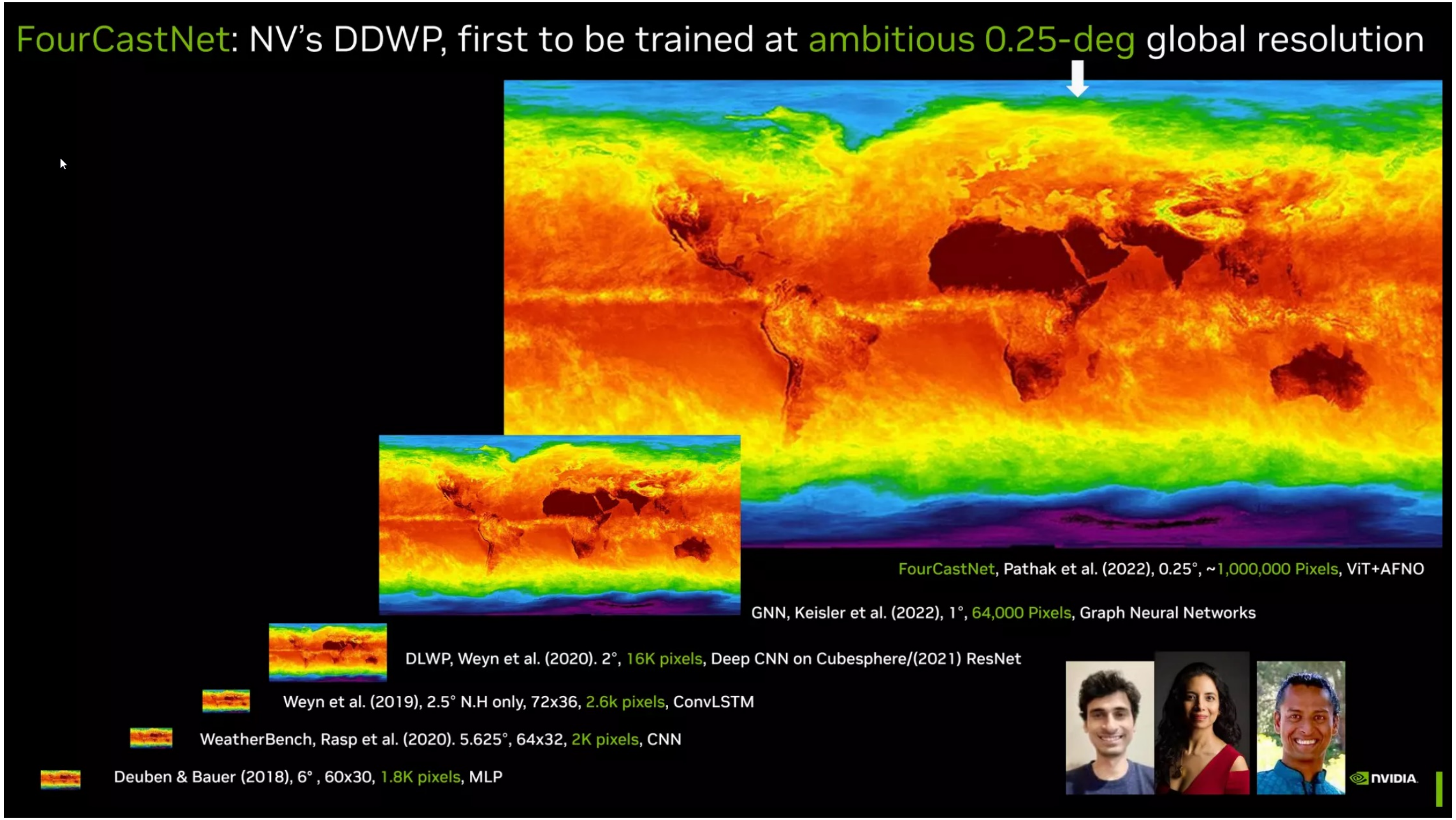
Running Data-Driven NWP

Jesper Dramsch & Gert Mertes

Scientist for Machine Learning

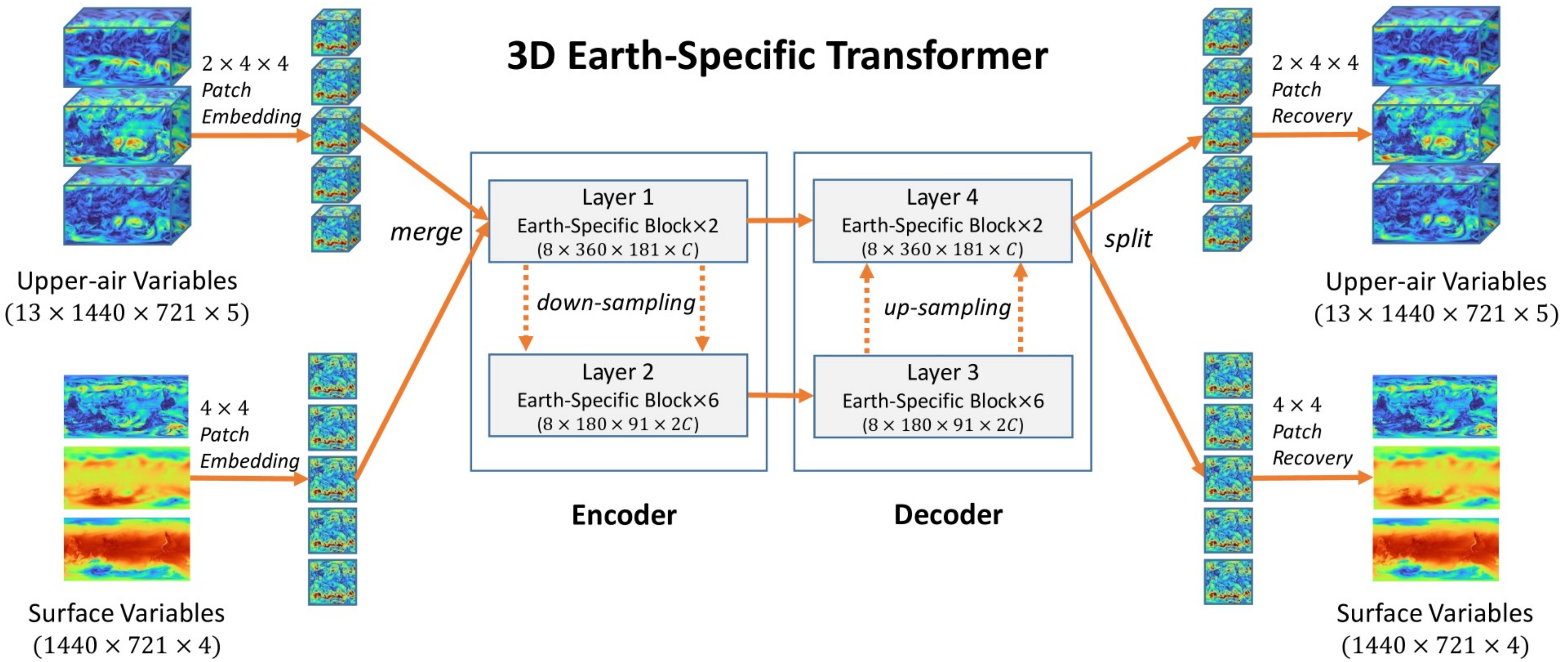
