

# Running Data-Driven NWP

Jesper Dramsch & Gert Mertes

Scientist for Machine Learning

[Jesper.Dramsch@ecmwf.int](mailto:Jesper.Dramsch@ecmwf.int)

[Gert.Mertes@ecmwf.int](mailto:Gert.Mertes@ecmwf.int)



# FourCastNet

FourCastNet: NV's DDWP, first to be trained at ambitious 0.25-deg global resolution

FourCastNet, Pathak et al. (2022), 0.25°, ~1,000,000 Pixels, VIT+AFNO

GNN, Keisler et al. (2022), 1°, 64,000 Pixels, Graph Neural Networks

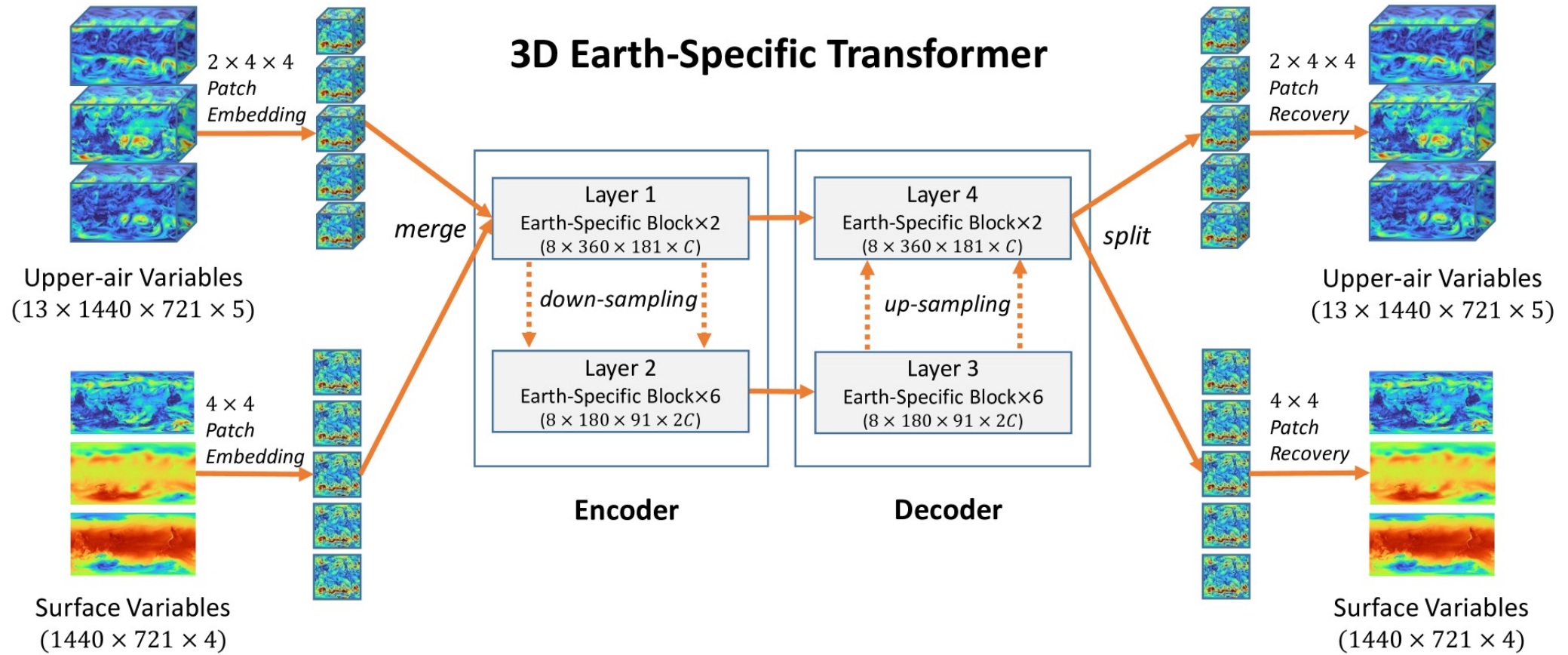
DLWP, Weyn et al. (2020). 2°, 16K pixels, Deep CNN on Cubesphere/(2021) ResNet

