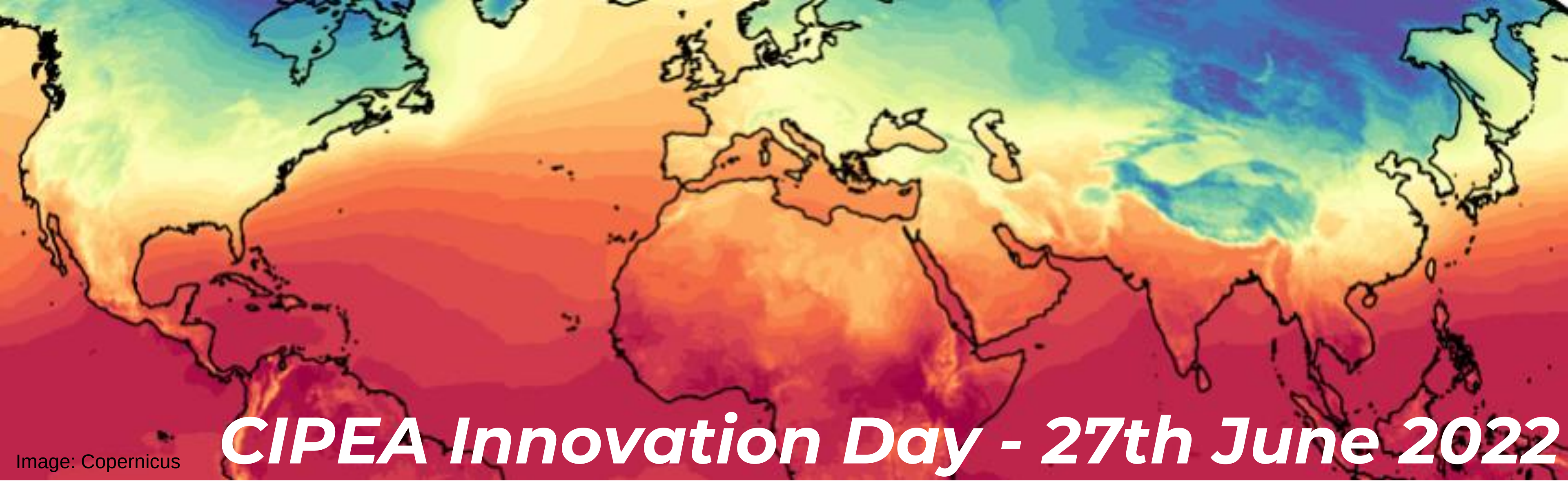


Image: Copernicus

CIPEA Innovation Day - 27th June 2022

CERN

JÜLICH Forschungszentrum

Kavli Institute for Theoretical Physics
University of California, Santa Barbara

EMP^2:
Environmental Modeling and Prediction Platform

A. Ferrari, ***I. Luise***, *C. Lessig, M. Schultz, M. Sonnewald,*
A. Subramanian, S. Vallecorsa

Climate and societal changes

UNITED NATIONS

Earth is facing a planetary emergency:

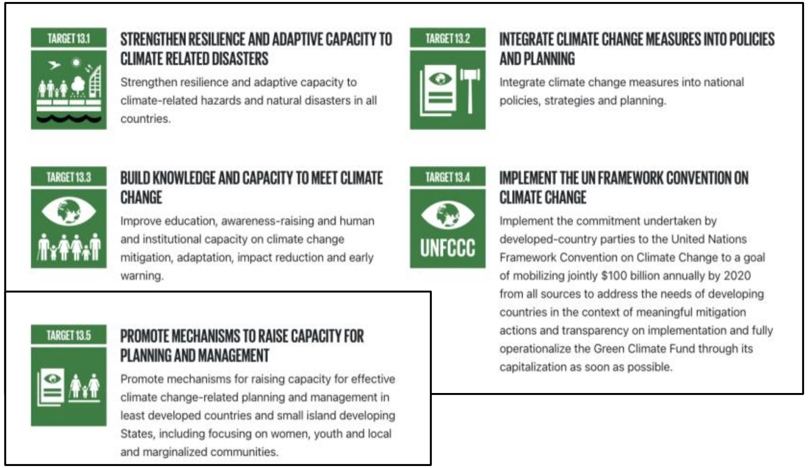
⇒ climate is heating up too quickly
⇒ loss of habitat for more than 1 million species
⇒ pollution continues to poison water, air and land

Science

IPCC
INTERGOVERNMENTAL PANEL ON climate change

Global scale ⇔ *Local scale*

Policy making



TARGET 13.1
STRENGTHEN RESILIENCE AND ADAPTIVE CAPACITY TO CLIMATE RELATED DISASTERS
Strengthen resilience and adaptive capacity to climate-related hazards and natural disasters in all countries.

