ESMO
Earth System Modelling and Observations

WCRP
World Climate Research Programme

Machine Learning in Earth System Modelling and Observations

Douglas Rao (NC State University)
Working Group on Observations for Researching Climate (WGORC)

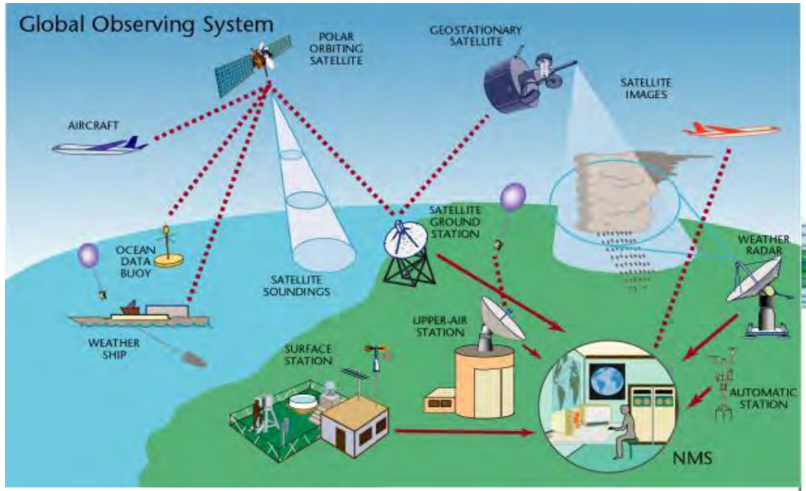
WMO Workshop on Needs and Application of Climate Data Management to Support the State of the Climate Reporting

23–26 June 2025, Astana, Kazakhstan

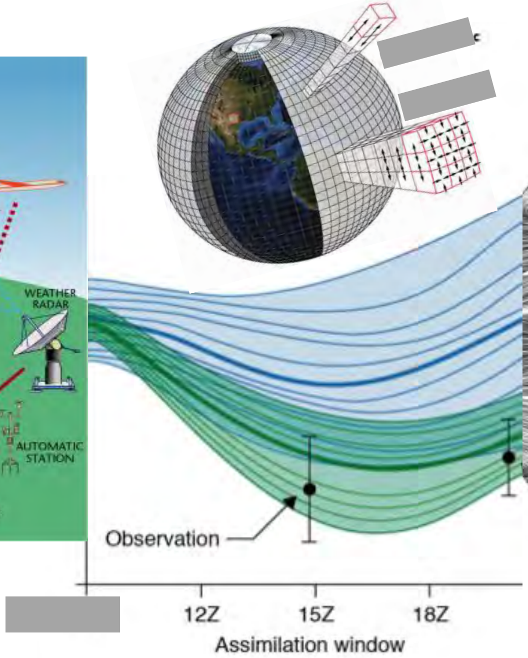
What is ESMO?

- A new WCRP Core project that coordinates and advances all WCRP **observations**, **data assimilation** and **modelling**
- Works on three objectives:
  - **Advancing predictions and projections of the Earth system** on time scales from weeks to centuries and furthering model-observation integrated frameworks.
  - **Improve monitoring, understanding, and attribution of Earth system changes and impacts** with robust uncertainty quantification through the synthetic use of models and observations.
  - **Advancing and harnessing emerging technologies** in modelling and observations.

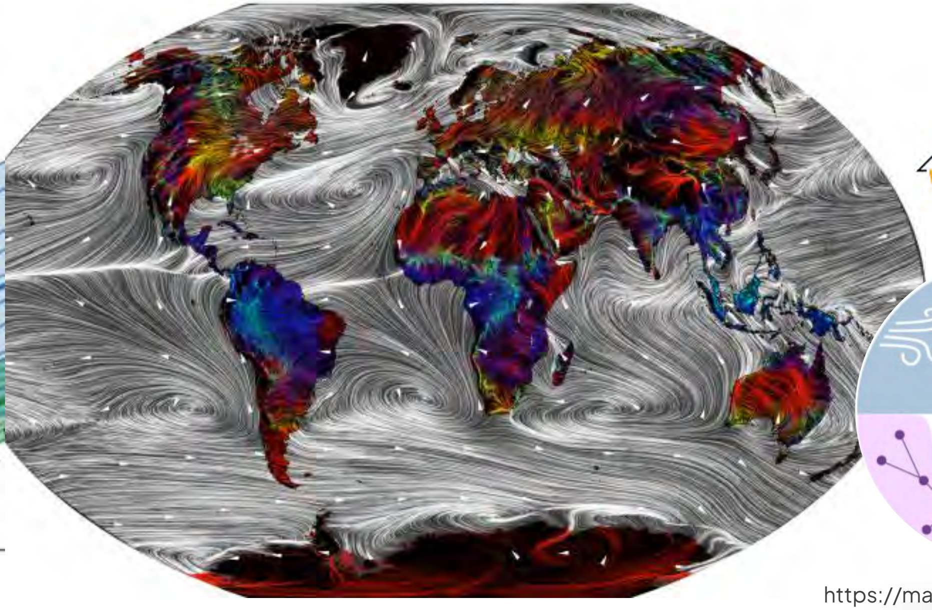
What is ESMO?



WMO Global Observing system



Cambridge University Press Climate Model and ECMWF Data Assimilation schematic



Kling and Ackerly, 2020

https://machineclimate.de/

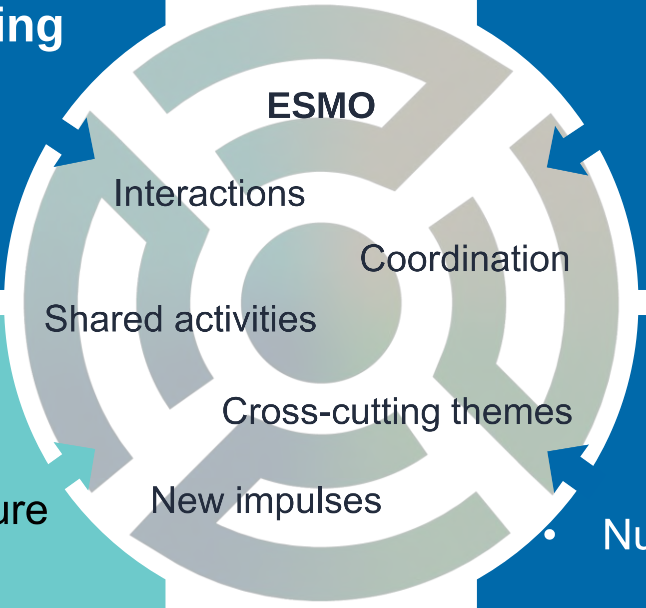
Modelling Community in WCRP

Working Group on Coupled Modelling (WGCM)

