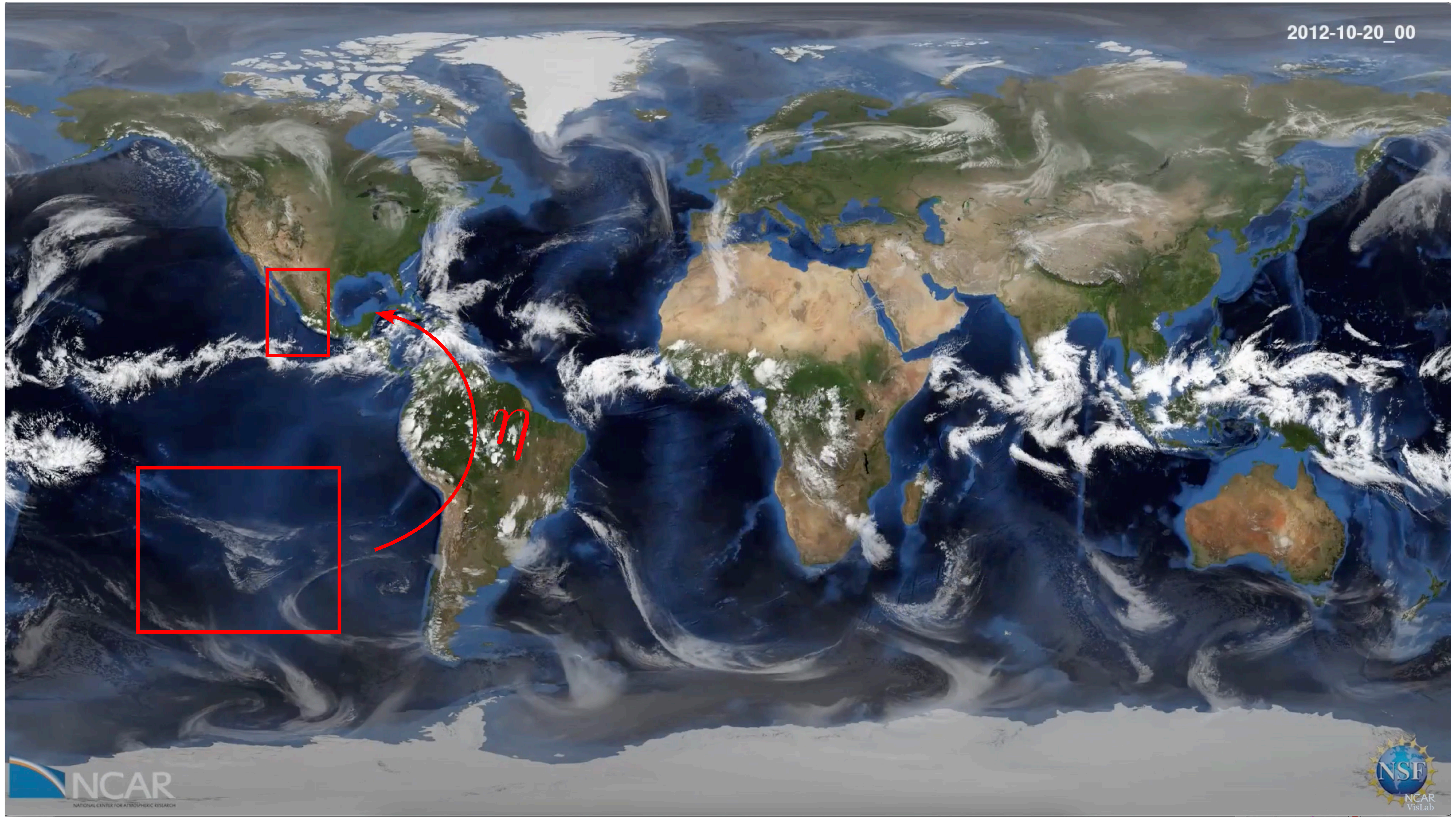
Scientific insight



Scientific insight



Scientific insight



Scientific insight

- Spatio-temporal representation learning:
  - Learn representation that describe finite time spatio-temporal interactions across scales
  - Capture interactions that are difficult to describe with classical approaches
  - Transformer neural networks allow for simple interpretability

Why machine learning?

- Only incomplete description of physical processes in the atmosphere
  - No (effective) models for cloud formation, interaction with biosphere, ...
  - Most models only provide infinitesimal information
  - Very large number of *interacting* scales (1 m to 10^8 m)

=> Machine learning based on observations to capture phenomena and interactions not well described so far