TARGET 13.2
INTEGRATE CLIMATE CHANGE MEASURES INTO POLICIES AND PLANNING
Integrate climate change measures into national policies, strategies and planning.

TARGET 13.3
BUILD KNOWLEDGE AND CAPACITY TO MEET CLIMATE CHANGE
Improve education, awareness-raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning.

TARGET 13.4
UNFCCC
IMPLEMENT THE UN FRAMEWORK CONVENTION ON CLIMATE CHANGE
Implement the commitment undertaken by developed-country parties to the United Nations Framework Convention on Climate Change to a goal of mobilizing jointly $100 billion annually by 2020 from all sources to address the needs of developing countries in the context of meaningful mitigation actions and transparency on implementation and fully operationalize the Green Climate Fund through its capitalization as soon as possible.

TARGET 13.5
PROMOTE MECHANISMS TO RAISE CAPACITY FOR PLANNING AND MANAGEMENT
Promote mechanisms for raising capacity for effective climate change-related planning and management in least developed countries and small island developing States, including focusing on women, youth and local and marginalized communities.

**Concrete actions start at local scale:
in small communities & cities**

Local communities & climate change:

Concrete example:

A local administration needs to select a location for a wind farm

Data should be collected at a potential site for 2–3 years

Analyse data on wind speed, flow constance..

final decision

!

How to meet the best decision with the lowest impact on the planet?

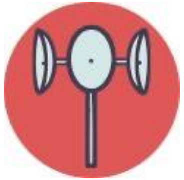
Classical models are too complex and expensive for local communities

⇒ data driven models are cheaper and faster

Local communities & climate change: digital Twin

Concrete example:

A local administration needs to select a location for a wind farm



Environmental model for predicting winds (global model)

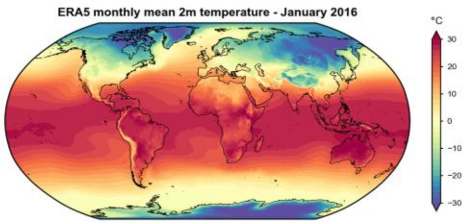
Geographic Informations of the chosen sites

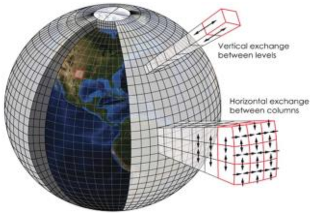
wind speed forecasting for each site

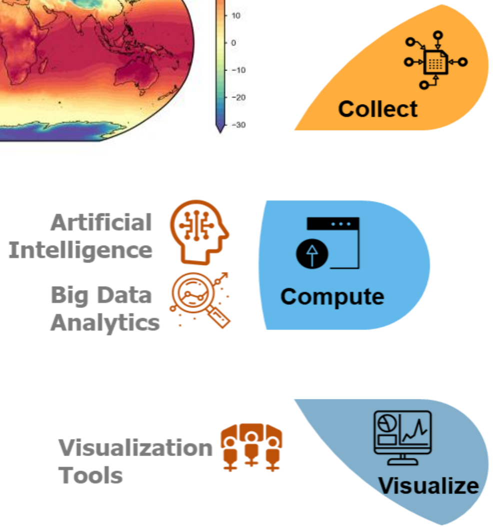
Digital models:
enable efficient, outcome focused decision making

New solution:
Digital Twin models

Digital Twin: main ingredients







Data availability
⇒ observational data for the process to be modelled

Accuracy
⇒ mathematical model of the observed process needs to be accurate

Efficient computational infrastructure
⇒ need supercomputing infrastructure for accurate prediction

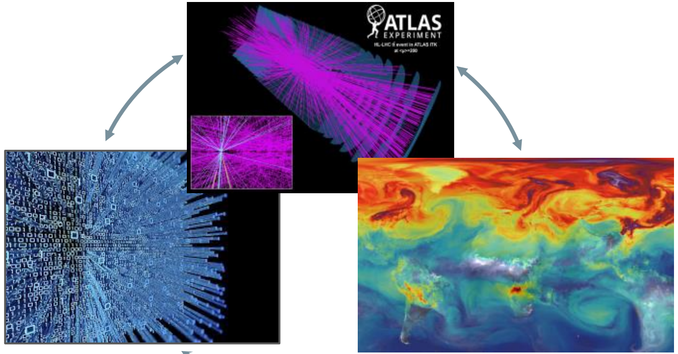
Visualisation and Accessibility
⇒ targeted user friendly interface of the model

