Funded by the European Union

Coordinated by ECMWF

Weather Generator

Image modified from: link

The WeatherGenerator

A foundation model approach for Weather and Climate

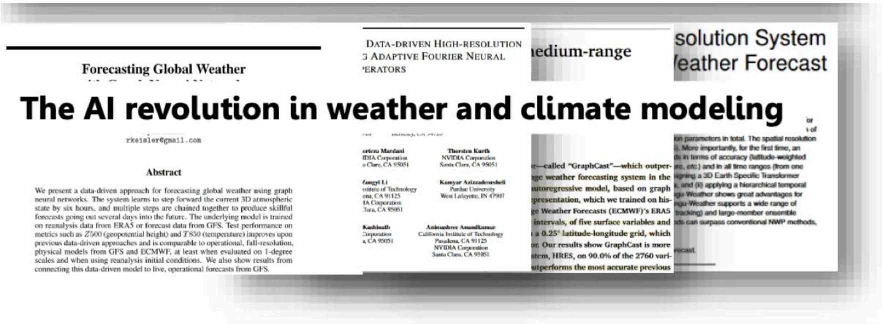
Ilaria Luise – ECMWF

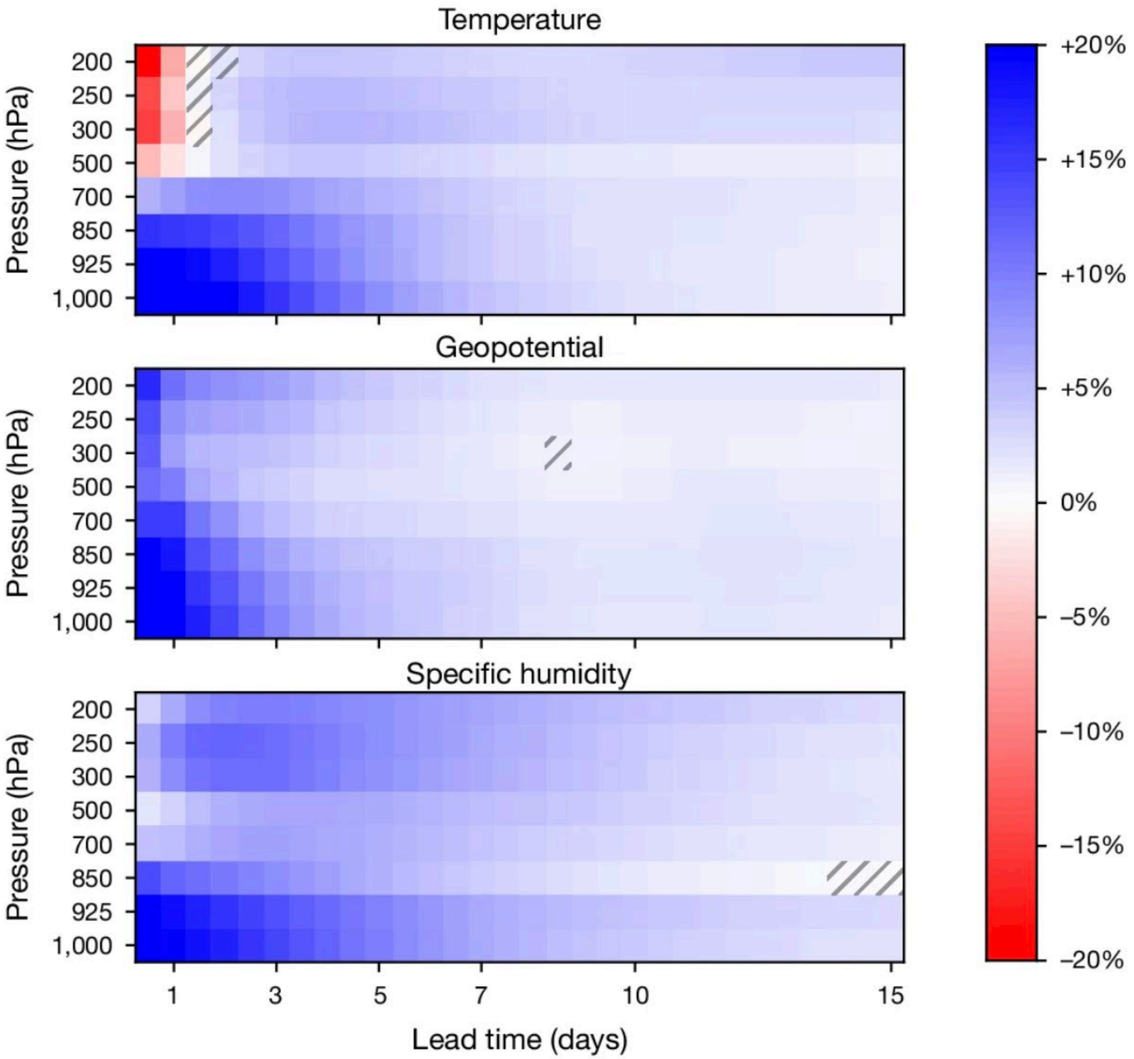
On behalf of the WeatherGenerator Collaboration

ESA-ECMWF Workshop – 13th April 2026, Bologna

Introduction

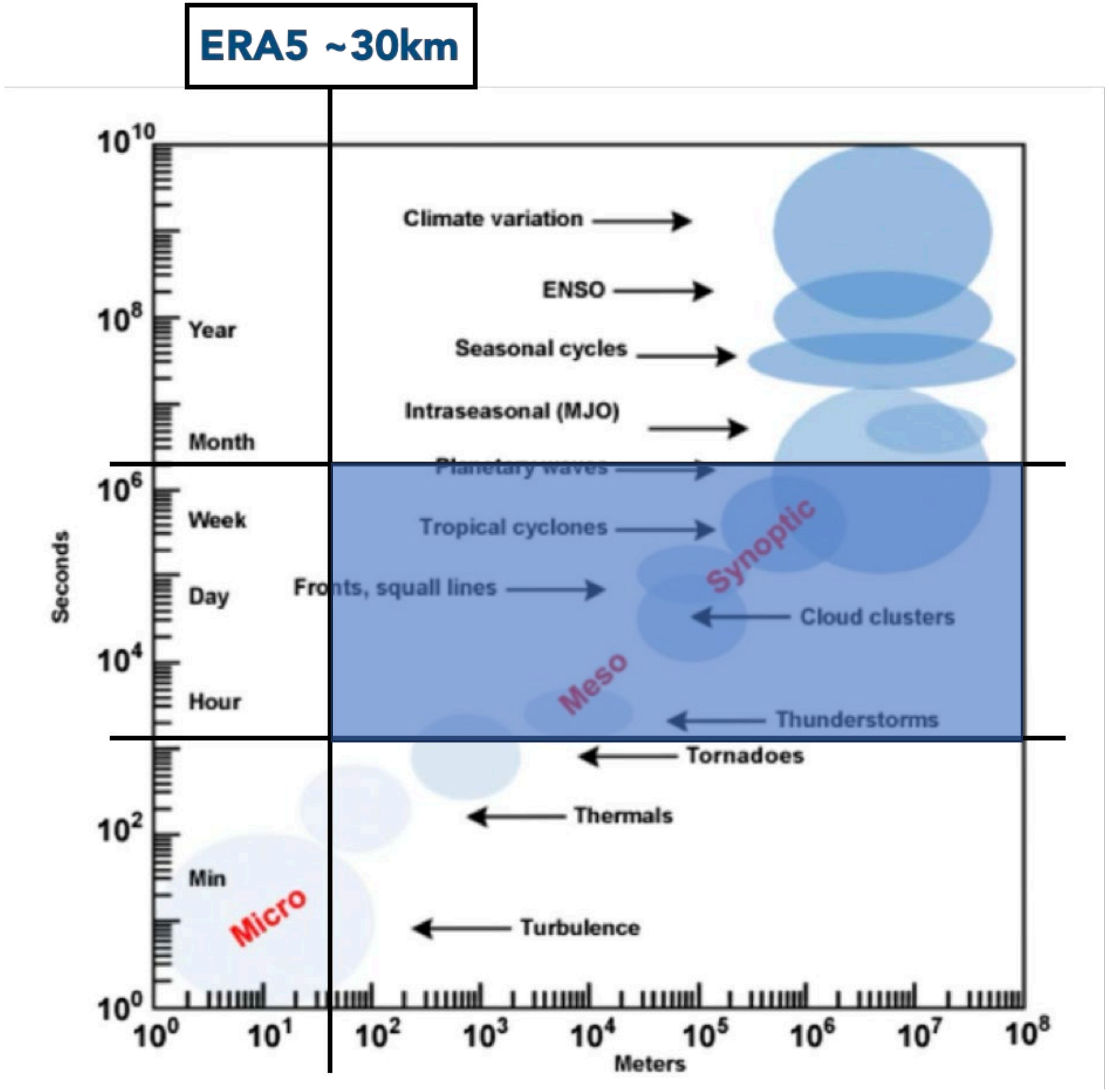
Exciting times for data driven weather modelling





GenCast: Probabilistic weather forecasting with machine learning, Nature 2024

But…



We are **spanning "only" part of the phase space**

still far from seamless ML models (but efforts ongoing e.g. BRIS)