- Development and review of coupled climate models

Working Group on Numerical Experimentation (WGNE)

- Development of Earth system models (design, implementation, error diagnosis, revisions)



Coupled Model Intercomparison Project (CMIP)

- Understanding of past, present and future climate changes
- Assessment of model performance

Working Group on Subseasonal to Interdecadal Prediction (WGSIP)

- Numerical Experimentation for S2I variability and predictability

Observational Community in WCRP

Topic-specific observational groups in core projects

GSOP
- CLIVAR Global Synthesis and Observations Panel

GDAP
- GEWEX Data and Analysis Panel

GASS & GLASS
- GEWEX Global Atmospheric System Studies
- GEWEX Global Land-Atmosphere System Stuc

APARC activities on
- Stratospheric ozone
- Temperature Trends

CliC activities on
- Sea Ice Processes
- Permafrost Carbon Network

GAW
- Global Atmosphere Watch Programme

GCOS
- Global Climate Observing System

GOOS
- Global Ocean Observing System

CEOS/CGMS WG Climate
- Committee for Earth Observation Satellites / Coordination Group of Meteorological Satellites

External Partners

ESMO Objectives

Advancing predictions and projections of the Earth system

Improving monitoring, understanding and attribution of climate system changes and impacts

Advancing and harnessing emerging technologies

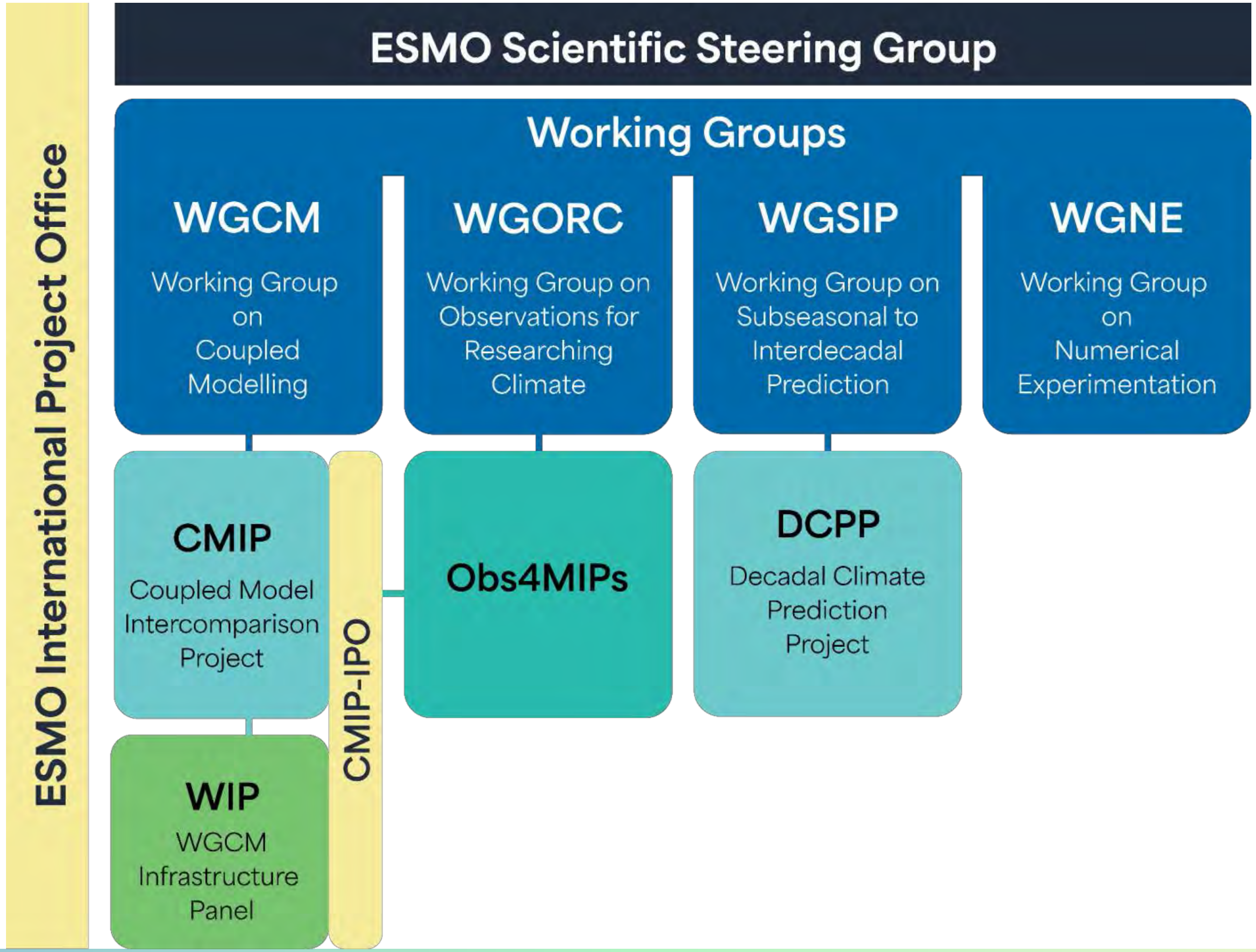
Requires an **integrated** and **consistent framework** combining global Earth system observations, data assimilation and modelling.

These objectives will be addressed in collaboration with the WCRP core projects, Lighthouse Activities and Working Groups.

Working Group on Observations for Researching Climate (WGORC)

- Co-chairs:
  - Douglas Rao (NC State University)
  - Amy Doherty (Met Office)
- ESMO SSG Liaison:
  - Claire MacIntosh (ESA ECSAT)
  - Alison Cobb (ECMWF)

- Coordinate research and development at the intersection of observation and modeling interface.