Weyn et al. (2019), 2.5° N.H only, 72x36, 2.6k pixels, ConvLSTM

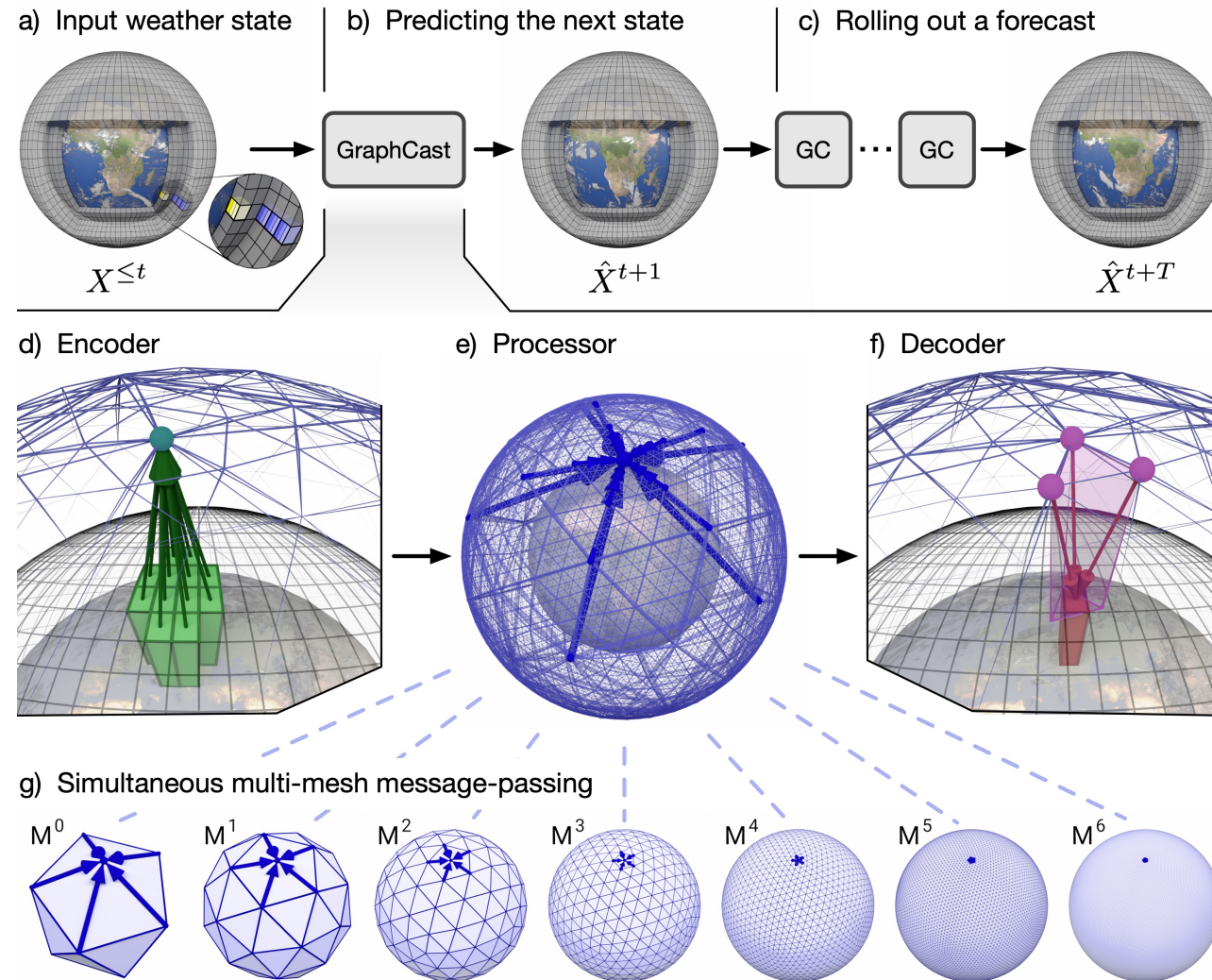
WeatherBench, Rasp et al. (2020). 5.625°, 64x32, 2K pixels, CNN

Deuben & Bauer (2018), 6°, 60x30, 1.8K pixels, MLP

# Pangu Weather



# DeepMind GraphCast



# AIFS v0.21

Live from Jan 2024.

