

# SciML for Weather Forecasting



SPL

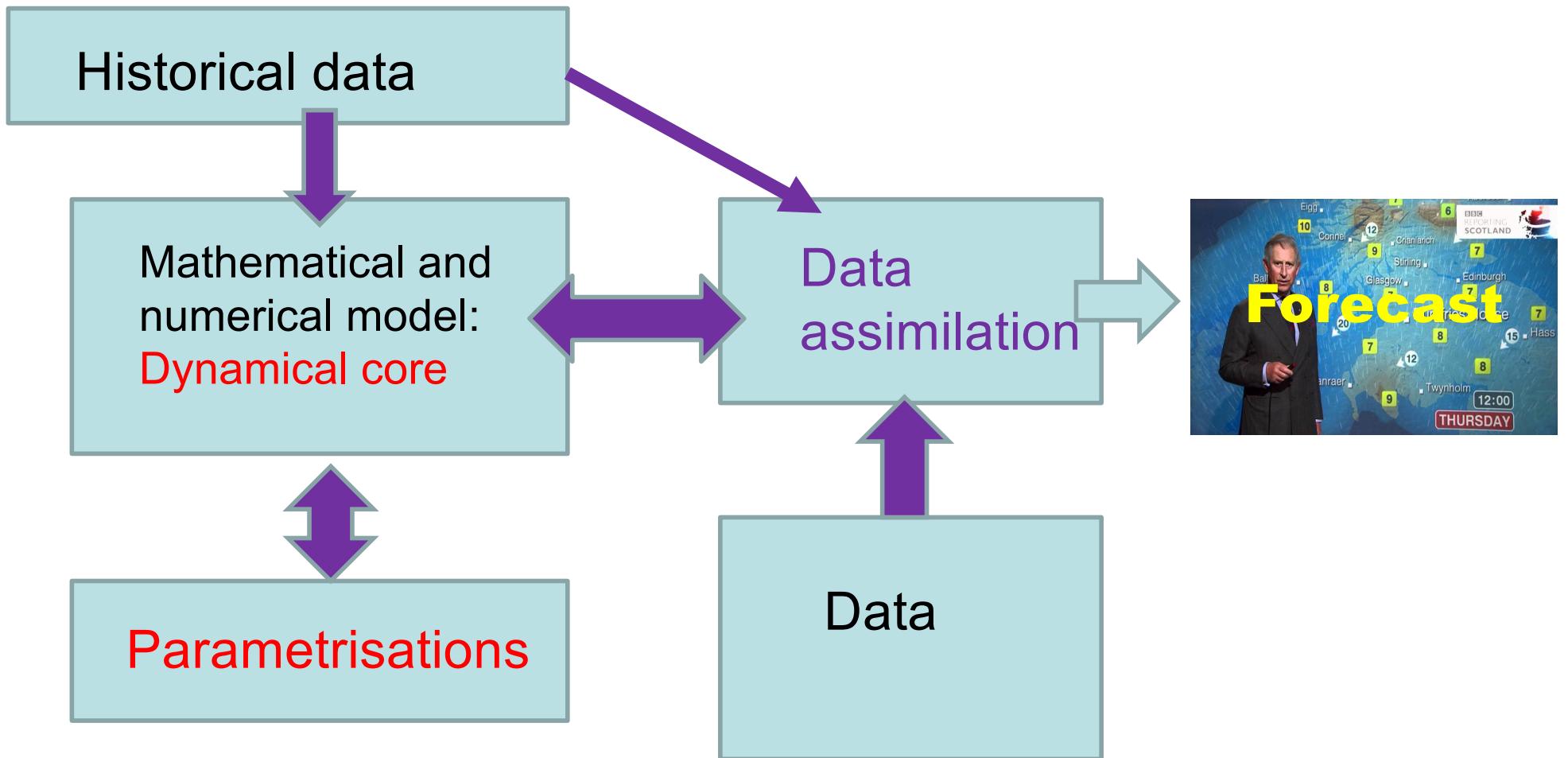


UNIVERSITY OF  
**BATH**

## Some papers

- GraphCast (2023)
- Aardvark (2025)
- Morcrette, B et. al. :Scale-Aware Parameterization of Cloud Fraction and Condensate for a Global Atmospheric Model Machine-Learned From Coarse-Grained Kilometer-Scale Simulations

