Machine learning for weather and climate prediction

Peter Dueben

Head of the Earth System Modelling Section

ECMWF | EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

MAELSTROM

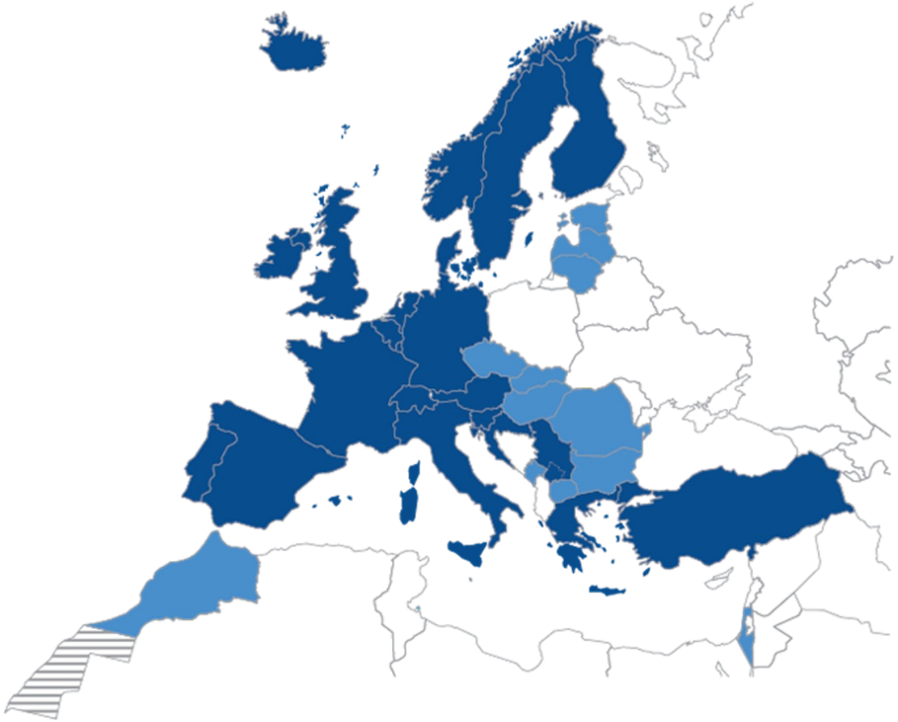
esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

Funded by the European Union

The strength of a common goal

The ESIWACE and MAELSTROM projects have received funding from the European Union under grant agreement No 823988 and 955513.

What is the European Centre for Medium-Range Weather Forecasts (ECMWF)?

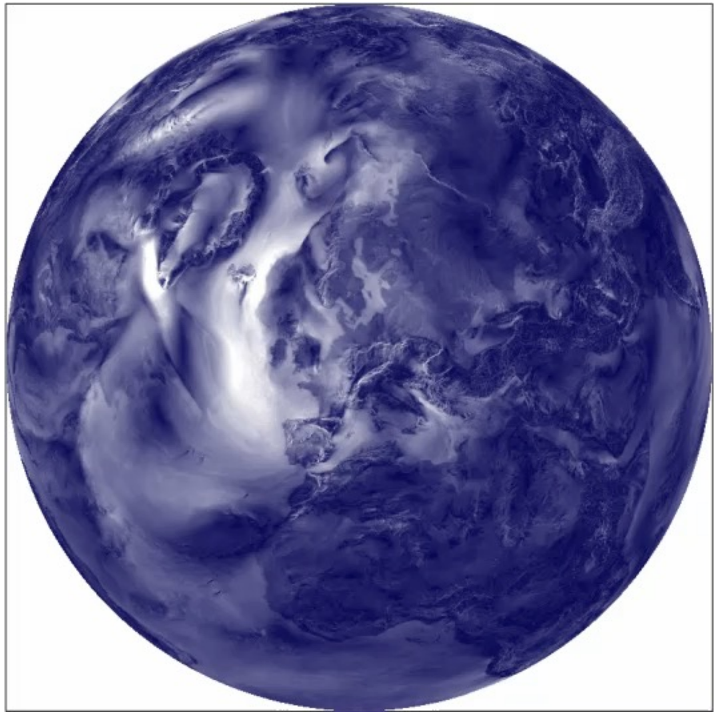


www.ecmwf.int

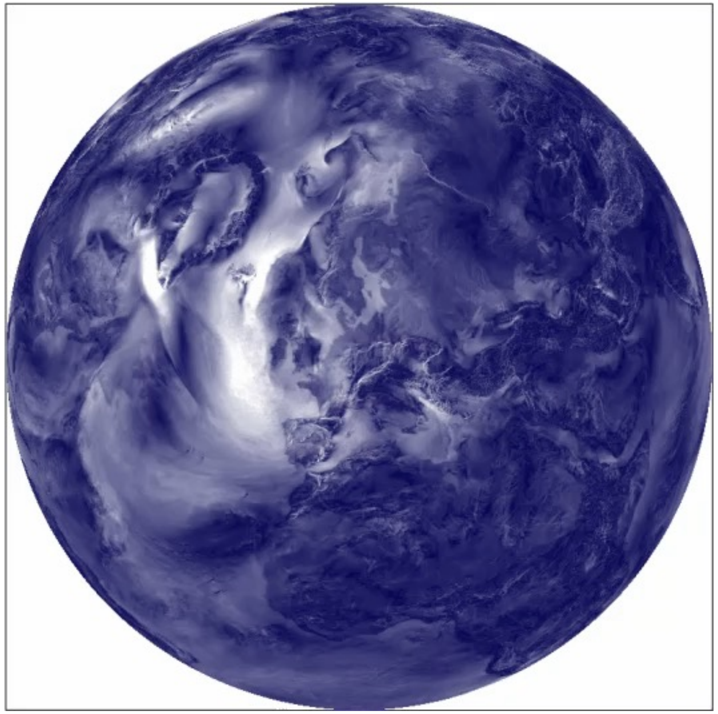
- Research institute.
- 24/7 operational weather service for medium-range, monthly and seasonal forecasts.
- Independent, intergovernmental organisation supported by 34 states.
- Based in Reading, Bologna and Bonn; ≈350 member of staff.
- Home of two supercomputers.
- Home of the Integrated Forecast System (IFS).