Image: Huawei Enterprise

The EMP^2 project:

Goal: Proof-of-concept for a data-driven digital twin model of the atmosphere for environmental applications using 70 years of observational data

CERN scientists



Computer scientists

Earth scientists

Multidisciplinarity

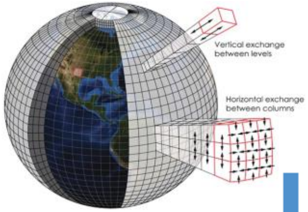
- big-data handling
- ML expertise
- Earth science expertise

Local scale applications

exascale machine learning

Environmental Modeling Core → **Applications** → **Visual interface**

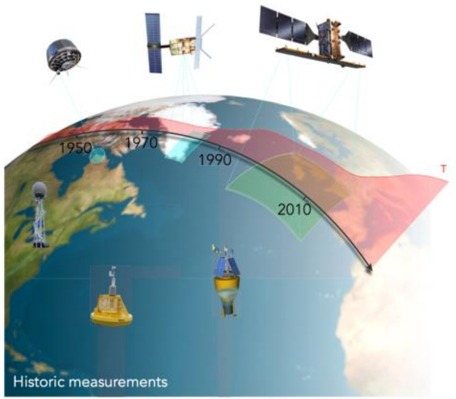
Global scale applications

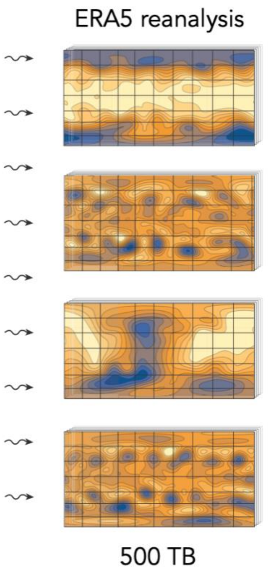


IPCC
INTERGOVERNMENTAL PANEL ON climate change

Learning from observations

Use 500 TB of publicly available Earth data observations

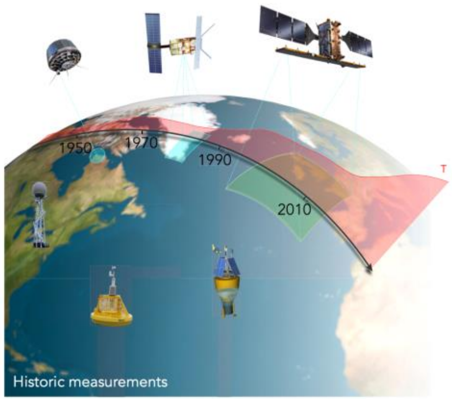




70 years of ERA5
=
lots of data about short to medium term phenomena

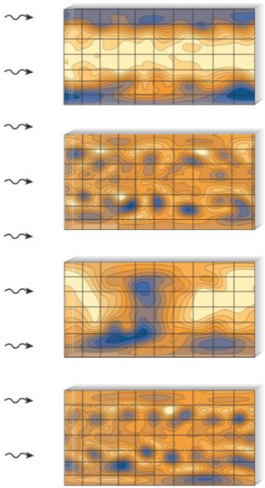
Learning from observations

Exascale training on supercomputer using newly introduced machine learning architectures