# Weather forecasting process



## Areas where ML is used in weather forecasting

1. Post-processing results eg. down scaling
2. Parametrisations eg. clouds
3. Dynamical core/physics: GraphCast
4. End-to-end/data only: Aardvark

# Complex interrelated processes described by differential equations

Basic equations: **Navier-Stokes** which describe the weather

$$\frac{Du}{Dt} + 2f \times u + \frac{1}{\rho} \nabla p + g = \nu \nabla^2 u,$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0,$$

$$C \frac{DT}{Dt} - \frac{RT}{\rho} \frac{D\rho}{Dt} = \kappa_h \nabla^2 T + S_h + LP,$$

$$\frac{Dq}{Dt} = \kappa_q \nabla^2 q + S_q - P,$$

$$p = \rho RT.$$

Motion

Density

Temperature

Moisture

Pressure

For **climate** add in ice, CO<sub>2</sub>, ocean currents, vegetation, ...

Traditional methods ie. a highly accurate PDE solver  
(FEM/Finite Volume/Spectral) plus data assimilation (pre and post the forecast where possible)

Lead to ....

Data sets you can use to train Sci ML systems

- Reanalysis:

ERA5 (1979-2017 Weather predictions based on HRES forecasts at ECMWF)

- UKCP18. (Climate till 2050, based on HadGEM3)

## Anniversary

[Charts](#)[Datasets](#)[Quality of our forecasts](#)[About our forecasts](#)[Access to forecasts](#)

## ECMWF Reanalysis v5 (ERA5)

[New Datasets search](#)

ERA5 is the fifth generation ECMWF atmospheric reanalysis of the global climate covering the period from January 1940 to present. ERA5 is produced by the Copernicus Climate Change Service (C3S) at ECMWF.

ERA5 provides hourly estimates of a large number of atmospheric, land and oceanic climate variables. The data cover the Earth on a 31km grid and resolve the atmosphere using 137 levels from the surface up to a height of 80km. ERA5 includes information about uncertainties for all variables at reduced spatial and temporal resolutions.

ERA5 is available on:

- Single levels
- Pressure levels:

1000/975/950/925/900/875/850/825/800/775/750/700/650/600/550/500/450/400/350/300/250/225/200/175/150/125/100/70/50/30/20/10/7/5/3/2/1

- Potential temperature levels:

### Explore this dataset:

[Climate Data Store >](#)

Access to the data portals and their features depends on [who you are](#)

[View licence](#)

**Big claim** weather forecasting papers which make use of these data sets

GraphCast (2023)

Aardvark (2025)

Methodology: GNN/Neural operator based methods

Trained on ERA5 data and satellite data

Very rapid prediction times, but expensive to train

Firefox File Edit View History Bookmarks Tools Window Help

zim FNO H Copy Untitled Untitled Microsoft Lawrence Your p Untitled Seattle How c Inside Dee X

Wed 28 May 08:50:32

2210.0 Neural FNO H Copy Untitled Untitled Microsoft Lawrence Your p Untitled Seattle How c Inside Dee X

water discussion

www.nature.com/articles/d41586-023-03552-y

nature

View all journals Search Log in

Explore content About the journal Publish with us Subscribe Sign up for alerts RSS feed

nature > news > article

NEWS | 14 November 2023

# DeepMind AI accurately forecasts weather – on a desktop computer

The machine-learning model takes less than a minute to predict future weather worldwide more precisely than other approaches.

By [Carissa Wong](#)

[Twitter](#) [Facebook](#) [Email](#)

You have full access to this article via **JISC Springer Compact (publishing + reading)**.