Numerical weather predications

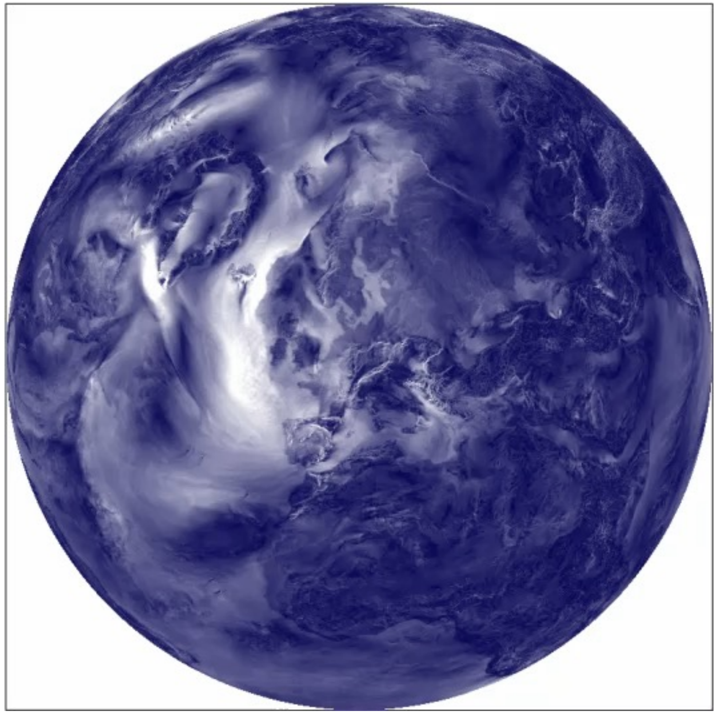
Special thanks to Simon Lang



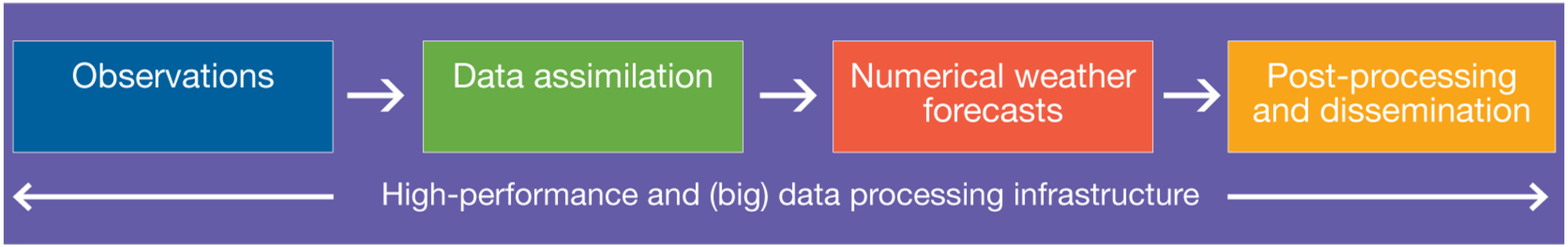
Control forecast



Perturbed forecast 1

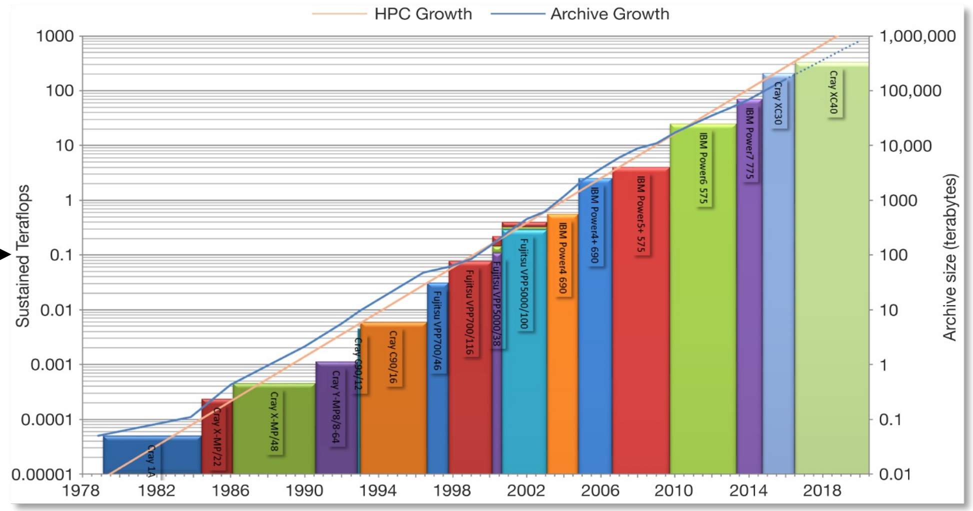


Perturbed forecast 2

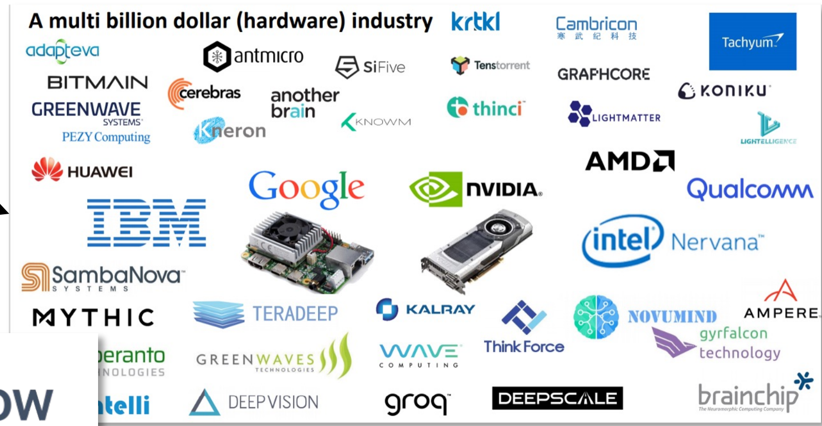


Machine Learning – Why did it start in 2018?

Increase in data volume



New computing hardware



New machine learning software

TensorFlow
Keras
PyTorch

Increase in knowledge

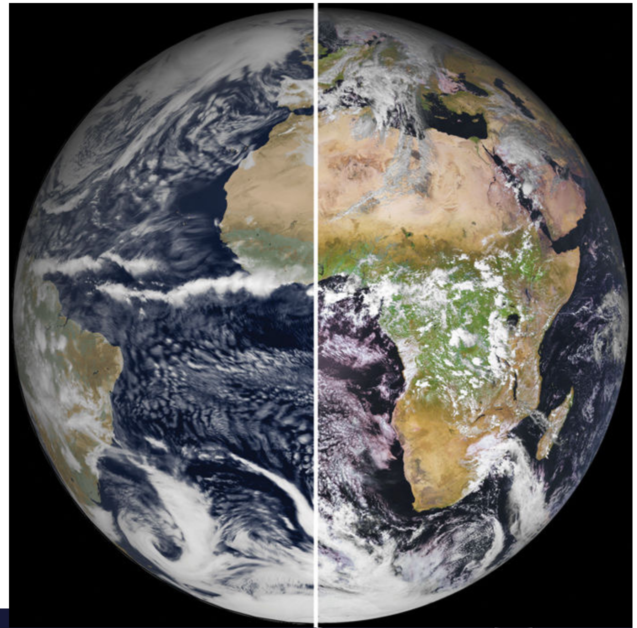
Machine Learning – Why in Earth System modelling

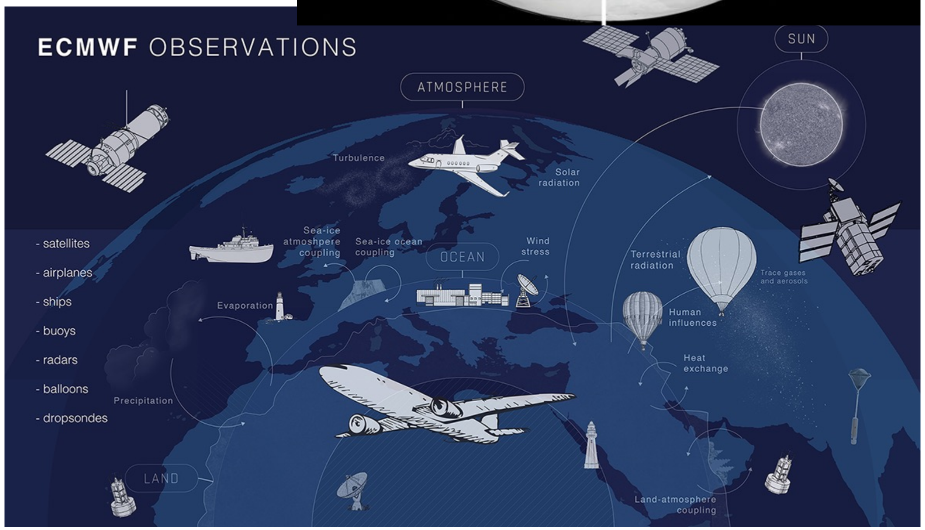
Inputs → machine learning → Outputs

Earth system science is difficult as the Earth system is huge, complex and chaotic, and as the resolution of our models is limited

However, we have a huge amount of observations and Earth system data

- ➢ **There are many application areas for machine learning in Earth system science**





And there is more to come in Destination Earth

Check for updates

comment

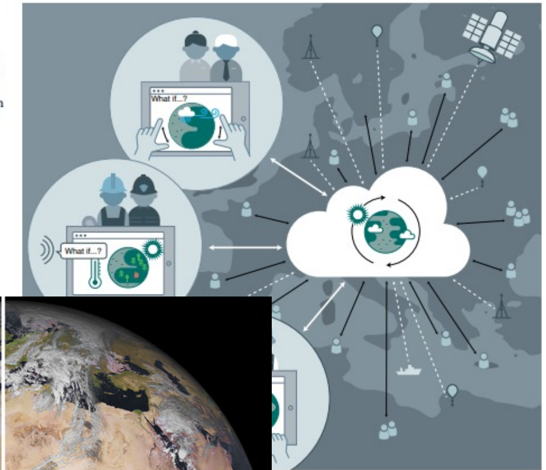
A digital twin of Earth for the green transition

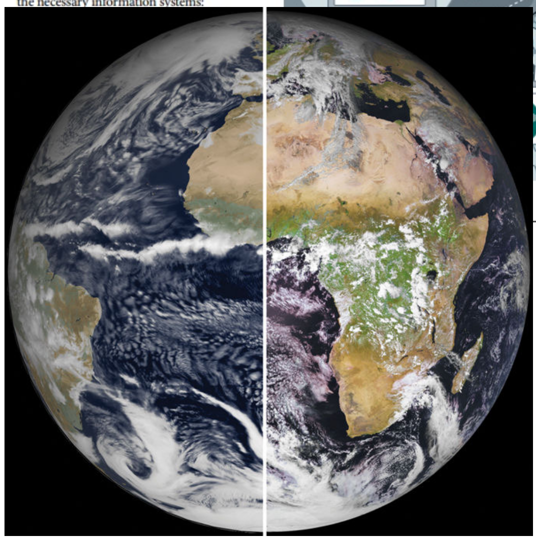
For its green transition, the EU plans to fund the development of digital twins of Earth. For these twins to be more than big data atlases, they must create a qualitatively new Earth system simulation and observation capability using a methodological framework responsible for exceptional advances in numerical weather prediction.

Peter Bauer, Bjorn Stevens and Wilco Hazeleger

The European Union (EU) intends to become climate neutral by 2050, and the set of policies designed to bring about this green transition — the European Green Deal — was announced in December 2019 (ref. [1]). Accompanied by €1 trillion of planned investment, Green Deal policies aim to help the world's second-largest economy sustainably produce energy, develop carbon-neutral fuels and advance circular products in energy-intensive industrial sectors with zero waste and zero pollution.