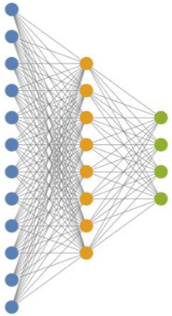


70 years of ERA5
=
lots of data about short to medium term phenomena

ERA5 reanalysis



500 TB

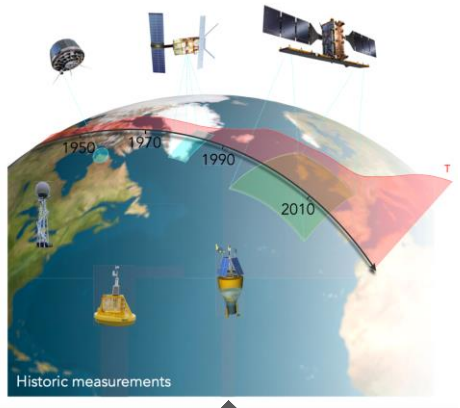


large scale machine learning

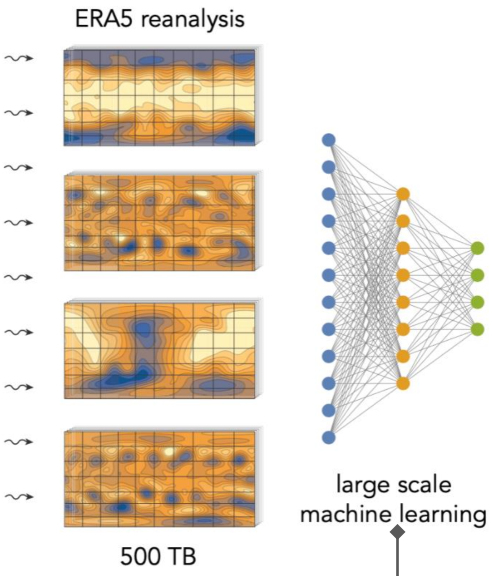
Transformers ML architectures
(scalable to large datasets, adaptable)

Learning from observations

Outcome: global scale representation learning of the atmosphere dynamics

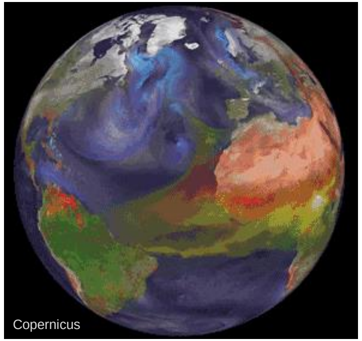


70 years of ERA5
=
lots of data about short to medium term phenomena



Transformers ML architectures
(scalable to large datasets, adaptable)

learned spatio-temporal representation of atmospheric dynamics

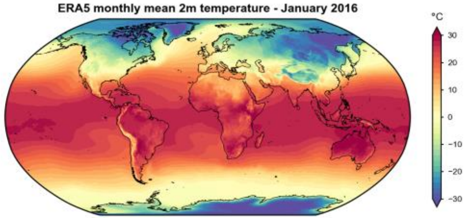


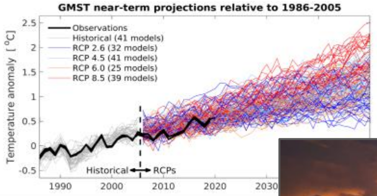
First ML-based global environmental core model of the atmosphere in 4D

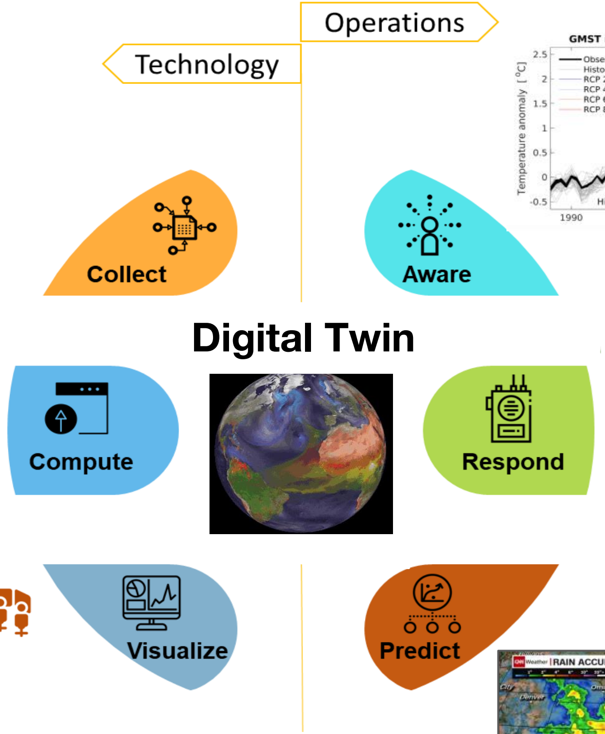
JÜLICH Forschungszentrum

Environmental modelling and societal challenges

Use the model as backbone for a set of global and local applications







Artificial Intelligence

Big Data Analytics

Visualization Tools

First proof-of-concept of a global machine-learning based environmental model trained on petabytes of observational data.

Image: Huawei Enterprise

Applications: local scale program

Environmental Modeling Core

partial dataset training

Occupational Health & Safety and Environmental Protection Unit (HSE)

⇒

Validation:
Use data from the CERN meteorological stations to generate predictions for ***CERN environmental monitoring applications***.