Cell Signaling TECHNOLOGY Cancer Signaling Pathways & Diagrams Download

May 28



file:///Users/mascjb/Downloads/science.adi2336.pdf

1417 (2 of 6) - | + 110% ▾

**RESEARCH | RESEARCH ARTICLE**

**Fig. 1. Model schematic.**

(A) The input weather state(s) are defined on a  $0.25^\circ$  latitude-longitude grid comprising a total of  $721 \times 1440 = 1,038,240$  points. Yellow layers in the close-up pop-out window represent the five surface variables, and blue layers represent the six atmospheric variables that are repeated at 37 pressure levels ( $5 + 6 \times 37 = 227$  variables per point in total), resulting in a state representation of 235,680,480 values.

(B) GraphCast predicts the next state of the weather on the grid. (C) A forecast is made by iteratively applying GraphCast (GC) to each previous predicted state, to produce a sequence of states that represent the weather at successive lead times.

(D) The encoder component of the GraphCast architecture maps local regions of the input (green boxes) into nodes of the multimesh graph representation (green, upward arrows that terminate in the green-blue node).

(E) The processor component updates each multimesh node using learned message-passing (heavy blue arrows that terminate at a node).

(F) The decoder component maps the processed multimesh features (purple nodes) back onto the grid representation (red, downward arrows that terminate