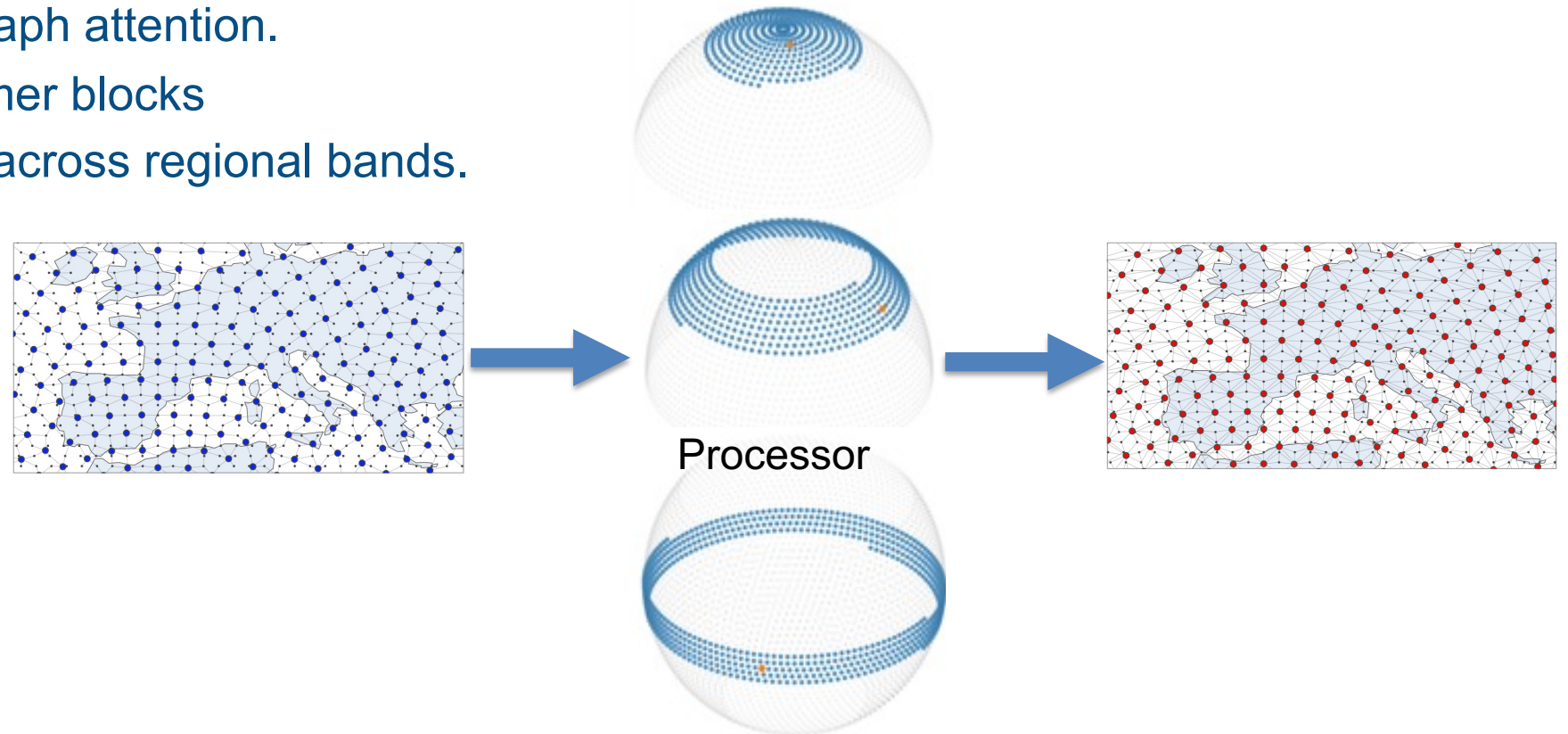
**Resolution 0.25 degrees** (4x finer)

New architecture.

Encoder/decoder: graph attention.

Processor: Transformer blocks

attention across regional bands.



# AI-Models Plugins for FOSS Data-Driven NWP



The collage displays four arXiv paper abstracts:

- Pangu-Weather: A 3D High-Resolution Model** (arXiv:2211.02556) by Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, Qi Tian. Submitted on 3 Nov 2022.
- FourCastNet: A Global Data-driven High-resolution Neural Operators** (arXiv:2202.11214) by Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashes Zongyi Li, Kamyar Azizzadenesheli, Pedram Hassanzadeh, Karthik Kashinath, Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wilmers, Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Hollingsworth, Battaglia. Submitted on 24 Dec 2022 (v1), last revised 4 Aug 2023 (this version, v21).
- GraphCast: Learning skillful medium-range weather forecasts** (arXiv:2212.12794) by Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wilmers, Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Hollingsworth, Battaglia. Submitted on 24 Dec 2022 (v1), last revised 4 Aug 2023 (this version, v21).
- Spherical Fourier Neural Operators: Learning Stable Dynamics on the Sphere** (arXiv:2306.03838) by Boris Bonev, Thorsten Kurth, Christian Hundt, Jaideep Pathak, Maximilian Baust, Karthik Kashinath, Anima Anandkumar. Submitted on 6 Jun 2023.

> pip install ai-models-panguweather

> ai-models panguweather

ONNX for model weights

> pip install ai-models-fourcastnet

> ai-models fourcastnet

PyTorch for code and model weights

> ai-models graphcast

Jax for code and model weights

> pip install ai-models-fourcastnetv2

> ai-models fourcastnetv2

PyTorch for code and model weights

# Operationalising Data-Driven Numerical Weather Prediction



# pip install ai-models

The screenshot shows the GitHub repository page for `ecmwf-lab/ai-models`. The page is viewed in a private browser window. The main content area displays the repository's README, which includes the following sections:

- ai-models**: The `ai-models` command is used to run AI-based weather forecasting models. These models need to be installed independently.
- Usage**: Although the source code `ai-models` and its plugins are available under open sources licences, some model weights may be available under a different licence. For example some models make their weights available under the CC-BY-NC-SA 4.0 license, which does not allow commercial use. For more informations, please check the license associated with each model on their main home page, that we link from each of the corresponding plugins.
- Prerequisites**: Before using the `ai-models` command, ensure you have the following prerequisites:
  - Python 3.10 (it may work with different versions, but it has been

The right sidebar of the repository page shows the following information:

- last month**: + 30 releases
- Packages**: No packages published
- Contributors** (3):
  - `b8raout`
  - `floriankrb` Florian Pinault
  - `gmertes` Gert Mertes
- Languages**: Python 100.0%

