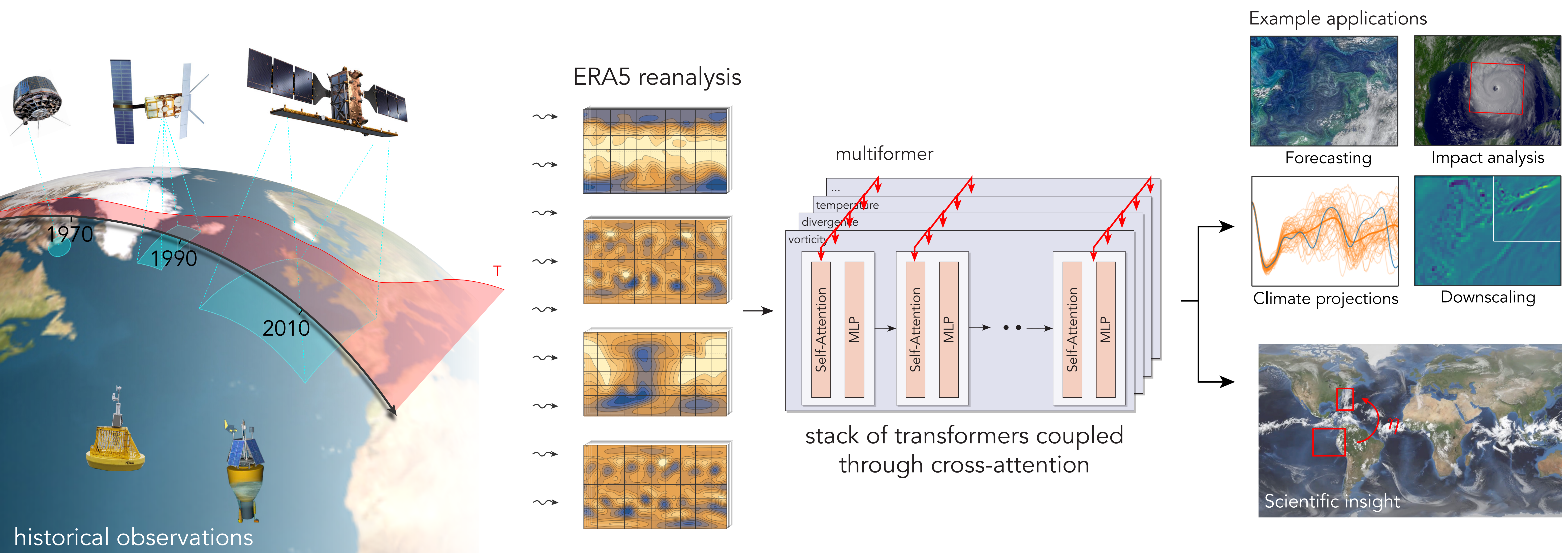


AtmoRep Large Scale Representation Learning of Atmospheric Dynamics

christian.lessig@ovgu.de

Christian Lessig (University of Magdeburg), Ilaria Luise (CERN), Martin Schultz (Jülich Supercomputing Center), et al.



Can we train **one** neural network model that encapsulates all Earth system dynamics by self-supervised training on large amounts of spatio-temporal observations?

Motivation

- Availability of petabytes of unlabelled observational and quasi-observational data
 - Data contains critical information, e.g. about unresolved process and their feedbacks to coarser scales
- Self-supervised, large scale representation learning allows one to make use of this data and amortizes training costs
 - Methodology has led to breakthroughs in natural language processing and computer vision (e.g. GPT-3)

Benefits

- Pre-trained network can be used with small computational costs for a wide range of applications
 - Highly compact representation of ERA5 with O(GB)
 - Better performance than directly training for application
 - Amortize training costs on very large data sets
 - Weather forecasting, climate projections, downscaling, ...
- Possible new scientific insights by accessing the spatio-temporal interactions encoded in the network (e.g. attention)