**A Input weather state**

**B Predict the next state**

**C Roll out a forecast**

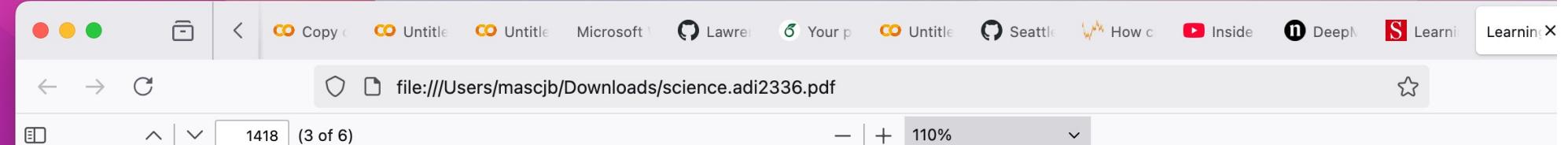
**D Encoder**

**E Processor**

**F Decoder**

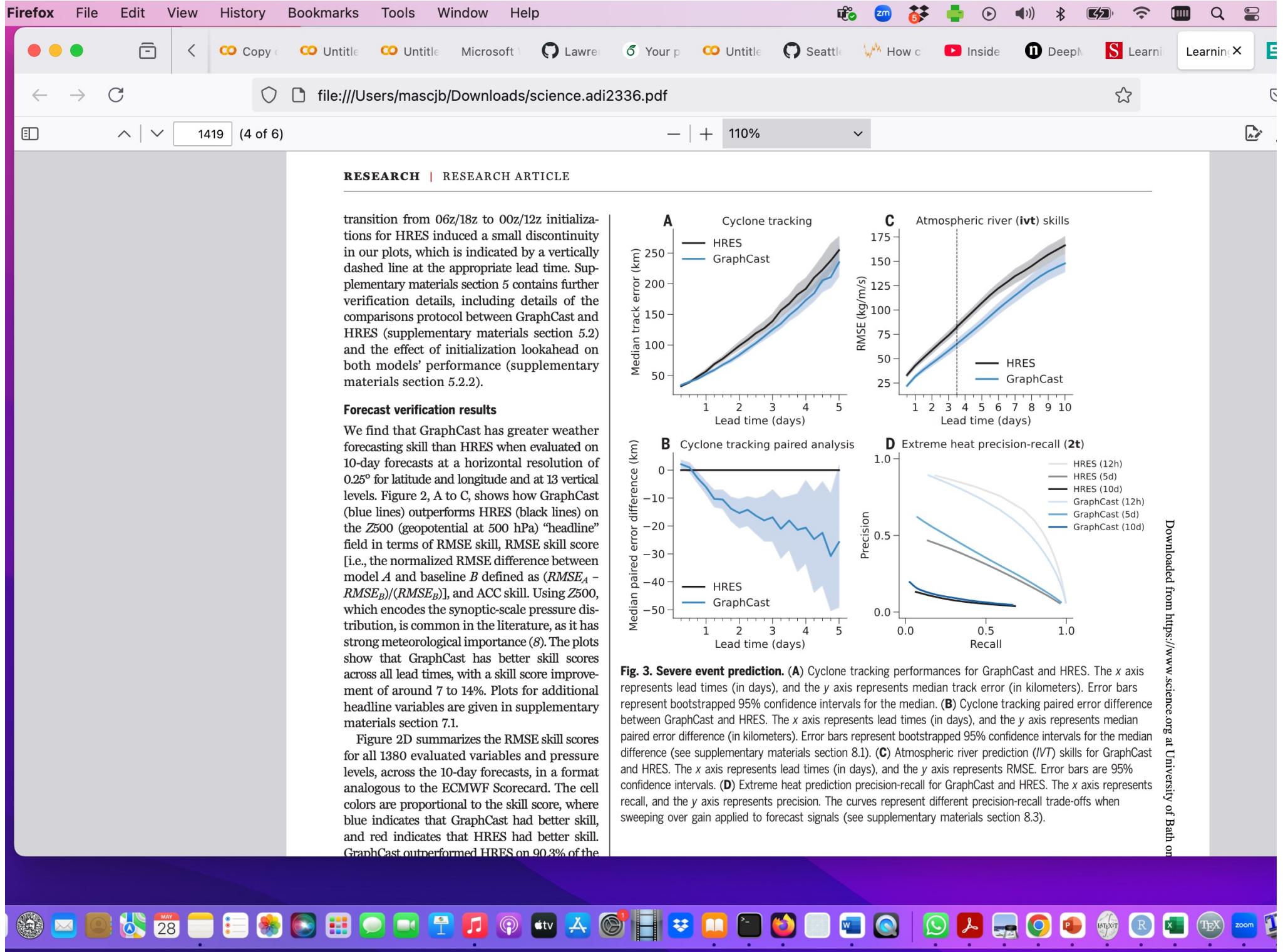
**G Simultaneous multi-mesh message-passing**

Downloaded from https://www.science.org at University of Bath on Ma



**Fig. 2. Global skill and skill scores for GraphCast and HRES in 2018.**  
**(A)** RMSE skill (y axis) for GraphCast (blue lines) and HRES (black lines), on Z500, as a function of lead time (x axis). Error bars represent 95% confidence intervals. The vertical dashed line represents 3.5 days, which is the last 12-hour increment of the HRES 06z/18z forecasts. The black line represents HRES, where lead times earlier and later than 3.5 days are from the 06z/18z and 00z/12z initializations, respectively. **(B)** RMSE skill score (y axis) for GraphCast versus HRES, on Z500, as a function of lead time (x axis). Error bars represent 95% confidence intervals for the skill score. We observe a discontinuity in GraphCast's curve because skill scores up to 3.5 days are computed between GraphCast (initialized at 06z/18z) and HRES's 06z/18z initialization, whereas skill scores

after 3.5 days are computed with respect to HRES's 00z/12z initializations. **(C)** ACC skill (y axis) for GraphCast (blue lines) and HRES (black lines), on Z500, as a function of lead time (x axis). **(D)** Scorecard of RMSE skill scores for GraphCast, with respect to HRES. Each subplot corresponds to one variable:  $U$ ,  $V$ ,  $Z$ ,  $T$ ,  $Q$ ,  $2T$ ,  $10U$ ,  $10V$ , and  $MSL$ . The rows of each heatmap correspond to the 13 pressure levels (for the atmospheric variables), from 50 hPa at the top to 1000 hPa at the bottom. The columns of each heatmap correspond to the 20 lead times at 12-hour intervals, from 12 hours on the left to 10 days on the right. Each cell's color represents the skill score, as shown in (B), where blue represents negative values (GraphCast has better skill) and red represents positive values (HRES has better skill).