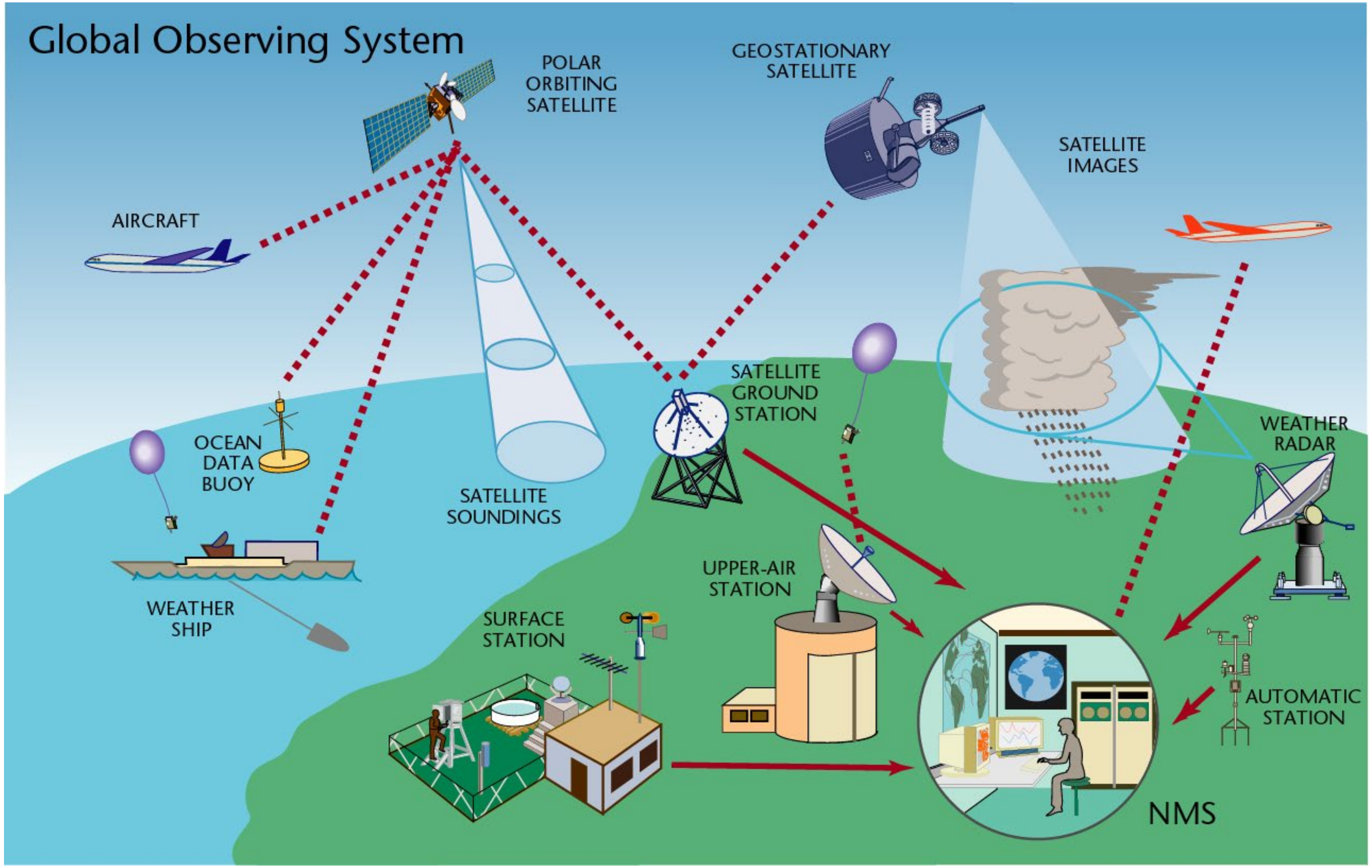
We suffer from the **(known) biases in ERA5**

We have **lots of other datasets available e.g. at higher resolution,** that are *not currently used* in our models

Can we go beyond?

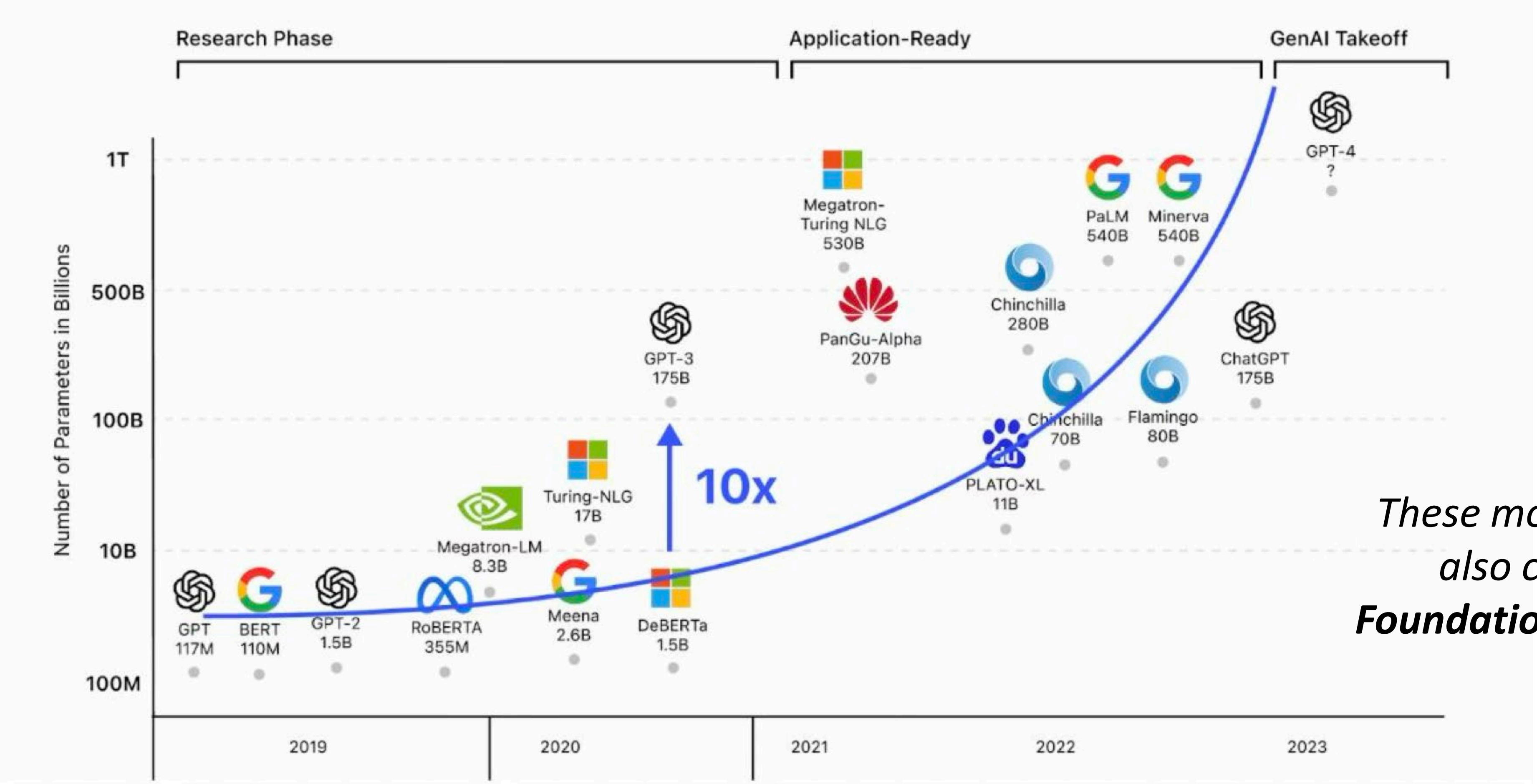
Towards a multi-resolution data driven model for Earth System science

We have lots of data that capture different scales



... and large scale ML can exploit very large data

- Models with trillions of parameters that can **fit very large dataset** and fulfill **multiple-tasks**



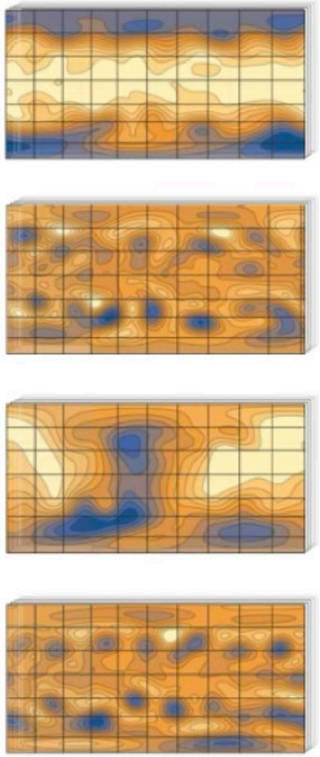
These models are also called ***Foundation Models***

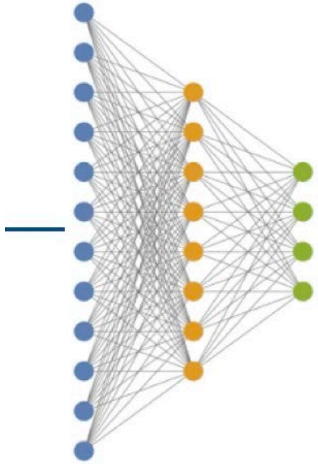
Can we use these ideas in Earth system science?

