Earth system models of the future

Peter Dueben

Head of the Earth System Modelling Section

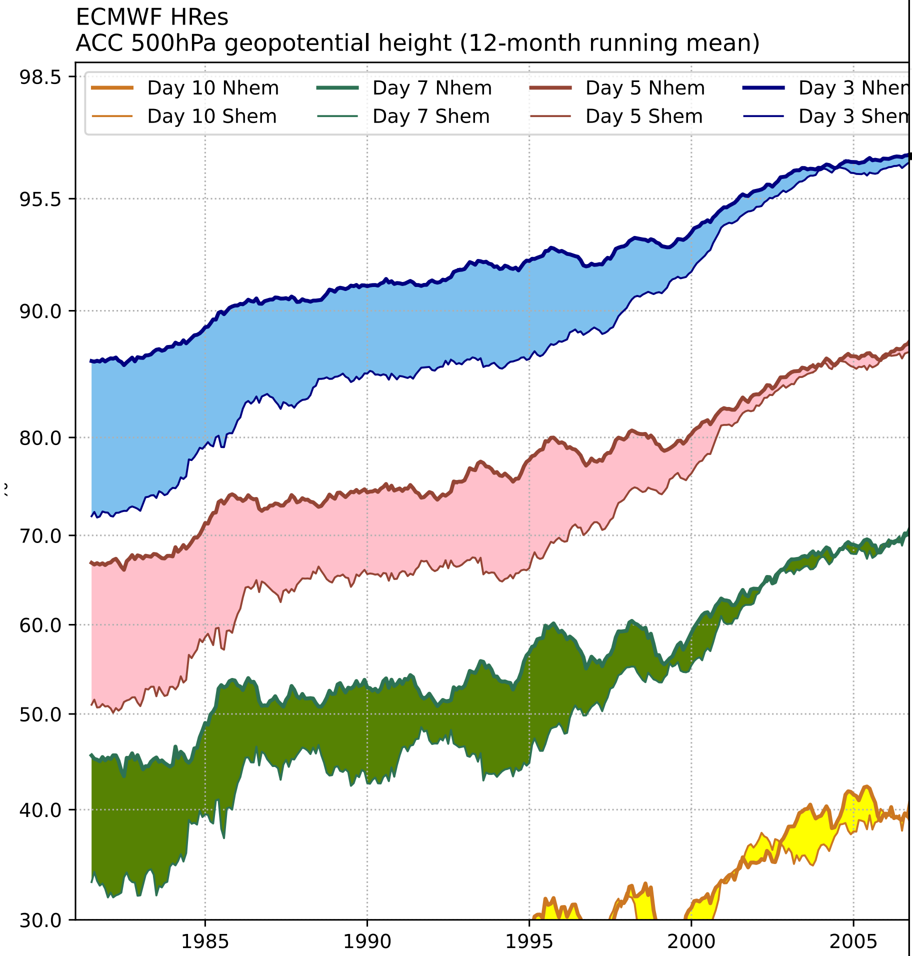
The strength of a common goal

Funded by the European Union

The MAELSTROM and ESiWACE projects have received funding from the EuroHPC-Joint Undertaking under grant agreement No 955513 and 101093054.

Earth system modelling is currently experiencing disruptive changes offering great opportunities.

1980-2020: The quiet revolution



ECMWF

nature

Explore content | About the journal | Publish with us | Subscribe

nature > review articles > article

Published: 02 September 2015

The quiet revolution of numerical weather prediction

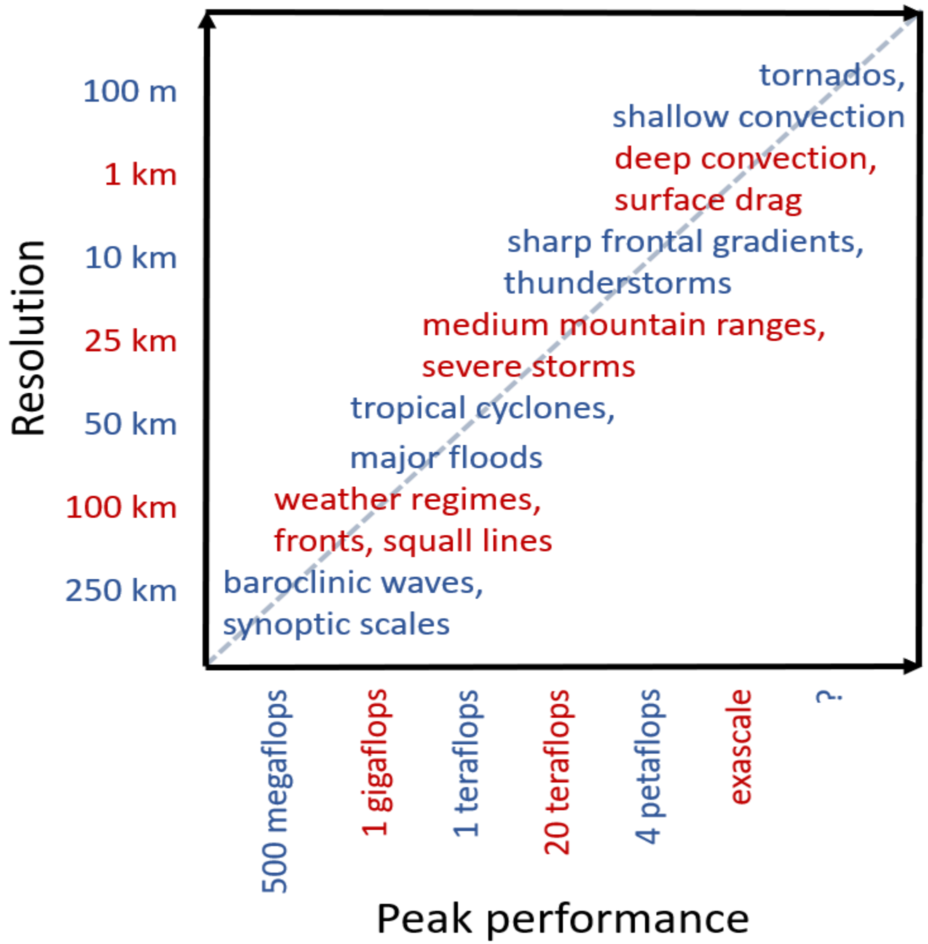
Peter Bauer, Alan Thorpe & Gilbert Brunet

Nature **525**, 47–55 (2015) | Cite this article

Abstract

Advances in numerical weather prediction represent a quiet revolution because they have resulted from a steady accumulation of scientific knowledge and technological advances over many years that, with only a few exceptions, have not been associated with the aura of fundamental physics breakthroughs. Nonetheless, the impact of numerical weather prediction is among the greatest of any area of physical science. As a computational problem, global weather prediction is comparable to the simulation of the human brain and of the evolution of the early Universe, and it is performed every day at major operational centres across the world.

Km-scale models for better predictions

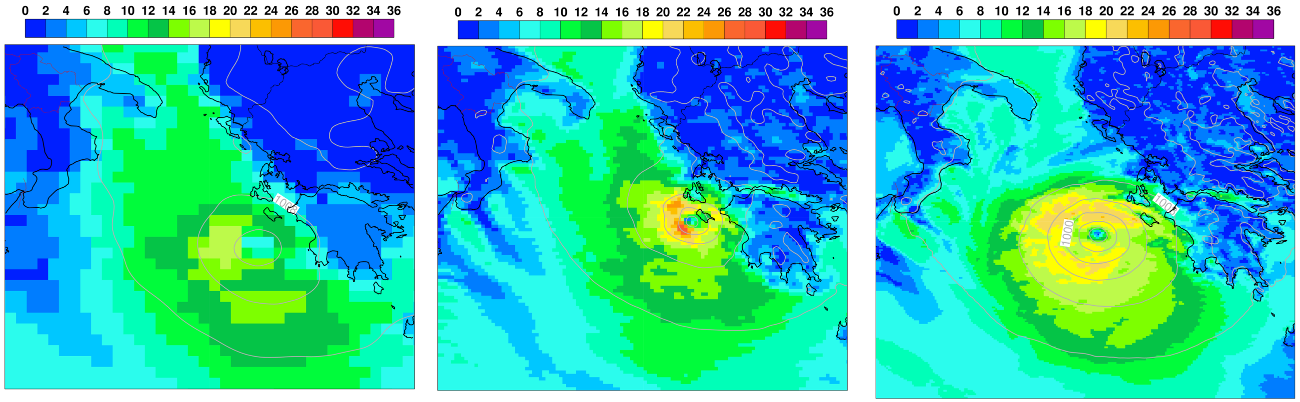


Adapted from Neumann, Dueben et al. Phil Trans A 2018

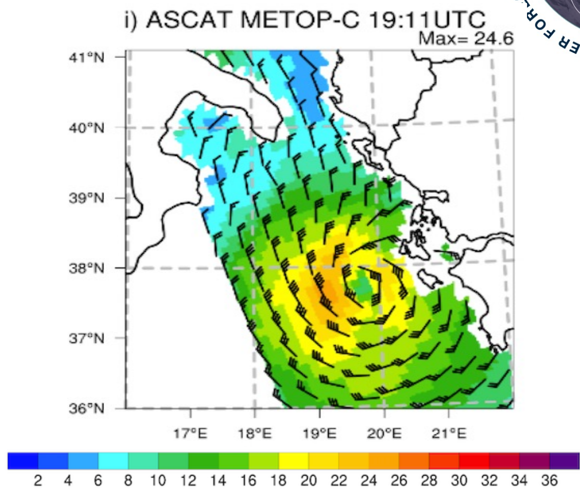
- More resolution – more skill
- Better representation of topography and gravity wave drag
- Explicit representation of convection (“storm-resolving” models)
- Eddy resolving oceans + tides
- Same resolution as satellite measurements

Km-scale models are great -- see Medicane Ianos (Sep 2020)

Surface wind speed (T+66h, m/s)

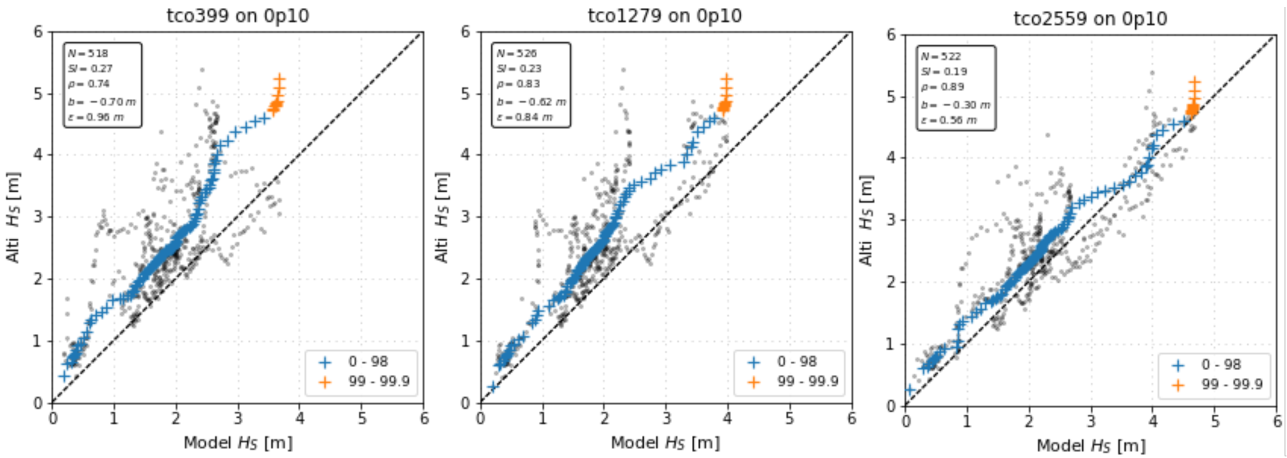


IFS 48r1 29 km **IFS 48r1 9 km** **IFS 48r1 4.5 km**



Observations

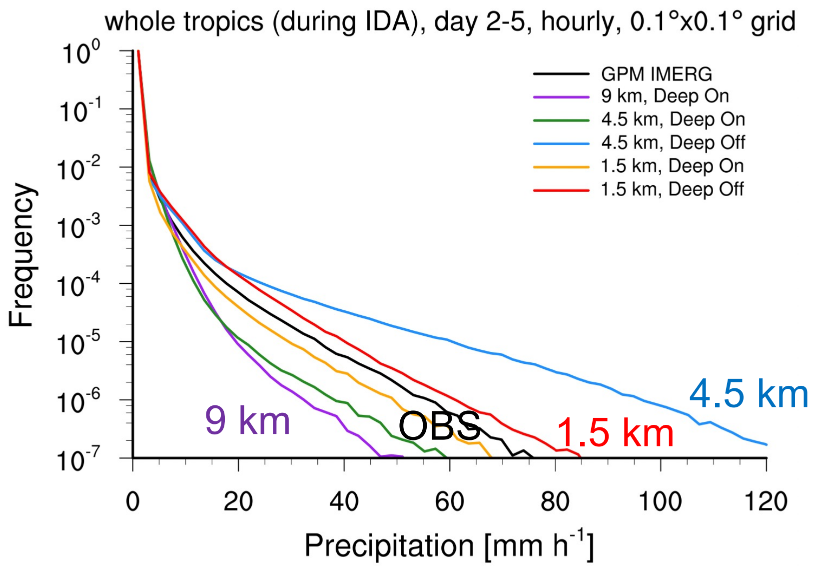
Waves (T+44h – T+84h) : Model vs. Observations (altimeter)