GraphCast's forecast skill and efficiency compared with HRES shows that MLWP methods are now competitive with traditional weather forecasting methods. Additionally, GraphCast's performance on severe event forecasting, which it was not directly trained for, demonstrates its robustness and potential for downstream value. We believe this marks a turning point in weather forecasting, which helps open new avenues to strengthen the breadth of weather-dependent decision-making by individuals and industries by making cheap prediction more accurate and accessible as well as suitable for specific applications. With 36.7 million parameters, GraphCast is a relatively small model by modern ML standards, chosen to keep the memory footprint tractable. And while HRES is released on 0.1 resolution, 137 levels, and up to 1-hour time steps, GraphCast operates on 0.25° latitude and longitude resolution, 37 vertical levels, and 6-hour time steps, because of the ERA5 training data's native 0.25° resolution and engineering challenges in fitting higher-resolution data on hardware. Generally, GraphCast should be viewed as a family of models, with the current version being the largest we can practically fit under current engineering constraints, but which have the potential to scale much further in the future with greater computer resources and higher-resolution data

## Criticisms of Graph-Cast

- Heavily based on a very skillfully created data set
- Makes predictions based on the data set, NOT on real (eg. satellite) data
- Suffers from **spectral bias**, so not resolving small scale phenomena well. Good for (smooth) pressure, not good for (localized) precipitation and fog.
- Norm used to compare with HRES is **flattering to smoother predictions**

Predicts the wrong weather at the right time, rather than the right weather at the wrong time

# 2025: Aardvark. Predictions from satellite data only

The screenshot shows a Firefox browser window with a purple sidebar on the left. The main content area displays a blog post from <https://www.turing.ac.uk/blog/project-aardvark-reimagining-ai-weather-prediction>. The post is titled "Project Aardvark: reimagining AI weather prediction" and discusses the potential of machine learning for weather prediction across different regions. The post includes a "Learn more" button and a sidebar with publication details and related programmes.

Firefox File Edit View History Bookmarks Tools Window Help

Mon 26 May 18:21:32

2210.0 Neural FNO H Copy Untitled Untitled Microsoft Lawrence G-ada Untitled Seattle Gary L Mail - How c T Pro X

https://www.turing.ac.uk/blog/project-aardvark-reimagining-ai-weather-prediction

The Alan Turing Institute

Home Events News About us Research Skills People Opportunities Partner with us Contact us

Home + Blog

# Project Aardvark: reimagining AI weather prediction

From the Global South to the Arctic, can machine learning-enabled weather prediction better protect communities and economies?

Learn more ↓

Thursday 20 Mar 2025

Filed under  
New research

Related programmes  
Environment and Sustainability

May 26

Authors

Icons for various applications like Mail, Calendar, Photos, and others are visible at the bottom of the screen.

## Assessing opportunities and challenges

Aardvark reimagines current weather prediction methods, offering a range of opportunities.

Alongside requiring less computing power, Aardvark is fast. Traditional forecasts can take hours to produce on a supercomputer whereas, once trained, Aardvark can create forecasts within minutes and can be run on a desktop computer.

There are highly promising signs of Aardvark's accuracy too. Globally, Aardvark is already as accurate as America's Global Forecast System (GFS), but it is only using about 10% of the available data to make its forecasts, meaning that further improvements in accuracy should be possible. We are excited to see what happens as we increase the amount of data and optimise Aardvark end-to-end to provide more accurate forecasts. This new paradigm could replace the traditional numerical approach in developing countries.

A streamlined system like this could also play a significant role in democratising access to advanced forecasting tools, empowering developing or data-sparse countries to build capacity and create bespoke weather forecasting systems that previously would have required large teams to operate, deploy and maintain.

There are of course challenges and it's important to acknowledge that, whilst machine learning weather tools are moving at great pace, this is still an experimental technology that will require rigorous evaluation over a period of time.

Weather prediction tools must accurately predict all types of weather, and extremes like hurricanes and floods are especially important. Unfortunately, rare events like these are less represented in the training data, meaning that AI systems may struggle more on these phenomena.

We also need to ensure we account for our changing climate, which could render models trained on past data less accurate.

Early signs suggest that we can rise to these challenges.