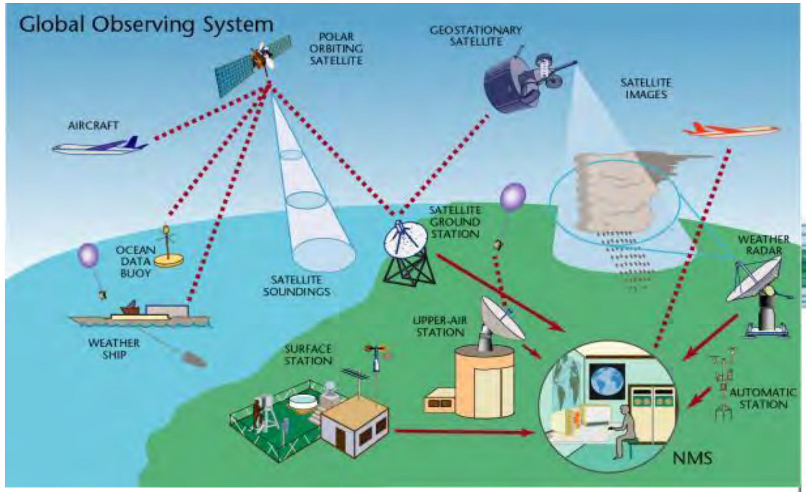
Machine Learning in Earth System Modeling & Observation