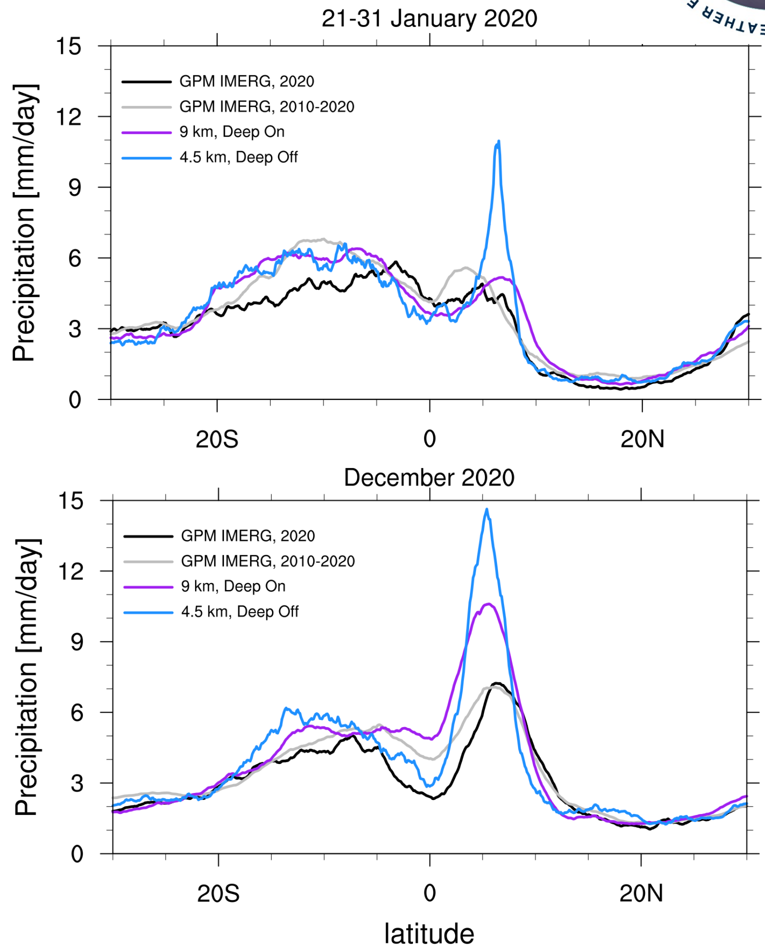


So let's push km-scale models to operations!? – Not out of the box

Scientific developments will be needed to make the most of km-scale models.

Precipitation (extremes, MCS, etc)





So let's push km-scale models to operations!?

Climate is changing,
→ we need better models now

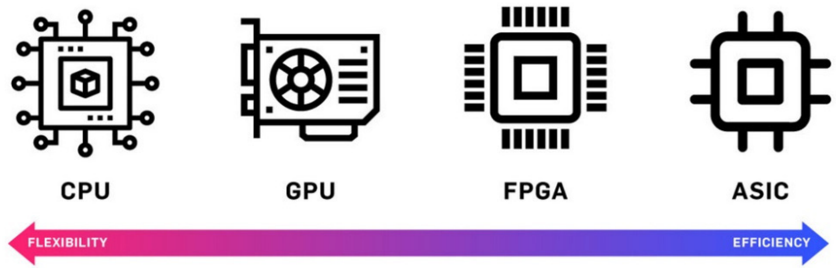
Compute power?
9 km → 1 km → Factor 9^3 = 729 compute power

Moore's law is the observation that the number of transistors in an integrated circuit doubles about every two years.
→ 2^9 = 512 → Let's wait for 18 years?

Data and storage?
9km: 6,599,680 points x 137 levels x10 variables
→ 9 billion points → > 0.5 TB

1.5km: 256,800,000 points x 137 levels x 10 variables
→ 352 billion points → > 20 TB

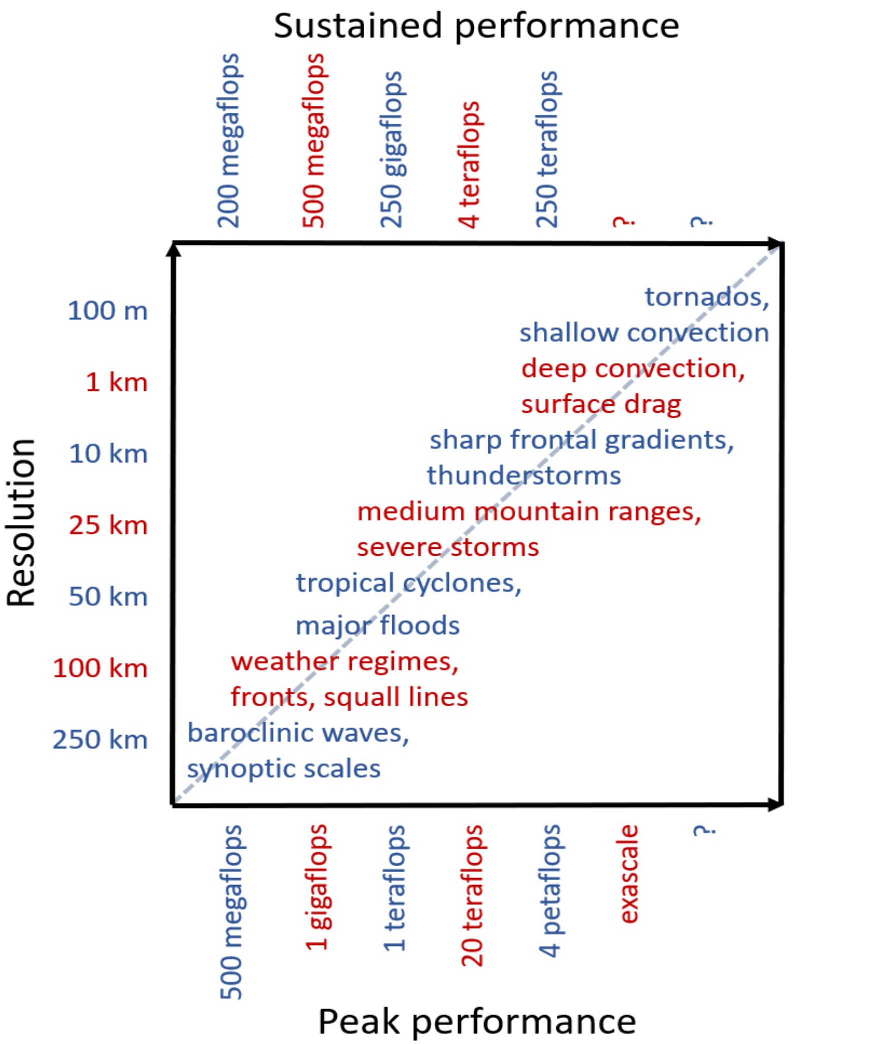
Uff…



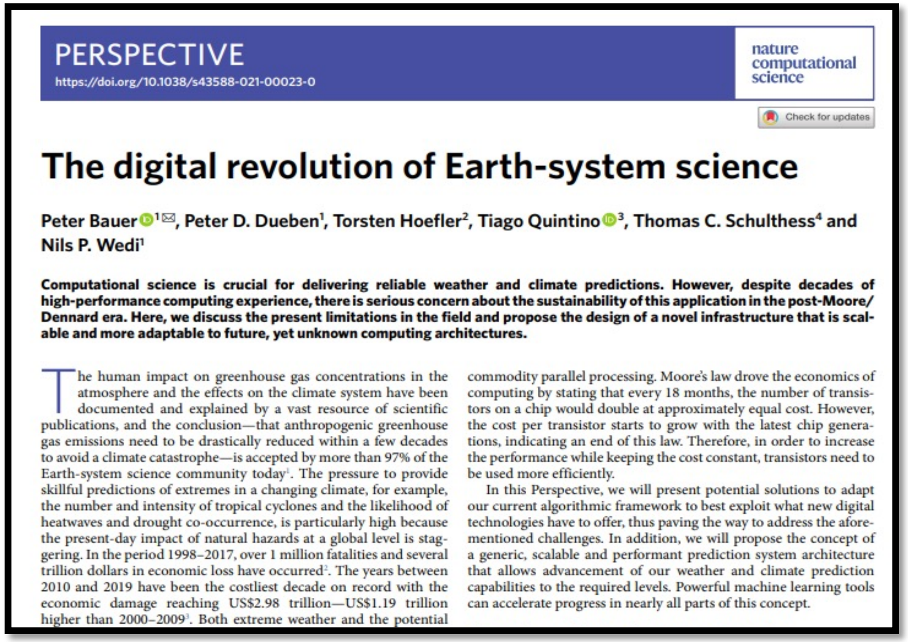
Source: venturebeat.com

- Individual processors will not be faster
  → Parallelisation / power consumption
- Hardware will be more heterogeneous
  → CPUs / GPUs / FPGAs / ASICs
- Machine learning has strong impact on hardware development
  → High floprate at low precision

2015-today: The digital revolution



Adapted from Neumann, Dueben et al. Phil Trans A 2018



PERSPECTIVE

https://doi.org/10.1038/s43588-021-00023-0

nature computational science

Check for updates

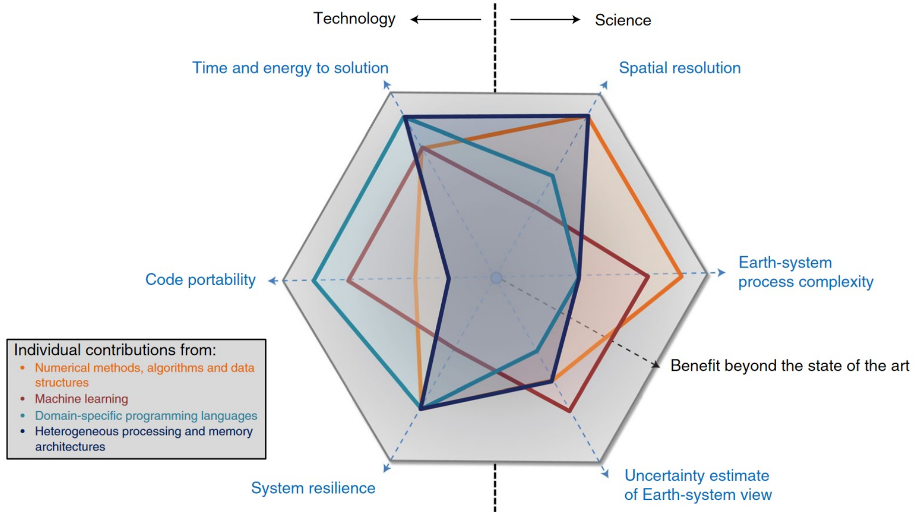
The digital revolution of Earth-system science

Peter Bauer[1], Peter D. Dueben[1], Torsten Hoefler[2], Tiago Quintino[3], Thomas C. Schulthess[4] and Nils P. Wedi[1]

Computational science is crucial for delivering reliable weather and climate predictions. However, despite decades of high-performance computing experience, there is serious concern about the sustainability of this application in the post-Moore/Dennard era. Here, we discuss the present limitations in the field and propose the design of a novel infrastructure that is scalable and more adaptable to future, yet unknown computing architectures.