Local communities in the Geneva area, e.g. regional administrations

⇒

Application:
E.g. understand the most effective ***placement of a wind park*** and the potential energy output.

unitar
United Nations Institute for Training and Research
UNOSAT

UN satellite agency (UNOSAT) at CERN, insurance companies

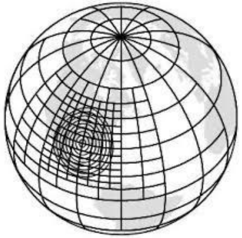
⇒

Application:
Improve the accuracy of the ***risk assessment*** and intervention plans, e.g. for ***flooding or wildfires***.

Applications: Global scale impact

Environmental Modeling Core

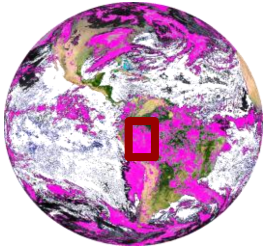
full dataset training



Improve super-resolution of simulations
⇒ Model downscaling

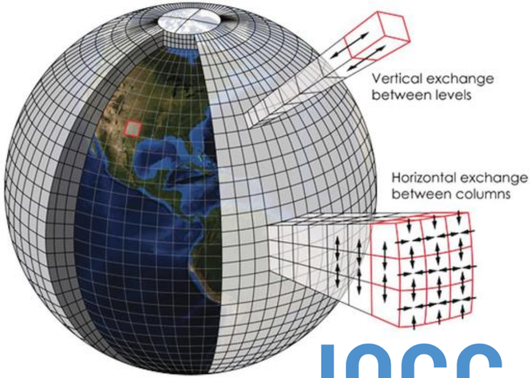
Improve weather forecasts
⇒ e.g. earlier and more accurate **prediction** of extreme events

Improve classification of extreme events ⇒ e.g. inform risk management and policy decisions



Disaster mapping + aid
⇒ shifting local regions; link between climate and, e.g., outbreaks, food security (longer term)

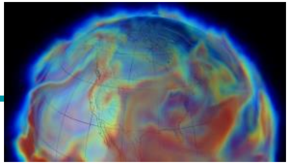
improve alert systems and/or better inform risk management and policy decisions through future IPCC related studies



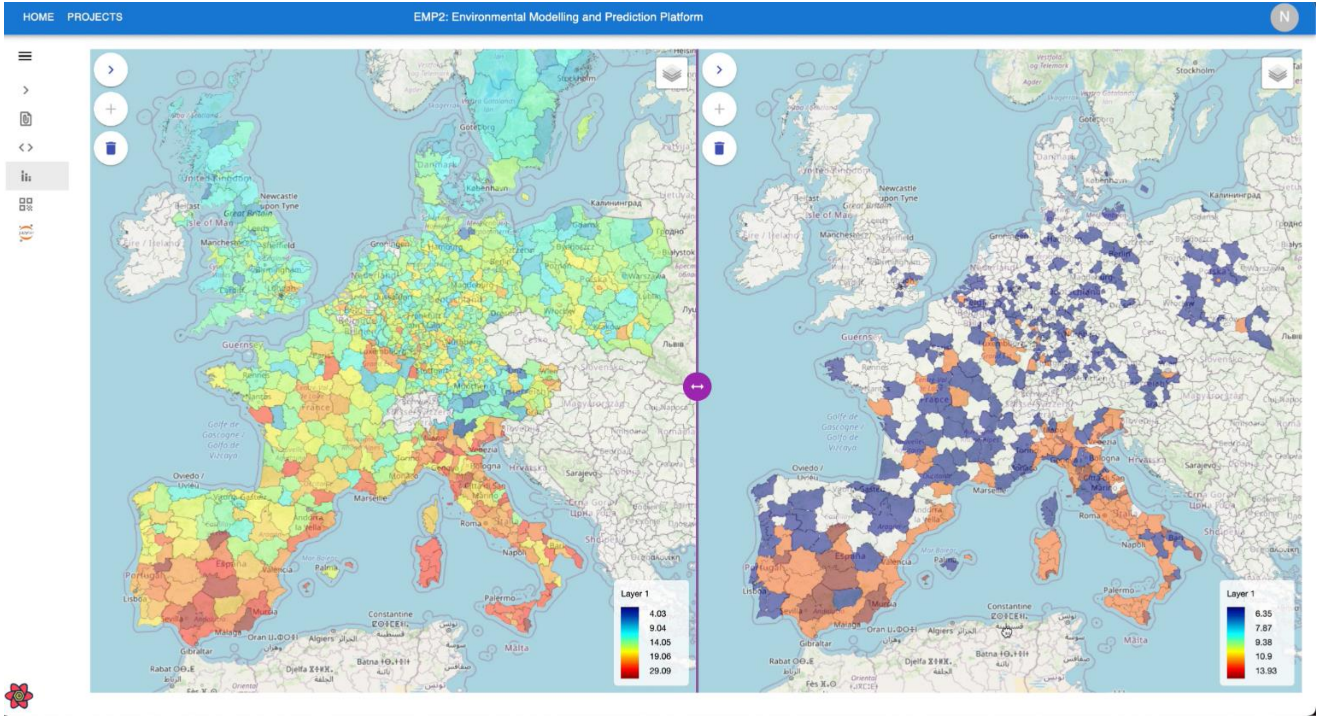
IPCC
INTERGOVERNMENTAL PANEL ON climate change