Current data driven models (the majority)

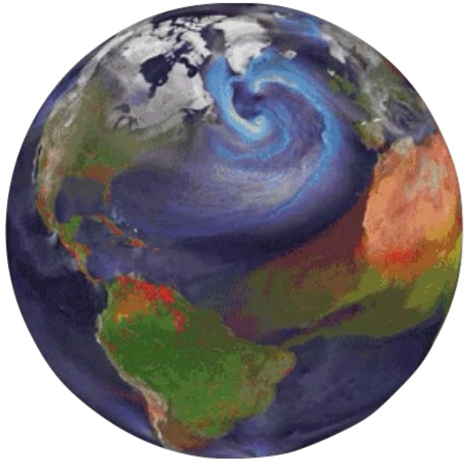
AIFS

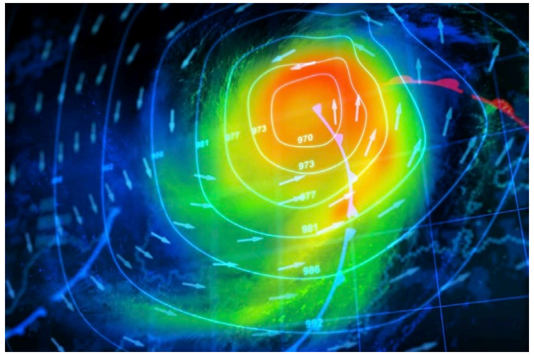
ERA5 reanalysis





model trained and tested for forecasting

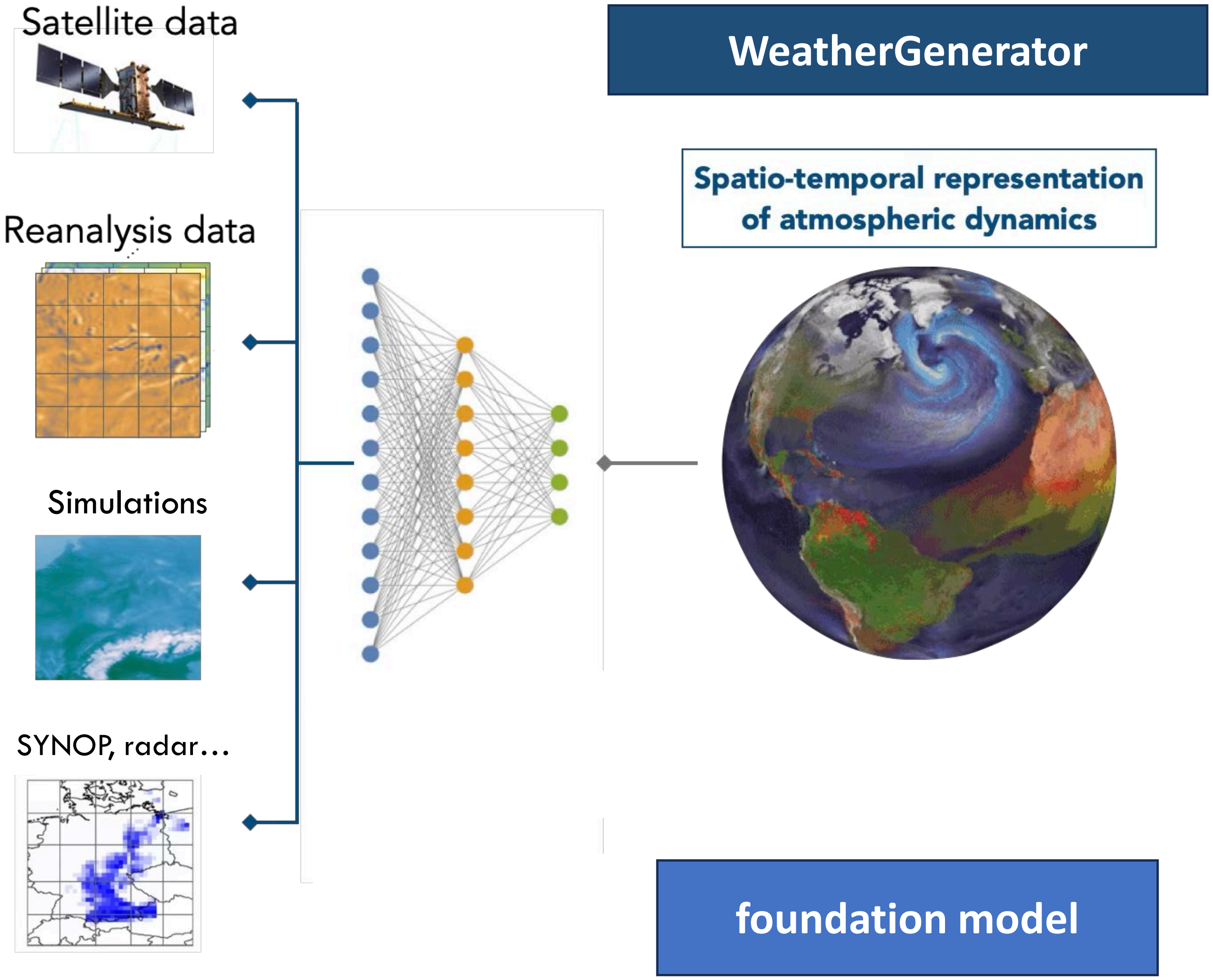




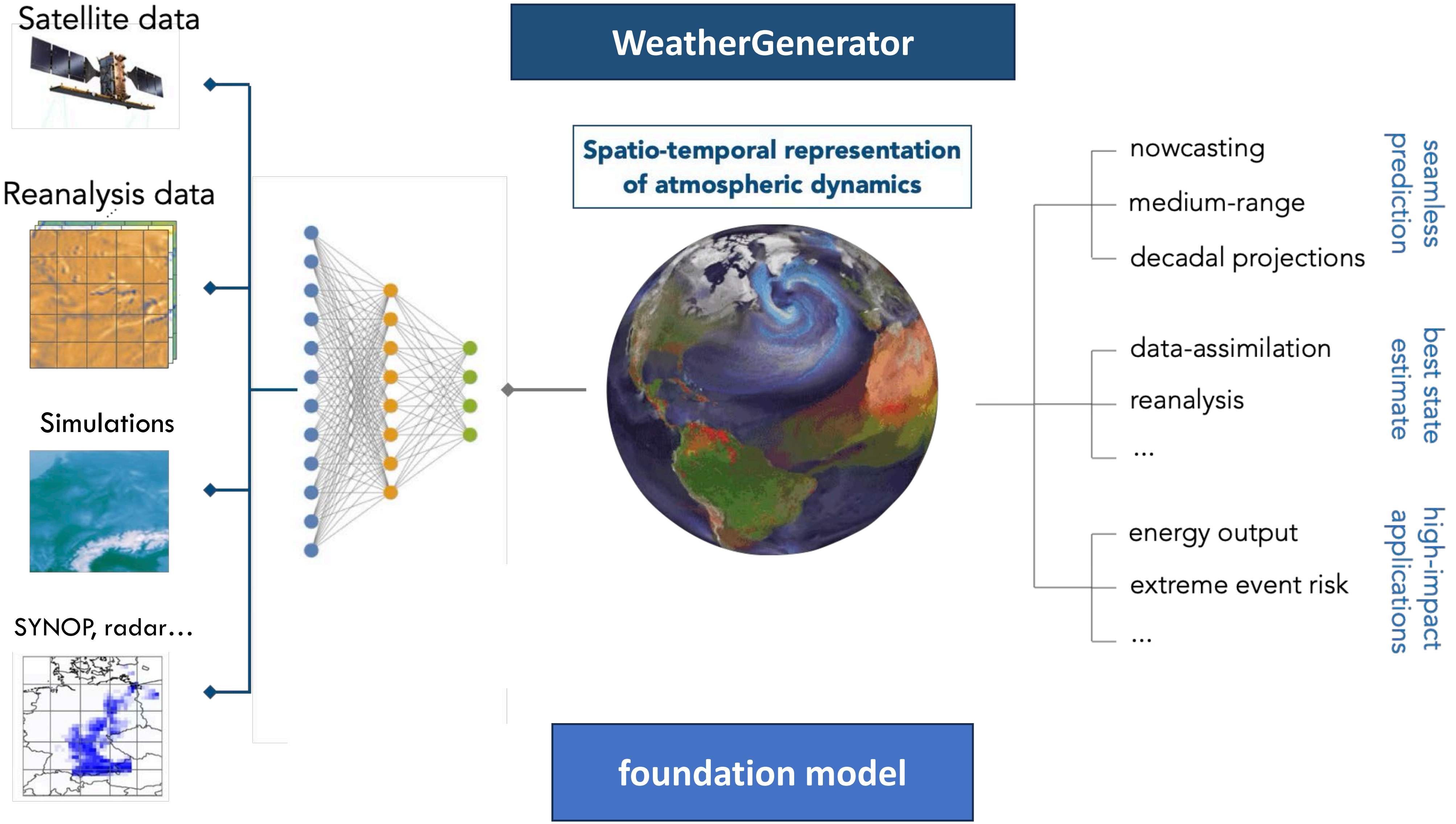
- Medium range forecasting
- Hurricane tracking
- …