A key element of the Green Deal is its dependence on the 'digital transformation' — an openly accessible and interoperable European dataspace as a central hub for informed decision making. The EU identified two landmark actions to support the necessary information systems:





change in the physical content of Earth system models is making them amenable to approaches that harmonize the physical laws they encode with ever more extensive observations to provide the best possible estimate of the state of our planet. Hence, digital twins must focus exactly on how best to realize this convergence of the modelling and observation worlds.

A methodological framework for the twin's architecture already exists in the form of data assimilation, which numerical weather prediction has developed with success over decades[10]. Data assimilation combines data from different observational sources with physical Earth system model simulations to derive an estimate of the state

PERSPECTIVE
https://doi.org/10.1038/s43588-021-00023-0
nature computational science

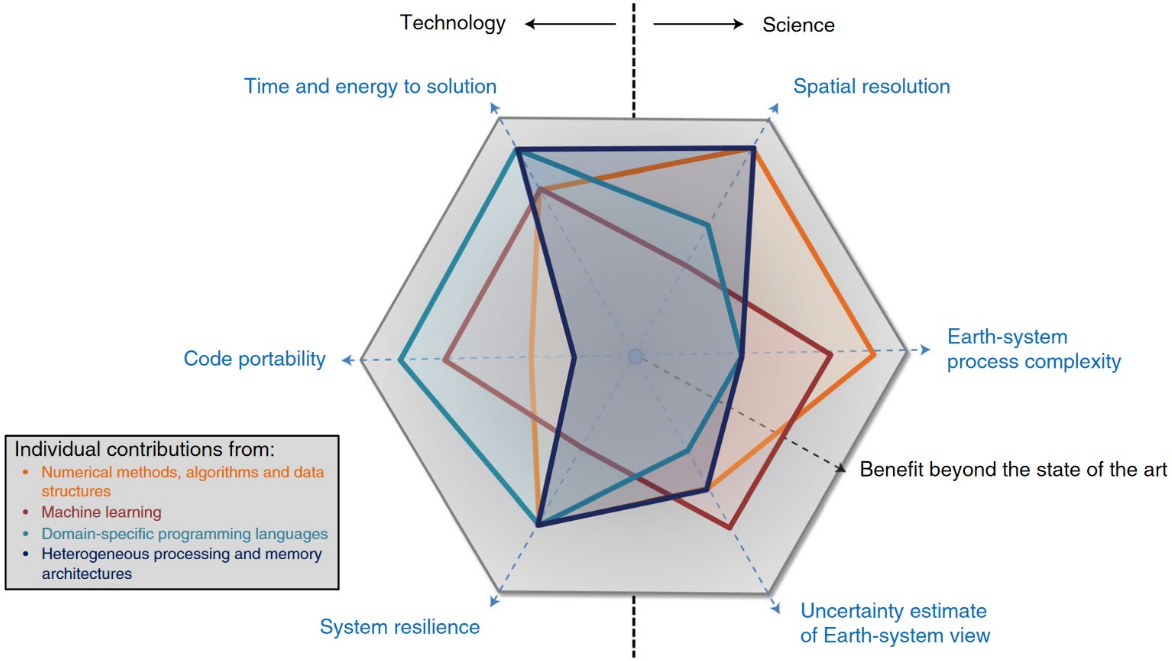
Check for updates

The digital revolution of Earth-system science

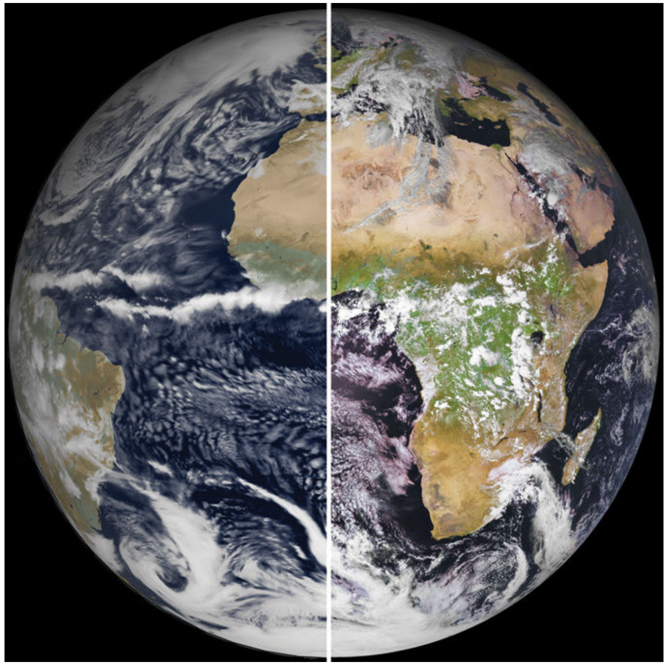
Peter Bauer[1]✉, Peter D. Dueben[1], Torsten Hoefler[2], Tiago Quintino[3], Thomas C. Schulthess[4] and Nils P. Wedi[1]

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 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.



Phase I: 2018-2021 – Explore the space



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

Speed up simulations and green computing
- 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
- …

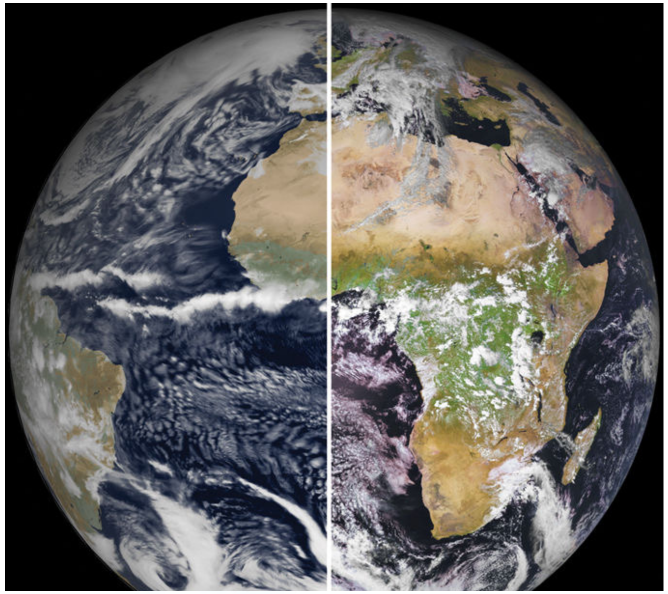
Link communities
- Health – e.g. for predictions of risks
- Energy – e.g. for local downscaling
- Transport – e.g. to combine weather and IoT data
- Pollution – e.g. to detect sources
- Extremes – e.g. to predict wild fires
- …

Phase II: 2022-2026 – Hybrid machine learning towards operational use

We already had great success stories when using machine learning in operational weather and climate predictions in:

- Some decision trees in ensemble post-processing…
- Some neural networks in observation operators…

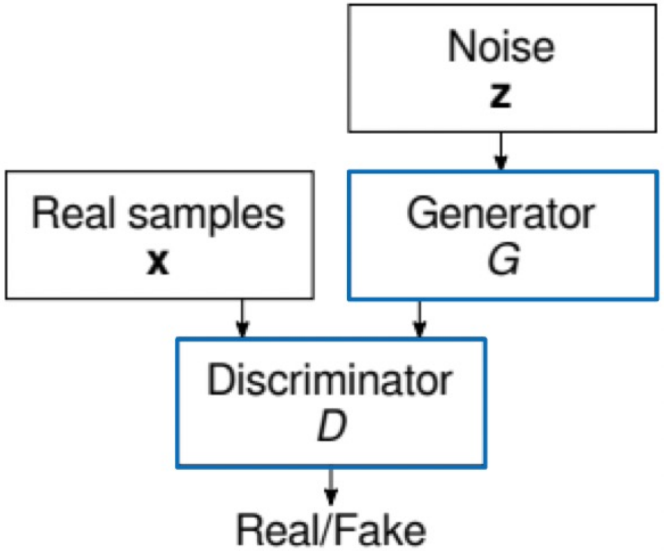
→ We will need to do more and fast!