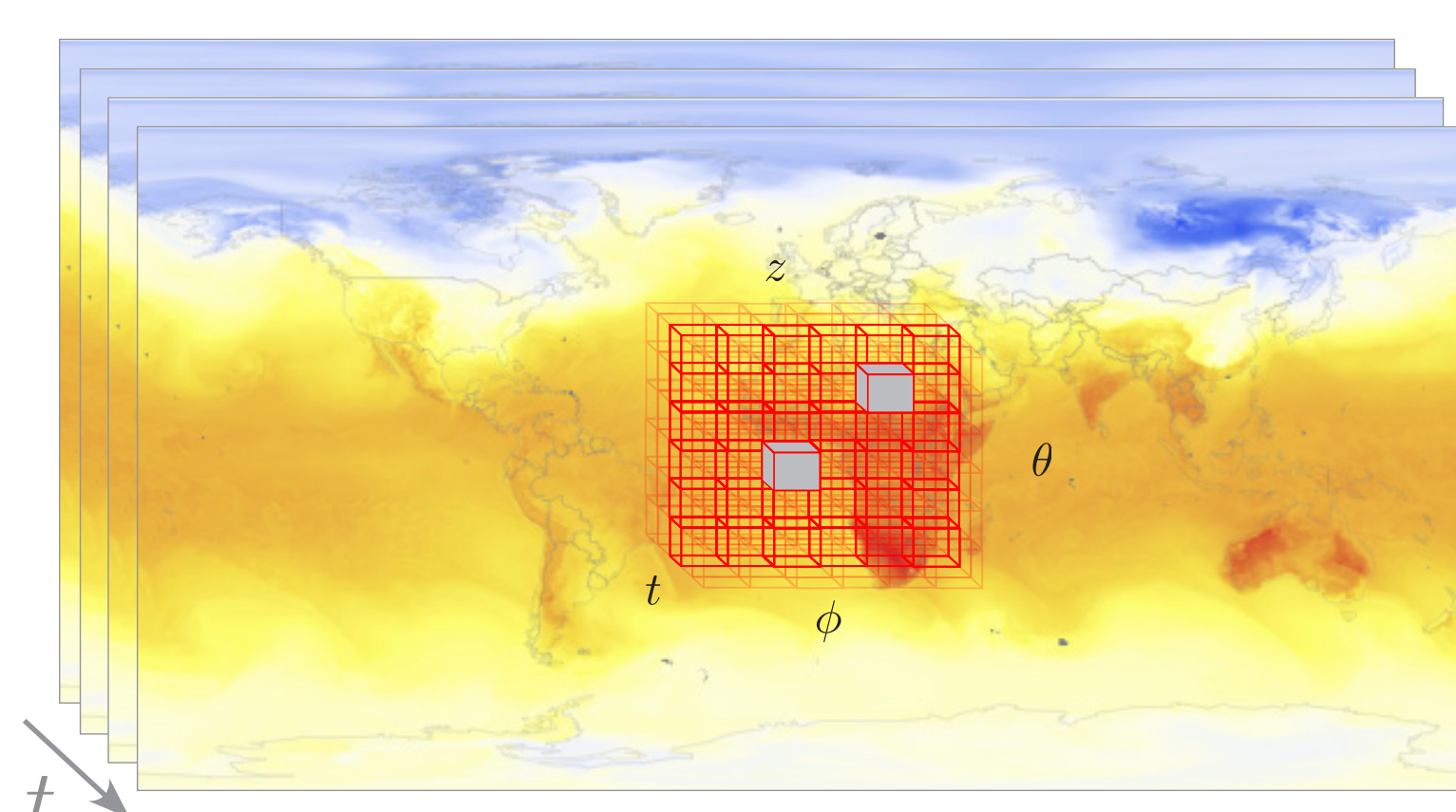
Proof-of-concept: train one transformer neural network on O(PB) of ERA5 reanalysis data

Multiformer

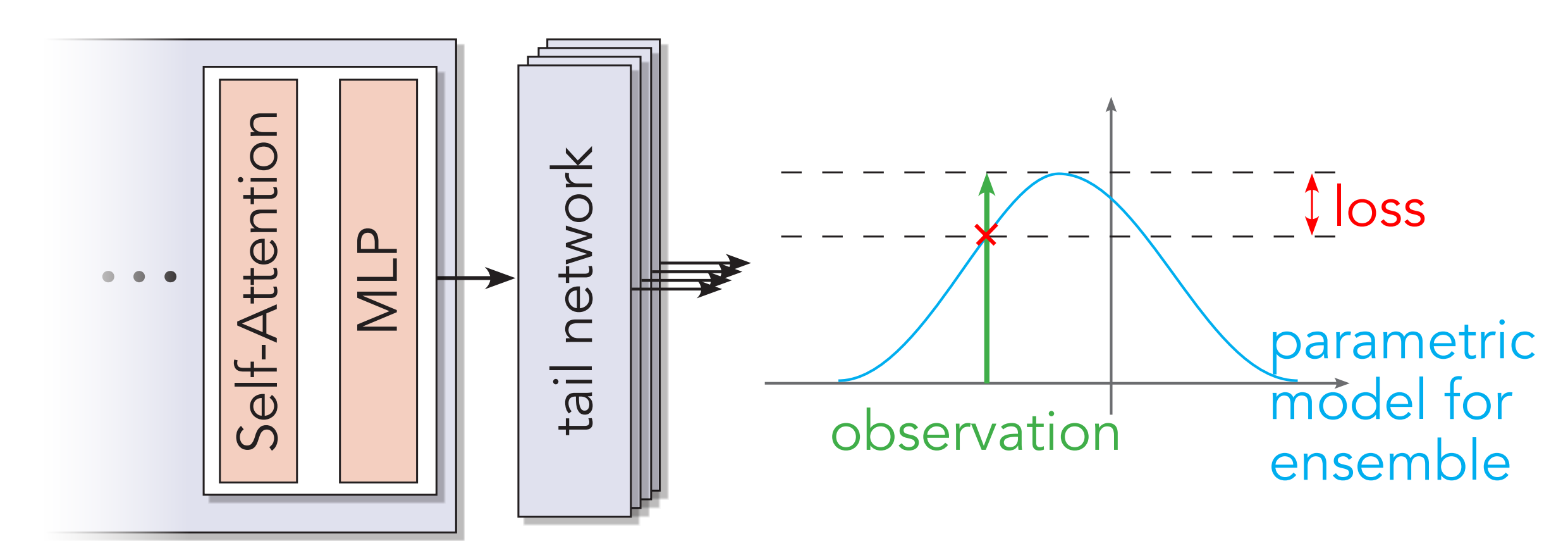
- Transformer-based architecture
 - Scales well to very large datasets
 - Local network applied to neighbourhood in space-time
- One transformer per physical field and possibly vertical level
 - Respects different properties of fields, coupling through attention
 - Fields can easily be added and removed