Jesper.Dramsch@ecmwf.int
Gert.Mertes@ecmwf.int

FourCastNet



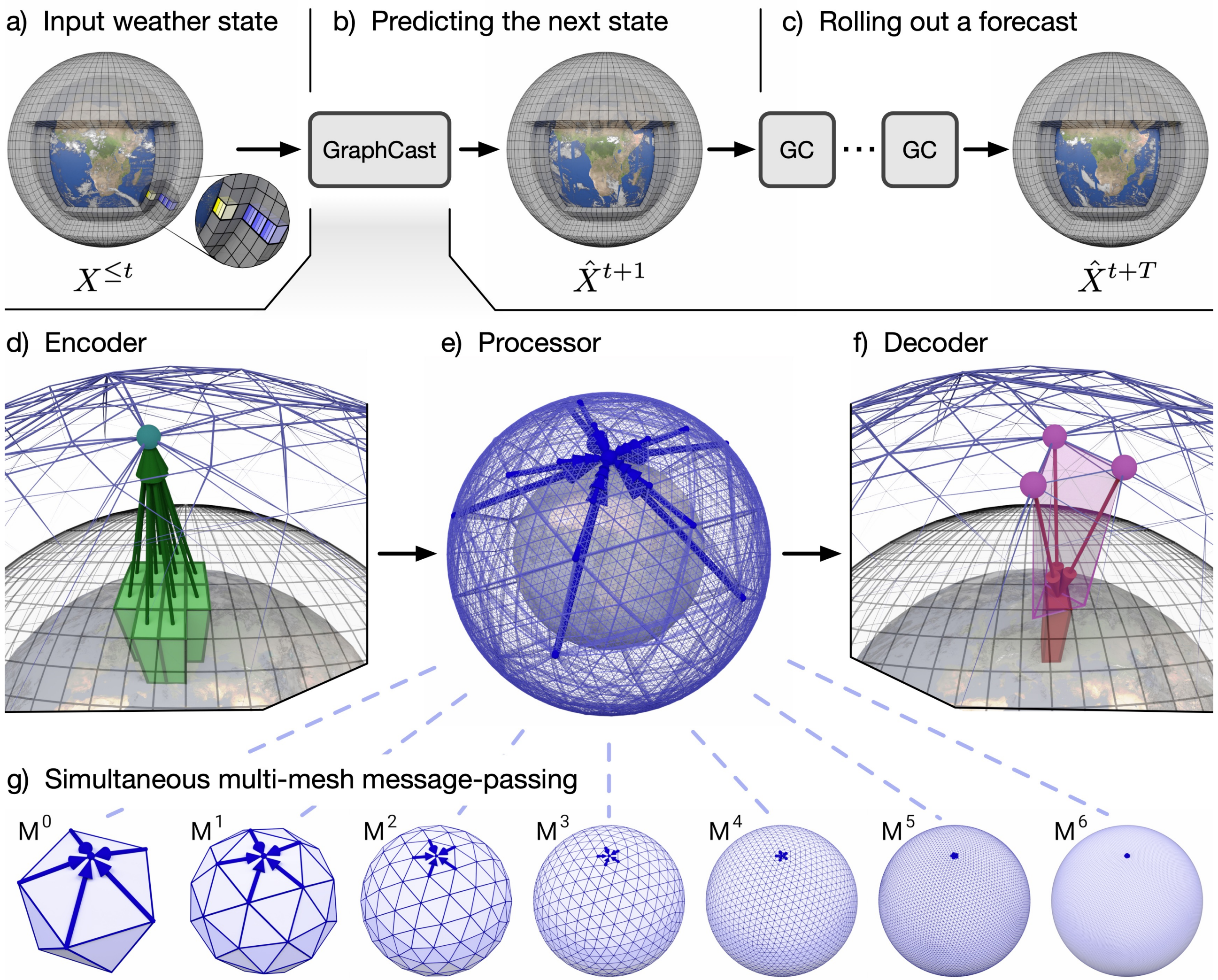
https://ar5iv.labs.arxiv.org/html/2202.11214