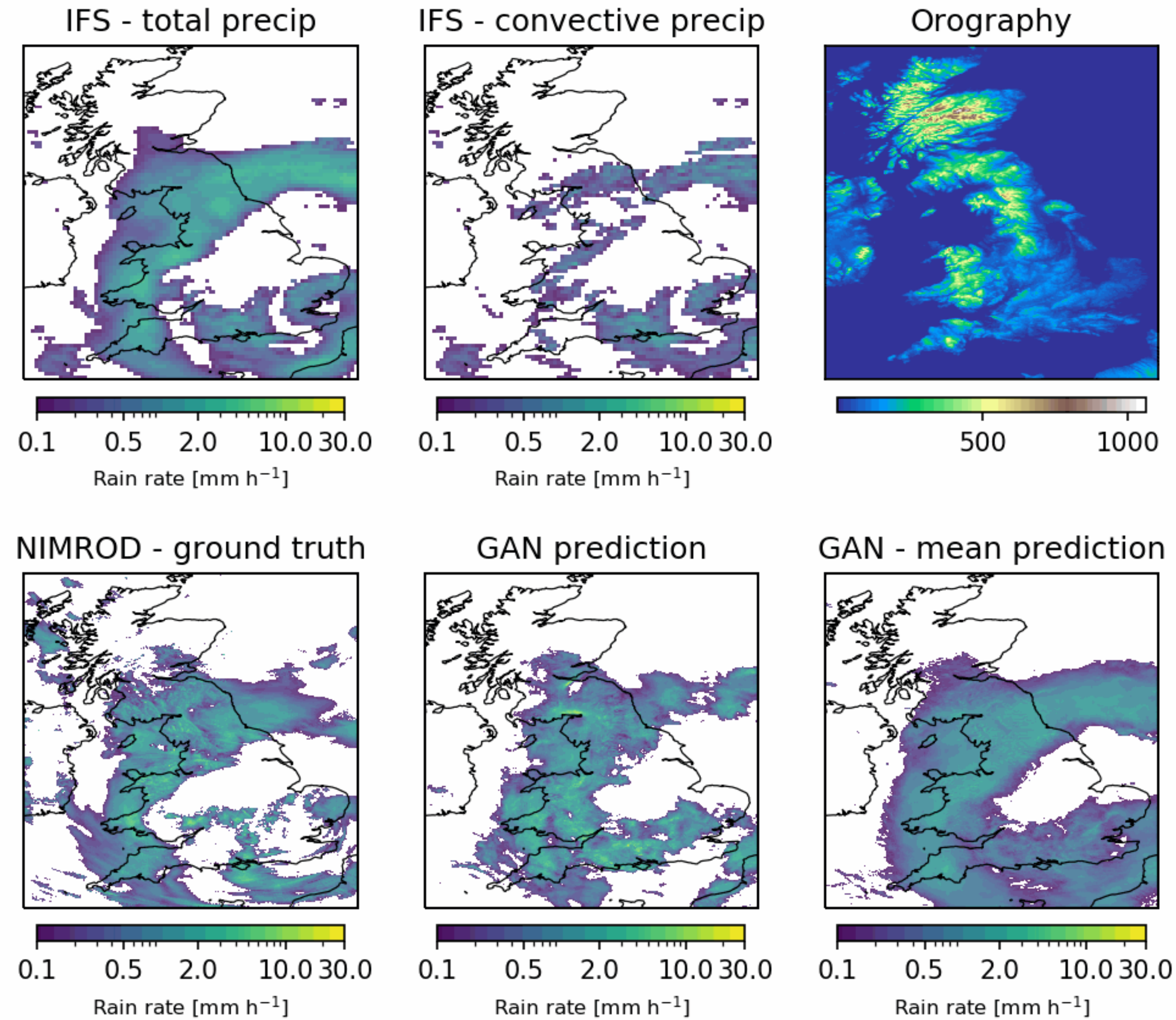
Most promising candidates for machine learning applications for operational use:

- Parametrisation emulation
- Observation operators
- Post-processing pointwise/ensembles/S2S
- Parameter optimisation
- Online bias correction

Science example 1: Downscaling with Generative Adversarial Networks

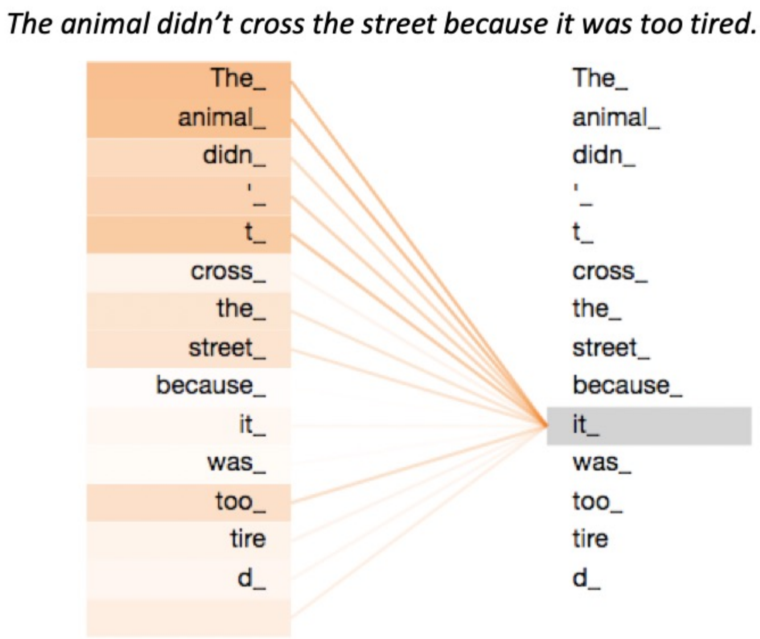


Input: IFS Model Simulation fields on coarse (9 km) grid
Output: Precipitation observation on fine (1 km) grid



Harris, McRae, Chantry, Dueben, Palmer JAMES 2022

Science example 2: Transformer networks for ensemble post-processing



https://jalammar.github.io/illustrated-transformer/

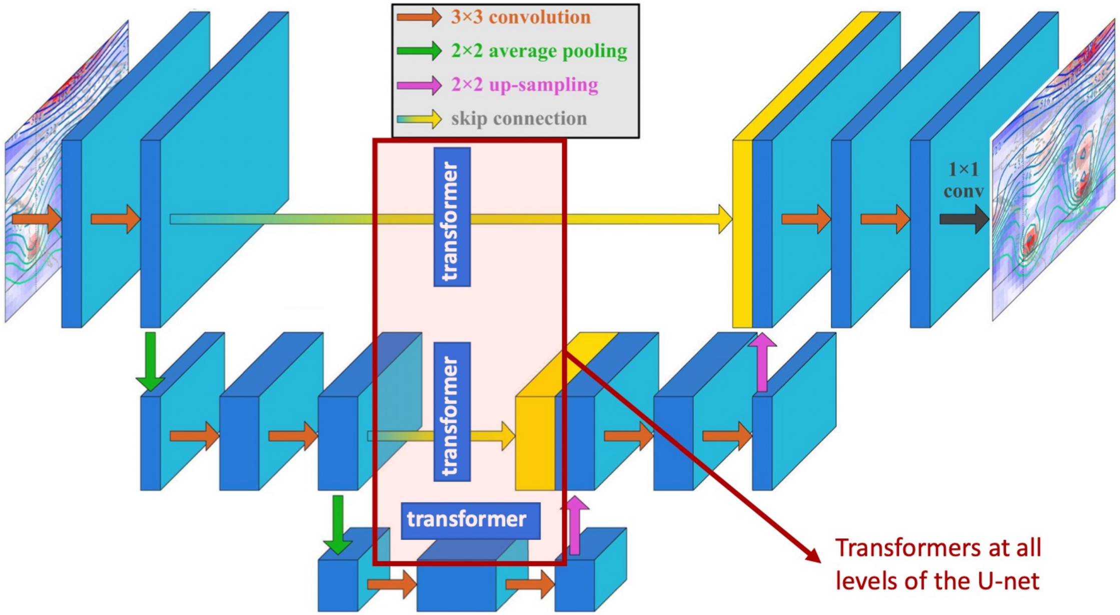
Let's use transformers in the ensemble dimension following the work of Tobias Finn

"Self-Attentive Ensemble Transformer: Representing Ensemble Interactions in Neural Networks for Earth System Models." *arXiv preprint arXiv:2106.13924*

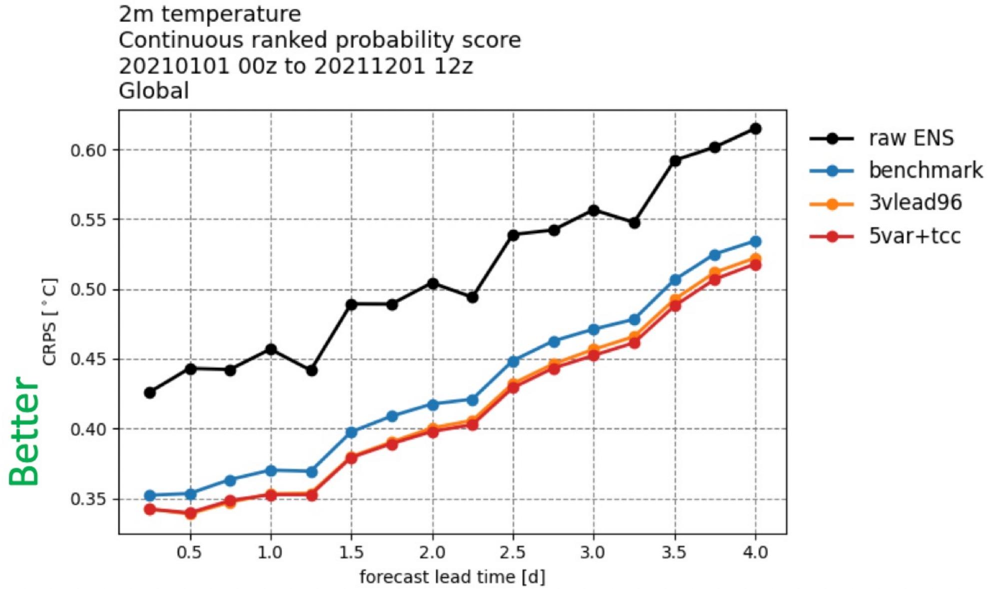
Let's test this for hindcast ensembles in a collaboration between Microsoft and ECMWF