The human impact on greenhouse gas concentrations in the atmosphere and the effects on the climate system have been documented and explained by a vast resource of scientific publications, and the conclusion—that anthropogenic greenhouse gas emissions need to be drastically reduced within a few decades to avoid a climate catastrophe—is accepted by more than 97% of the Earth-system science community today[1]. The pressure to provide skillful predictions of extremes in a changing climate, for example, the number and intensity of tropical cyclones and the likelihood of heatwaves and drought co-occurrence, is particularly high because the present-day impact of natural hazards at a global level is staggering. In the period 1998–2017, over 1 million fatalities and several trillion dollars in economic loss have occurred[2]. The years between 2010 and 2019 have been the costliest decade on record with the economic damage reaching US$2.98 trillion—US$1.19 trillion higher than 2000–2009[3]. Both extreme weather and the potential commodity parallel processing. Moore's law drove the economics of computing by stating that every 18 months, the number of transistors on a chip would double at approximately equal cost. However, the cost per transistor starts to grow with the latest chip generations, indicating an end of this law. Therefore, in order to increase the performance while keeping the cost constant, transistors need to be used more efficiently.

In this Perspective, we will present potential solutions to adapt our current algorithmic framework to best exploit what new digital technologies have to offer, thus paving the way to address the aforementioned challenges. In addition, we will propose the concept of a generic, scalable and performant prediction system architecture that allows advancement of our weather and climate prediction capabilities to the required levels. Powerful machine learning tools can accelerate progress in nearly all parts of this concept.



EU’s Destination Earth (DestinE) initiative

Towards a Digital Twin Earth

simulations

(ECMWF IFS 1.4 km)

observations

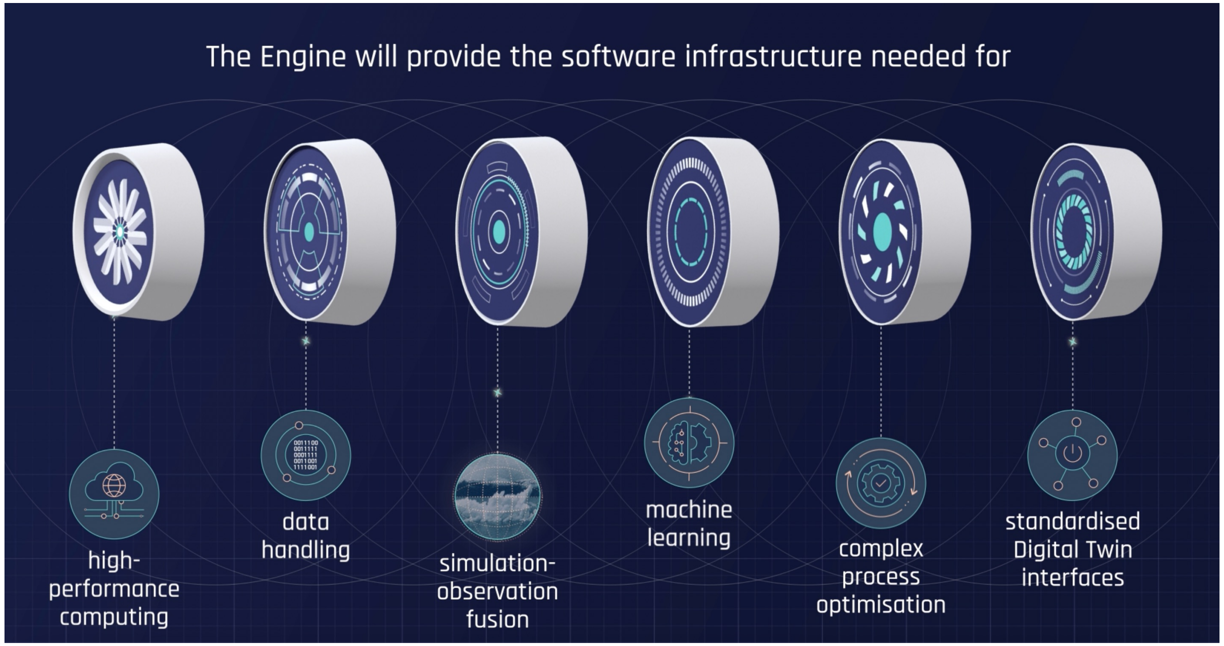
Funded by the European Union

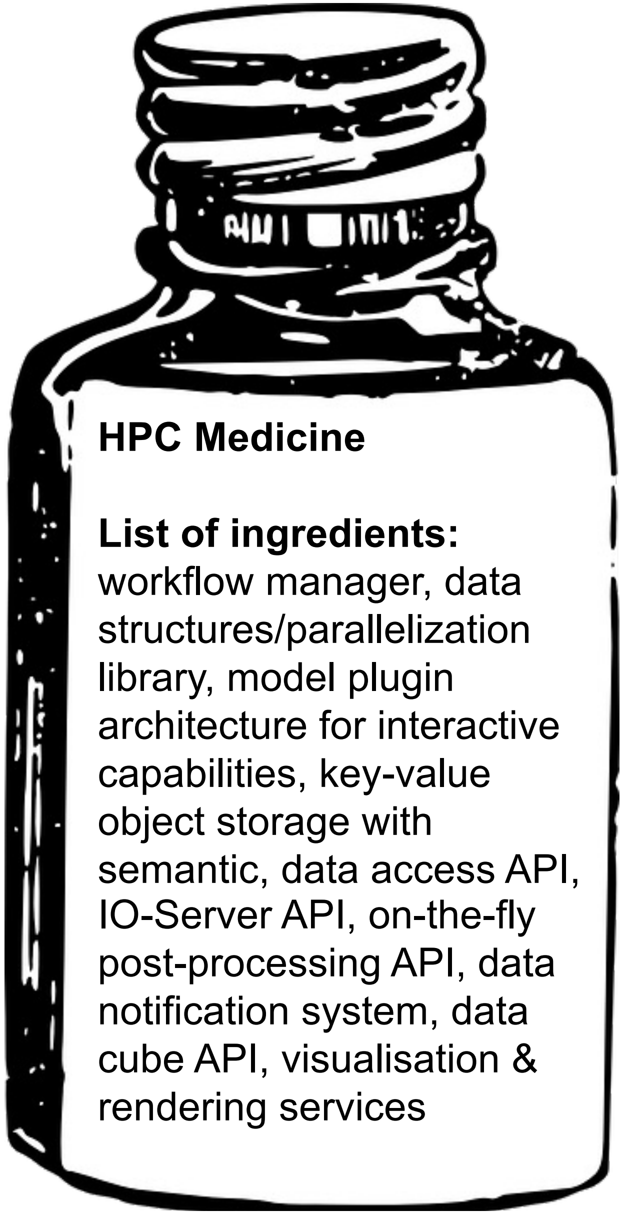
DestinE's Digital Twin Engine

Framework for Digital Twin Workflows

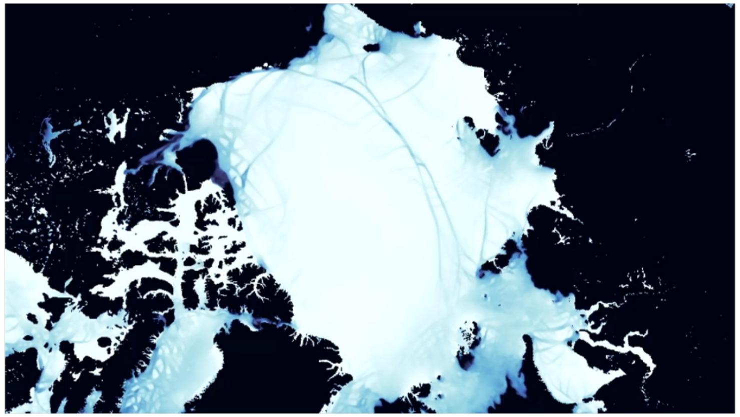
- High Performance Computing adaptation / Digital Twin optimisation
- IO and data workflows
- Software management, controlling workflows, cloud environments
- Visualization

A Game Engine type framework but for Earth Systems…

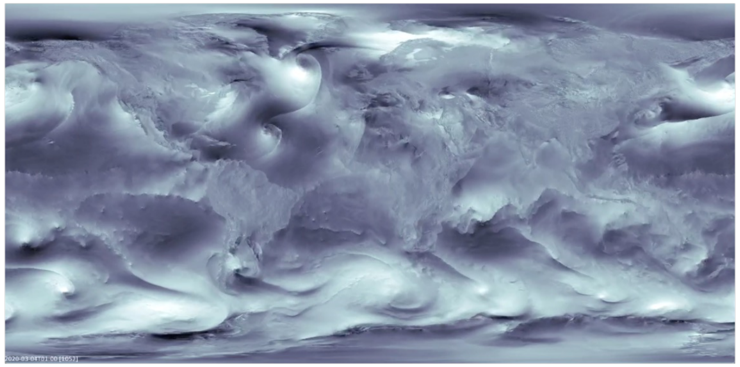




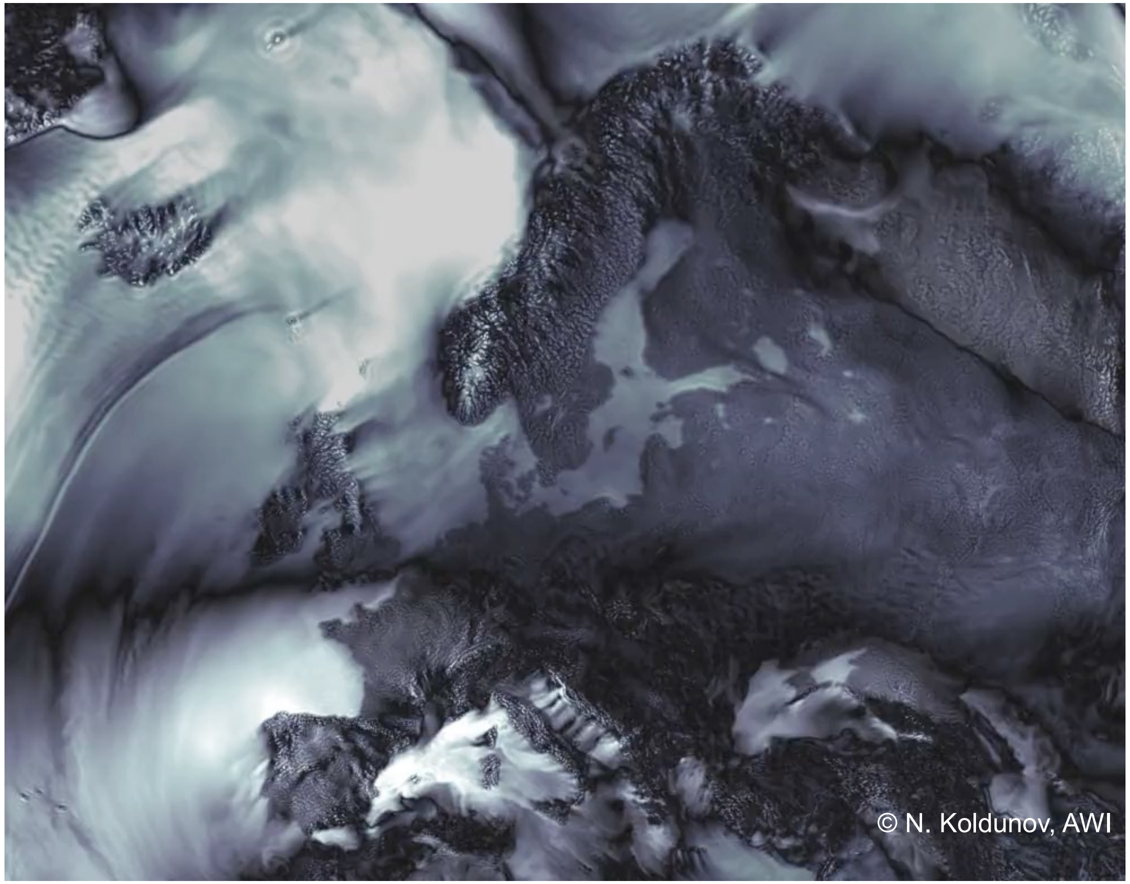
2015-today: The digital revolution to allow for km-scale models



More realistic at local scale



More realistic at global scale



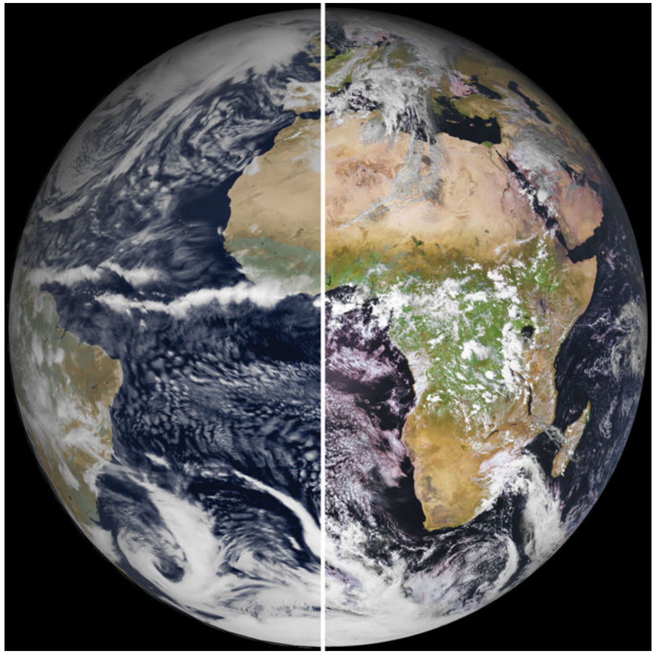
Better results via a coupled model system

Global km-scale models will improve realism of simulations significantly and are now becoming available.