Collaboration with Climate Scientists within the Climate21 group

Visual Interface: CS4OD



@



Platform for open data
⇒ accessibility of data from different communities, based on FAIR principles and data sharing best practices

Platform for collaborations
⇒ Facilitate scalable linking, collection, sharing, and storage of environmental data from open access repositories, crowd-sourced projects, institutional repositories

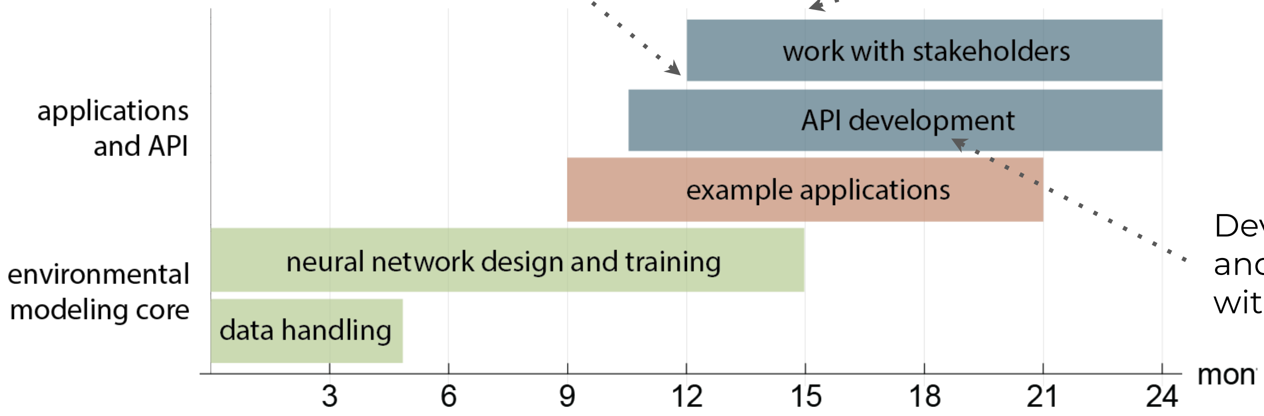
Platform for data exploration
⇒ Facilitate data exploration and results visualisation to diverse stakeholders

Timeline & Resources

Ideally: first results available already in summer 2023.

~12 months: First **proof-of-concept on partial dataset** & model reliability tested on local applications

~15 months: First results on **full dataset**
+ Start testing on global applications



Develop **Visual Interface** and finalise applications with stakeholders

Dedicated effort and funds from the CIPEA programme:

- Estimated at the level of **one graduate researcher for the duration of the program (2 years)**
  - Development of the environmental modeling core and integration in CS4OD
  - Validation and application of the models on the use cases (local and global)

Summary

⇒ **Final goal: develop a proof-of-concept of Environmental Digital Twin** for policy making

⇒ Multidisciplinary project involving CERN, climate scientists and computer scientists

⇒ **Modeling and prediction core** based on large scale machine learning architecture trained on supercomputing infrastructure.

⇒ First proof-of-concept of large scale training on 70 years of observational data!

⇒ Exploit existing ongoing work like **interTwin project** and **CS4OD** within the CERN IT Department

⇒ Test on a **set of applications** relevant to **addressing climate change at a local and global scale!**

Thank you!

Questions?