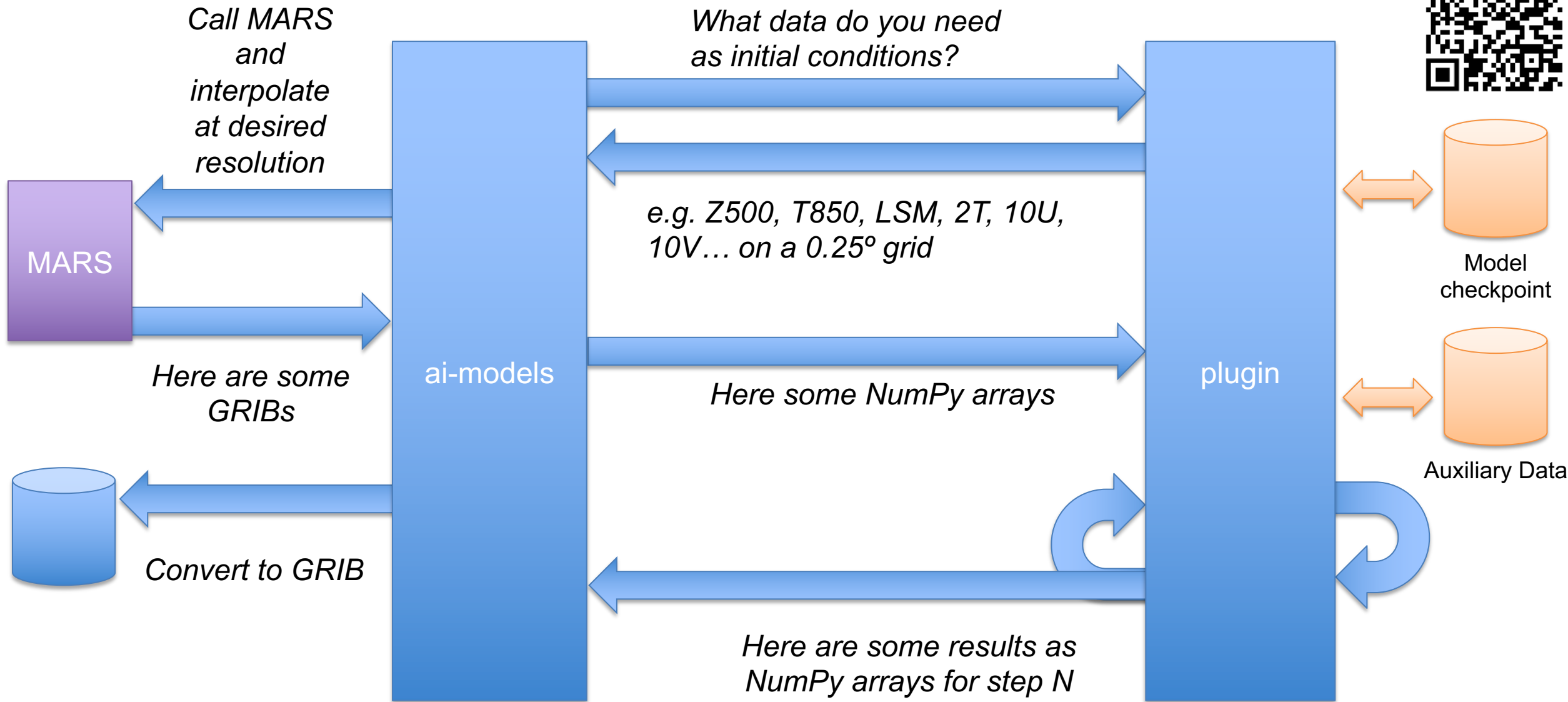


# Running 10-day forecasts in Minutes on ATOS

```
2023-09-03 13:25:00,810 INFO Writing results to panguweather.grib.
2023-09-03 13:25:00,810 INFO Loading pressure fields from MARS
2023-09-03 13:25:02,350 INFO Loading surface fields from MARS
2023-09-03 13:25:02,476 INFO ONNXRuntime providers: ['CUDAExecutionProvider', 'CPUExecutionProvider']
2023-09-03 13:25:02,476 INFO Using device 'GPU'. The speed of inference depends greatly on the device.
2023-09-03 13:25:20,438 INFO Loading /usr/local/apps/ai-models/0.24/assets/panguweather/pangu_weather_24.onnx: 18 seconds.
2023-09-03 13:25:37,420 INFO Loading /usr/local/apps/ai-models/0.24/assets/panguweather/pangu_weather_6.onnx: 16 seconds.
2023-09-03 13:25:37,420 INFO Model initialisation: 36 seconds
2023-09-03 13:25:37,420 INFO Starting inference for 40 steps (240h).
2023-09-03 13:25:40,575 INFO Done 1 out of 40 in 3 seconds (6h), ETA: 2 minutes 6 seconds.
2023-09-03 13:25:42,718 INFO Done 2 out of 40 in 2 seconds (12h), ETA: 1 minute 43 seconds.
2023-09-03 13:25:44,851 INFO Done 3 out of 40 in 2 seconds (18h), ETA: 1 minute 34 seconds.
2023-09-03 13:25:47,196 INFO Done 4 out of 40 in 2 seconds (24h), ETA: 1 minute 30 seconds.
```