Machine learned weather models

Explore the space of machine learning for weather and climate modelling

Improve understanding
- Fuse information content from different datasources
- Unsupervised learning
- Causal discovery
- AI powered visualisation
- Uncertainty quantification
- …



Speed up simulations
- Emulate model components
- Port emulators to heterogeneous hardware
- Use reduced numerical precision and sparse machine learning
- Optimise HPC and data workflow
- Data compression
- …

Improve models
- Learn components from observations
- Correct biases
- Quality control of observations and observation operators
- Feature detection
- …

Link communities
- Health – e.g. for predictions of risks
- Energy – e.g. for local downscaling
- Transport – e.g. to combine weather and IoT data
- Pollution – e.g. to detect sources
- Extremes – e.g. to predict wild fires
- …

Weather and climate modelling centres mostly explore the "hybrid space" coupling machine learning to conventional models.

Can we replace conventional Earth System models by deep learning?

If Hamlet were a musical comedy, rather than a tragedy, how would it have ended? Describe the plot.

If Hamlet were a musical comedy, rather than a tragedy, the ending would likely be much happier and more lighthearted. Instead of ending in a violent and tragic showdown, with the deaths of several major characters, the play would likely have a more comedic and satisfying resolution.

For example, the play could end with Hamlet successfully avenging his father's death, but then choosing to forgive and reconcile with his uncle, the new king. This would allow Hamlet to put aside his grudges and bitterness, and to find peace and happiness.

As part of this resolution, there could be a big, energetic finale song, where the characters sing and dance together, expressing their newfound sense of joy and unity. This could be a lively and upbeat number, with a catchy and infectious melody, and lyrics that reflect the themes and events of the play, while also expressing a sense of hope and optimism for the future.

Overall, a musical comedy version of Hamlet would likely have a more uplifting and optimistic ending, with the main character achieving a sense of closure and resolution, and moving on with his life in a positive and fulfilling way.

The New York Times

THE SHIFT

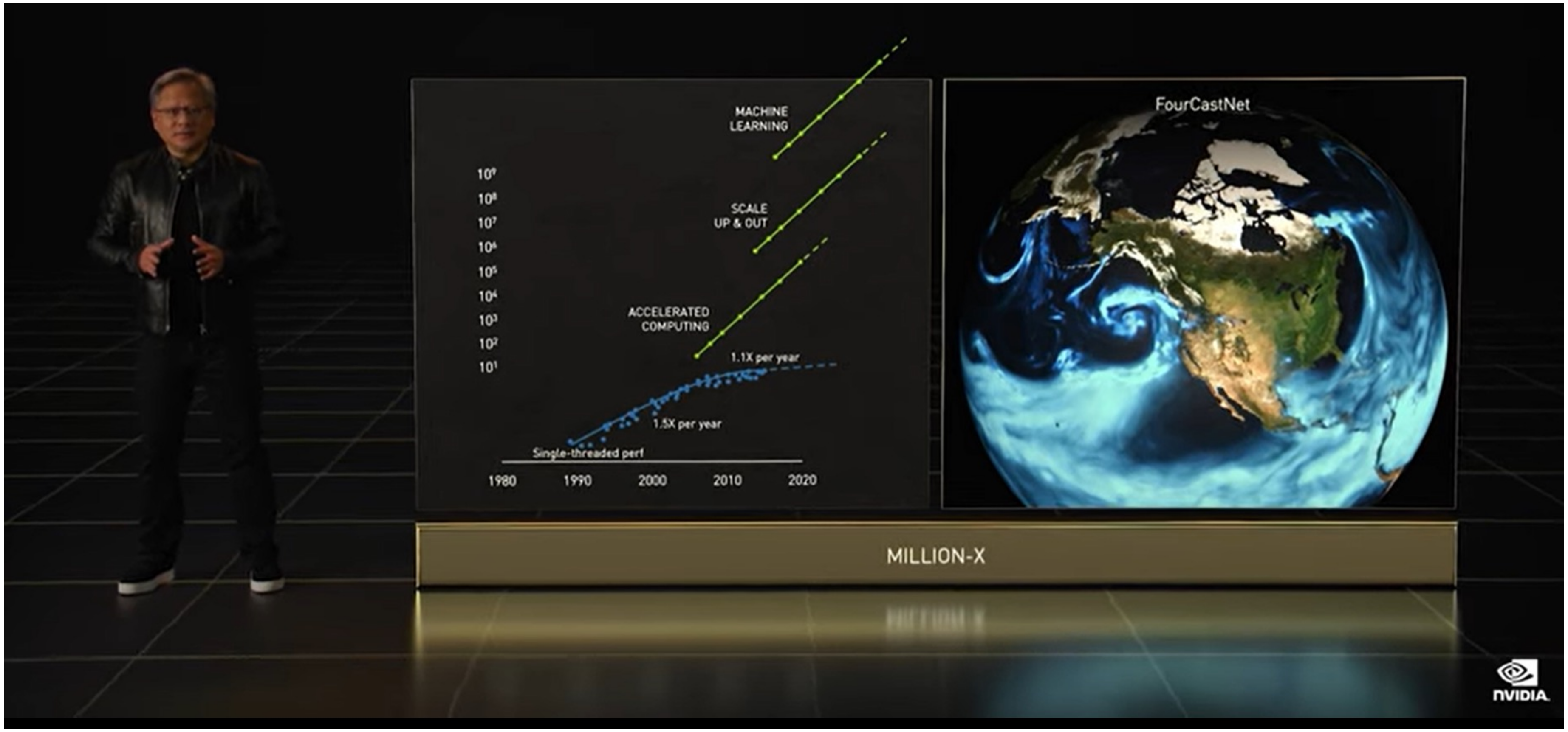
An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy.

"I won, and I didn't break any rules," the artwork's creator says.

Give this article 1.5K

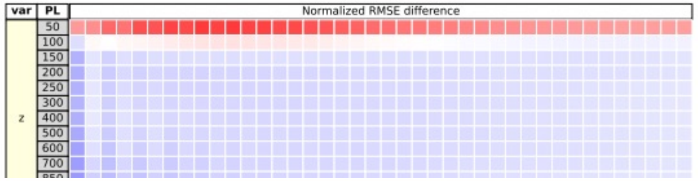
Jason Allen's A.I.-generated work, "Théâtre D'opéra Spatial," took first place in the digital category at the Colorado State Fair. via Jason Allen

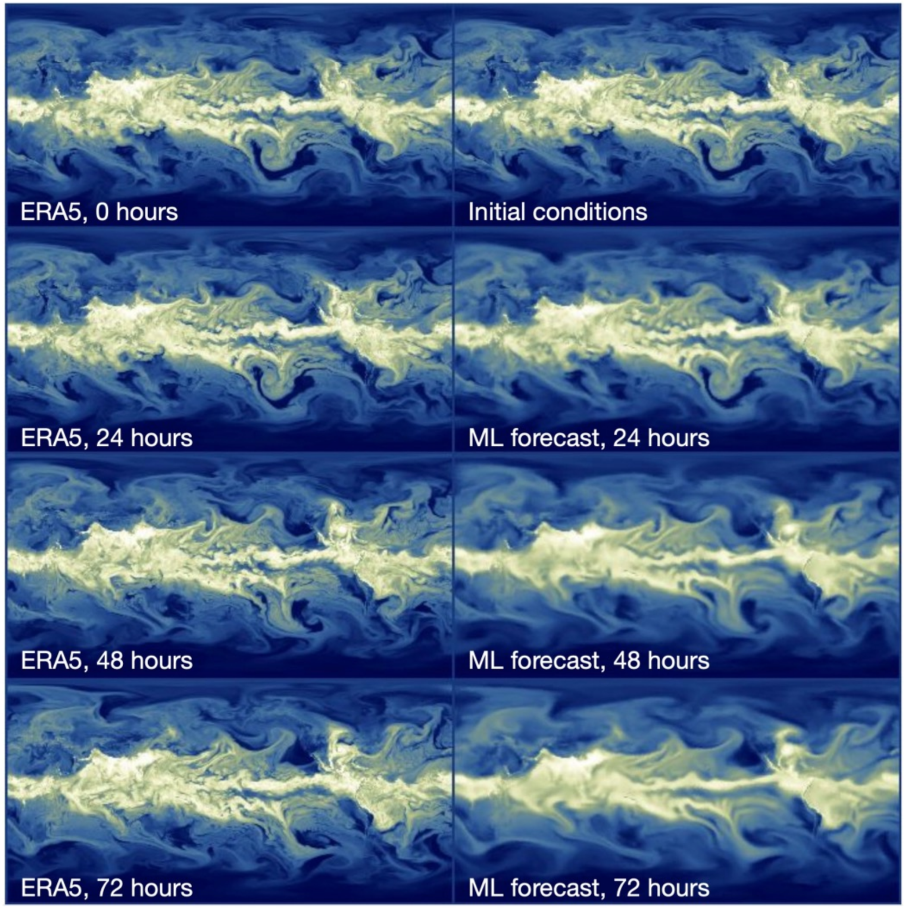
Can we replace conventional Earth System models by deep learning?



NIVIDA’s Earth-2 is coming with FourCastNet

2022-today: The machine learning revolution





GraphCast from Google/Deepmind and Fourcastnet from NVIDIA are beating conventional weather forecast model in deterministic scores and are orders of magnitudes faster.

But how do these models actually work?

They get the best results when using very large timesteps.

They are trained for a small Root Mean Square Error.
→ They smear out for large lead times.

Many questions remain:
Can the models extrapolate?
Can they represent extreme events?
Can they learn uncertainty?
Can they be trained from observations?
Can they represent physical consistency?