### What's next for Aardvark?



## Article

# End-to-end data-driven weather prediction

<https://doi.org/10.1038/s41586-025-08897-0>

Received: 10 July 2024

Accepted: 12 March 2025

Published online: 20 March 2025

Open access

Check for updates

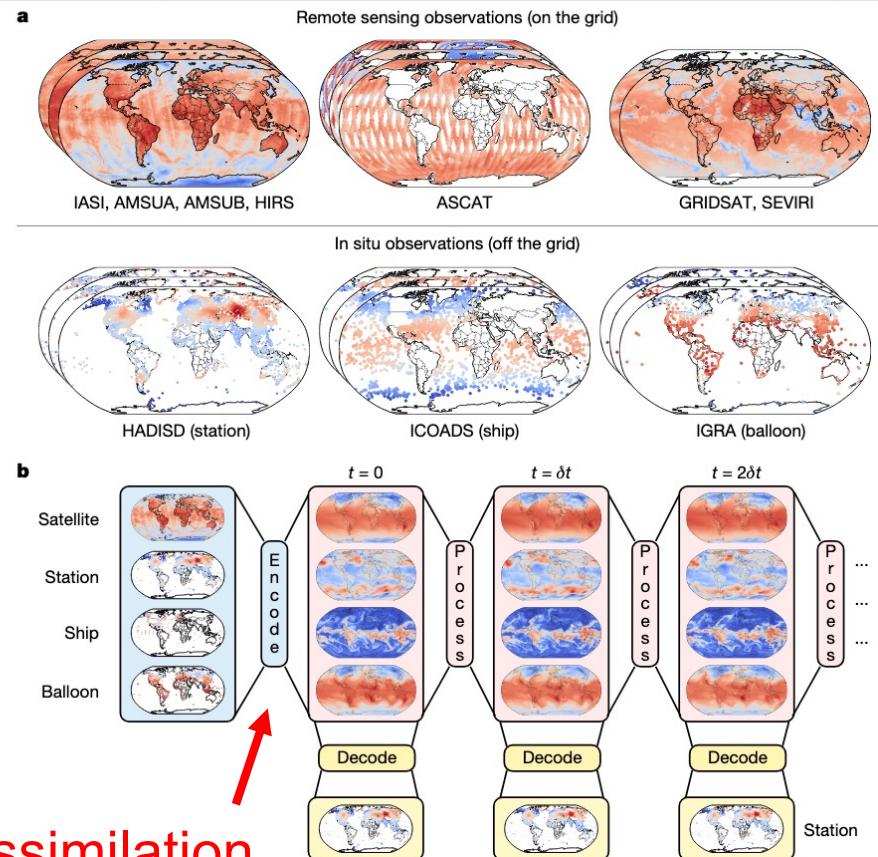
Anna Allen<sup>1,11</sup>, Stratis Markou<sup>2,11</sup>, Will Tebbutt<sup>2,9</sup>, James Requeima<sup>3</sup>, Wessel P. Bruinsma<sup>4</sup>, Tom R. Andersson<sup>5,10</sup>, Michael Herzog<sup>6</sup>, Nicholas D. Lane<sup>1</sup>, Matthew Chantry<sup>7</sup>, J. Scott Hosking<sup>5,8</sup> & Richard E. Turner<sup>2,8</sup>

Weather prediction is critical for a range of human activities, including transportation, agriculture and industry, as well as for the safety of the general public. Machine learning transforms numerical weather prediction (NWP) by replacing the numerical solver with neural networks, improving the speed and accuracy of the forecasting component of the prediction pipeline<sup>1–6</sup>. However, current models rely on numerical systems at initialization and to produce local forecasts, thereby limiting their achievable gains. Here we show that a single machine learning model can replace the entire NWP pipeline. Aardvark Weather, an end-to-end data-driven weather prediction system, ingests observations and produces global gridded forecasts and local station forecasts. The global forecasts outperform an operational NWP baseline for several variables and lead times. The local station forecasts are skilful for up to ten days of lead time, competing with a post-processed global NWP baseline and a state-of-the-art end-to-end forecasting system with input from human forecasters. End-to-end tuning further improves the accuracy of local forecasts. Our results show that skilful forecasting is possible without relying on NWP at deployment time, which will enable the realization of the full speed and accuracy benefits of data-driven models. We believe that Aardvark Weather will be the starting point for a new generation of end-to-end models that will reduce computational costs by orders of magnitude and enable the rapid, affordable creation of customized models for a range of end users.