EMP²

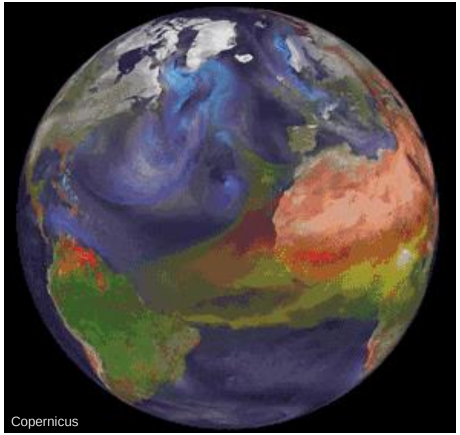
Backup

Learning from observations

First proof-of-concept of a global machine-learning based environmental model trained on petabytes of observational data.

First example of large-scale spatio-temporal representation learning (3D space+time) of atmospheric dynamics.

EMP²



environmental modeling core

Single environmental modeling core as **common backbone** for a wide range of applications.

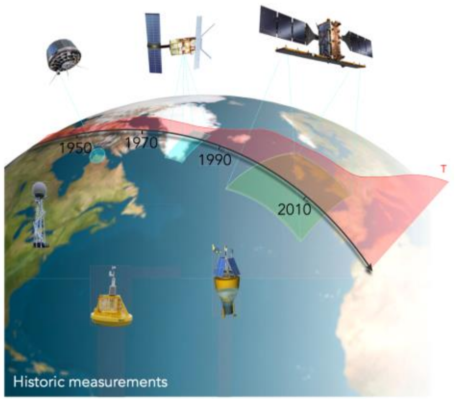
++++++

Global coverage

Flexible task-specific adaptation of the model realised, for example, through smaller and simpler neural networks

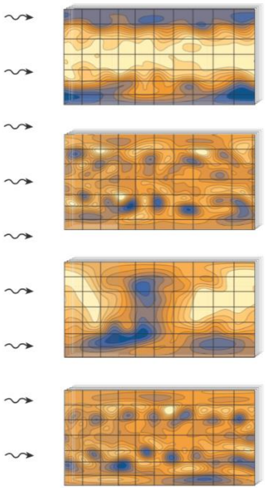
Learning from observations

Use the model as backbone for a set of global and local applications

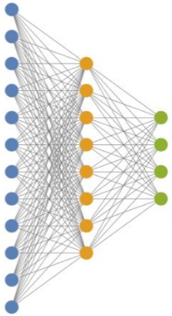


70 years of ERA5
=
lots of data about short to medium term phenomena

ERA5 reanalysis



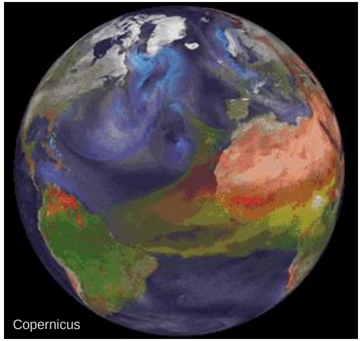
500 TB



large scale machine learning

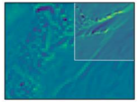
Transformers ML architectures
(scalable to large datasets, adaptable)