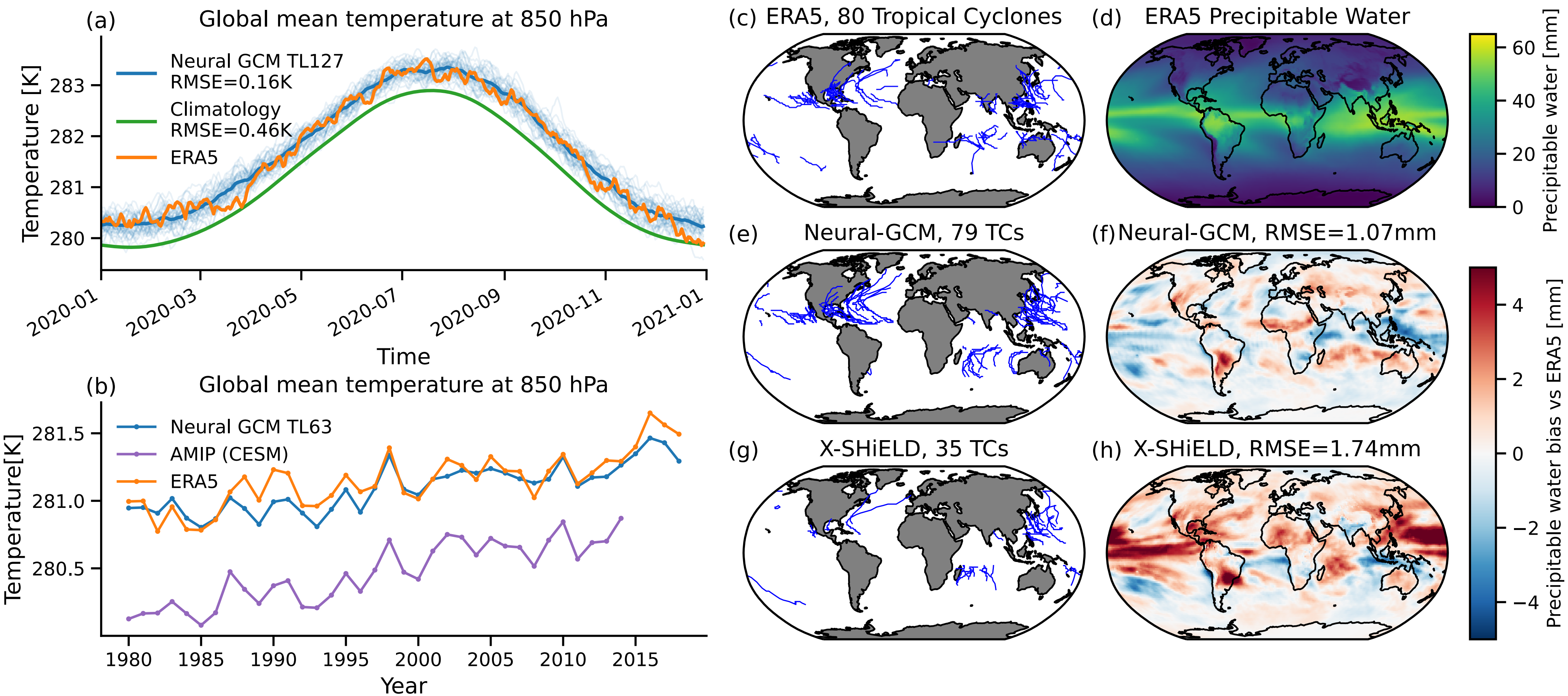
Images from Keisler (2022)

2022-today: The machine learning revolution

What machine learned models can and cannot do?

- Conventional models will not be replaced by machine learning models entirely.
- Within the next couple of years most weather predictions will come from machine learning models.
- Machine learning will be the perfect glue between models and observations.
- Km-scale models will make a difference for the generation of training datasets.
- Machine learning models have also potential for climate projections despite the extrapolation problem.

2022-today: The machine learning revolution



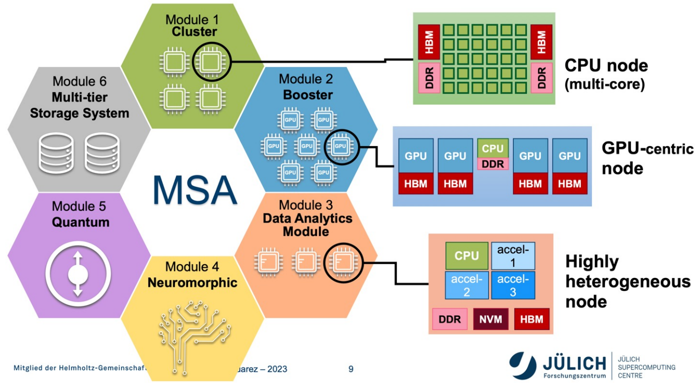
Machine learned models can now also do AMIP simulations.

Kochokov et al. @Google in preparation

And Foundation Models will enter the domain…

Dude, when does he finally start to talk about correctness and reproducibility?

Change of gear



Slide from Estela Suarez

Workmode of 2010:

- A single scientist can understand the whole Earth system model
- Earth system models consist of 100,000 lines of Fortran Code
- Code is shared via tarballs, data is stored locally
- Models run on CPUs and Moore's law is still working

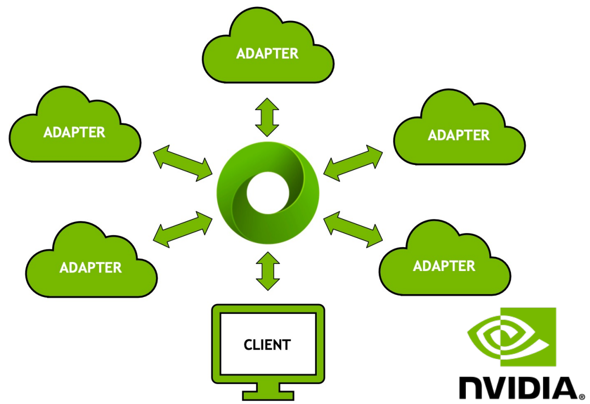
Workmode of 2020:

- A single scientist cannot understand the whole Earth system model anymore
- A team of software developers is needed to use heterogeneous hardware
- Models start to run on GPUs, Moore's law is dying
- Data is stored locally but meta information is available online
- Online code repositories are used to control quality and share model code

Tim Palmer's A380 comparison

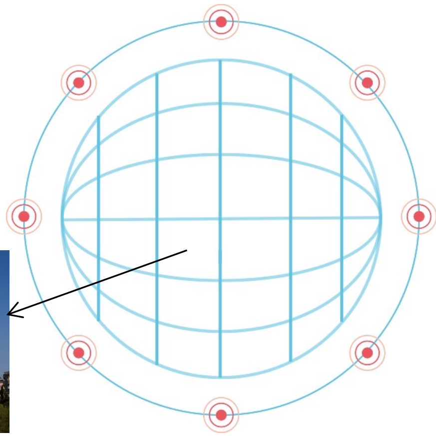
Workmode of 2030:

- Machine learning models of 2,000 lines of Python code compete with conventional models
- There are hundreds of models and many of them with specific tasks
- HPC is federated
- Data is federated



Current Systems

Earth System
models & observations



Limited resolutions

Small-scall processes not represented

Separation of earth system & impact sector models

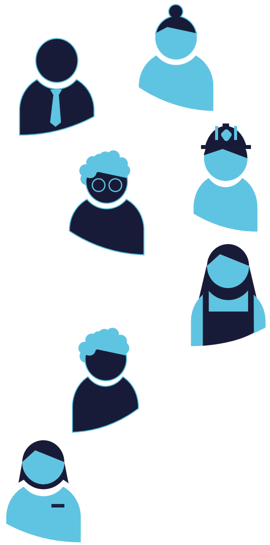
Impact sectors

Limited ways to change experiment design or output

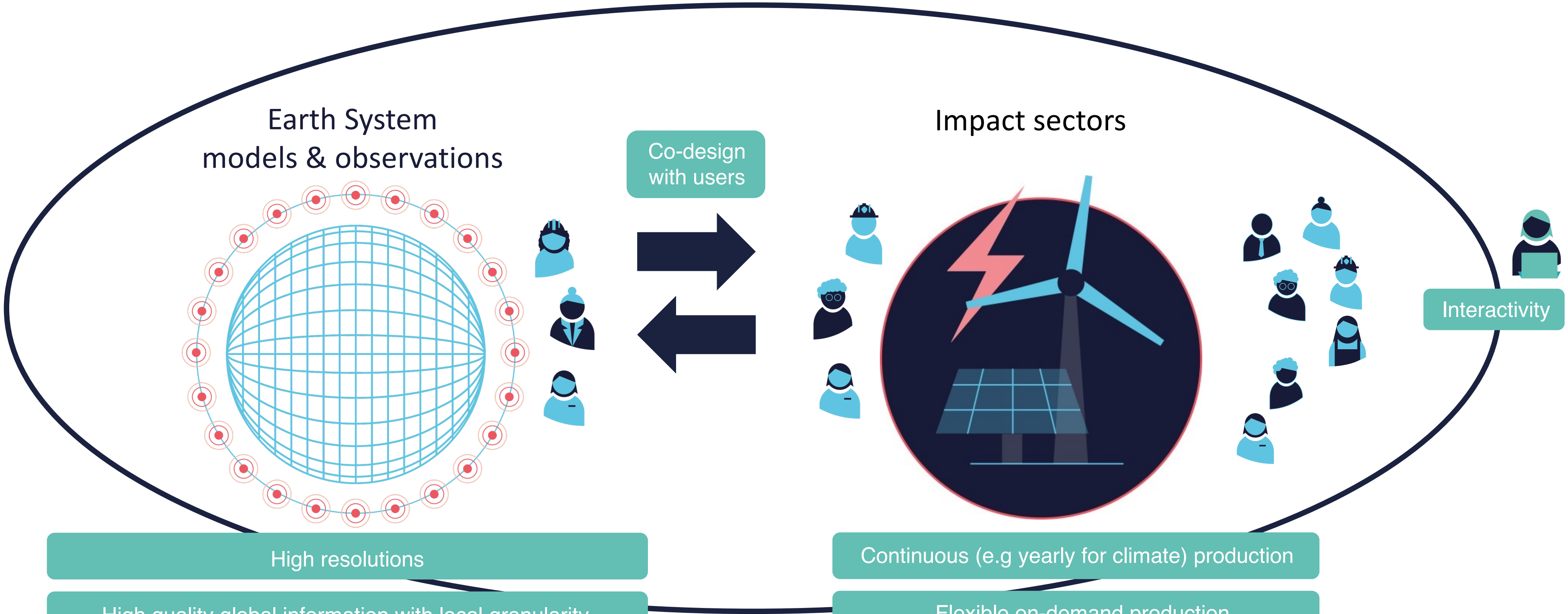
Limited versatility to access and interact with the data

Pre-defined configurations & simulation run times (every 7 years for climate!)

Users



DestinE builds Digital Twins of the Earth



- High resolutions
- High quality global information with local granularity
- Bringing impact sector models WITHIN the same workflow

- Continuous (e.g yearly for climate) production
- Flexible on-demand production
- Data streaming and tailored data access

Reproducibility – So much room for pessimism

- In 2020, the IFS ran in production on ECMWF's supercomputer and on a handful other computers for research. In 2024, the IFS will generate semi-operational predictions on several of the EuruHPC supercomputers as part of DestinE.

- Today's high-end models use several tools for portability (OpenACC, HIP, CUDA, DSLs such as GT4Py, Loki and Psyclone…).

- Machine learning models add new dimensions of complexity:
  - Retrained models can be very different
  - Use of pre-trained models
  - Use of transfer learning
  - Use of Foundation models in the future
  - Need to define and keep track of dataset and data manipulations

Make machine learning developments comparable via benchmark datasets

Benchmark datasets include:

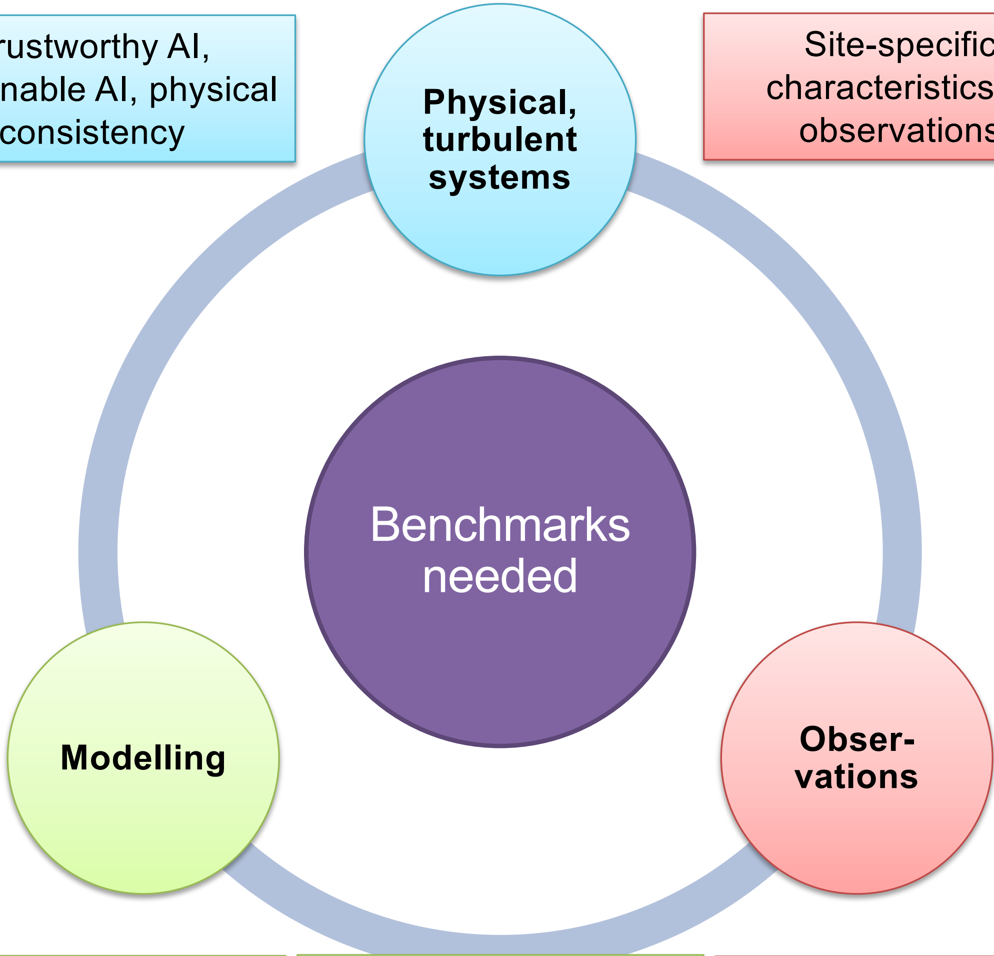
- A problem statement
- Data that is available online
- Python code or Jupyter notebooks
- A reference machine learning solution
- Quantitative evaluation metrics
- Visualisation, diagnostics and robustness tests
- Computational benchmarks