✓ You: Actions from last meeting. 1. Im...



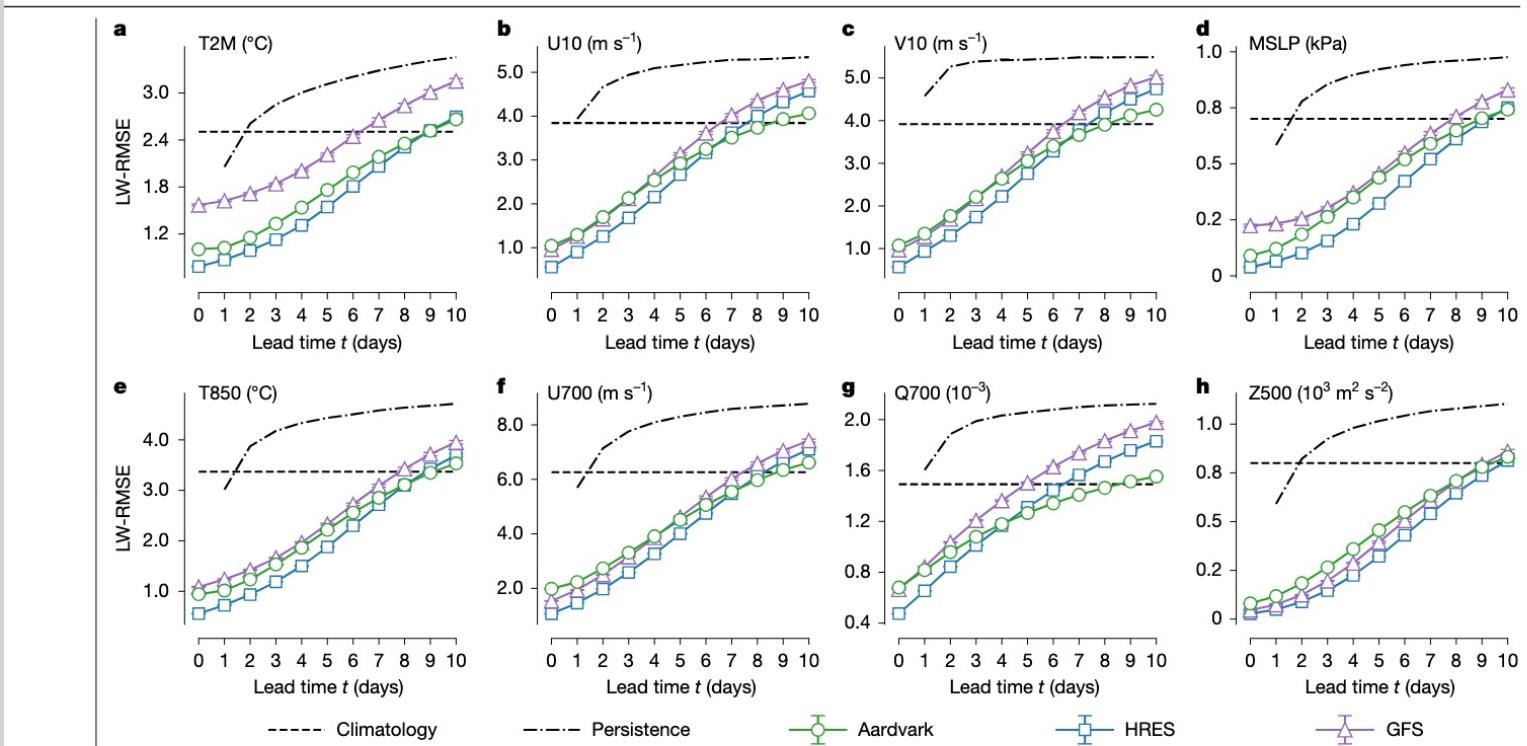
## Article



## Data assimilation

**Fig. 1 | Data and operation of Aardvark Weather.** **a**, Different data sources leveraged in Aardvark. The input data consist of observations from remote sensing instruments (top row), which we pre-grid before passing to the model, as well as in situ observations from land and marine observation platforms and radiosondes (bottom row). Each of these data modalities contains several observational variables, of which we selected a subset here for the purposes of illustration. Here we show remote sensing data<sup>40–45</sup>, after performing our gridding step, and raw in situ data<sup>46–48</sup>. Note that the colours in all six plots are meant for illustration purposes. The remote sensing data also include a range of metadata about the measurements, omitted here for simplicity. White areas

indicate regions of missing data, which must be handled by the encoder module of Aardvark. **b**, Aardvark at deployment time. First, an encoder module uses raw observations as input to estimate the initial state of the atmosphere across key variables at  $t = 0$ . Next, a processor module ingests the estimated state to produce a forecast at the next lead time  $t = \delta t$ . Forecasts at subsequent lead times are produced autoregressively. Finally, a decoder module is applied to the on the grid states to produce off the grid predictions. The modular design of Aardvark allows for pretraining on large high-quality ERA5 reanalysis data<sup>34</sup>. In this figure, the displayed data are the training data used to train each module of Aardvark from the aforementioned sources.



**Fig. 2 | Gridded global forecast performance for selected variables.**

**a–h**, Latitude-weighted RMSE using ERA5 (ref. 34) reanalysis data as the ground truth, on the held-out test year (2018), for the four surface variables: 2-m temperature (**a**; T2M), 10-m eastward wind (**b**; U10), 10-m northward wind (**c**; V10) and mean sea level pressure (**d**; MSLP), as well as four headline upper-atmosphere variables: temperature at 850 hPa (**e**; T850), eastward wind at 700 hPa (**f**; U700), specific humidity at 700 hPa (**g**; Q700) and geopotential at 500 hPa (**h**; Z500) as