Quality Control

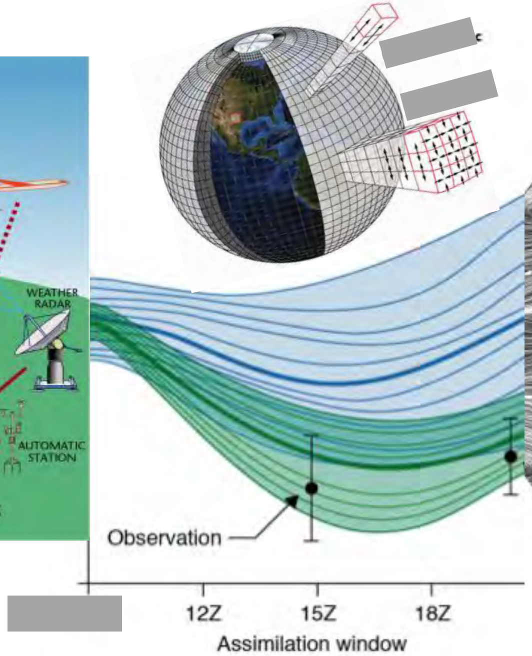
Data fusion

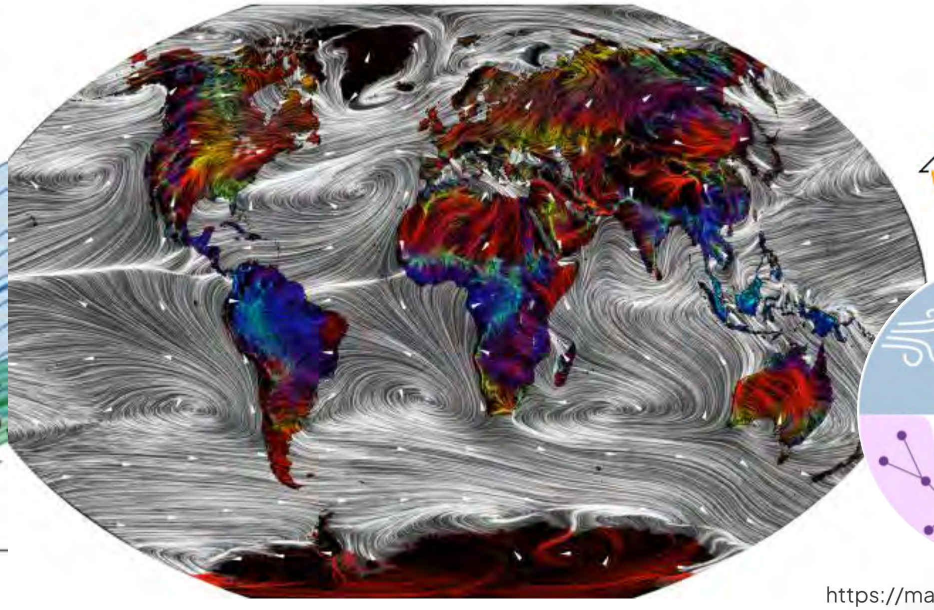
Assimilation

Emulation



WMO Global Observing system





https://machineclimate.de/

Cambridge University Press Climate Model and ECMWF Data Assimilation schematic

Kling and Ackerly, 2020

Reconstruction

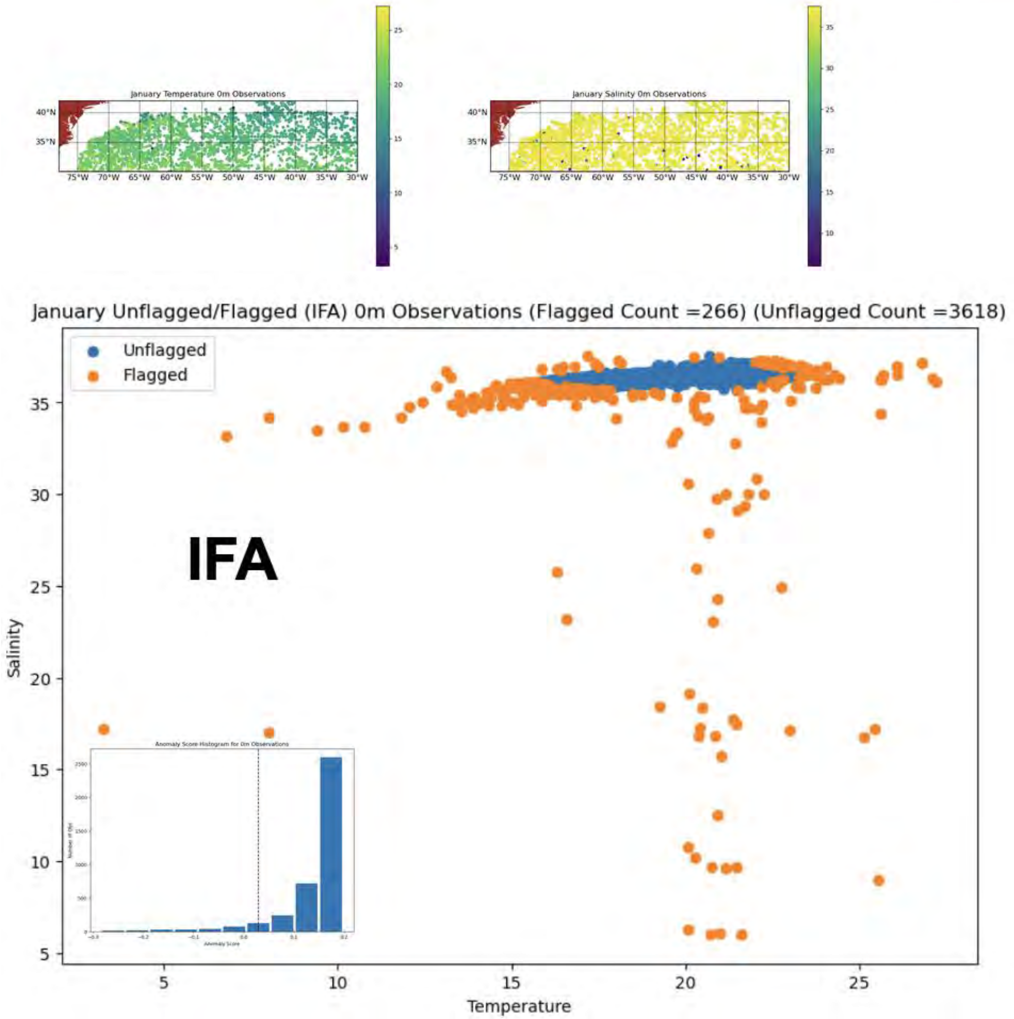
Parameterization

Data-driven simulation

ML for Data Quality Assurance

Source: Jim Reagan

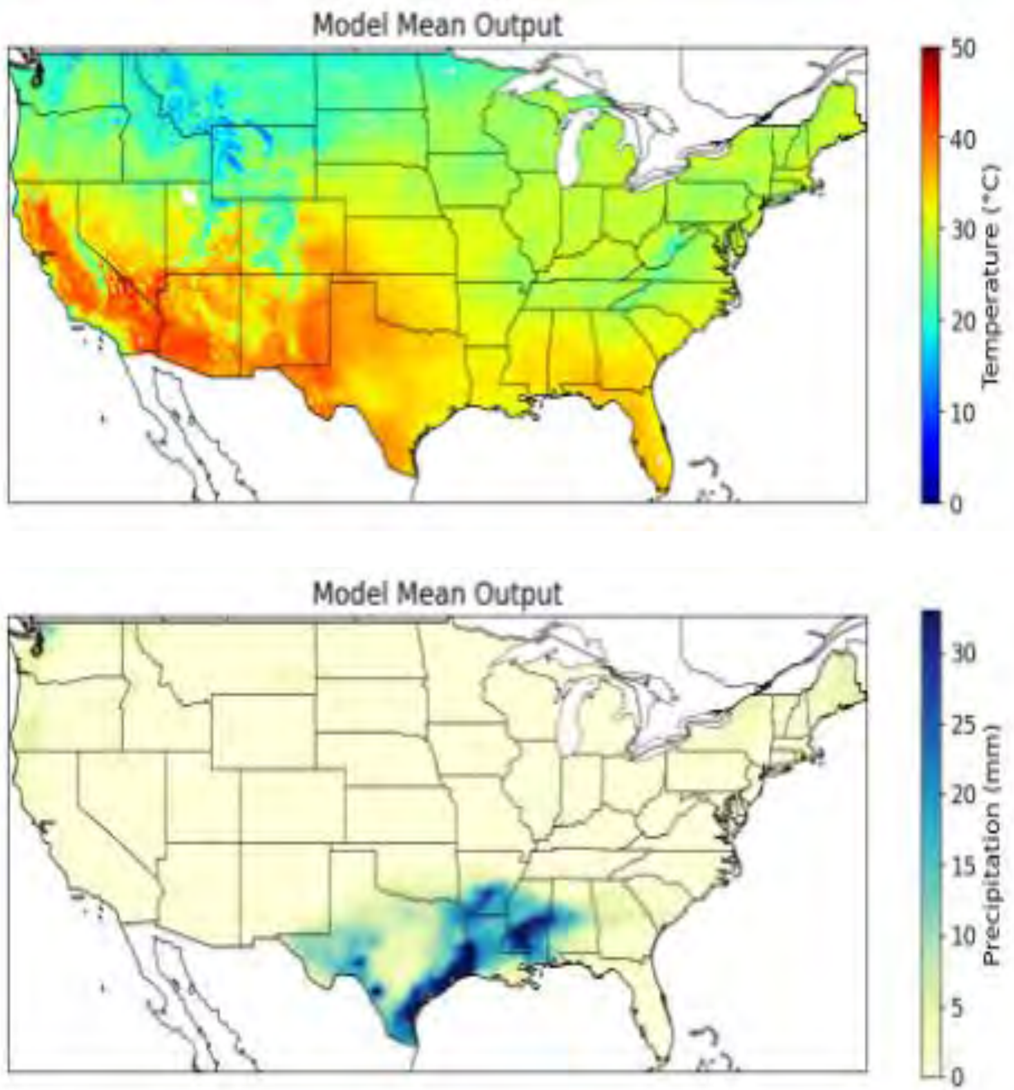
- Machine learning is skillful in learning from data patterns and identifying anomalous data in time series.
- Common use of ML for QA
  - Anomaly detection
  - Change point detection
  - Bias correction
- Example pilot use cases:
  - US Climate Reference Network
  - Ocean profiles data



ML for Data Reconstruction

Source: Douglas Rao

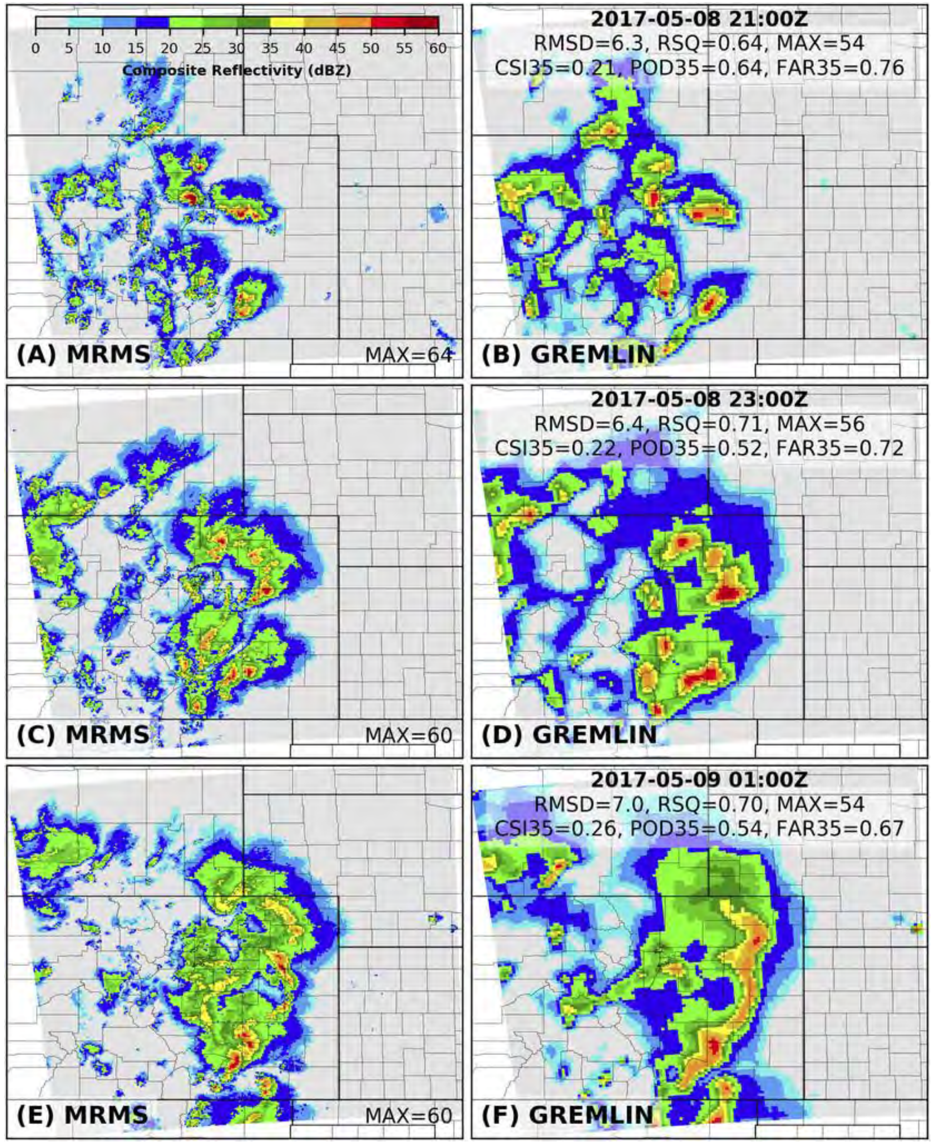
- ML's ability to represent nonlinear relationship in data could benefit climate data reconstruction.
- Example use cases:
    - NOAA GlobalTemp (v6.0): https://doi.org/10.25921/rzxg-p717
    - ERSST (v6.0): https://doi.org/10.1175/JCLI-D-23-0707.1
    - High-resolution gridded climate dataset (right figure)