Benchmark datasets are useful because:

- They allow a quantitative evaluation of machine learning approaches
- They reduce data access and help scientists to get access to relevant data
- They allow for a separation of concerns between domain sciences and machine learning experts
- They allow for a separation of concerns between domain sciences and HPC experts

Missing machine learning benchmark datasets for atmospheric sciences

- Transfer learning
- Online and reinforcement learning
- Training of machine learning tools in a changing climate
- Unstructured grids on the sphere
- Multi-scale interactions in space and time
- Dataflow and handling of huge datasets
- Trustworthy AI, explainable AI, physical consistency

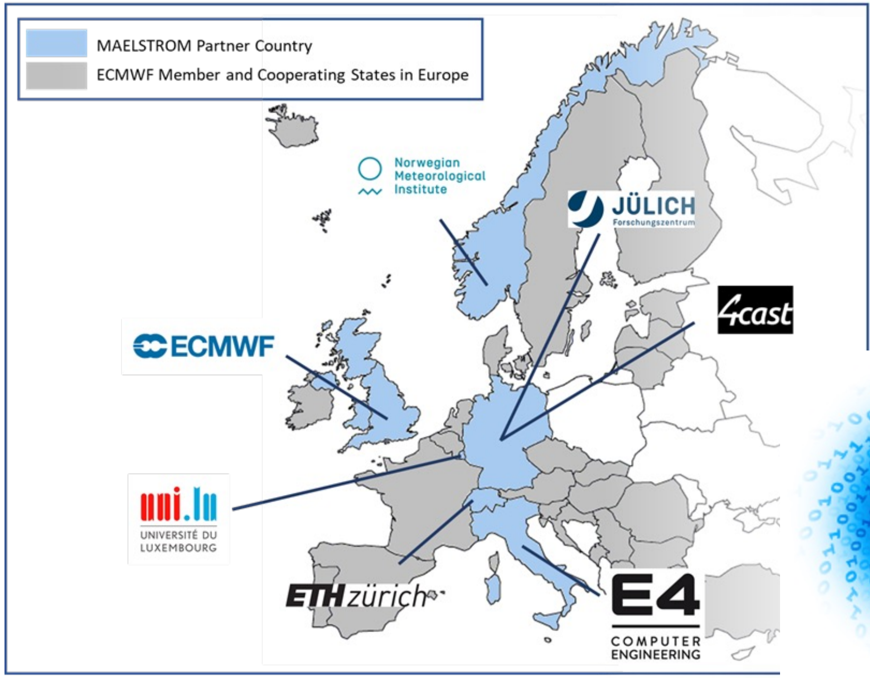


- Hybrid modelling and coupling
- Uncertainty quantification and representation
- Weather and climate predictions based on machine learning
- The emulation of model components
- Post-processing and down-scaling

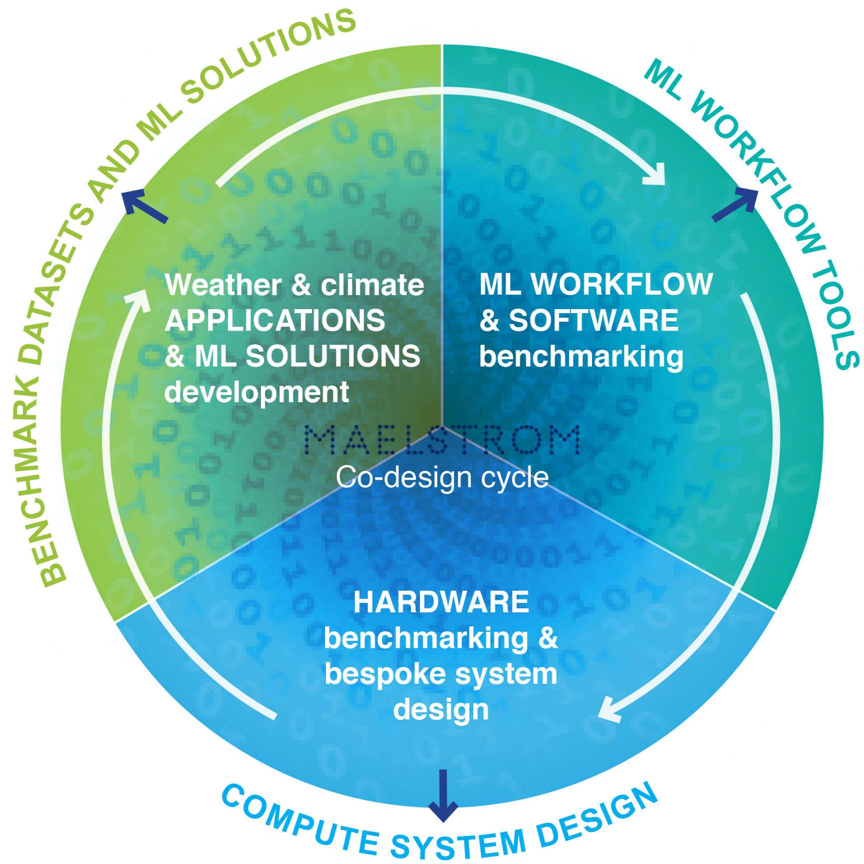
- Site-specific characteristics of observations
- Composite distributions of observations
- Extreme value predictions
- Missing data and irregular spacing of monitoring sites
- Auto correlation and periodic patterns
- Physical constraints
- Data anonymity
- Fusion of diverse datasets
- Air-quality applications
- Nowcasting applications
- Uncertainty estimates
- Feature detection

Dueben, Schultz, Chantry, Gagne, Hall, McGovern AIES 2022

Learn how to use machine learning at scale → The MAELSTROM project



MAELSTROM



MAchinE Learning for Scalable meTeoROlogy and cliMate

The **datasets** have been published! https://www.maelstrom-eurohpc.eu/content/docs/uploads/doc6.pdf

https://www.maelstrom-eurohpc.eu/ @MAELSTROM_EU

Data is open, diagnostics are at hand – A new way to verify correctness

WeatherBench 2

Search docs

- WeatherBench 2 Evaluation Quickstart
- WeatherBench 2 Data Guide
- Command line scripts
- Distributed computing using Beam on GCP
- Official Evaluation
- Submit a new model to WeatherBench 2
- Init vs Valid Time Conventions
- API docs

/ Why WeatherBench?

Edit on GitHub

WeatherBench 2

Why WeatherBench?

WeatherBench 2 is a framework for evaluating and comparing data-driven and traditional numerical weather forecasting models. WeatherBench consists of:

- Publicly available, cloud-optimized ground truth and baseline datasets. For a complete list, see this page.
- Open-source evaluation code. See this quick-start to explore the basic functionality or the API docs for more detail. Since high-resolution forecast files can be large, the WeatherBench 2 code was written with scalability in mind. See the command-line scripts based on Xarray-Beam and this guide for running the scripts on GCP using DataFlow.
- A website displaying up-to-date scores of many of the state-of-the-art data-driven and physical approaches.
- A paper describing the rationale behind the evaluation setup.

WeatherBench 2 has been built as an evolving tool for the entire community. For this reason, we welcome any feedback (ideally, submitted as GitHub issues) or contributions. If you would like your model to be part of WeatherBench, check out this guide.

WeatherBench 2: A benchmark for the next generation of data-driven global weather models

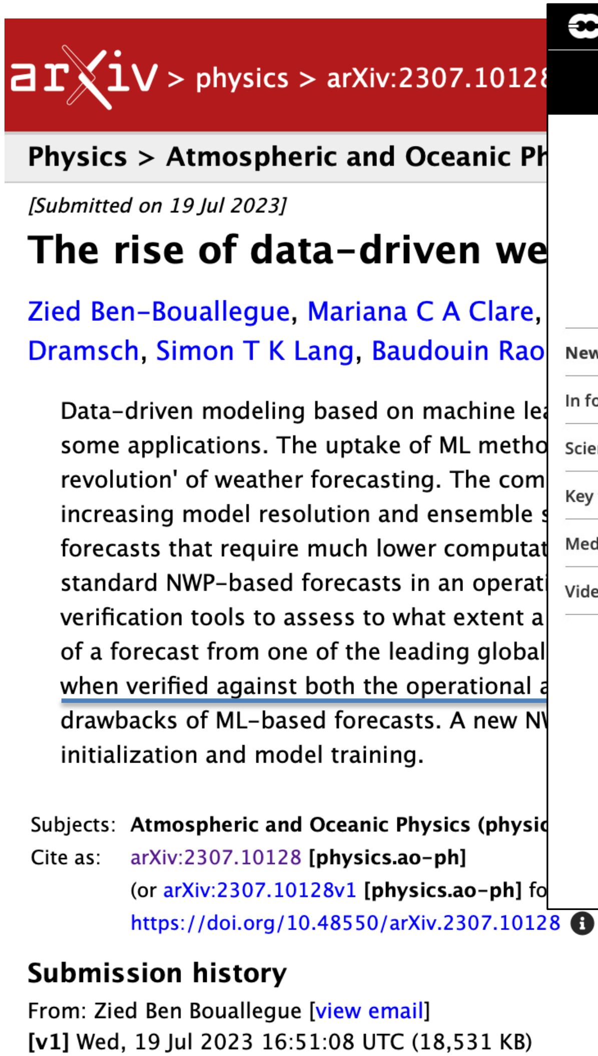
Stephan Rasp[1,*], Stephan Hoyer[1], Alexander Merose[1], Ian Langmore[1], Peter Battaglia[2], Tyler Russell[1], Alvaro Sanchez-Gonzalez[2], Vivian Yang[1], Rob Carver[1], Shreya Agrawal[1], Matthew Chantry[3], Zied Ben Bouallegue[3], Peter Dueben[3], Carla Bromberg[1], Jared Sisk[1], Luke Barrington[1], Aaron Bell[1], and Fei Sha[1]

[1]Google Research
[2]Google DeepMind
[3]European Centre for Medium-Range Weather Forecasts
[*]Corresponding author: srasp@google.com

Abstract

WeatherBench 2 is an update to the global, medium-range (1–14 day) weather forecasting benchmark proposed by Rasp et al. (2020), designed with the aim to accelerate progress in data-driven weather modeling. WeatherBench 2 consists of an open-source evaluation framework, publicly available training, ground truth and baseline data as well as a continuously updated website with the latest metrics and state-of-the-art models: `https://sites.research.google/weatherbench`. This paper describes the design principles of the evaluation framework and presents results for current state-of-the-art physical and data-driven weather models. The metrics are based on established practices for evaluating weather forecasts at leading operational weather centers. We define a set of headline scores to provide an overview of model performance. In addition, we also discuss caveats in the current evaluation setup and challenges for the future of data-driven weather forecasting.