Training

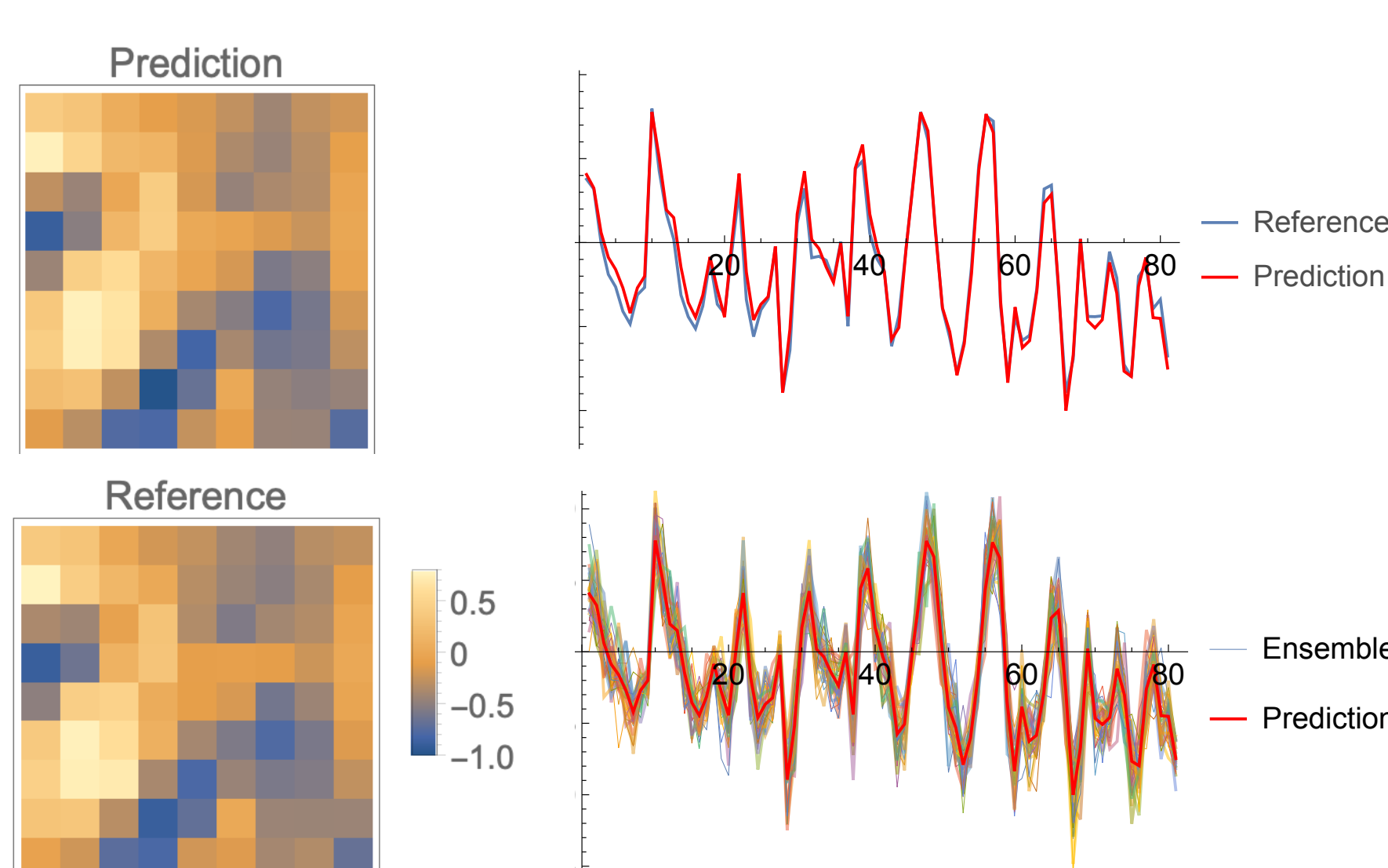
Self-supervised training with spatio-temporal extension of BERT masked language model:



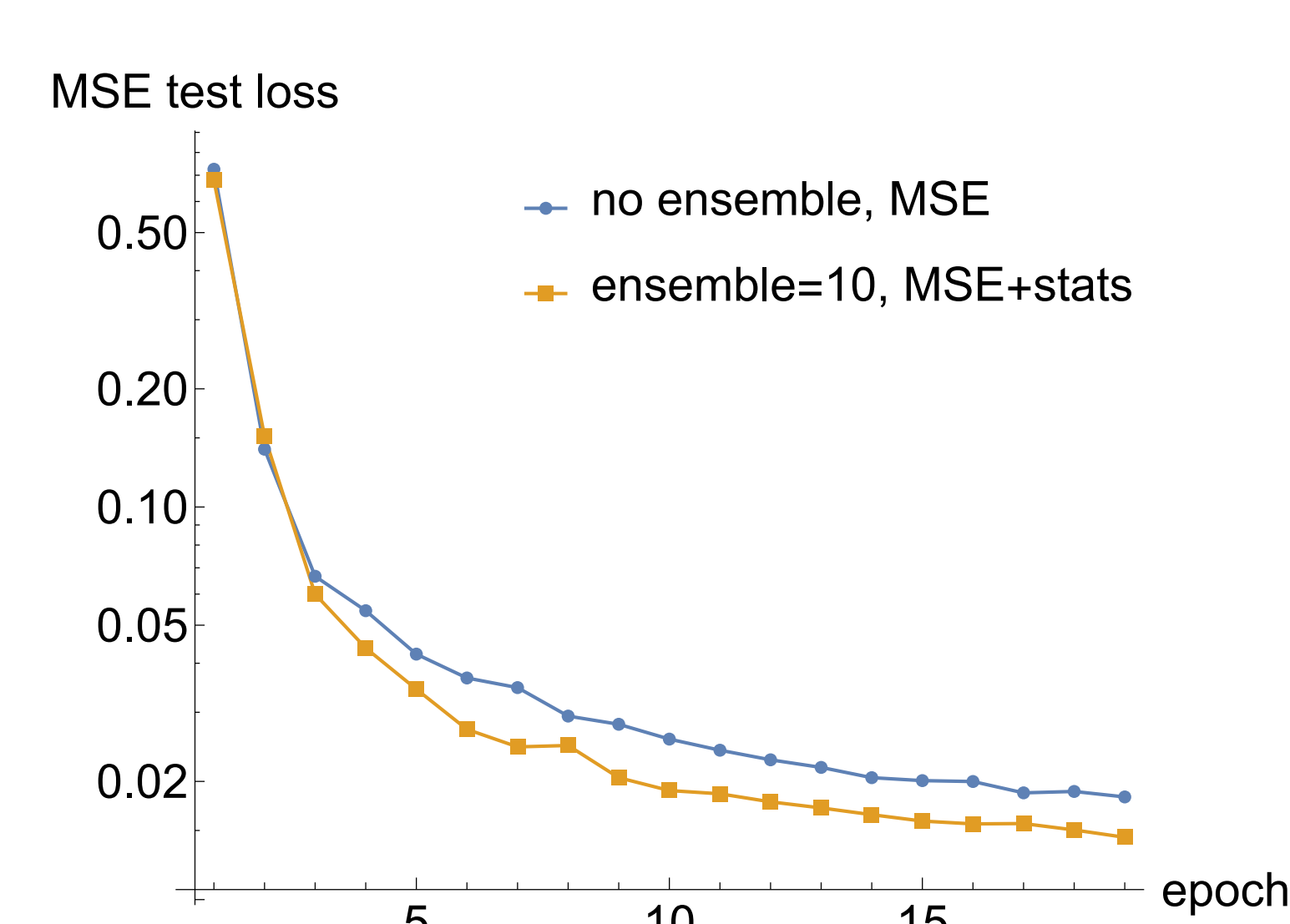
Training with ensemble of tail networks to learn statistical representation of quasi-chaotic atmospheric dynamics and improve training behaviour:



Example for predictions



Ensemble loss



Zero-shot forecasting performance