supervised learning

The WeatherGenerator in a nutshell



The WeatherGenerator in a nutshell



Scientific potential: multiple applications

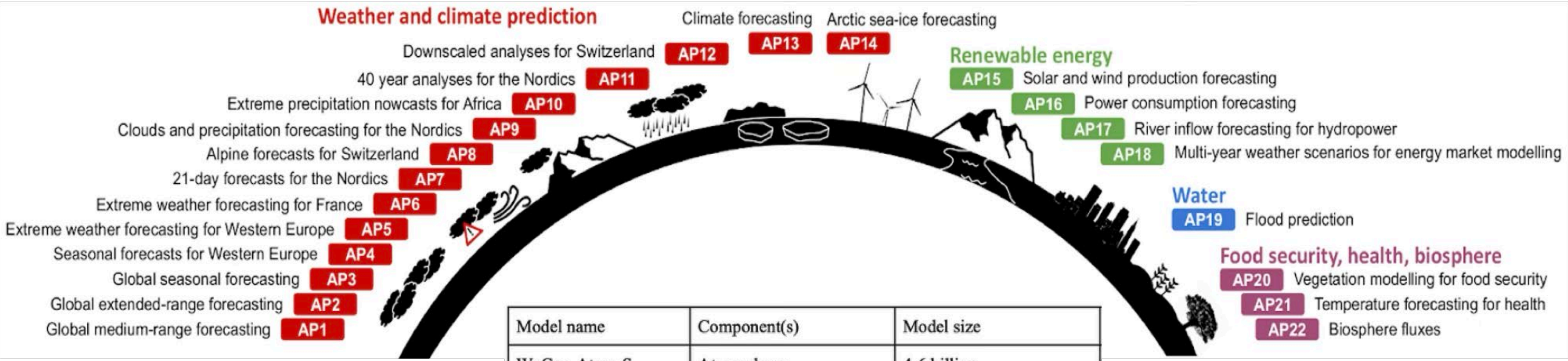
large latent space

Input → **Weather Generator** → **Output**

Input		**Output**		**Application**
Satellite and synoptic observation	→	ERA (35km) and CERRA (5km)	→	Data assimilation
IFS initial conditions (9km)	→	IFS state in the future (9km)	→	Global weather forecast
IFS forecast (9km)	→	OPERA (2km)	→	Model downscaling
DestinE Extremes Twin (4km)	→	synoptic observations	→	Post-processing
Sentinel-1	→	AROME (2km)	→	Observation operator
DestinE Climate Twin (4km)	→	ECOSTRESS (100m)	→	Vegetational water stress predictor
A date	→	IFS state	→	A climate emulator

…

WeatherGenerator downstream applications



Model name	Component(s)	Model size
WeGen-Atmo-S	Atmosphere	4-6 billion
WeGen-Atmo-M	Atmosphere	10-20 billion
WeGen-Atmo-L	Atmosphere	50-100 billion
WeGen-Land	Land	100 million
WeatherGenerator	Atmosphere, Land	4-6 billion/50-100 billion

Where do we stand?

The first step: a multi-resolution forecasting model

Overview

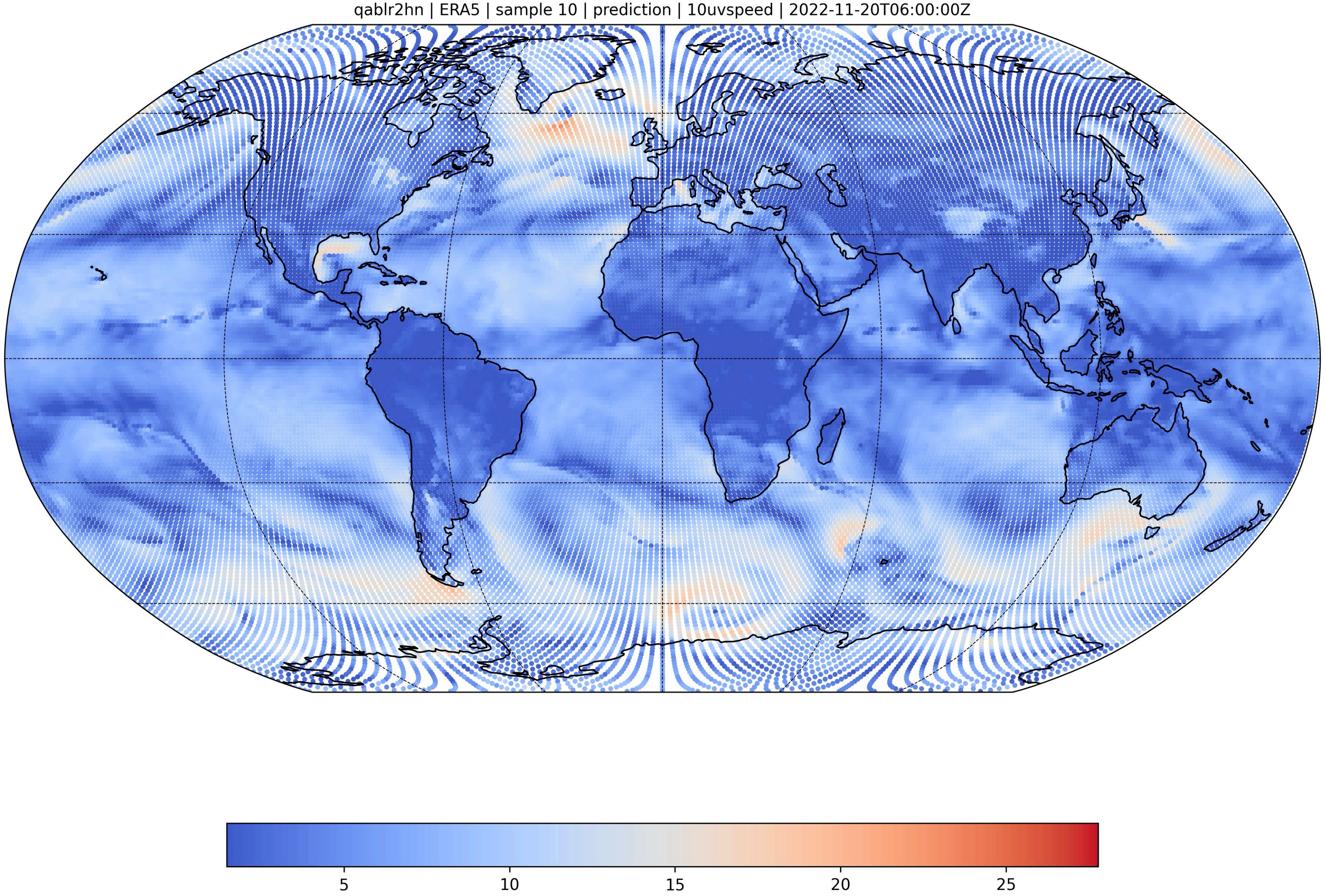
latent space that fuses input datasets into coherent, very rich representation of Earth system state

Current:
Transformer based forecasting engine

Future (WIP):
diffusion model to capture dynamic uncertainty

latent space allows us to decode arbitrary fields not in the input and training data