Data is open, diagnostics are at hand – A new way to verify correctness



arXiv > physics > arXiv:2307.1012

Physics > Atmospheric and Oceanic Ph

[Submitted on 19 Jul 2023]

The rise of data-driven we

Zied Ben-Bouallegue, Mariana C A Clare, Dramsch, Simon T K Lang, Baudouin Rao

Data-driven modeling based on machine lea
some applications. The uptake of ML metho
revolution' of weather forecasting. The com
increasing model resolution and ensemble s
forecasts that require much lower computat
standard NWP-based forecasts in an operati
verification tools to assess to what extent a
of a forecast from one of the leading global
when verified against both the operational a
drawbacks of ML-based forecasts. A new NV
initialization and model training.

Subjects: Atmospheric and Oceanic Physics (physic

Cite as: arXiv:2307.10128 [physics.ao-ph]

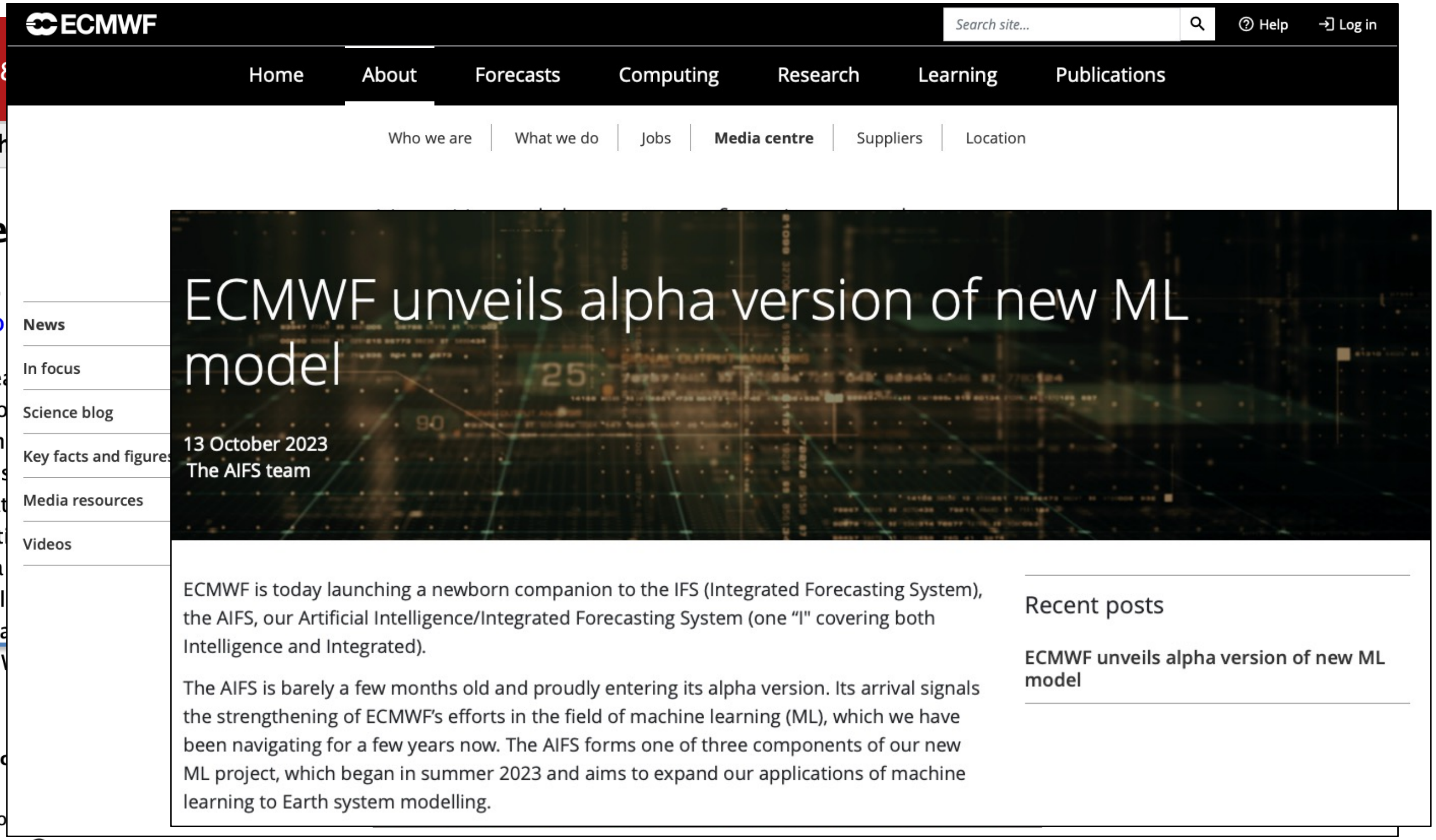
(or arXiv:2307.10128v1 [physics.ao-ph] fo

https://doi.org/10.48550/arXiv.2307.10128

Submission history

From: Zied Ben Bouallegue [view email]

[v1] Wed, 19 Jul 2023 16:51:08 UTC (18,531 KB)



ECMWF

Search site... Help Log in

Home About Forecasts Computing Research Learning Publications

Who we are | What we do | Jobs | Media centre | Suppliers | Location

News
In focus
Science blog
Key facts and figures
Media resources
Videos

ECMWF unveils alpha version of new ML model

13 October 2023
The AIFS team

ECMWF is today launching a newborn companion to the IFS (Integrated Forecasting System), the AIFS, our Artificial Intelligence/Integrated Forecasting System (one "I" covering both Intelligence and Integrated).

The AIFS is barely a few months old and proudly entering its alpha version. Its arrival signals the strengthening of ECMWF's efforts in the field of machine learning (ML), which we have been navigating for a few years now. The AIFS forms one of three components of our new ML project, which began in summer 2023 and aims to expand our applications of machine learning to Earth system modelling.

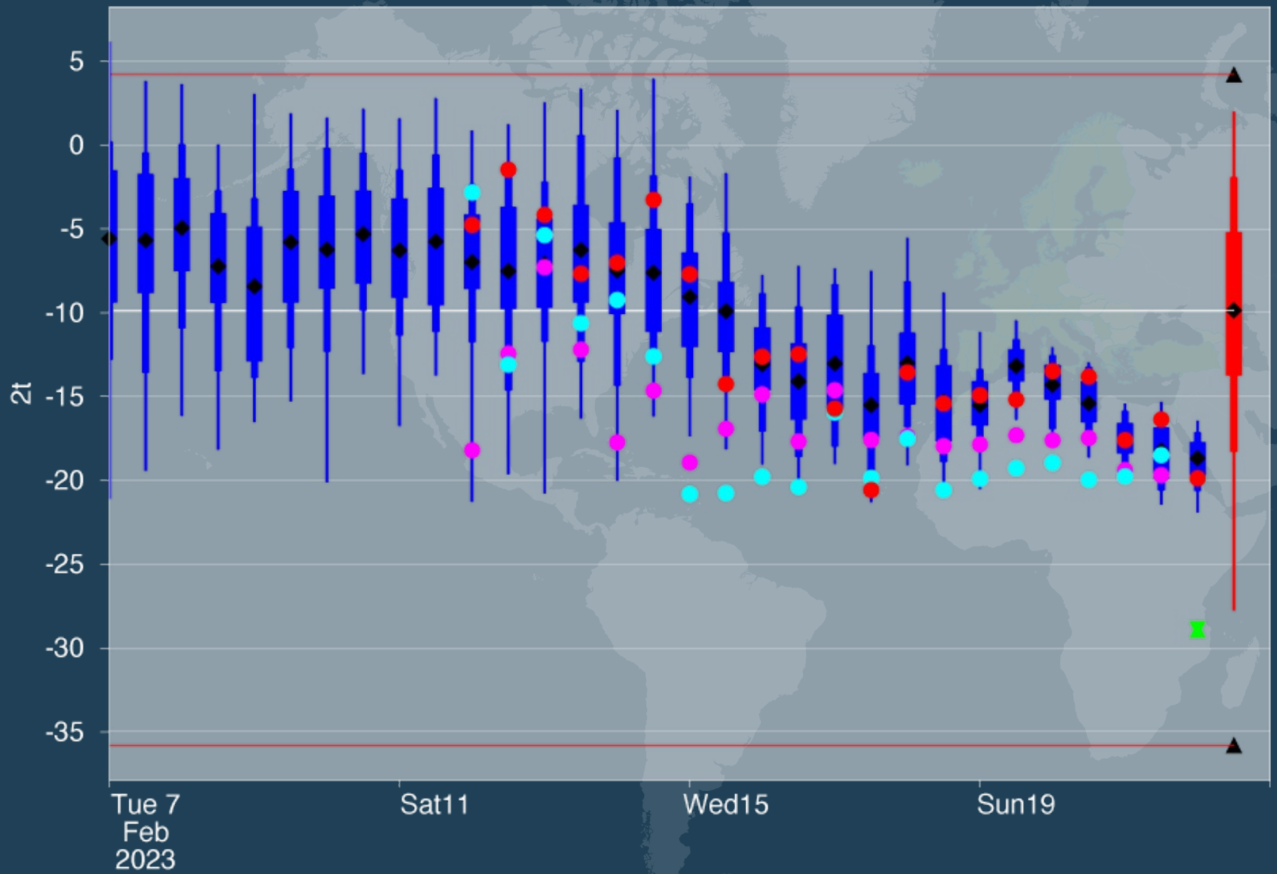
Recent posts

ECMWF unveils alpha version of new ML model

https://www.ecmwf.int/en/about/media-centre/news/2023/how-ai-models-are-transforming-weather-forecasting-showcase-data

https://www.ecmwf.int/en/about/media-centre/aifs-blog/2023/ECMWF-unveils-alpha-version-of-new-ML-model

What the forecasts are showing: **Severe Cold** / Sodankylä, Finland, 22 Feb 00UTC



To explore the ability of data-driven models to capture extreme events we examine a case study from Finland from earlier this year, when -29C was observed.

We find that Pangu and FourCastNet recognised the severity of this event earlier, however all models underestimated the temperature significantly, to a similar degree.

Observation – green hourglass
IFS HRES – red dot
IFS ENS - blue
Pangu – cyan dot
FourCastNet – magenta dot
Climatology – red box plot

Figure from
Zied Ben-Bouallegue

Data is open, diagnostics are at hand – A new way to verify correctness

Why was the approach of global machine learning models so successful?

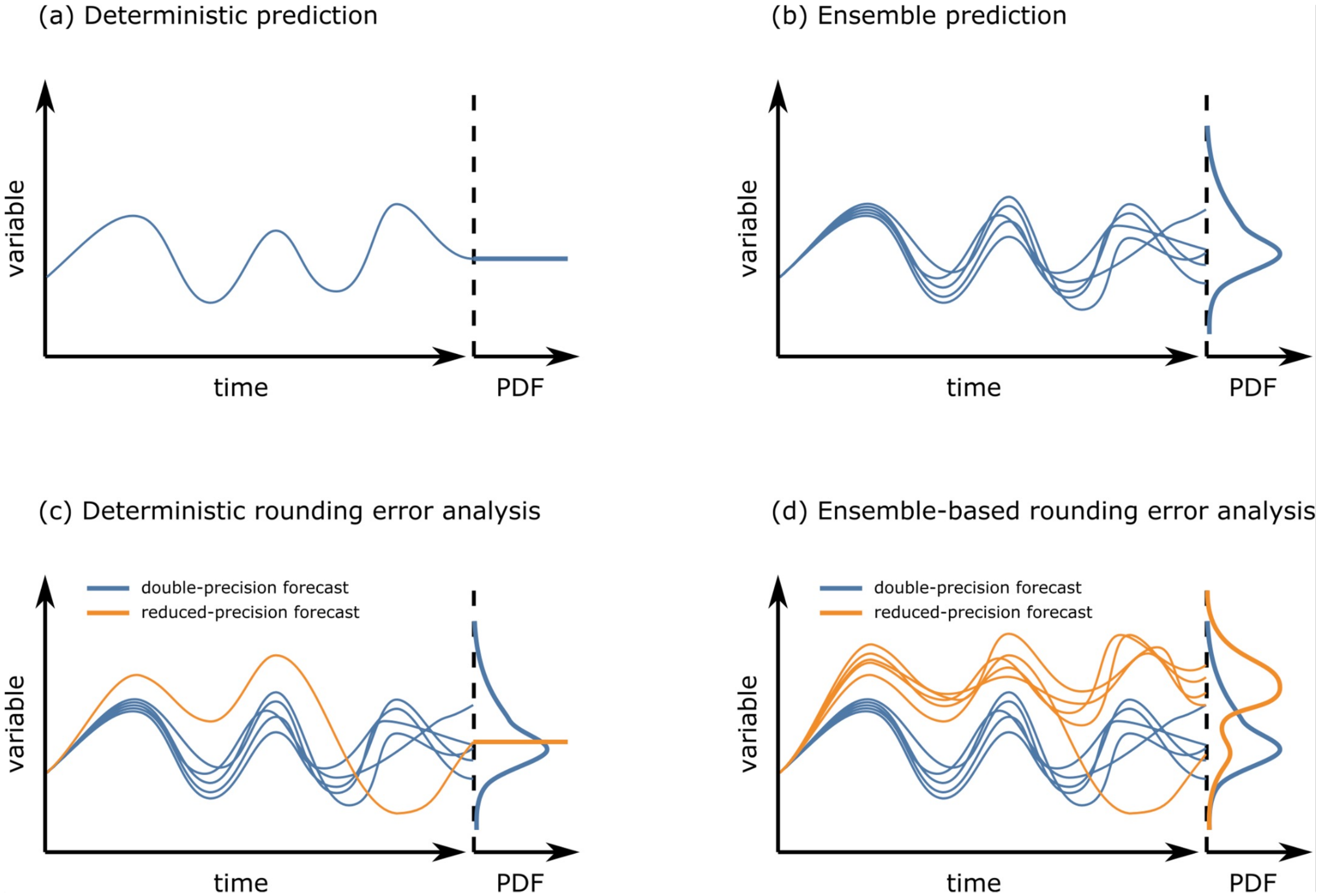
- Because there was a very large unified training dataset available with ERA5 from Copernicus.
- Open benchmark datasets are needed to allow for quantitative comparisons and to bridge communities.
- Km-scale models will make a difference for the generation of training datasets.