ML for Climate Data Fusion

- Machine learning can be used to bring different observations (e.g., *in situ* & satellite) together for better climate monitoring.
- Increasingly, ML has been used to generate synthetic data where traditional observations are unavailable.

Actual radar data Synthetic radar data



Continuum of Physics-based & ML-based modeling

Machine Learning

Hybrid Model

Physics-based Model

cBottle (2025)

ACE (2025)

NeuralGCM (2024)

Pangu Wx (2023)

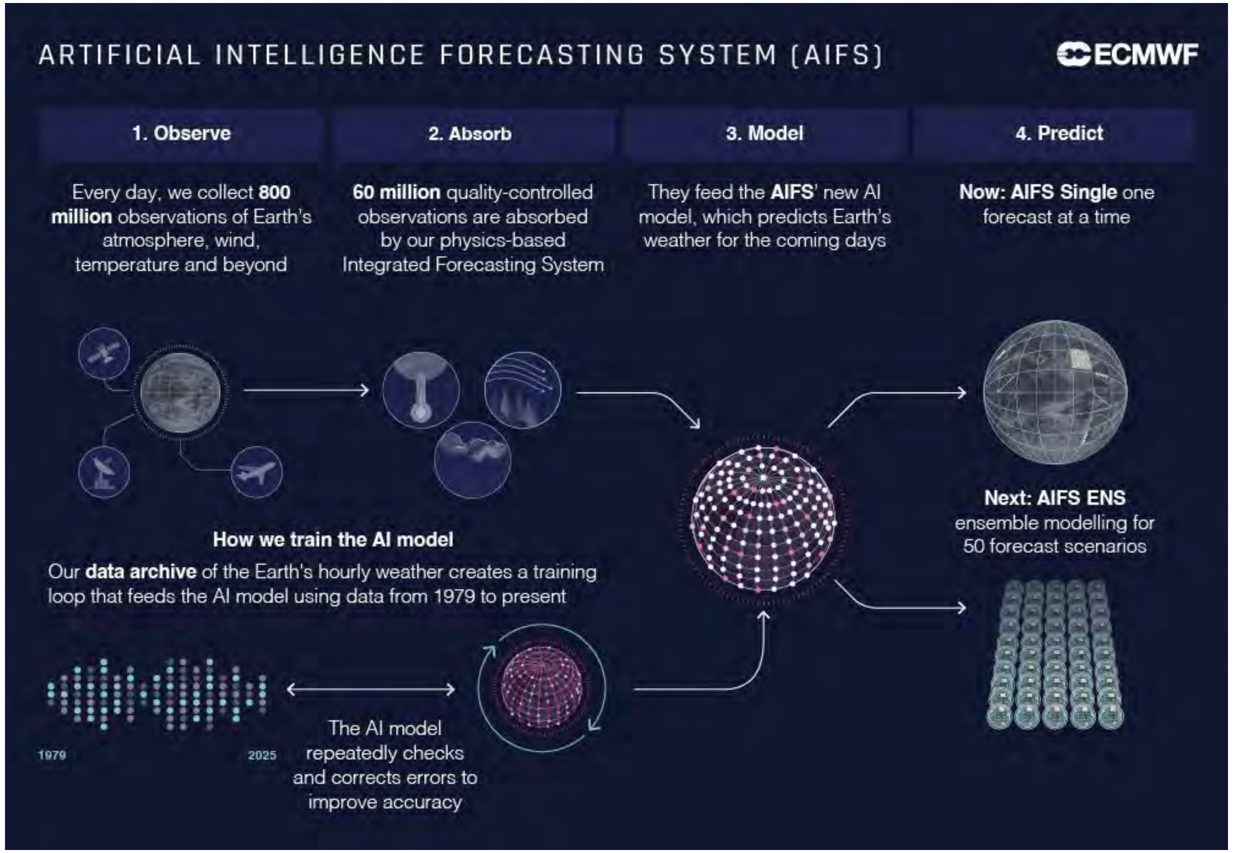
FourCastNet (2023)

GraphCast (2024)

ML-enabled weather and climate models are rapidly evolving and adopted by global modeling centers.