[...]

```
2023-09-03 13:27:05,223 INFO Done 38 out of 40 in 2 seconds (228h), ETA: 6 seconds.
2023-09-03 13:27:07,400 INFO Done 39 out of 40 in 2 seconds (234h), ETA: 4 seconds.
2023-09-03 13:27:09,587 INFO Done 40 out of 40 in 2 seconds (240h), ETA: 2 seconds.
2023-09-03 13:27:09,588 INFO Elapsed: 1 minute 32 seconds.
2023-09-03 13:27:09,588 INFO Average: 2 seconds per step.
```



# Different Frameworks and Checkpoint-files

ONNX for model weights

PyTorch for code and model weights

Jax for code and model weights

PyTorch for code and model weights

# Operational Data-driven NWP: Enabling Evaluation & Verification



# Operational Public Models

The screenshot displays the ECMWF Charts website interface. The browser address bar shows the URL: [https://charts.ecmwf.int/?facets={"Product type%3A\["Experimental%3AMachine learning models"\]%2](https://charts.ecmwf.int/?facets={). The page header includes the ECMWF logo and navigation links for Home, Charts, Help, and Log In. The main content area is titled "Charts catalogue" and features a search bar and several filter categories: Range (Medium (15 days), Extended (42 days), Long (Months)), Type (Forecasts, Verification), Component (Surface, Atmosphere), and Product type (High resolution forecast (HRES), Ensemble forecast (ENS), Combined (ENS + HRES), Extreme forecast index). The main display area shows three columns of experimental machine learning models, each with a weather map thumbnail, a title, and a brief description. The models are: 1) FourCastNet machine learning model: Experimental: 500 hPa geopotential height and 850 hPa temperature; 2) GraphCast machine learning model: Experimental: 500 hPa geopotential height and 850 hPa temperature; 3) Pangu-Weather machine learning model: Experimental: 500 hPa geopotential height and 850 hPa temperature. Below these, a second row of three smaller map thumbnails is visible.

Search products...

Range

- Medium (15 days)
- Extended (42 days)
- Long (Months)

Type

- Forecasts
- Verification

Component

- Surface
- Atmosphere

Product type

- High resolution forecast (HRES)
- Ensemble forecast (ENS)
- Combined (ENS + HRES)
- Extreme forecast index

Latest forecast

**(FourCastNet machine learning model: Experimental): 500 hPa geopotential height and 850 hPa temperature**

FourCastNet v2-small: a deep learning-based system developed by NVIDIA in collaboration with researchers at several US universities. It is initialised with ECMWF HRES analysis. FourCastNet operates at 0.25° resolution.

Latest forecast

**(GraphCast machine learning model: Experimental): 500 hPa geopotential height and 850 hPa temperature**

GraphCast (Google DeepMind): a deep learning-based system developed by Google DeepMind. It is initialised with ECMWF HRES analysis. GraphCast operates at 0.25° resolution.

Latest forecast

**(Pangu-Weather machine learning model: Experimental): 500 hPa geopotential height and 850 hPa temperature**

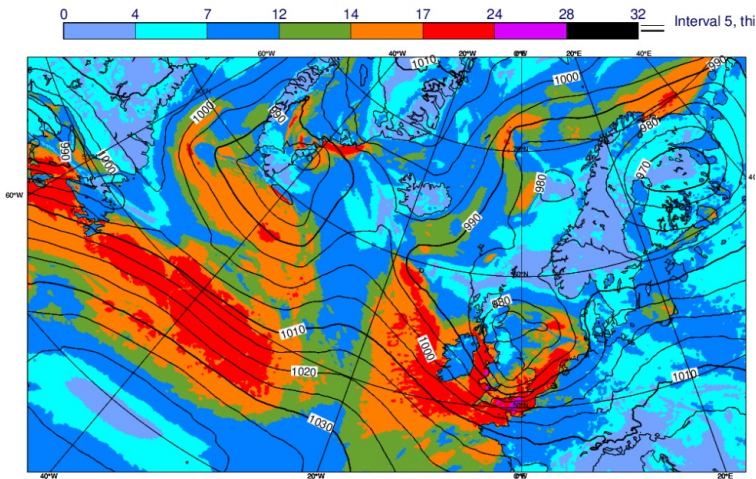
Pangu-Weather: a deep learning-based system developed by Huawei. It is initialised with ECMWF HRES analysis. Pangu-Weather operates at 0.25° resolution.