Ben Bouallegue, Weyn, Clare, Dramsch, Dueben, Chantry arXiv 2023

Science example 2: Transformer networks for ensemble post-processing



CRPS



In comparison to the ENS-10 benchmarks from https://arxiv.org/abs/2206.14786

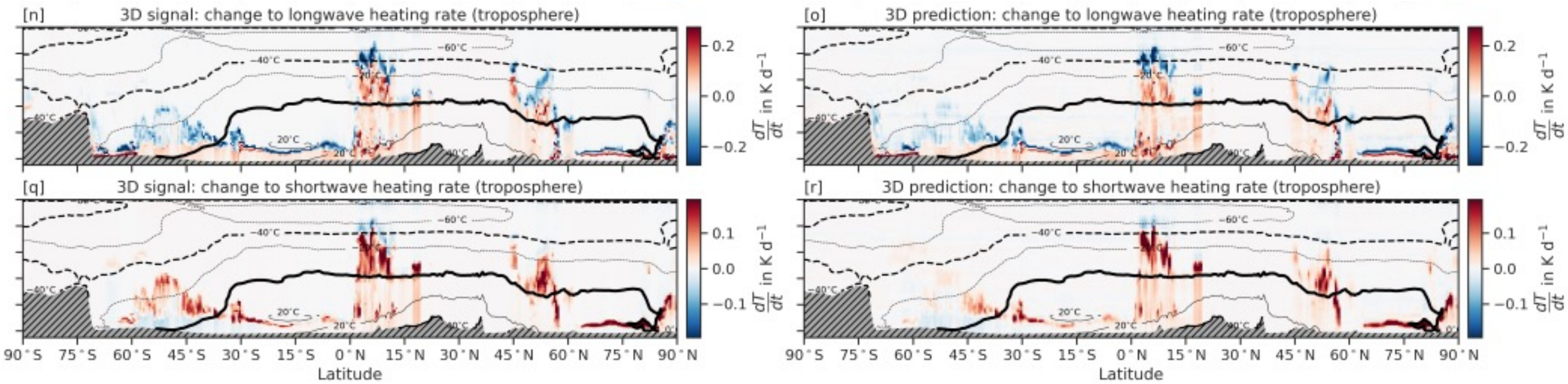
		Z500 $[\mathrm{m}^2\,\mathrm{s}^{-2}]$		T850 [K]		T2m [K]	
Metric	**Model**	5-ENS	10-ENS	5-ENS	10-ENS	5-ENS	10-ENS
CRPS	Raw	81.03	78.24	0.748	0.719	0.758	0.733
	EMOS	$79.08^{\pm 0.739}$	$81.74^{\pm 6.131}$	$0.725^{\pm 0.002}$	$0.756^{\pm 0.052}$	$0.718^{\pm 0.003}$	$0.749^{\pm 0.054}$
	MLP	$75.84^{\pm 0.016}$	$74.63^{\pm 0.029}$	$0.701^{\pm 2e-4}$	$0.684^{\pm 4e-4}$	$0.684^{\pm 6e-4}$	$0.672^{\pm 5e-4}$
	LeNet	$\mathbf{75.56}^{\pm 0.101}$	$\mathbf{74.41}^{\pm 0.109}$	$0.689^{\pm 2e-4}$	$0.674^{\pm 2e-4}$	$0.669^{\pm 7e-4}$	$0.659^{\pm 4e-4}$
	U-Net	$76.66^{\pm 0.470}$	$76.25^{\pm 0.106}$	$0.687^{\pm 0.003}$	$0.669^{\pm 0.009}$	$0.659^{\pm 0.005}$	$0.644^{\pm 0.006}$
	Transformer	$77.30^{\pm 0.061}$	$74.79^{\pm 0.118}$	$\mathbf{0.686}^{\pm 0.002}$	$\mathbf{0.665}^{\pm 0.002}$	$\mathbf{0.649}^{\pm 0.004}$	$\mathbf{0.626}^{\pm 0.004}$
PoET U-net Transformer			**73.97**		**0.650**		

Ben Bouallegue, Weyn, Clare, Dramsch, Dueben, Chantry arXiv 2023

Science example 3: Improve results via emulation

To represent 3D cloud effects for radiation (SPARTACUS) within simulations of the Integrated Forecast Model is four time slower than the standard radiation scheme (Tripleclouds)

Can we emulate the difference between Tripleclouds and SPARTACUS using neural networks?



Rel. Cost	Tripleclouds	SPARTACUS	Neural Network	Tripleclouds+Neural Network
	1.0	4.4	0.003	1.003

Meyer, Hogan, Dueben, Mason JAMES 2022

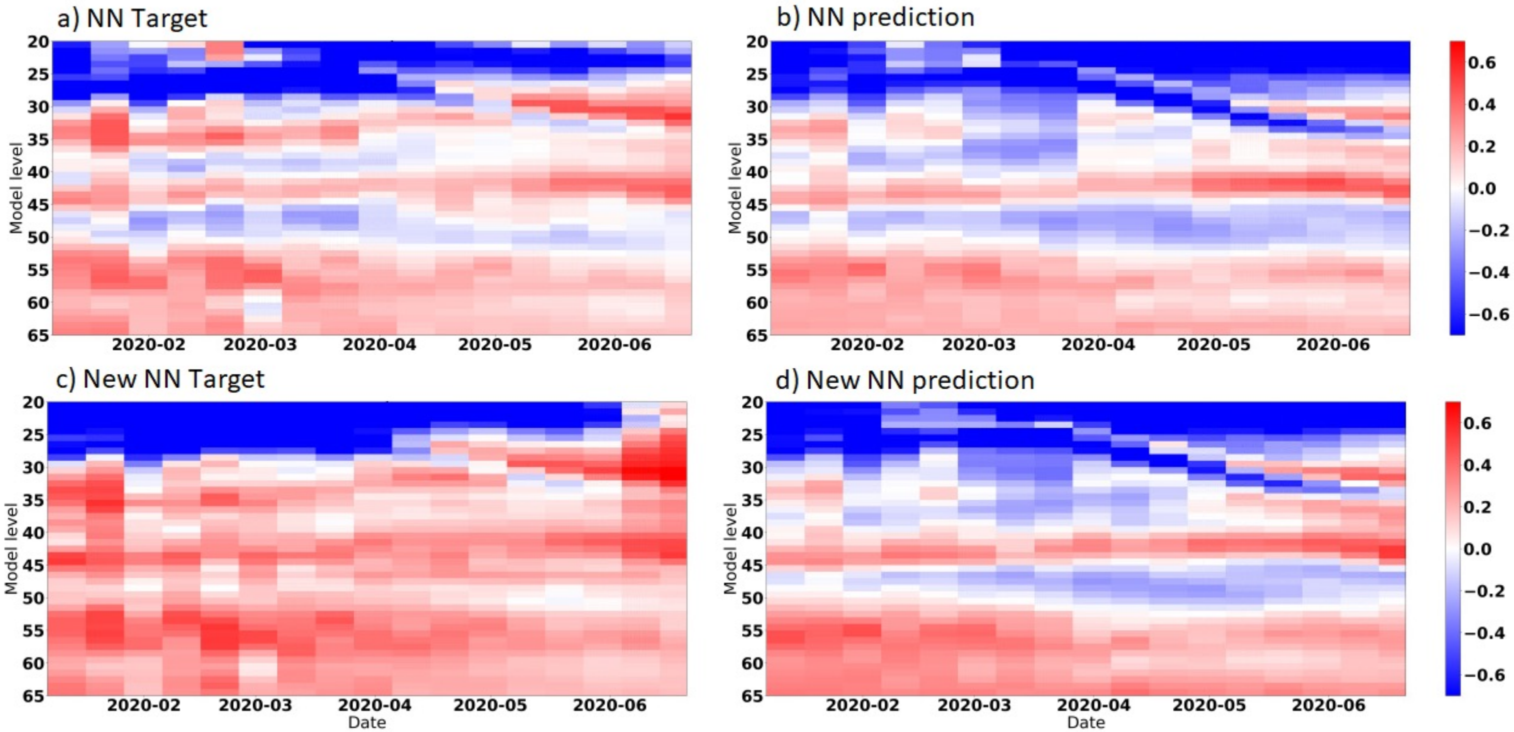
Learn how to combine operational models and machine learning

- During data-assimilation the model trajectory is “synchronised” with observations
- It is possible to learn model error when comparing the model with (trustworthy) observations

Approach: Learn model error from a direct comparison of the model trajectory and observations

Benefit: Correct for model error and understand model deficiencies

Question: What happens when the model is upgraded and the error pattern change?