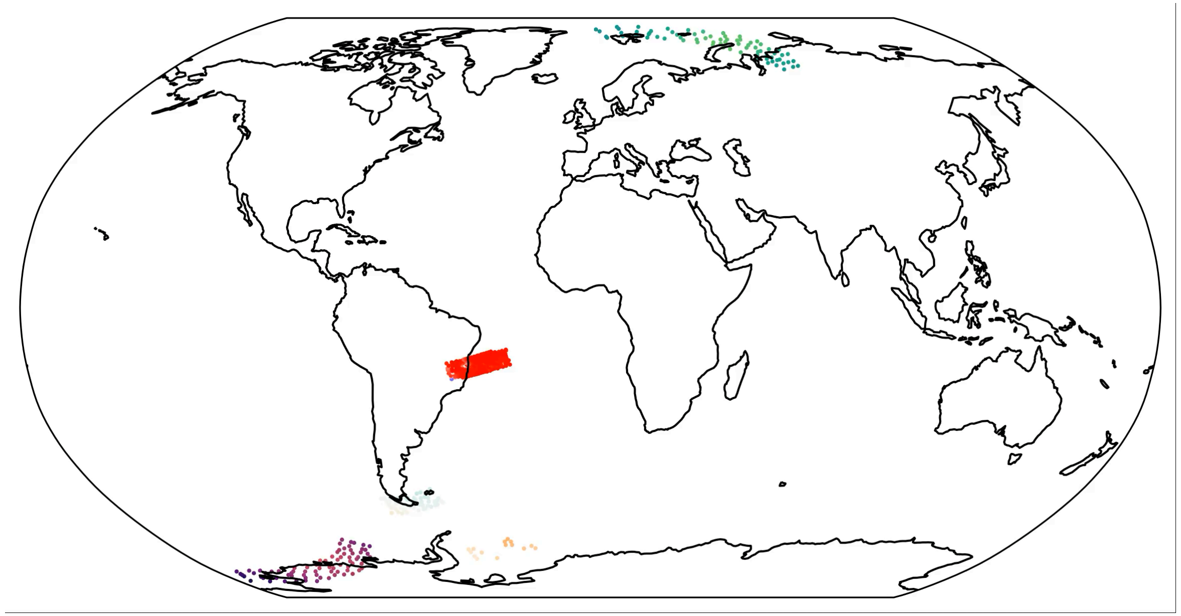
Overview



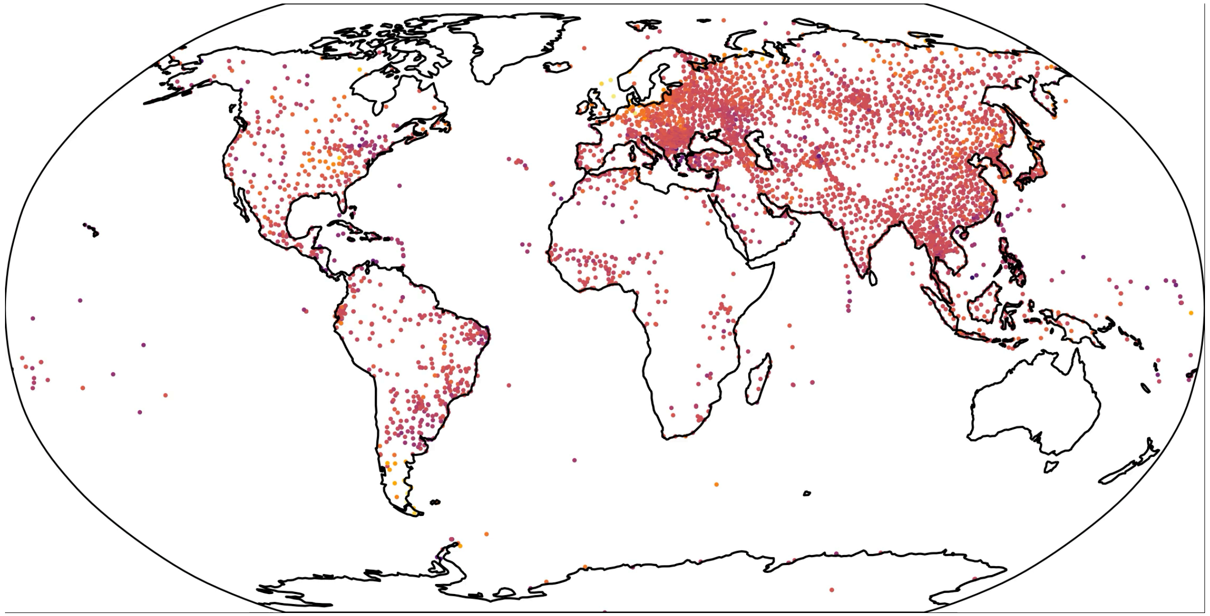
qablr2hn | ERA5 | sample 10 | prediction | 10uvspeed | 2022-11-20T06:00:00Z
5
10
15
20
25

Overview

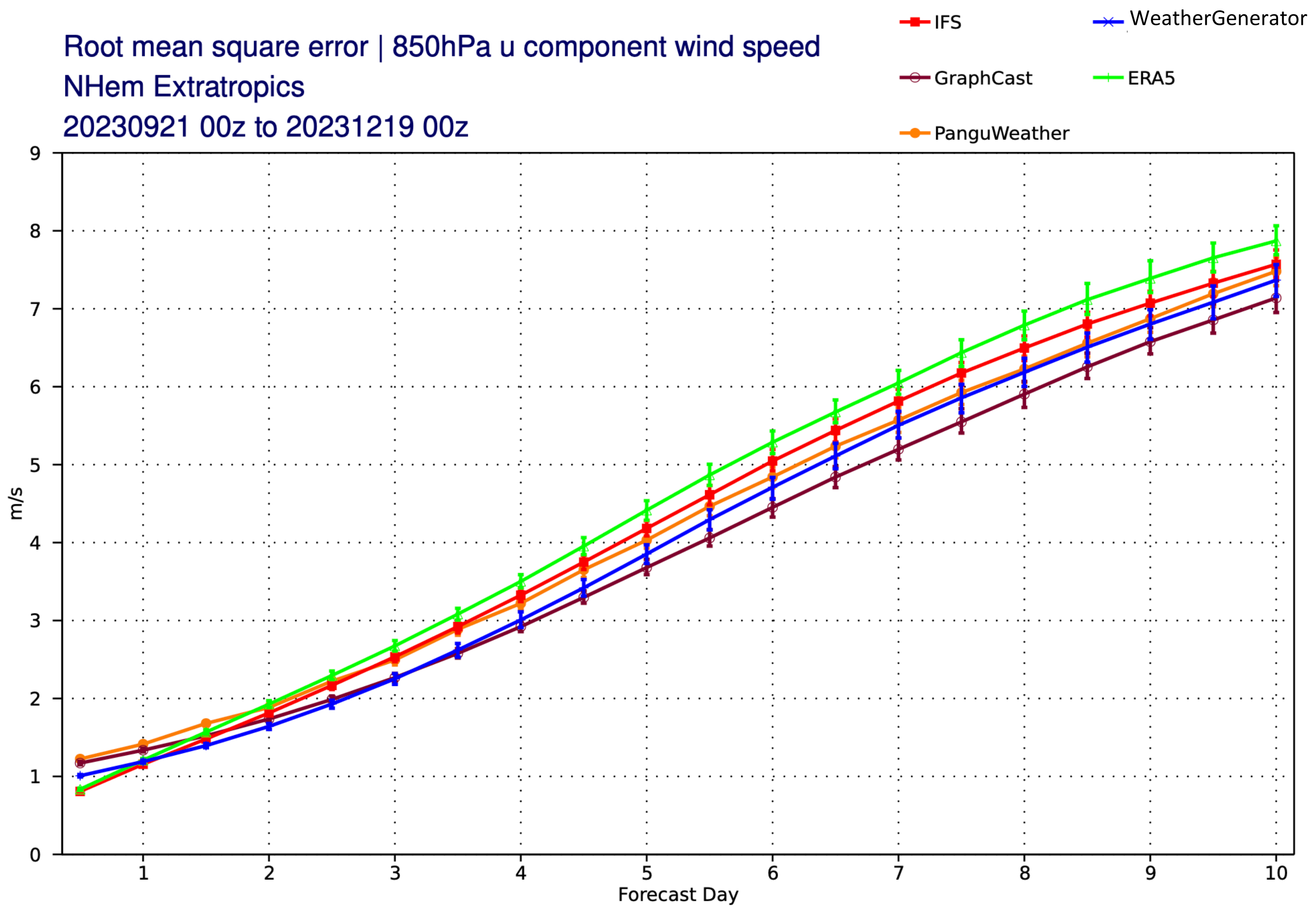
polar orbiting satellites
(infrared and microwave sounders)