# Evaluation Storm Eunice over UK 2022-02-16 00z + 60h

See ECMWF Newsletter 176

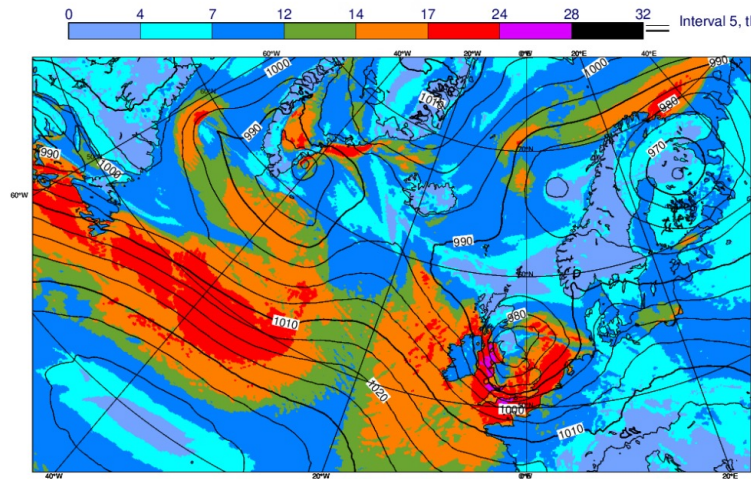
## Analysis

MSLP+WS 2022021612 Step: 0  
AN



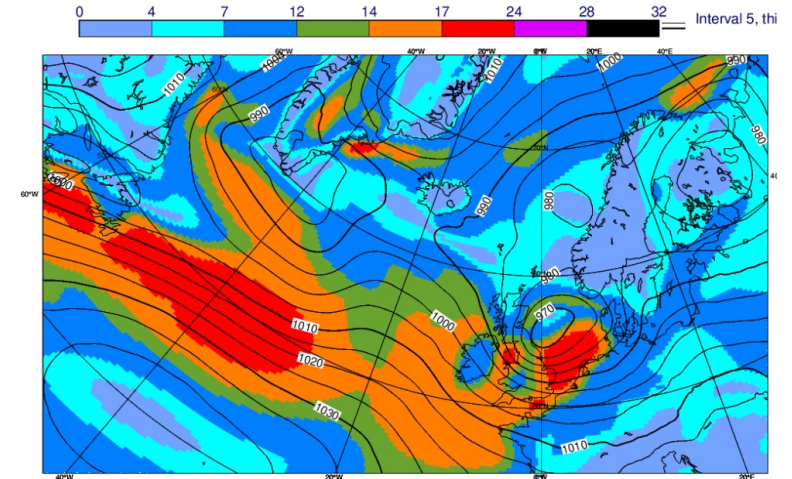
## HRES

MSLP+WS 2022021600 Step: 60  
HRES



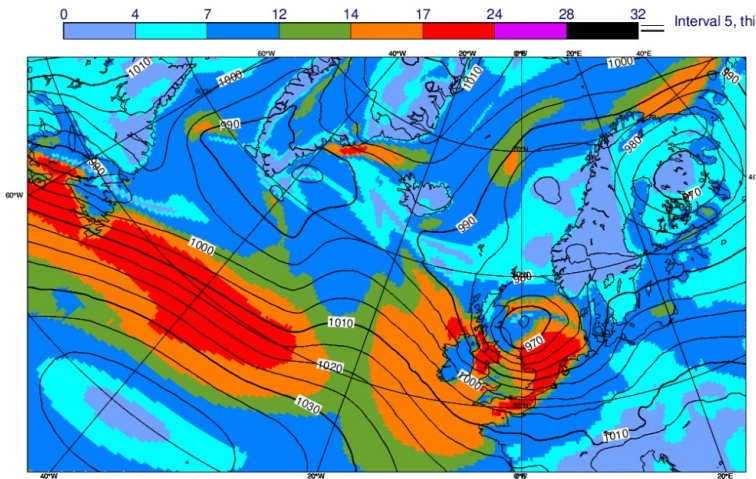
## Fourcastnet

MSLP+WS 2022021600 Step: 60  
i51c



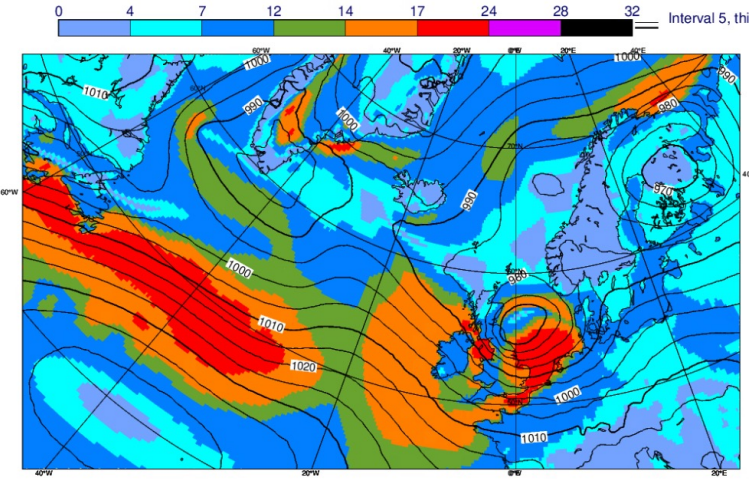
## PanguWeather

MSLP+WS 2022021600 Step: 60  
PanguWeather



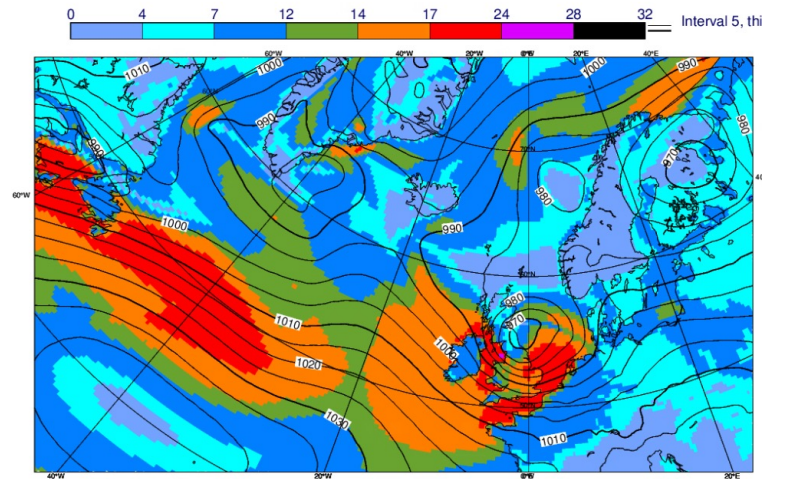
## Graphcast

MSLP+WS 2022021600 Step: 60  
i51d



## AIFS n320

MSLP+WS 2022021600 Step: 60  
i5e6



# Enables Research like this:

The screenshot shows a web browser displaying the arXiv page for the paper "The rise of data-driven weather forecasting" (arXiv:2307.10128). The browser's address bar shows the URL <https://arxiv.org/abs/2307.10128>. The page header includes the Cornell University logo and a message of gratitude from the Simons Foundation. The arXiv navigation bar shows the paper is in the "physics" category. The main content area features the title, authors (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), and a detailed abstract. The abstract discusses the potential of machine learning for weather forecasting, comparing ML-generated forecasts with standard NWP-based forecasts. On the right side, there are sections for "Access Paper" (with links for PDF, PostScript, and other formats), "References & Citations" (with links to NASA ADS, Google Scholar, and Semantic Scholar), and a "Bookmark" section. At the bottom, there is a "Get citation" button and a row of navigation tabs: "Bibliographic Tools", "Code, Data, Media", "Demos", "Related Papers", and "About arXivLabs".

Cornell University

We gratefully acknowledge support from the Simons Foundation, member institutions, and all contributors. [Donate](#)

arXiv > physics > arXiv:2307.10128

Search... All fields Search

Help | Advanced Search

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](https://arxiv.org/abs/2307.10128) [physics.ao-ph]  
(or [arXiv:2307.10128v1](https://arxiv.org/abs/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)

[Get citation](#)

Bibliographic Tools Code, Data, Media Demos Related Papers About arXivLabs

### Access Paper:

- Download PDF
- PostScript
- Other Formats

Current browse context: **physics.ao-ph**  
< prev | next >  
new | recent | 2307  
Change to browse by: [physics](#)

### References & Citations

- NASA ADS
- Google Scholar
- Semantic Scholar

[Export BibTeX Citation](#)

### Bookmark

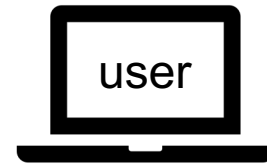
# Running ai-models anywhere



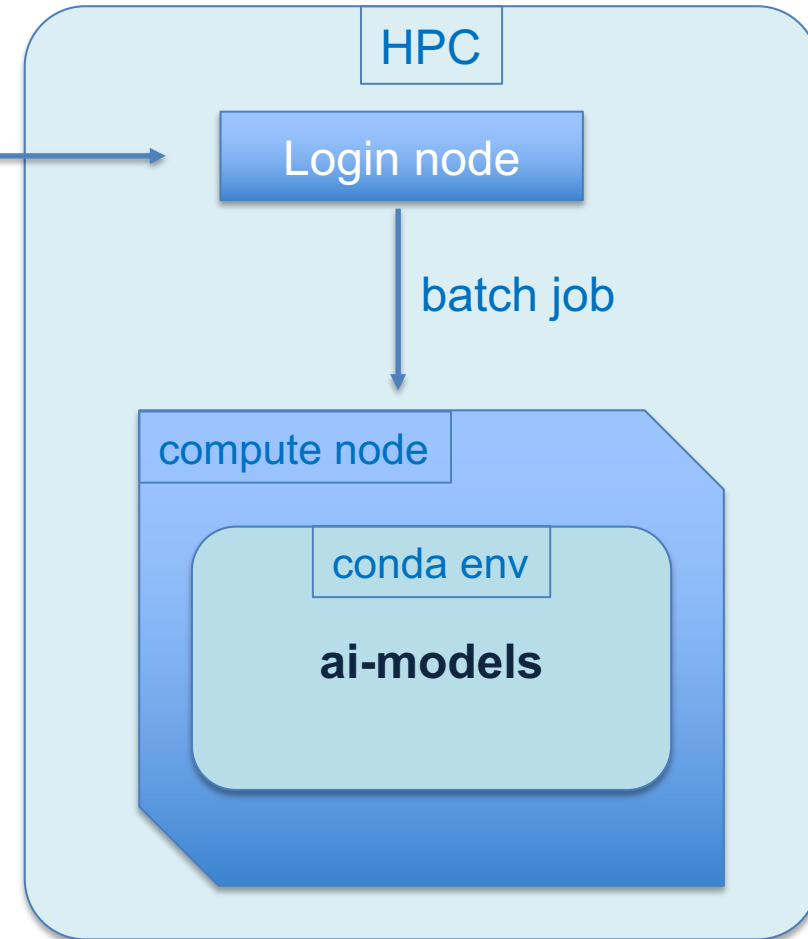


# ai-models in production

Typical workflow:



SSH

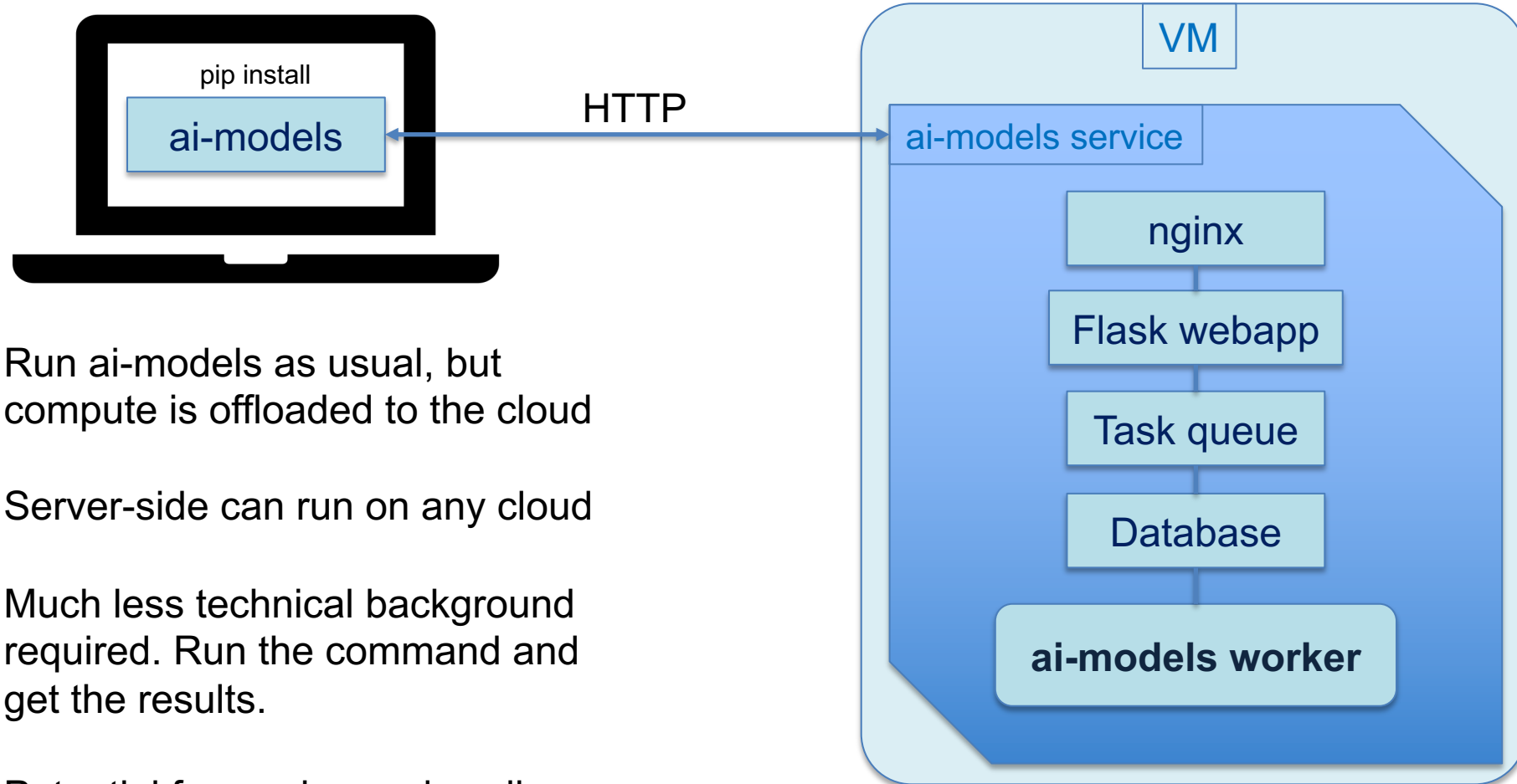


SSH to HPC -> submit job -> wait -> copy results back

- Need HPC access (MS have it, but others may not)
- Some technical background and setup required (batch scripts, SSH, Slurm, HPC architecture, conda env, ...)

# ai-models ✨in the cloud✨

Remote workflow:



- Run ai-models as usual, but compute is offloaded to the cloud
- Server-side can run on any cloud
- Much less technical background required. Run the command and get the results.
- Potential for on-demand scaling

```
(aifs) cloud-user@ml-prod ~ $
```

# ai-models on your toaster

if your toaster has a web browser

Live demo time!

## ai-models web

Model:

Date:

Time:

Lead time:

Token:

---