Laloyaux, Kurth, Dueben, Hall JAMES 2022

We need to make developments comparable via benchmark datasets for Phase II

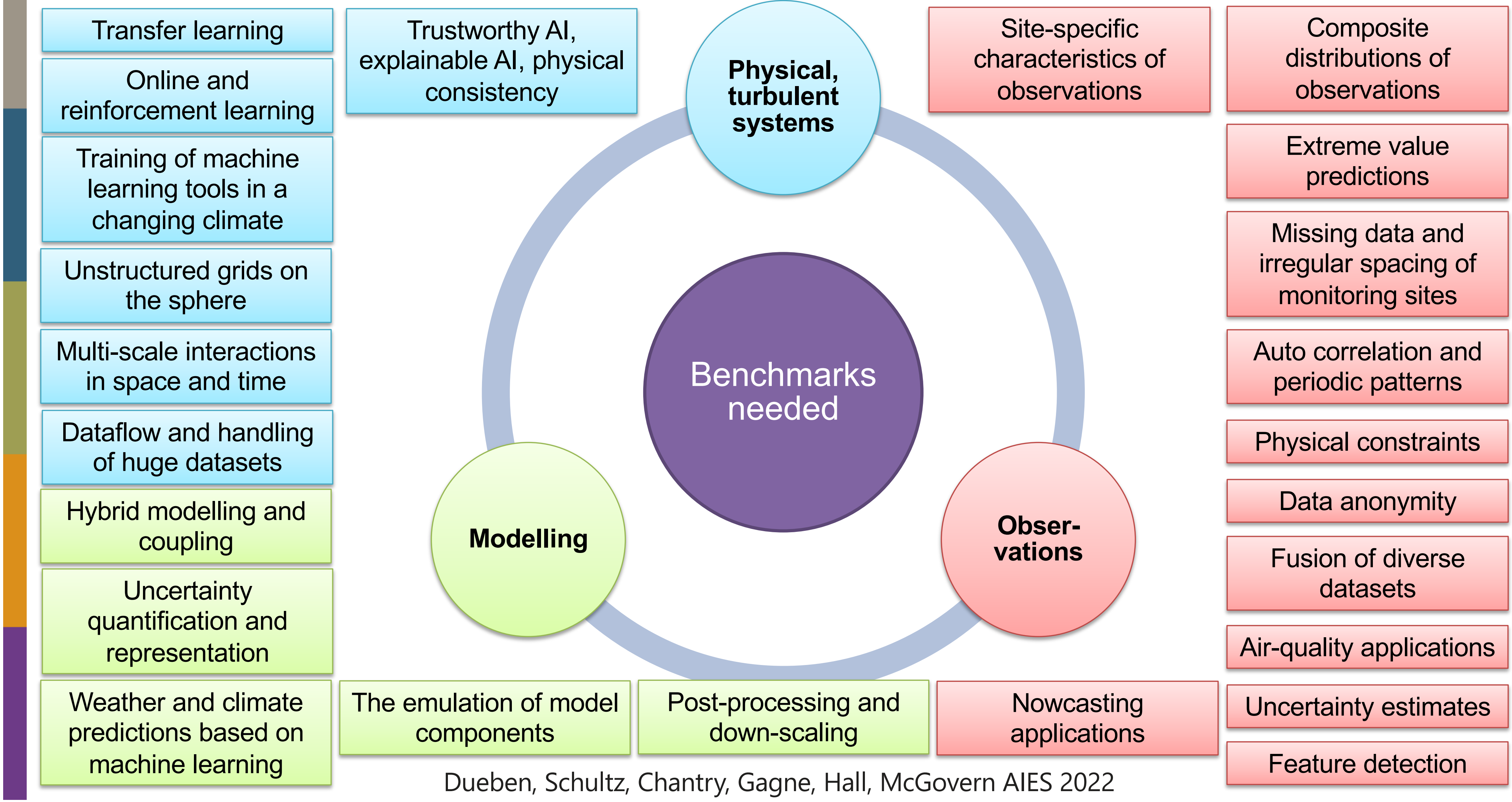
Benchmark datasets include:

- 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:

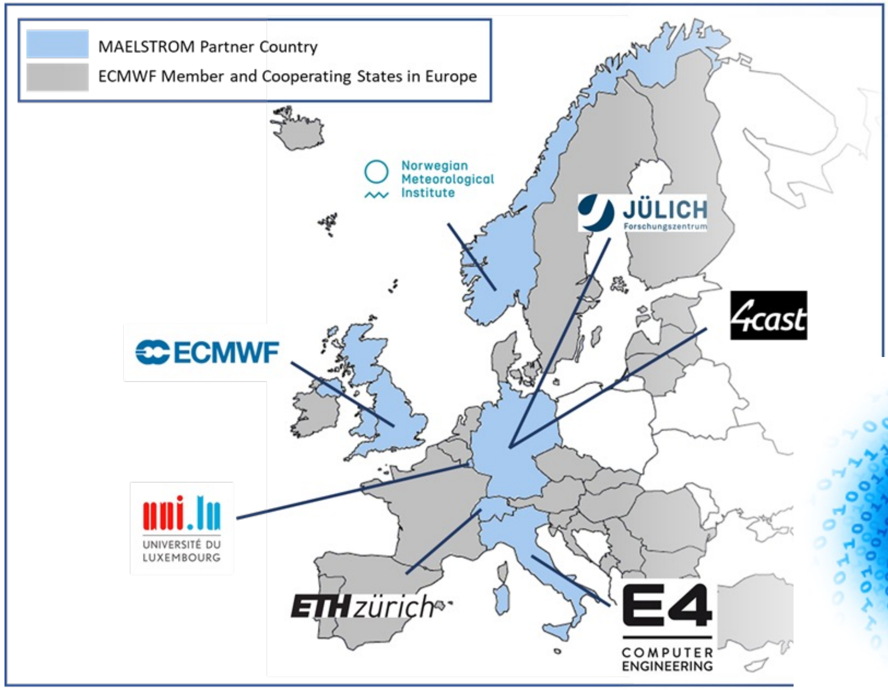
- They allow a quantitative evaluation of machine learning approaches
- They reduce data access and help scientists to get access to relevant data
- They allow for a separation of concerns between domain sciences and machine learning experts
- They allow for a separation of concerns between domain sciences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

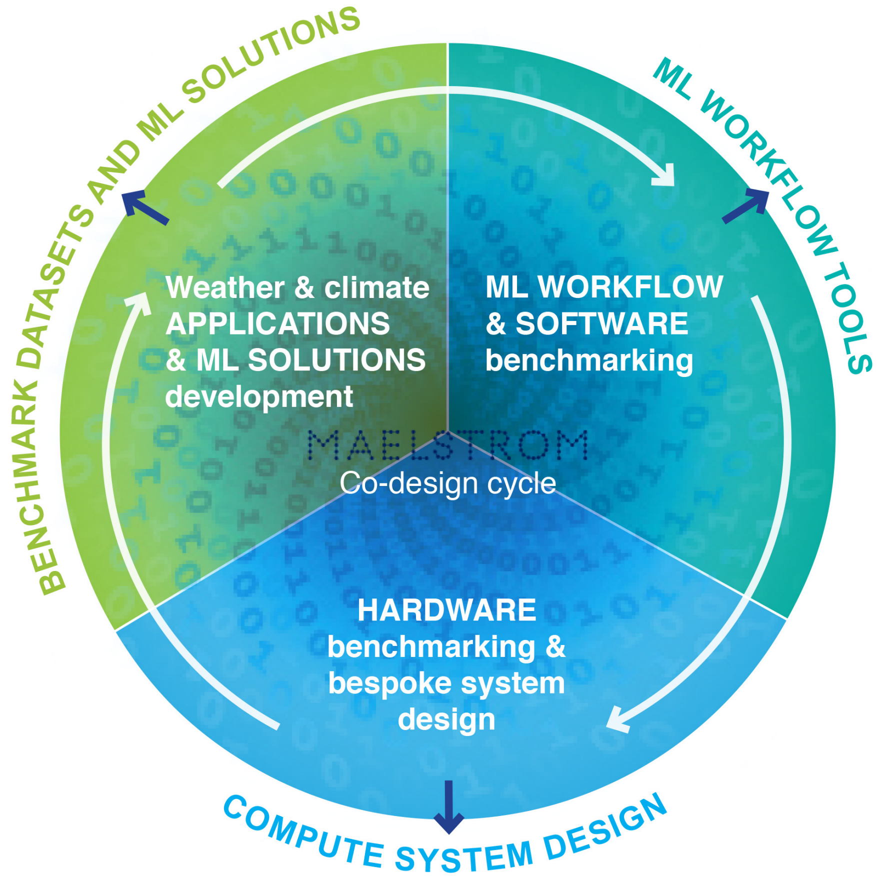
Missing machine learning benchmark datasets for atmospheric sciences



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

Our **datasets** have been published!