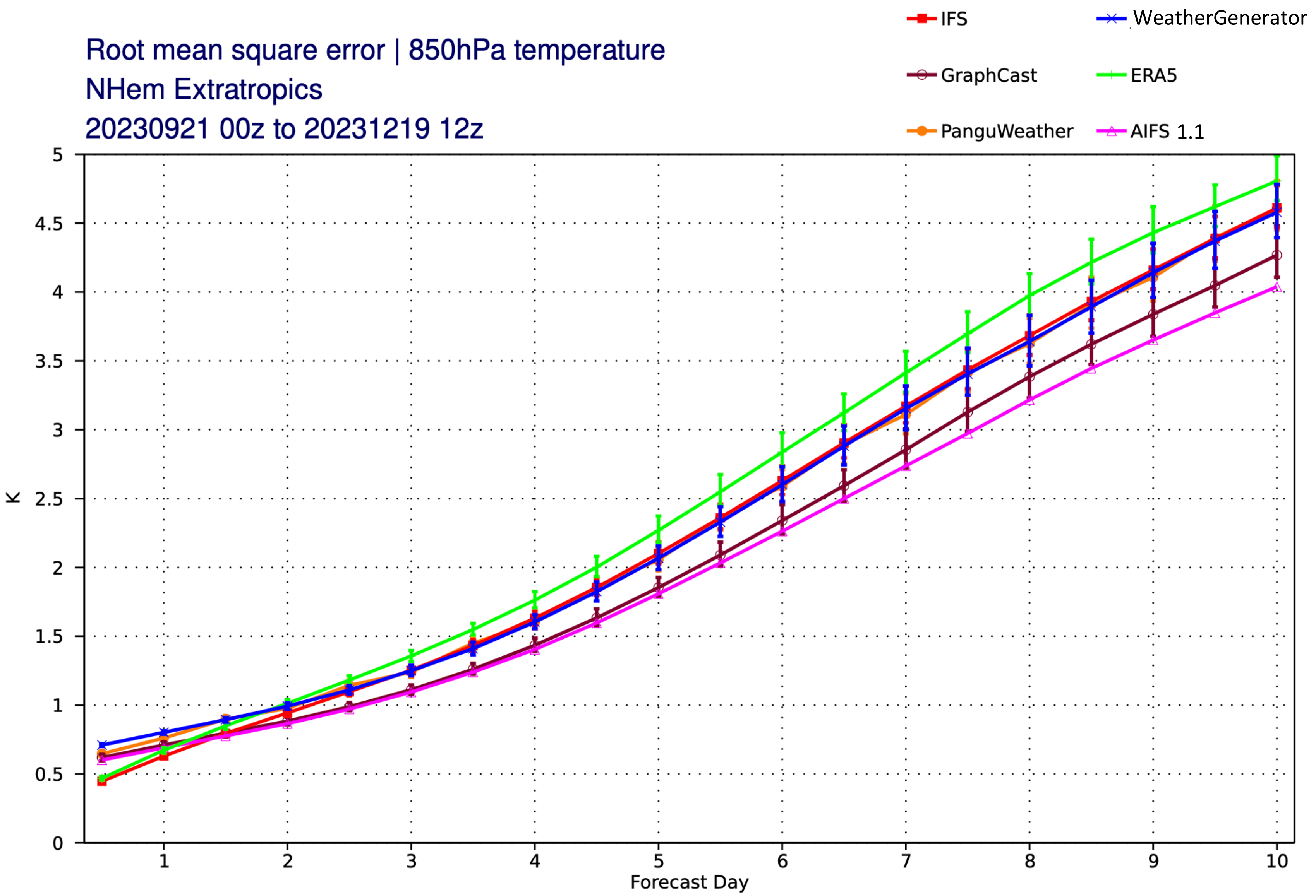
ground stations



Forecasting



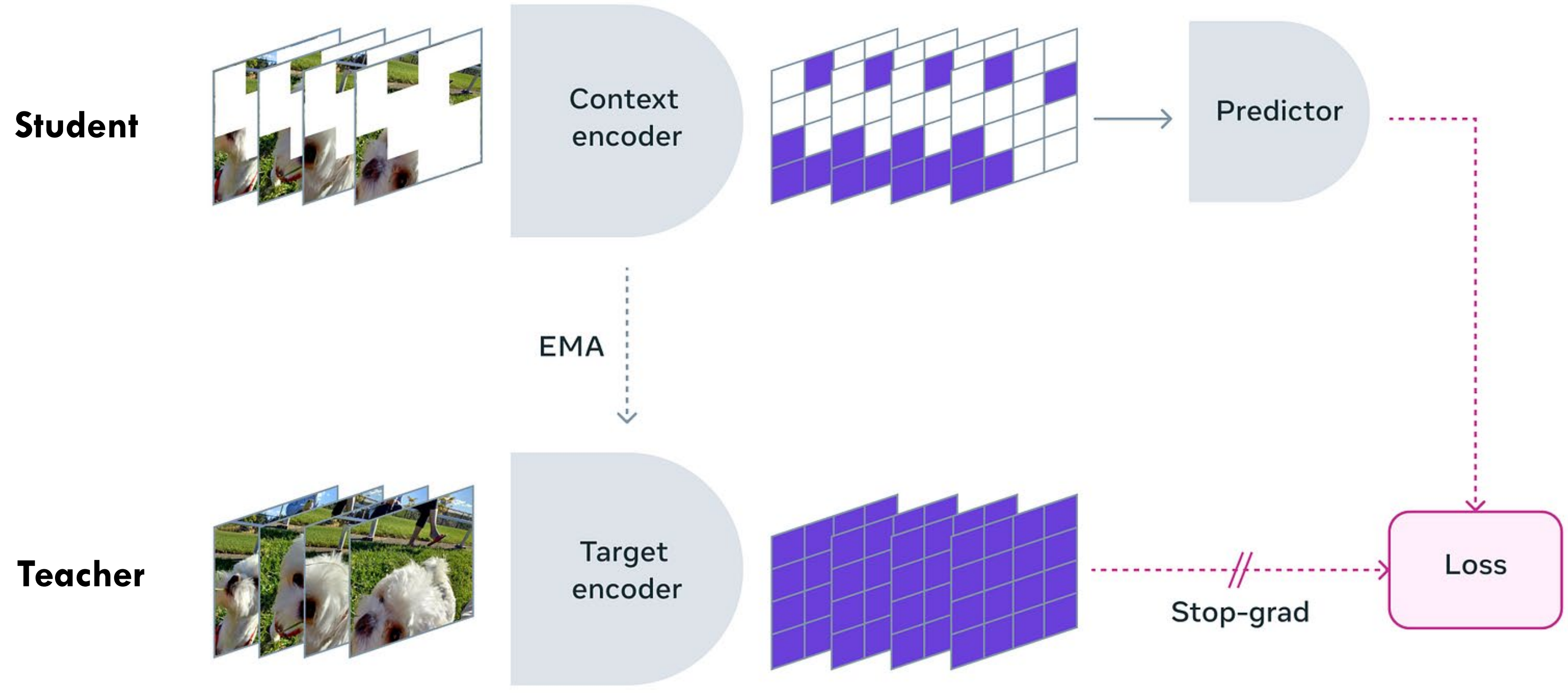
Forecasting



Latest developments

The next step: self supervised learning and training in latent space

The student-teacher training in latent space



Their weights are related: EMA = exponential moving average

Multimodal alignment
connects multiple datasets in the same space

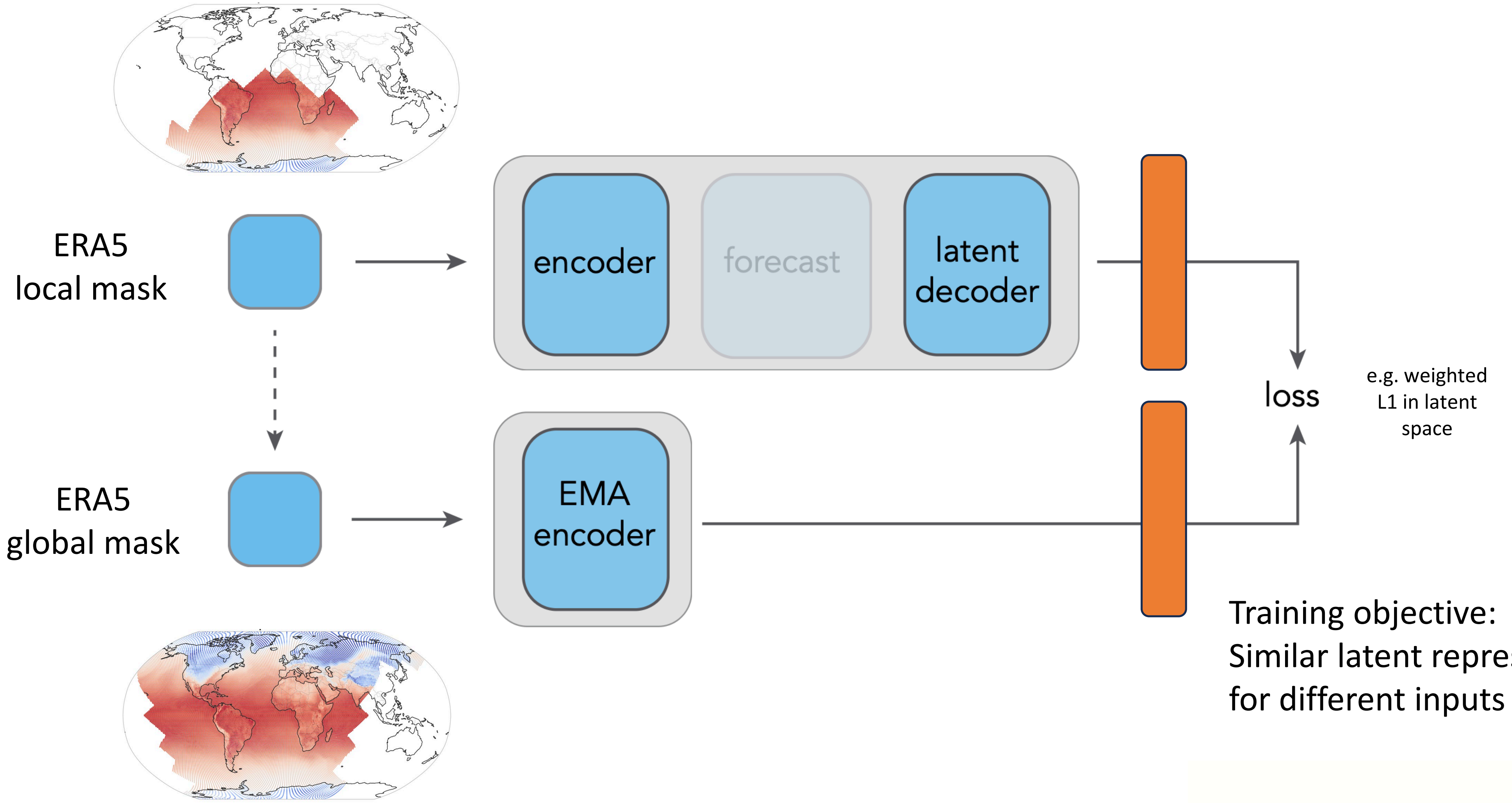
Transfer learning
reusable representations across tasks