ECMWF's AI forecasts become operational

25 February 2025

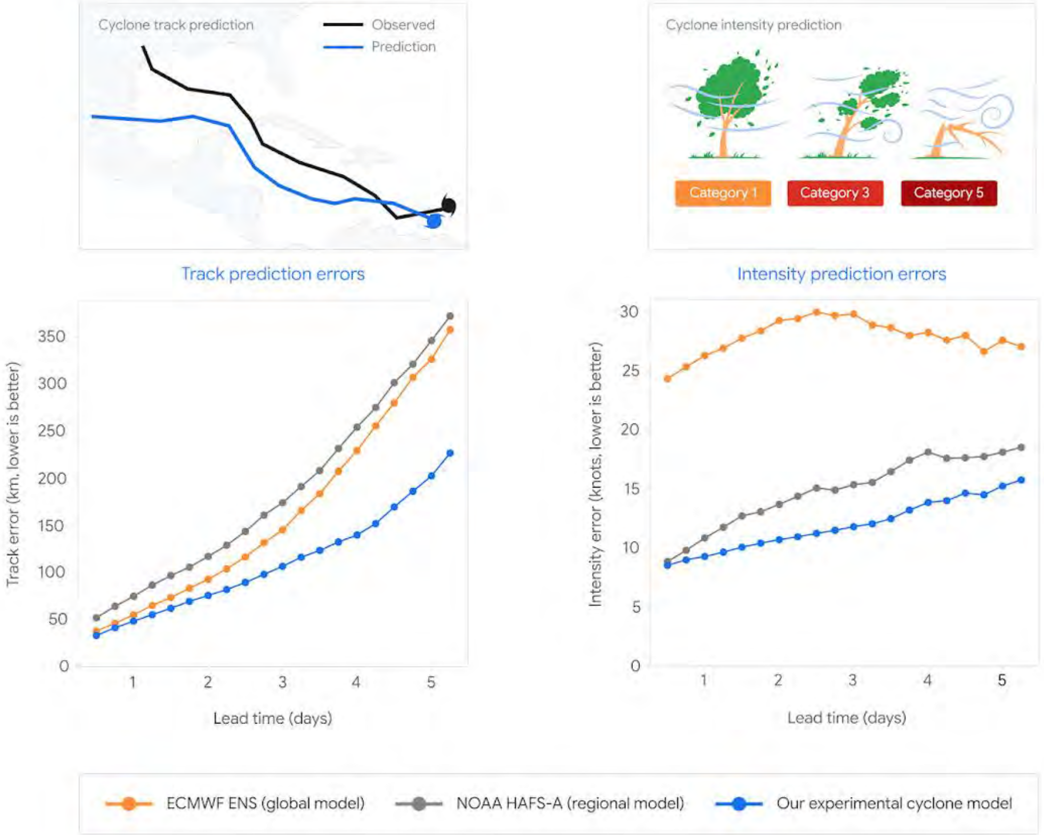


Source: ECMWF

RESEARCH

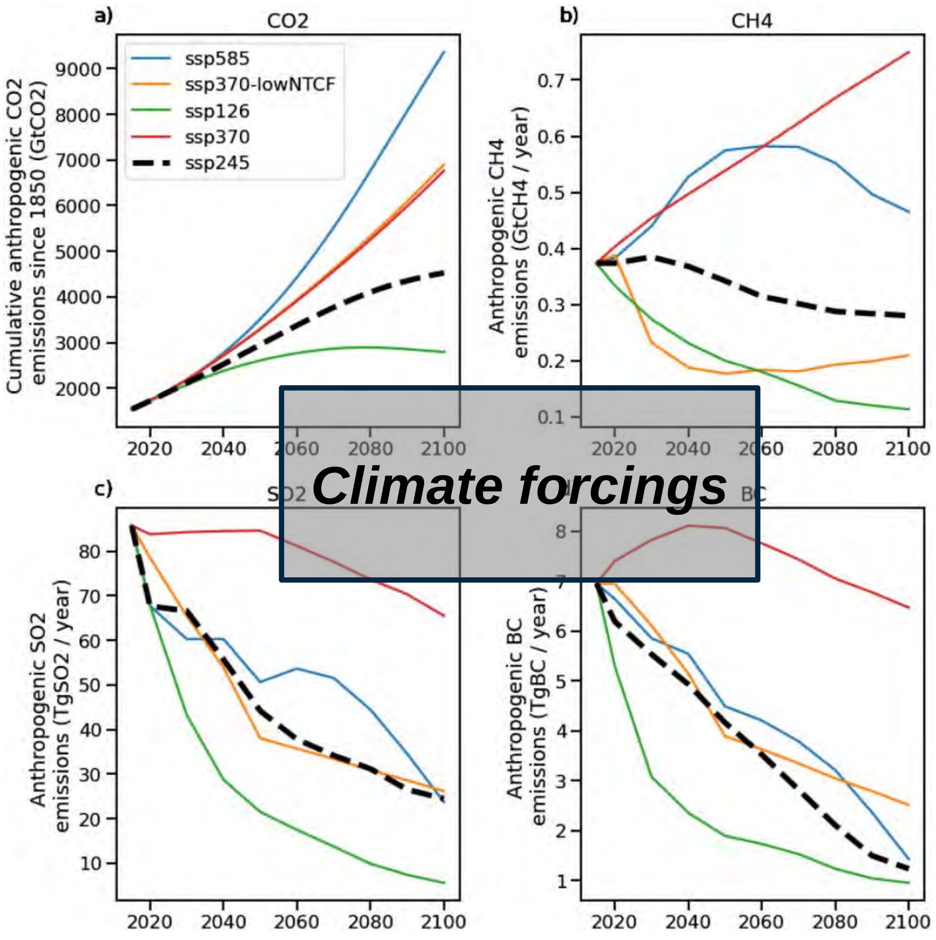
How we're supporting better tropical cyclone prediction with AI