MAELSTROM

Reproducibility – So much room for optimism

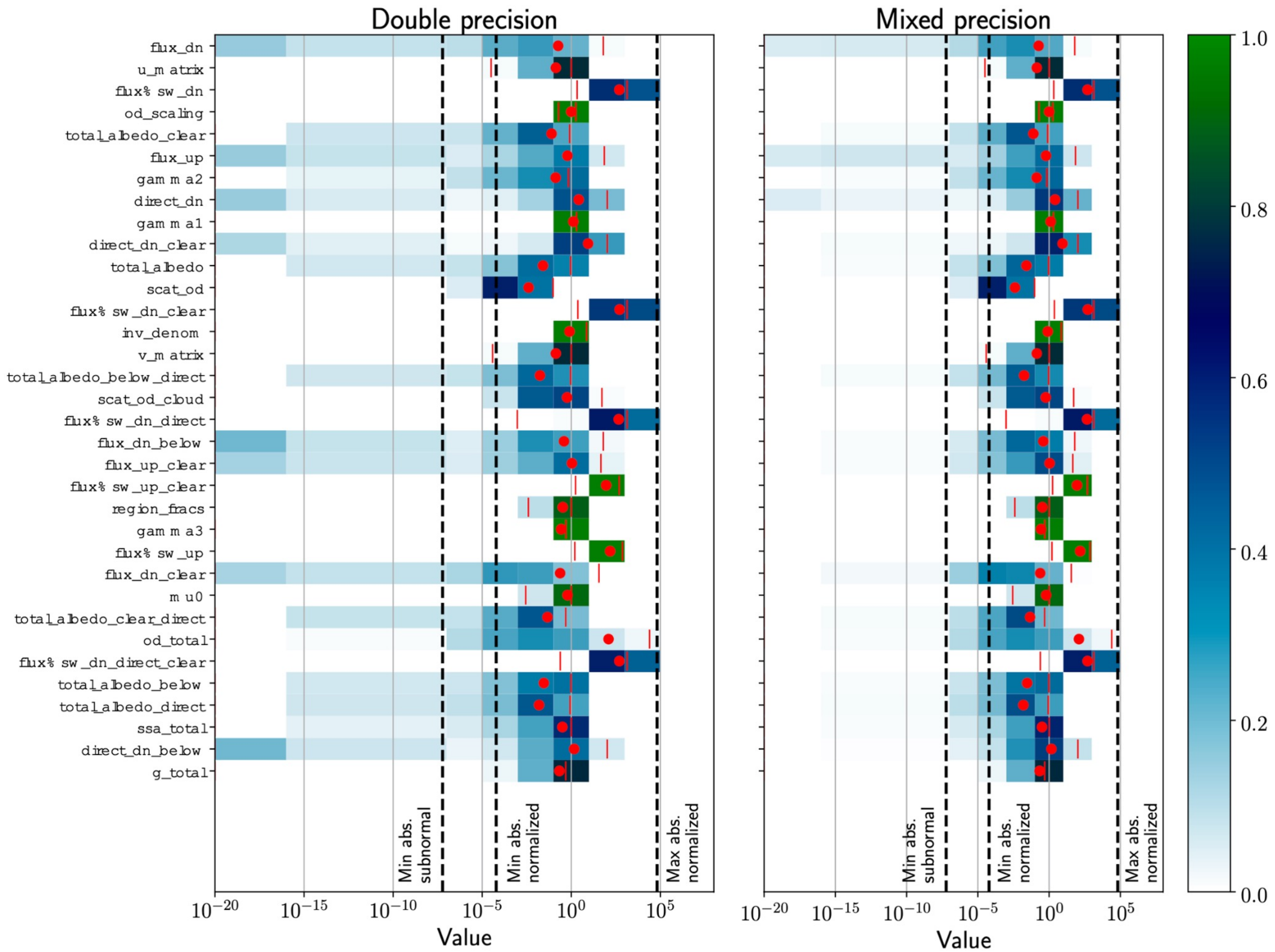
- Git is standard, CI tests are getting better, model code is unified via DSLs
- Many datasets are nowadays open and Journals require open data and source code
- Most machine learning models and training datasets are published with the papers as open source
- Bit-wise reproducibility may not be as important anymore when compared to the ability to fulfil loss functions and complex diagnostics as we approach a new era in model comparison
- Compute and data handling will (hopefully) be more unified as we approach federated data and computing

The best way to check correctness is via automated tools



Pershin, Chantry, Dueben, Hogan, Palmer ESS Open Archive 2023

The best way to check correctness is via automated tools



Pershin, Chantry, Dueben, Hogan, Palmer ESS Open Archive 2023

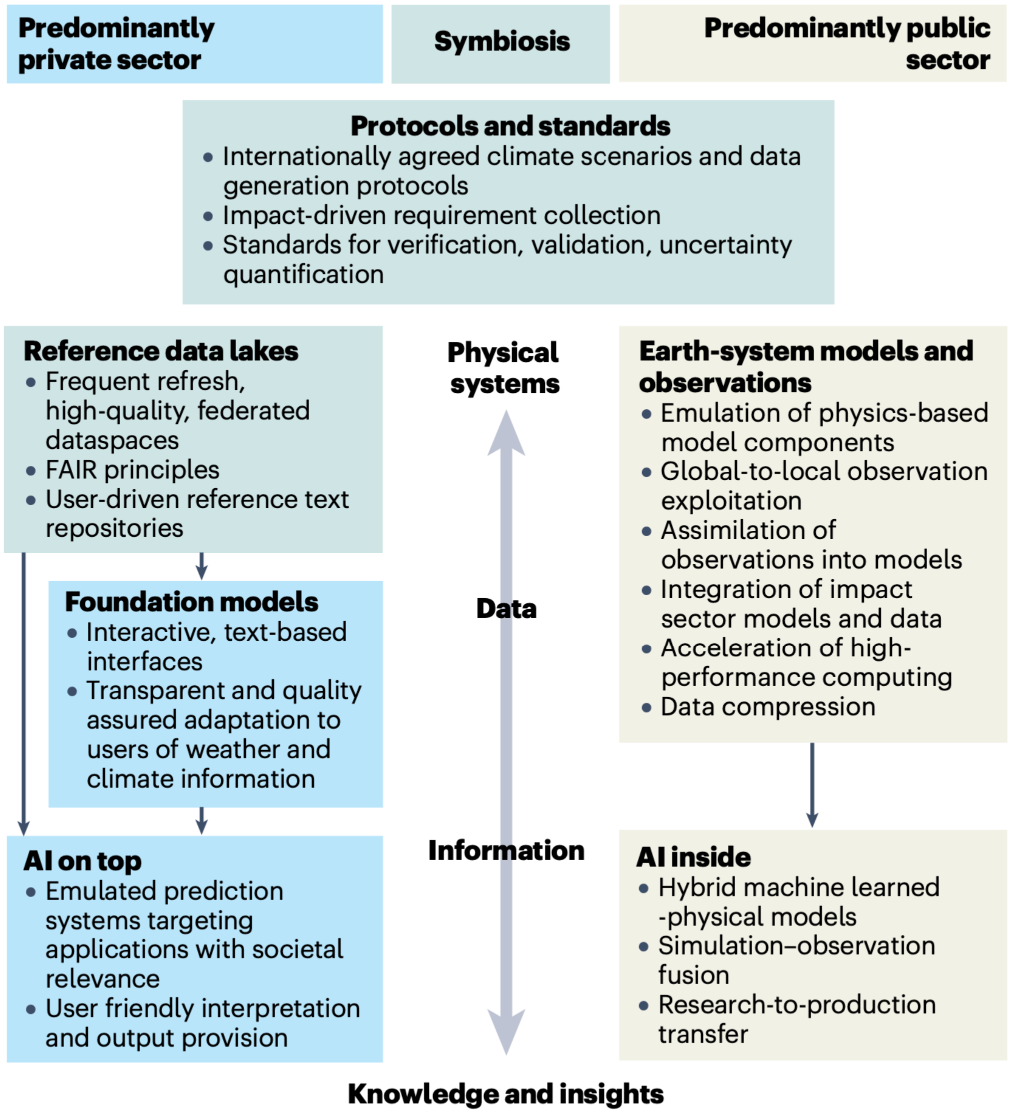
What is our aim today? – Imagine if…

- …all model output, reanalysis data and all observations from the past and presence would be available online via federated data storage and next to federated computing and with a uniform API and as uniform packages
- …there would be conventional Earth system models and machine learning models that can be used off-the-shelf to analyse and extend those datasets
- …all of these tools would be scalable and easy to use on laptops and supercomputers from various computing languages
- …there would be off-the-shelf tools to interpret physical reasoning and causality via unsupervised machine learning, to perform uncertainty quantification, and to perform state-of-the-art visualisation
- …there are off-the shelf machine learning solutions, and unit testing would be standard

We need to fight complexity and diversity of software with centralised infrastructure efforts and norms.

We need world-wide collaboration on data and infrastructure developments to achieve this.
First approaches already exist with Destination Earth, Earth-2 and EVE

How will machine learning for weather and climate evolve in a public/private partnership?



nature reviews earth & environment

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature › nature reviews earth & environment › comment › article

Comment | Published: 01 August 2023

Deep learning and a changing economy in weather and climate prediction

Peter Bauer, Peter Dueben, Matthew Chantry, Francisco Doblas-Reyes, Torsten Hoefler, Amy McGovern & Bjorn Stevens

Nature Reviews Earth & Environment **4**, 507–509 (2023) | Cite this article

638 Accesses | **34** Altmetric | Metrics

The rapid emergence of deep learning is attracting growing private interest in the traditionally public enterprise of numerical weather and climate prediction. A public–private partnership would be a pioneering step to bridge between physics- and data-based methods, and necessary to effectively address future societal challenges.

What have we learned?

The quiet revolution (1980-2020):

- Steady investment into Earth system modelling and Earth system observations made a difference.

The digital revolution (2015-today):

- Conventional models need to be made future proof via new software and hardware standards.
- Large scale efforts make km-scale models possible today and they will make a difference.

The machine learning revolution (2022-today):

- A PhD student can write a machine learning tool that can beat the best weather prediction model in the world based on hundreds of person years of developments.
- Data needs to be open and easy to use to make progress.

The next step: Models will be better, tools will be easier, and data/HPC will be federated

- To achieve this needs programmes such as Destination Earth, Earth-2 and EVE.

Many thanks! **Peter.Dueben@ecmwf.int** **@PDueben**

ECMWF | EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

The strength of a common goal