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

Phase X: A full machine learning model for weather and climate

ChatGPT:

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

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

For example, the play could end with Hamlet successfully avenging his father then choosing to forgive and reconcile with his uncle, the new king. This woul Hamlet to put aside his grudges and bitterness, and to find peace and happin

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

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

The New York Times

SUBSCRIBE F

THE SHIFT

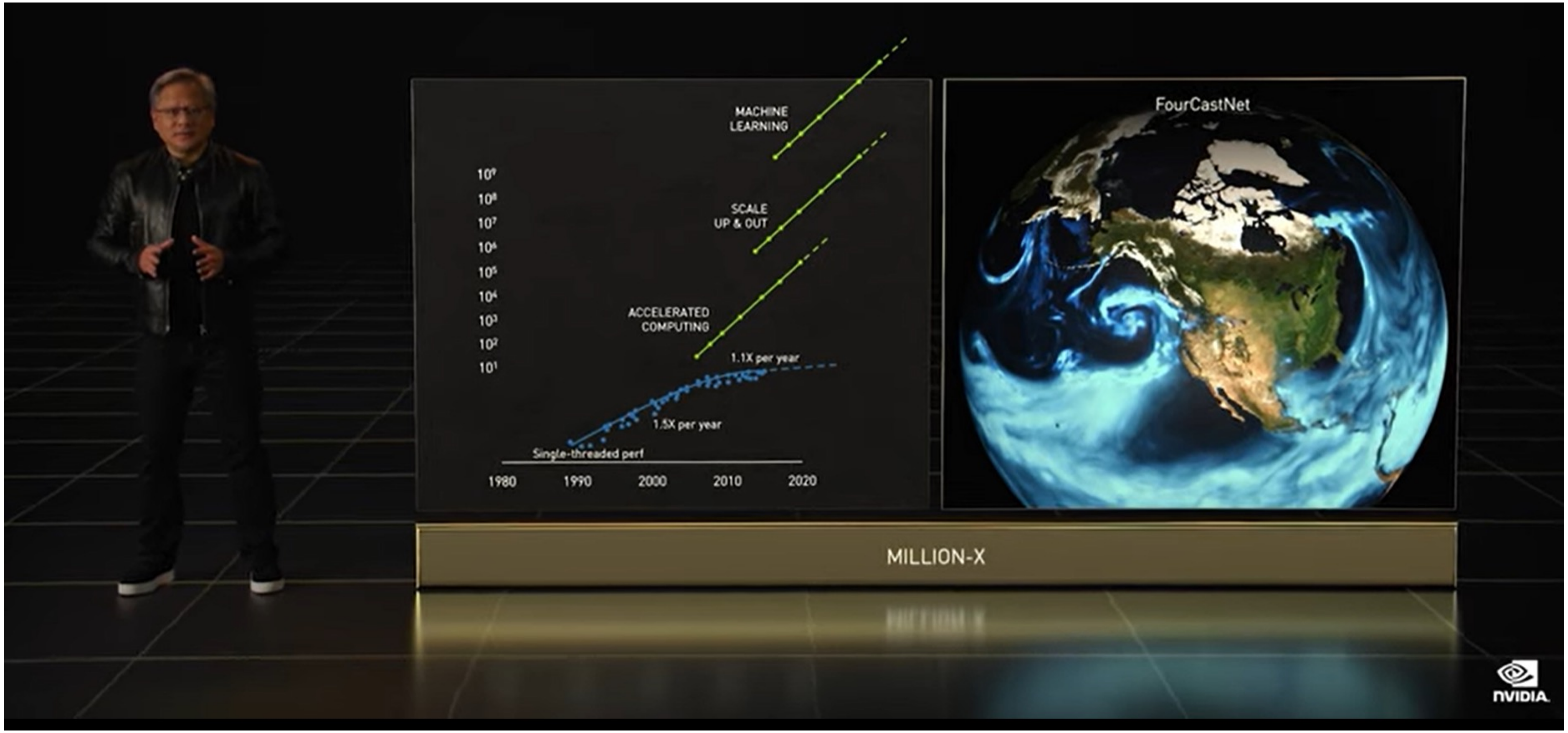
An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy.

"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

Phase X: A full machine learning model for weather and climate

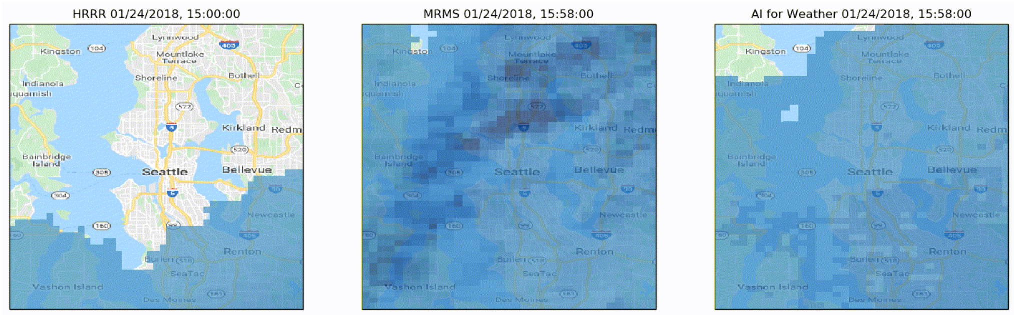


NIVIDA’s Earth-2 is coming with FourCastNet

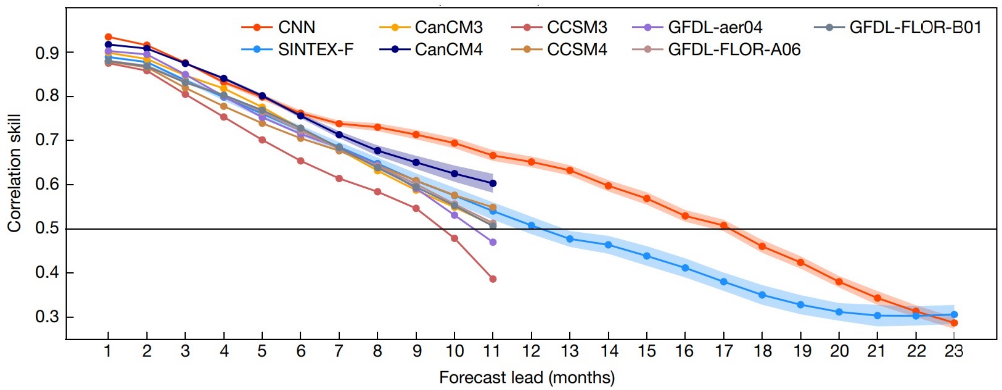
Can we replace conventional Earth System models by deep learning?

Nowcasting: E.g. via Google's MetNet or by Deepmind in Ravuri et al. *Nature* 2021

NOAA forecast   Ground truth   Machine learning:

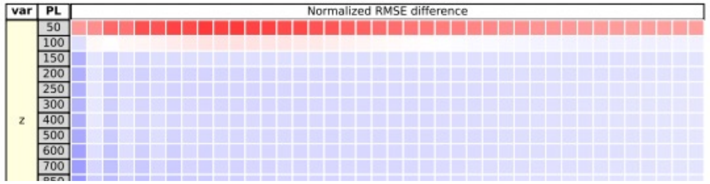


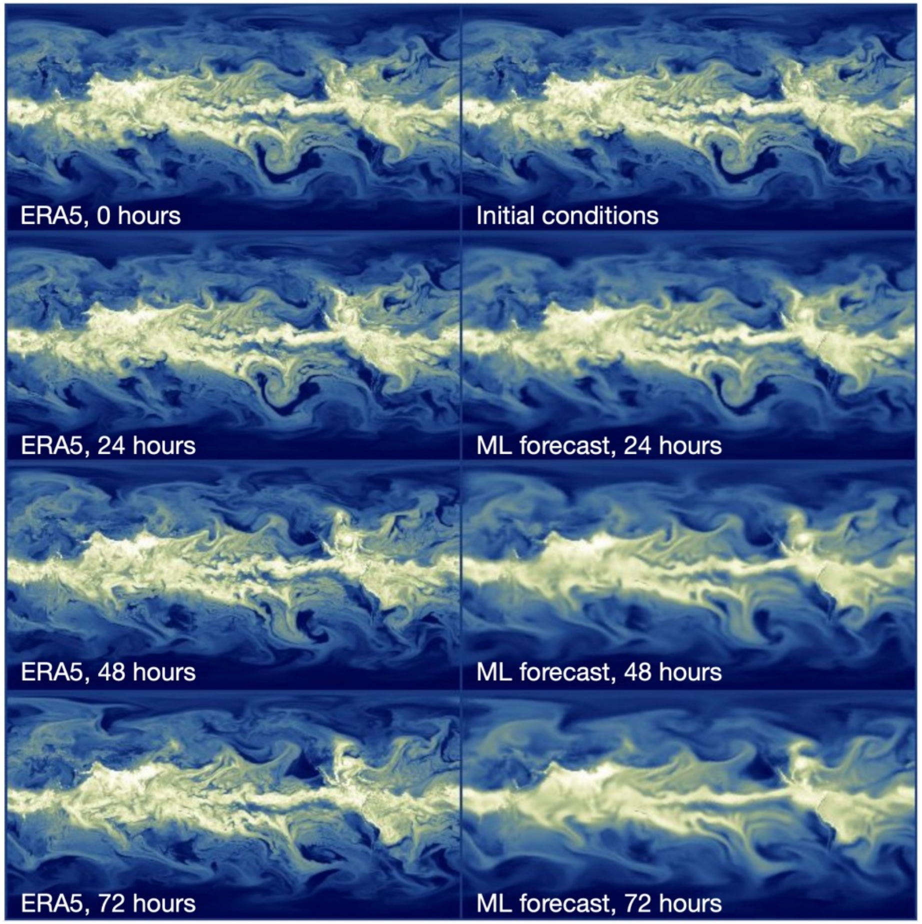
Deep learning for multi-year ENSO forecasts: E.g. Ham, Kim, Luo *Nature* 2019



And climate?