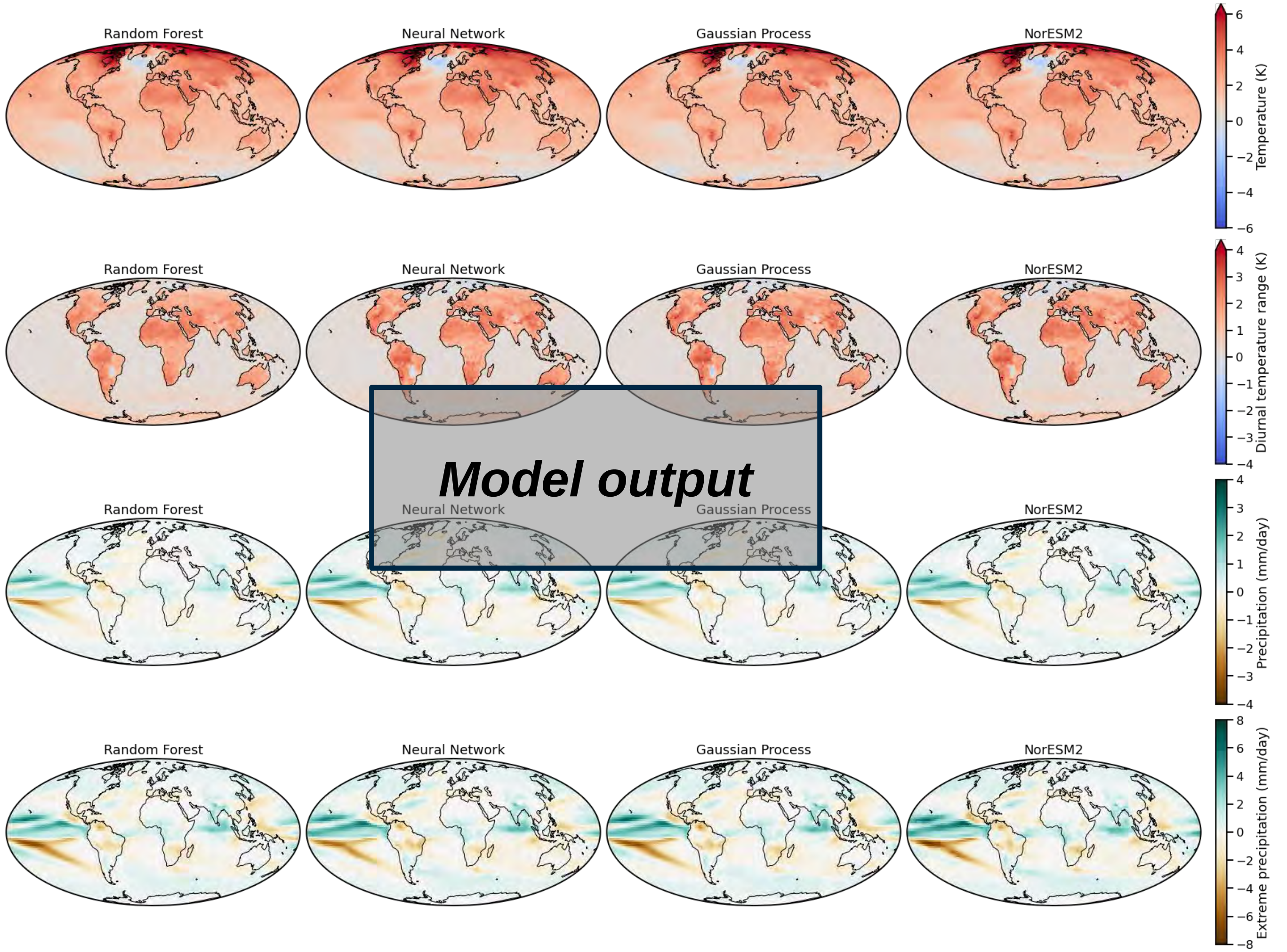
12 JUNE 2025



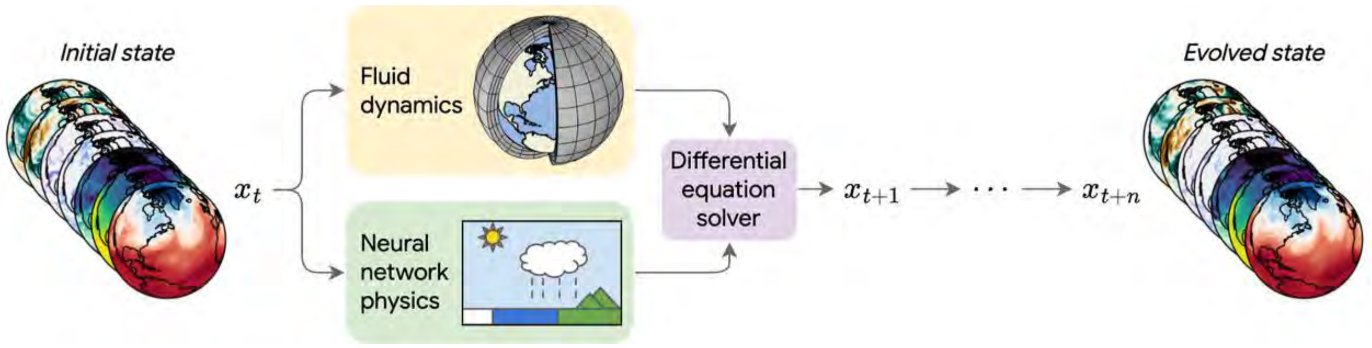
Source: Google Research

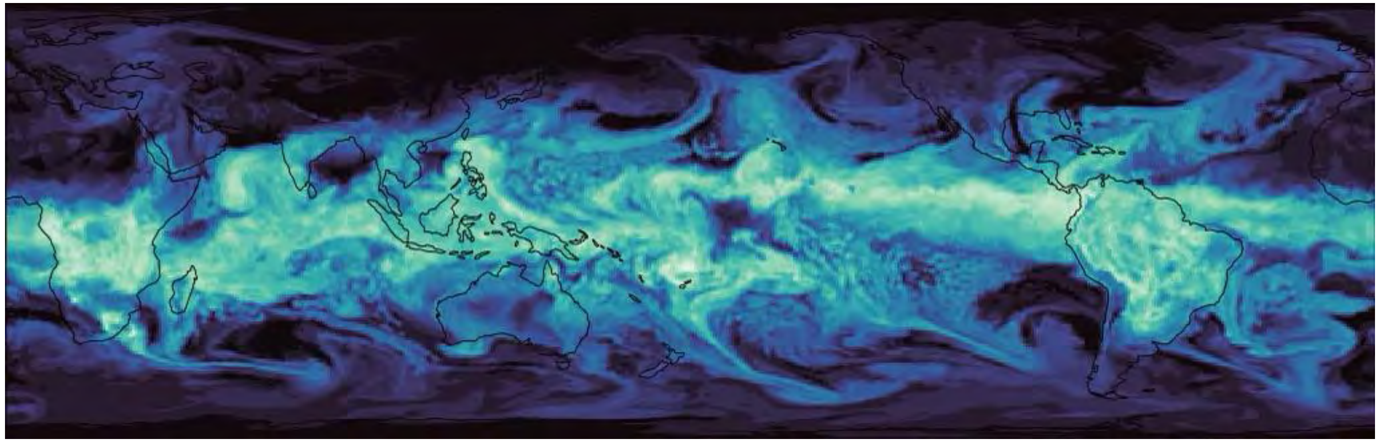
Climate Model Emulation

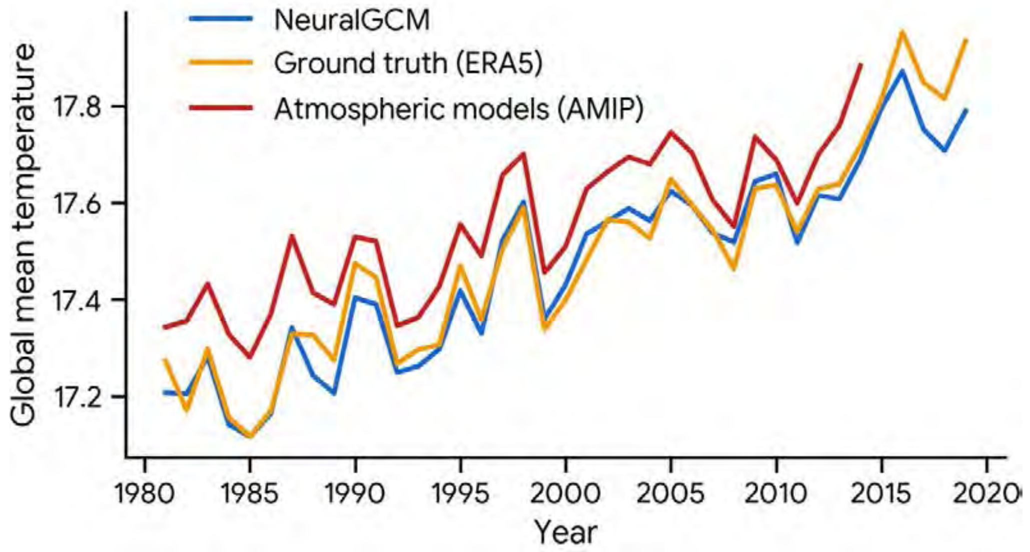




Data-driven & Hybrid Modeling

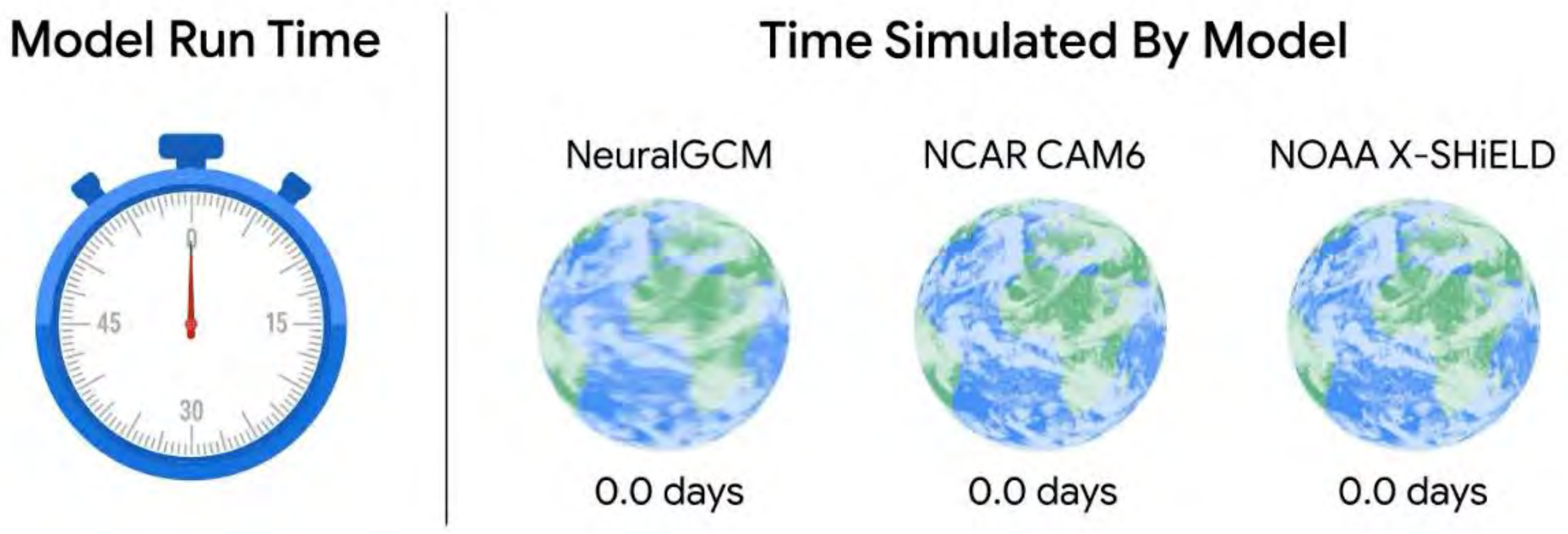






Source: Stephen Hoyer (2024)
https://doi.org/10.1038/s41586-024-07744-y

Data-driven & Hybrid Modeling

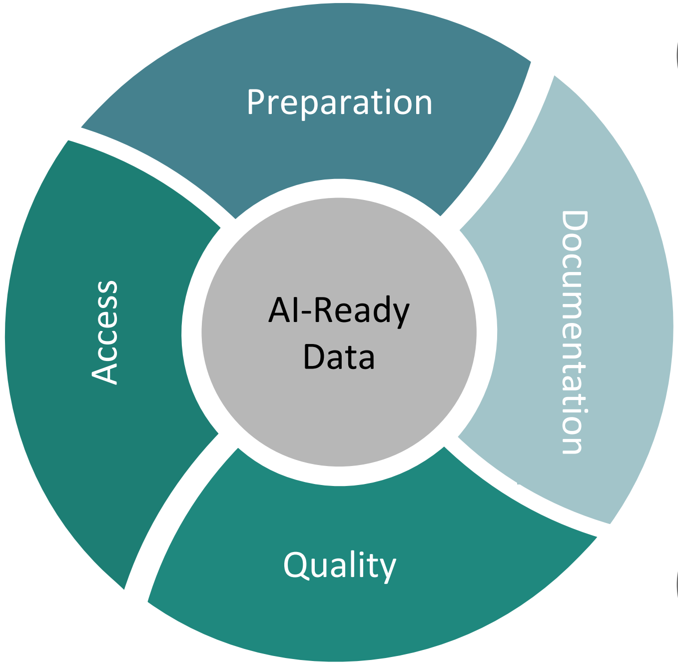


Source: Google Research
https://research.google/blog/fast-accurate-climate-modeling-with-neuralgcm/

Issues to Consider for Data Stewardship

- ML could and should be leveraged to improve the climate data for various applications.
- ***Governance*** of the ML-based climate data is challenged by the increasing volume, diversity, and producers.
- ***Reproducibility*** of ML-based climate data is still an issue.
- **Uncertainty** quantification of ML-based climate data is less mature.
- ML-based climate data is fundamentally limited by existing climate datasets because of the model training workflow.

Modernizing Data Management for AI/ML



Documentation enables users to understand the contents of a data set and should provide information and tools to ***increase data usability***.

Quality* information** determines the dataset's "fit-for-purpose***" for AI & affects the trustworthiness of AI applications.

Access affects the ***efficiency & reproducibility*** for AI R&D process.

Preparation help identify common ***data services and tools*** to reduce preprocessing burden for users including AI practitioners.

Earth Science Information Partners AI-ready data checklist (https://github.com/ESIPFed/data-readiness) is developed as a framework to assess the readiness level of open environmental data for AI applications.

ESMO

Earth System Modelling and Observations

WCRP

World Climate Research Programme

Thank you!

Contacts:

Douglas Rao (NC State University)

yrao5@ncsu.edu

wcrp-esmo.org