a function of lead time  $t$ . At lead time  $t = 0$ , Aardvark

predicted the initial atmospheric state from observational data alone. The error at  $t = 0$  corresponds to the error in the initial state. Note that HRES has a non-zero error at  $t = 0$  compared to ERA5 reanalysis. The HRES forecasts<sup>33</sup> we used have been conservatively re-gridded to prevent aliasing, and we performed the same operation on the GFS forecasts<sup>49</sup>. We report the mean performance of each system together with 98% confidence intervals in our estimate of the mean performance.

use a recurrent update in which the previous forecast is adjusted in light of new observations, similar to Kalman filter recursions in a Markov model. In principle, data assimilation accumulates information from observations across all past time steps. However, in practice, it has been estimated that the effective window size is as short as 4 days (ref. 25). Owing to the complexities of training recurrent neural networks, including the need for a spin-up period and gradient instabilities<sup>26</sup>, we opted

a way that mimics how it will be deployed. We started by pretraining the encoder module using raw observations as input and reanalysis data as targets. An advantage of this machine learning approach is that the model can learn to correct for biases in the input observations during training; therefore, no bias correction step was performed on the input data. We also pretrained the processor using reanalysis data for both inputs and targets and then fine-tuned the output of the

# ML Parametrisations

Augment a NWP forecast by using ML to give sub-grid scale level physics



## Example: CARAMEL Project: Cloud fraction prediction

Scale-Aware Parameterization of Cloud Fraction and Condensate for a Global Atmospheric Model  
Machine-Learned From Coarse-Grained Kilometer-Scale Simulations

Cyril Morcrette, Tobias Cave, Helena Reid, Joana da Silva Rodrigues, Teo Deveney, Lisa Kreusser,  
Kwinten Van Weverberg, and B

# Conclusions

Scientific Machine Learning is showing great promise in both making its own weather forecasts and in enhancing existing forecasting methods

We will have to see if the great recent claims are fully justified

This is a VERY rapidly growing field

Watch this space!!!!