learned spatio-temporal representation of atmospheric dynamics

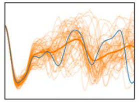


First ML-based global environmental core model of the atmosphere in 4D

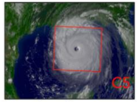
Applications



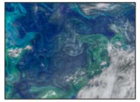
Downscaling



Predictability



Classification



Missing data

Disaster support

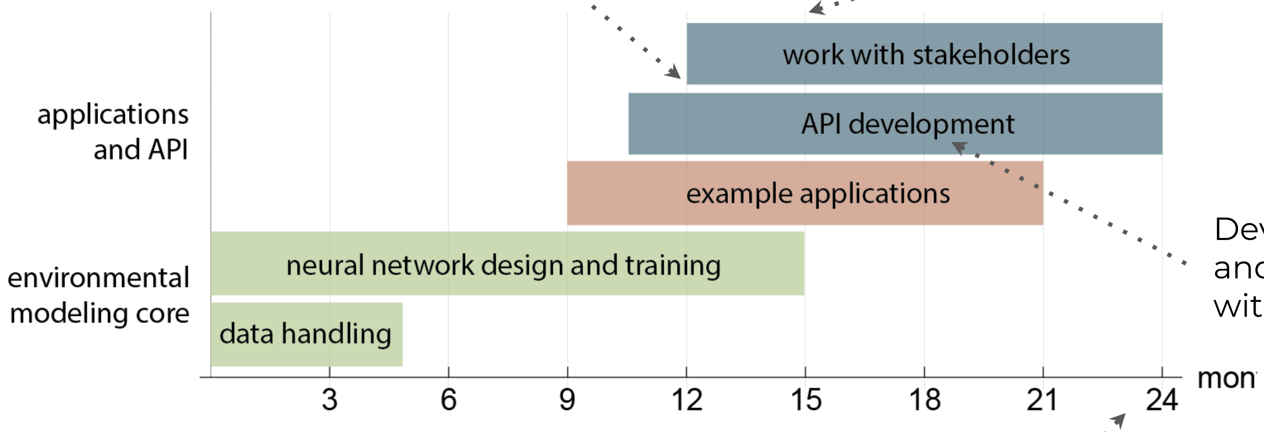
JÜLICH Forschungszentrum

Timeline & Resources

Ideally: first results available already in summer 2023.

~12 months: First **proof-of-concept on partial dataset** & model reliability tested on local applications

~15 months: First results on **full dataset**
+ Start testing on global applications



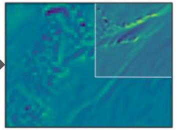
Develop **Visual Interface** and finalise applications with stakeholders

Longer timescale

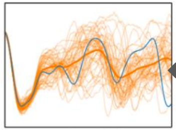
- Include ocean data or use raw data (e.g. **satellite data** from ESA, EUMETSAT)
- Broad impact applications: e.g. improve simulations, disaster mapping

Applications

Downscaling of atmospheric data, e.g. to assess the **local impact of climate change**.



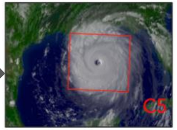
Downscaling



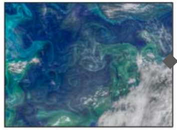
Predictability

Improve numerical **weather forecasting**, and error analysis in **existing weather and *climate simulation models***.

Classification of **extreme events**. Study how their ***statistics*** change over time and how it is affected by climate change.



Classification



Missing data

Interpolation of **sparse or *corrupted measurements***, e.g. airplane data and measurements occluded by clouds.

Improve disaster mapping and aid through increased dynamical understanding and targeted satellite monitoring

Satellite data

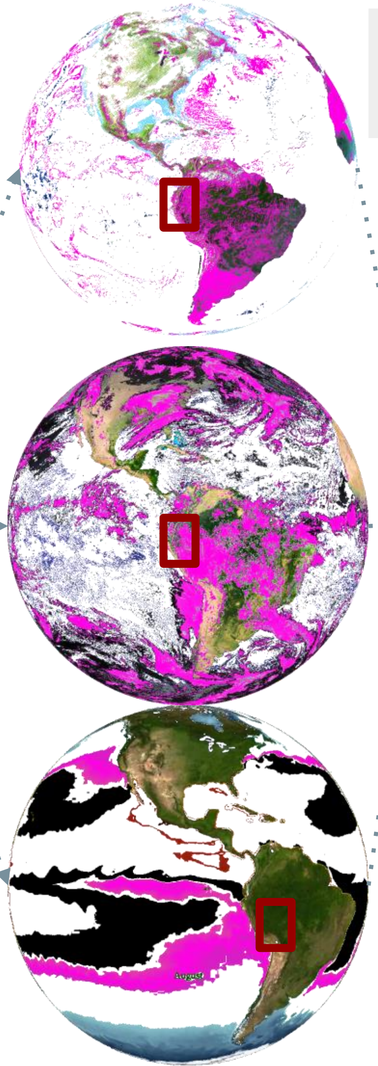
Longer term applications:

Improve disaster mapping and aid

through increased dynamical understanding and targeted satellite monitoring

AtmoRep: Learned spatio-temporal representation

Learn where to look to facilitate image processing and interpretation



Rainfall Distribution

Malaria outbreak support

The Team

CERN IT Department:

- Anna Ferrari
- Alberto Di Meglio
- Sofia Vallecorsa

Researchers:

- Christian Lessig - Magdeburg University
- Ilaria Luise - Stony Brook University, CERN
- Martin Schultz - Juelich Supercomputing Center
- Maike Sonnewald - Princeton University - Washington University
- Aneesh Subramanian - University of Colorado Boulder

PhD students and post-docs:

- Clara Betancourt - Juelich Supercomputing Center
- Bing Gong - Juelich Supercomputing Center
- Sebastian Hoffmann - Magdeburg University

Affiliated researchers: Claire Monteleoni (UC Boulder), Donata Giglio (UC Boulder), Abigail Bodner (NYU-Courant), Yi Deng (GeorgiaTech), Julien Brajard (NERSC)