Phase X: A full machine learning model for weather and climate





GraphCast from Google/Deepmind is beating conventional weather forecast model in deterministic scores.

But how do these models actually work?

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

They get the best results when using very large timesteps (6h vs. 600s) and a couple of the previous timesteps as input.
→ Implicit? Explicit?

They do not model the physical equations,
they learn to please the scores.

Can they extrapolate? Learn uncertainty? Learn from observations? Fill the state vector? Learn all important processes?

Images from Keisler (2022)

Phase X: A full machine learning model for weather and climate

arXiv > physics > arXiv:2307.10128

Search...
Help | Advanc

Physics > Atmospheric and Oceanic Physics

[Submitted on 19 Jul 2023]

The rise of data-driven weather forecasting

Zied Ben-Bouallegue, Mariana C A Clare, Linus Magnusson, Estibaliz Gascon, Michael Maier-Gerber, Martin Janousek, Mark Rodwell, Florian Pinault, Jesper S Dramsch, Simon T K Lang, Baudouin Raoult, Florence Rabier, Matthieu Chevallier, Irina Sandu, Peter Dueben, Matthew Chantry, Florian Pappenberger

Data-driven modeling based on machine learning (ML) is showing enormous potential for weather forecasting. Rapid progress has been made with impressive results for some applications. The uptake of ML methods could be a game-changer for the incremental progress in traditional numerical weather prediction (NWP) known as the 'quiet revolution' of weather forecasting. The computational cost of running a forecast with standard NWP systems greatly hinders the improvements that can be made from increasing model resolution and ensemble sizes. An emerging new generation of ML models, developed using high-quality reanalysis datasets like ERA5 for training, allow forecasts that require much lower computational costs and that are highly-competitive in terms of accuracy. Here, we compare for the first time ML-generated forecasts with standard NWP-based forecasts in an operational-like context, initialized from the same initial conditions. Focusing on deterministic forecasts, we apply common forecast verification tools to assess to what extent a data-driven forecast produced with one of the recently developed ML models (PanguWeather) matches the quality and attributes of a forecast from one of the leading global NWP systems (the ECMWF IFS). The results are very promising, with comparable skill for both global metrics and extreme events, when verified against both the operational analysis and synoptic observations. Increasing forecast smoothness and bias drift with forecast lead time are identified as current drawbacks of ML-based forecasts. A new NWP paradigm is emerging relying on inference from ML models and state-of-the-art analysis and reanalysis datasets for forecast initialization and model training.

Subjects: **Atmospheric and Oceanic Physics (physics.ao-ph)**
Cite as: arXiv:2307.10128 **[physics.ao-ph]**
(or arXiv:2307.10128v1 **[physics.ao-ph]** for this version)
https://doi.org/10.48550/arXiv.2307.10128

Submission history

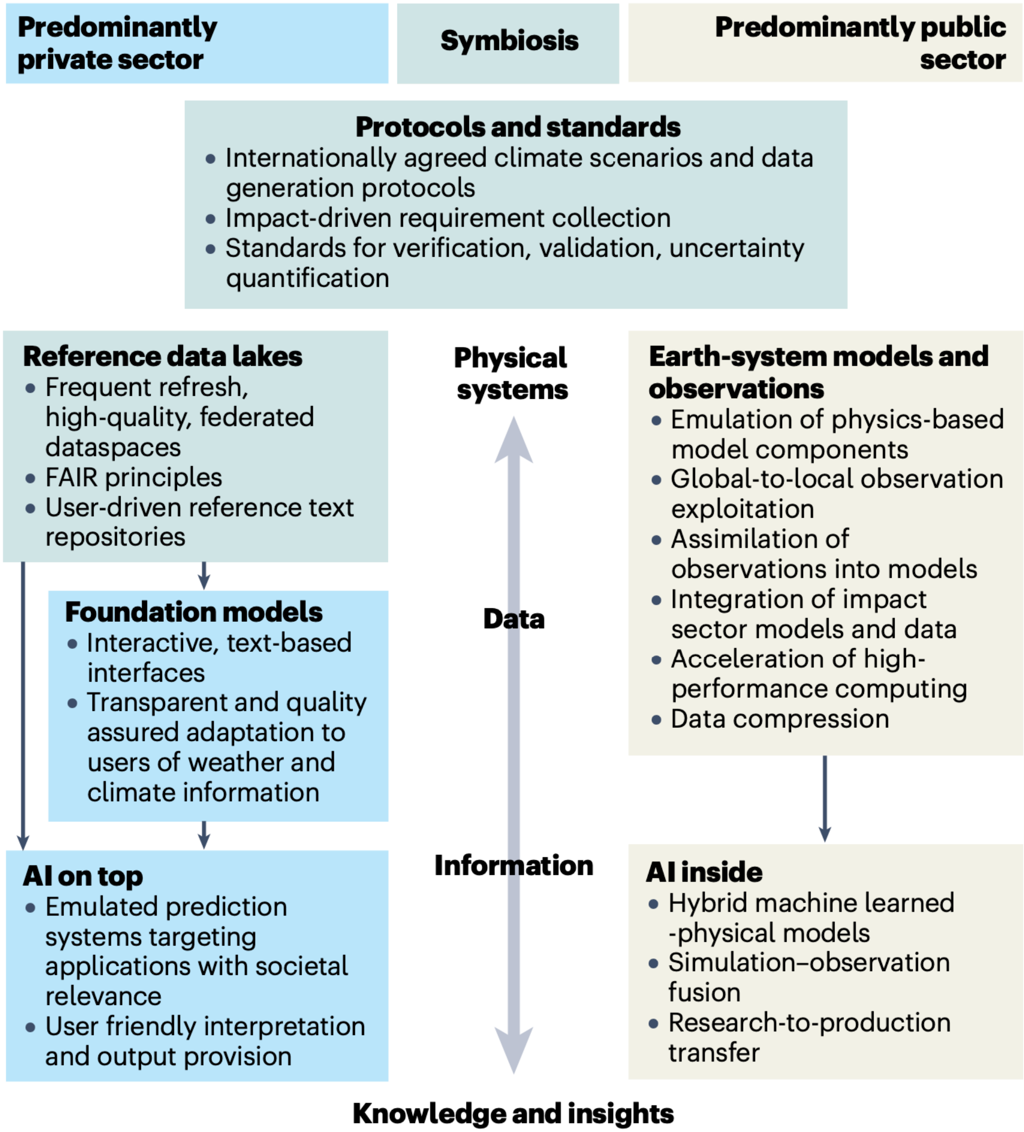
From: Zied Ben Bouallegue [view email]
[v1] Wed, 19 Jul 2023 16:51:08 UTC (18,531 KB)

Phase X: A full machine learning model for weather and climate

How will the forecast system of the future look like?

Conventional model	Hybrid model	Pure ML model
Met Services	Met Services	Technology companies, SMEs and Met Services
Development cycles of decades with 10-100 developers	Development cycles of years with 1-10 developers based on conventional models	Development cycles of years with 1-10 developers
Fortran/DSLs/bits of Python	Fortran/Python/TensorFlow/PyTorch	Python/TensorFlow/PyTorch
CPUs and GPUs, single precision, low sustained flop-rate	CPUs and GPUs, single precision, low sustained flop-rate	GPUs and ML accelerators, low precision, high sustained flop-rate
Needs significant data-handling infrastructure	Needs significant data-handling infrastructure	Needs many tera-bytes of training data
Developed and run at peta-scale	Developed and run at peta-scale	Training at peta-scale, inference on single node
Unbeatable in reliability and physical consistency, bad time-to-solution	Reasonable physical consistency, bad time-to-solution	Unbeatable in scores (deterministic/ensemble) and time-to-solution
Perform well in data assimilation	Unbeatable in data assimilation	Perform reasonable in data assimilation
Unbeatable in climate simulations and ocean modelling	Questionable in climate simulations	Unusable for climate simulations, geo-engineering, climate attribution, large-scale 1/100 years extreme events, ocean models

How will ML 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.

You want to learn more? – Have a look at our MOOC material

ECMWF Massive Open Online Course (MOOC) on Machine Learning in Weather & Climate:
https://lms.ecmwf.int/course/index.php?categoryid=1

40h of content, >9000 registered participants, 159 countries, 60 experts, 47 videos

What is the direction? – Imagine if…

- …we could collect and centralise most datasets of observations from the past and presence, as well as model output and reanalysis data
- …we would have mapping tools from any point in time and space to any point in time and space for all datasets available
- …we would have interpretation tools for physical reasoning including the extraction of physical laws and the understanding of causality
- …we would have a tool to estimate uncertainties of all datasets based on the interpretation of mappings between different datasources
- …all of these tools were scalable and easy to use from Python, Jupyter, Julia…

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

ECMWF | EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

The strength of a common goal