Better generalization
reduces noise and overfitting

Captures meaningful structure
encodes semantic and spatial features

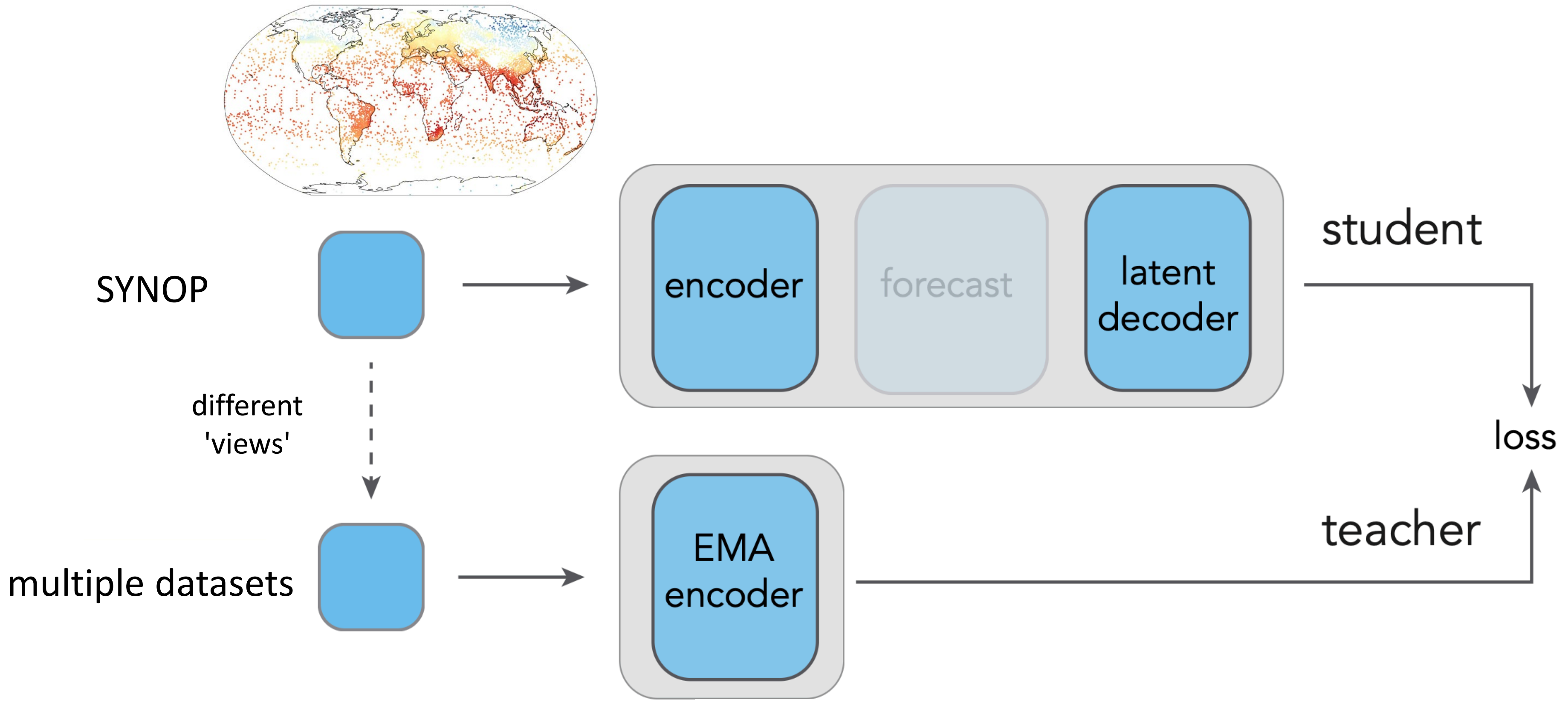
The training protocol

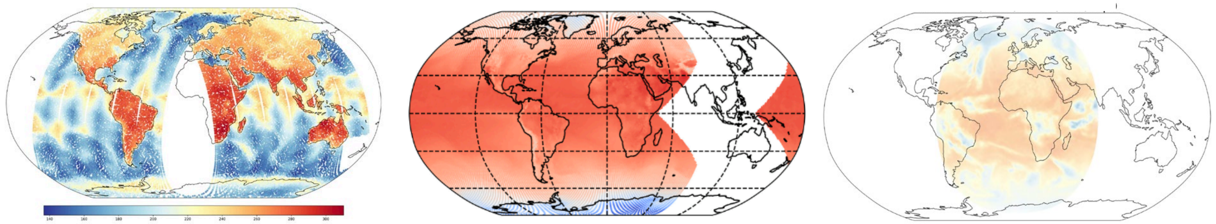
Learning in latent space



Training objective:
Similar latent representation
for different inputs

Why student-teacher for our data?



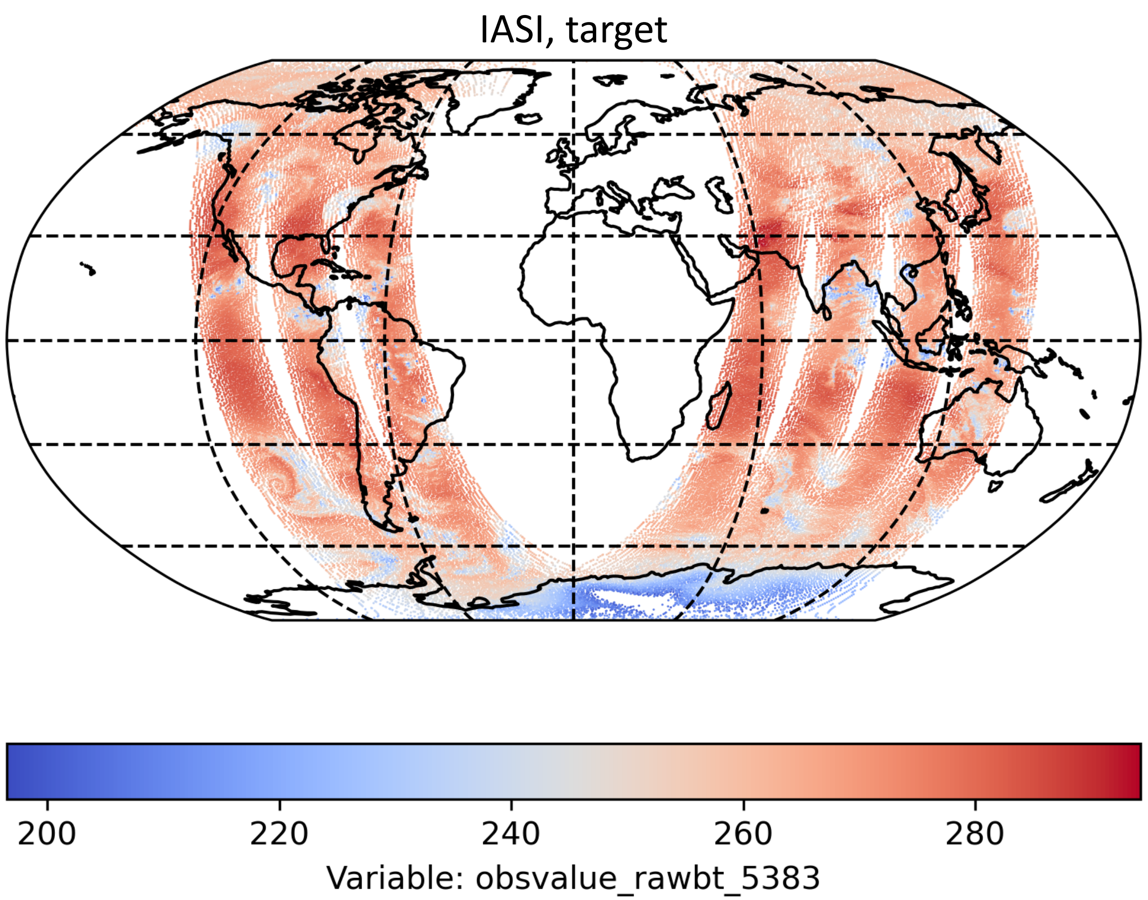


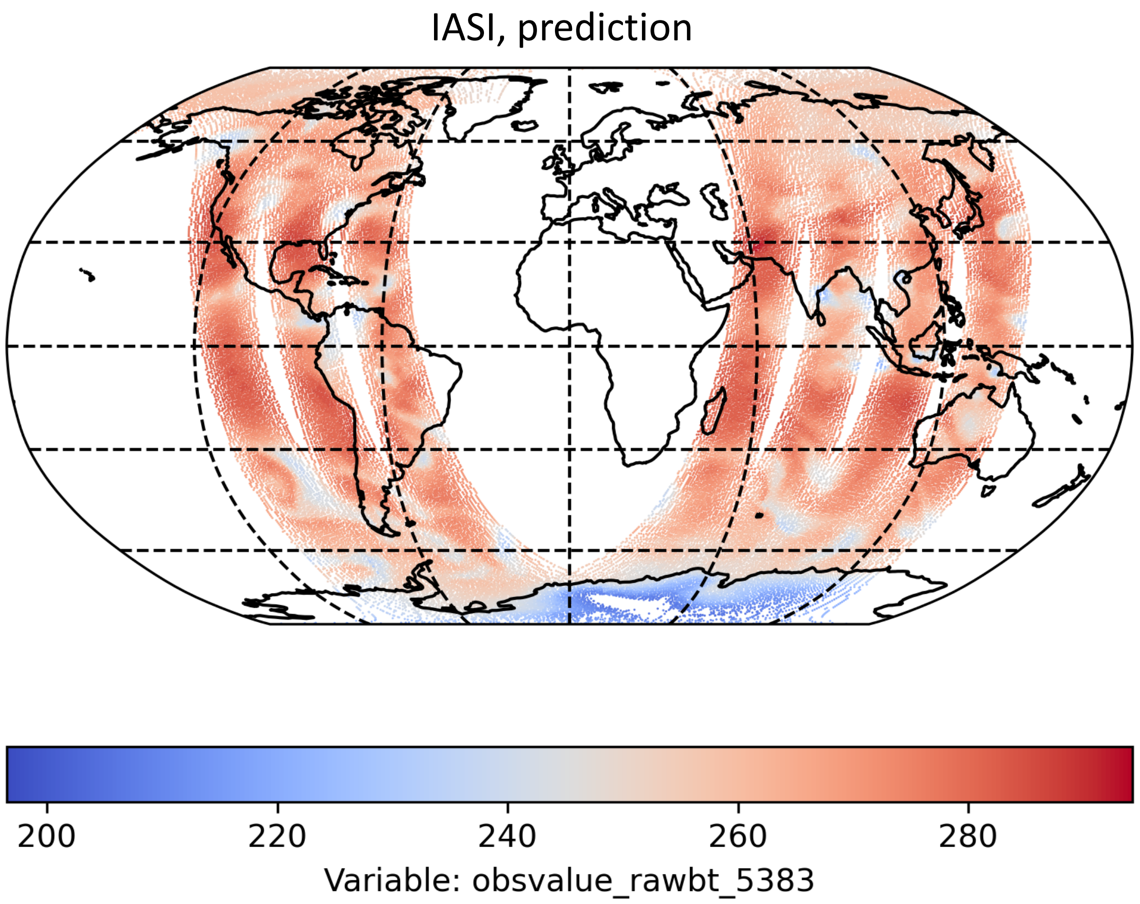
How can we tell if a latent space is good?

Pretraining–EMA Teacher

1. Decode on a different dataset never seen in training

e.g. Pretraining on ERA5 and decode IASI





Progress on catching up with the forecast model

2. Test on a different task (e.g. forecasting)

Progress on catching up with the forecast model

2. Test on a different task (e.g. forecasting)

Some thoughts

Nobody knows where ML models will stand in 5 years from now

- How many models do we need?
- How will we integrate them with current NWPs?

There is no guarantee that this approach will work better than having multiple models

but

We need to stretch the boundaries of what we think is possible with ML to shape the next generation of Earth system models

Funded by the European Union

Coordinated by ECMWF

Weather Generator

THANK YOU

www. weathergenerator.eu

JÜLICH Forschungszentrum

Norwegian Meteorological Institute

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and Water Management

METEO FRANCE
À VOS CÔTÉS, DANS UN CLIMAT QUI CHANGE

SMHI

Met Office

cmcc Centro Euro-Mediterraneo sui Cambiamenti Climatici

netherlands eScience center

buluttan Weather Intelligence

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

KAJO www. kajoservices .com

LT

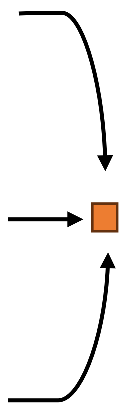
Statkraft

ETH zürich

MeteoSwiss

The WeatherGenerator project (grant agreement No101187947) is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.

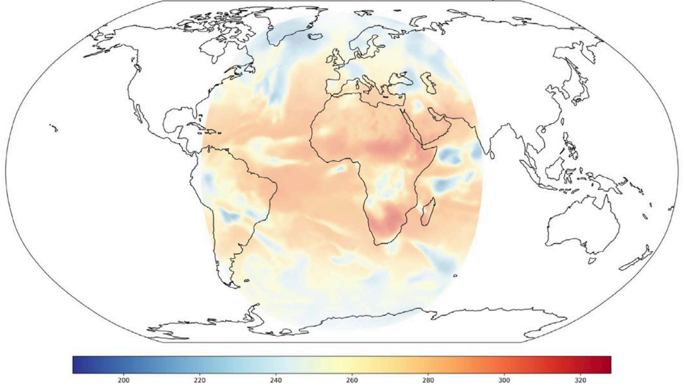
Intermezzo: the multi-resolution approach



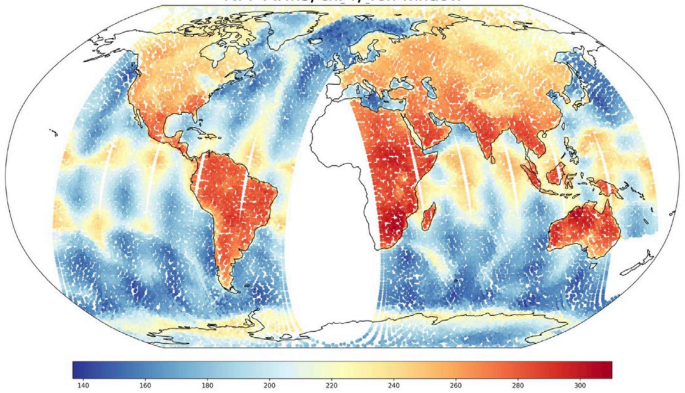
Preliminary results: multi-resolution model

Examples of generated predictions

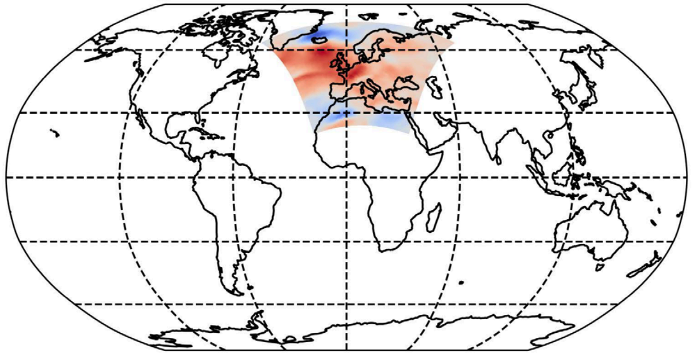
METEOSAT 11, SEVIRI, ch. IR 87, win 1, step 2



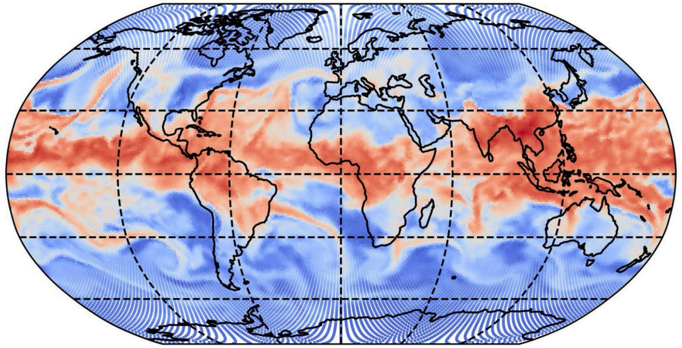
NPP-ATMS, ch. 1, 18h window

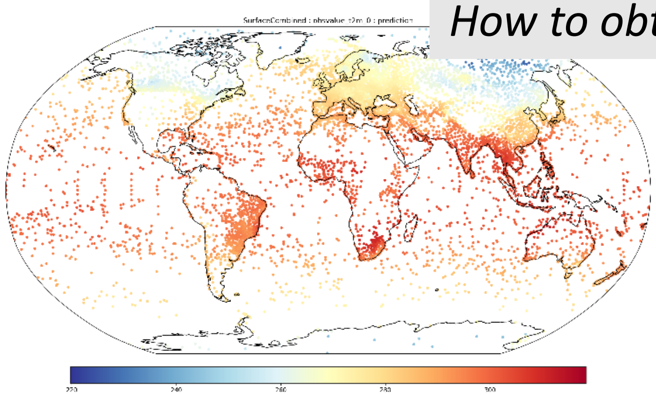


CERRA, u_850 : fstep = 001 (2022-10-01T06:00:00)



ERA5, q_850 : fstep = 001 (2022-10-01T06:00:00)





How to obtain SYNOP?

https://github.com/ecmwf/WeatherGenerator

Overview

Overview

Overview

latent space that fuses input datasets into coherent, very rich representation of Earth system state

Overview

Overview

latent space allows to decode arbitrary
fields not in the input and training data

Overview

Current:
Transformer based
forecasting engine

latent space allows to decode arbitrary
fields not in the input and training data

Funded by the European Union

Coordinated by ECMWF

Weather Generator

www. weathergenerator.eu

JÜLICH Forschungszentrum

Norwegian Meteorological Institute

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and Water Management

METEO FRANCE
À VOS CÔTÉS, DANS UN CLIMAT QUI CHANGE

SMHI

Met Office

cmcc Centro Euro-Mediterraneo sui Cambiamenti Climatici

netherlands eScience center

buluttan Weather Intelligence

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

KAJO www. kajoservices .com

LT

Statkraft

ETH zürich

MeteoSwiss

The WeatherGenerator project (grant agreement No101187947) is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.