Pangu Weather



https://ar5iv.labs.arxiv.org/html/2211.02556

DeepMind GraphCast



https://ar5iv.labs.arxiv.org/html/2212.12794

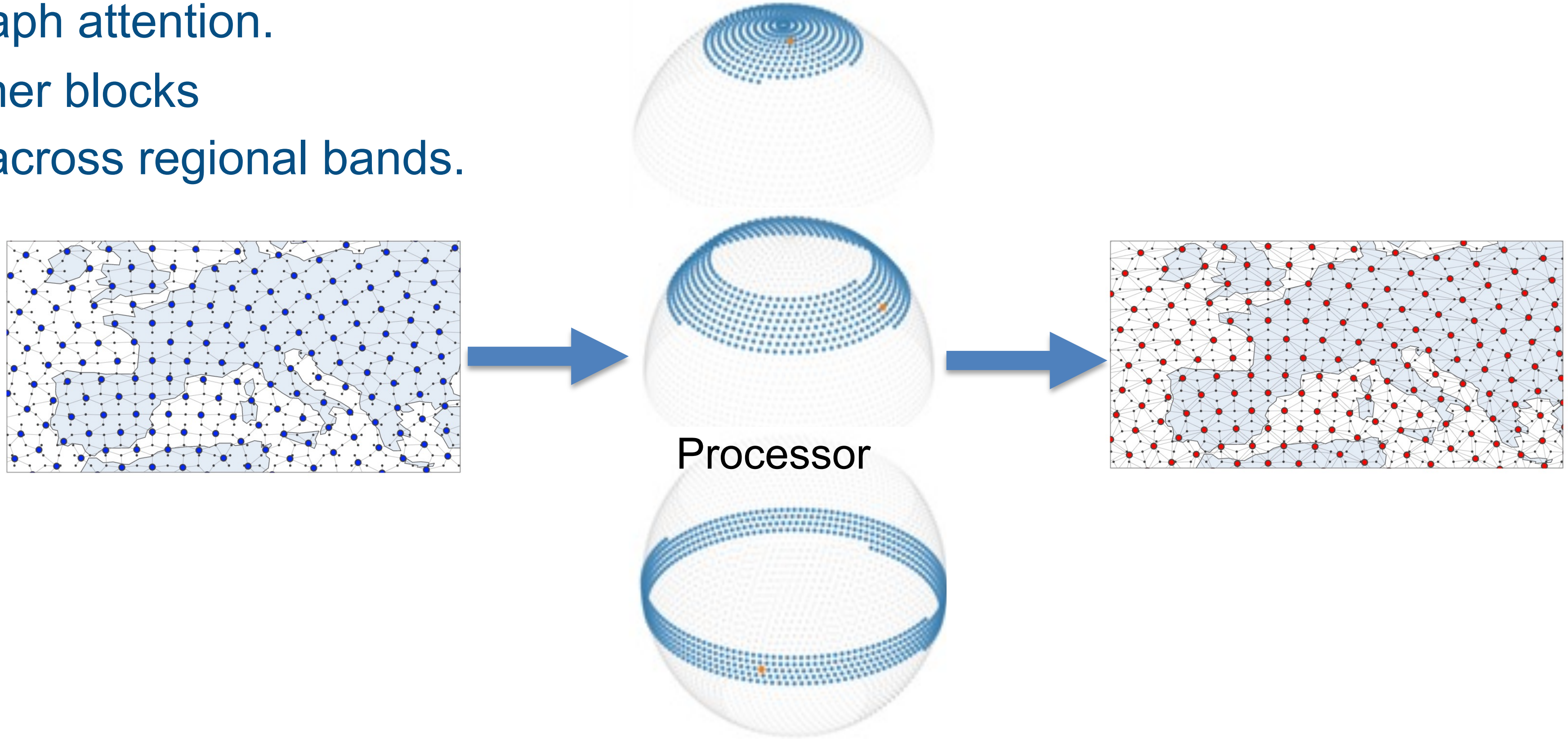
AIFS v0.21

Live from Jan 2024.

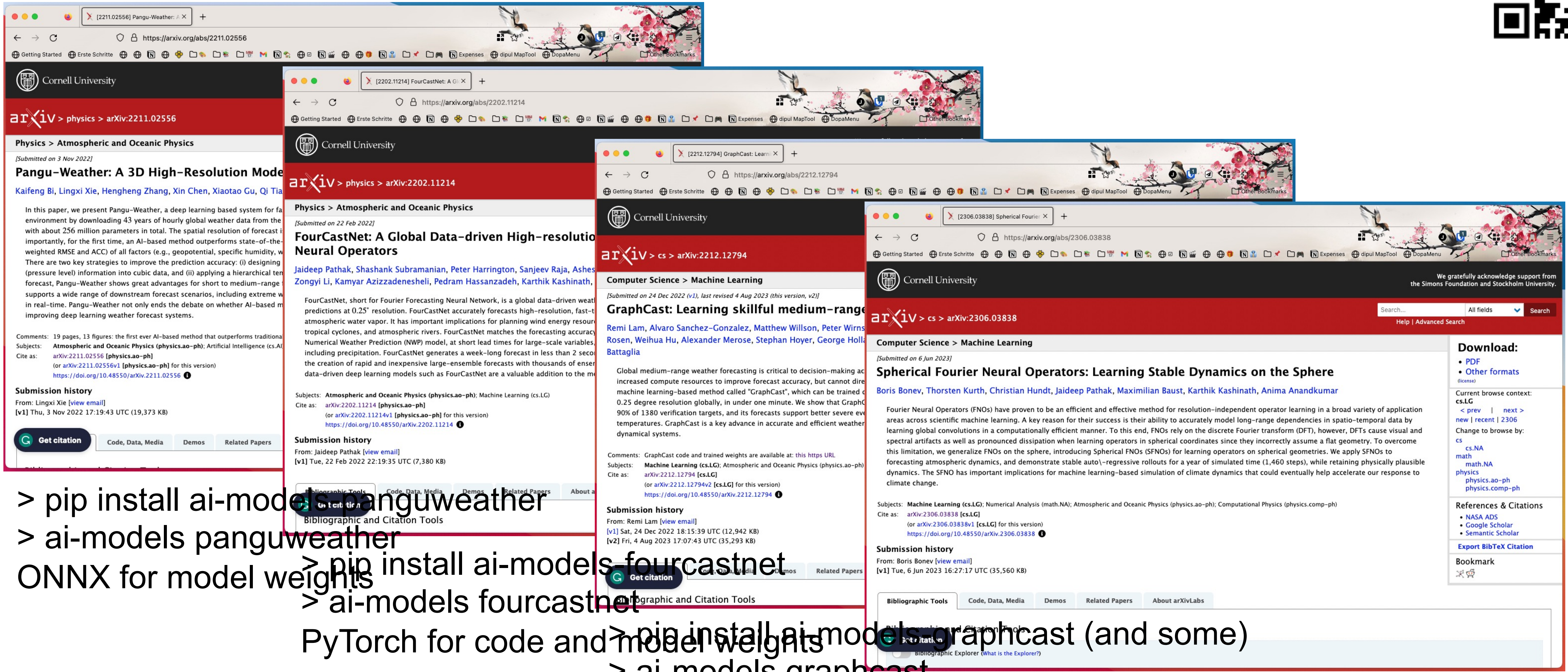
Resolution 0.25 degrees (4x finer)

New architecture.

- Encoder/decoder: graph attention.
- Processor: Transformer blocks
  attention across regional bands.



AI-Models Plugins for FOSS Data-Driven NWP



> pip install ai-models-panguweather
> ai-models panguweather
ONNX for model weights

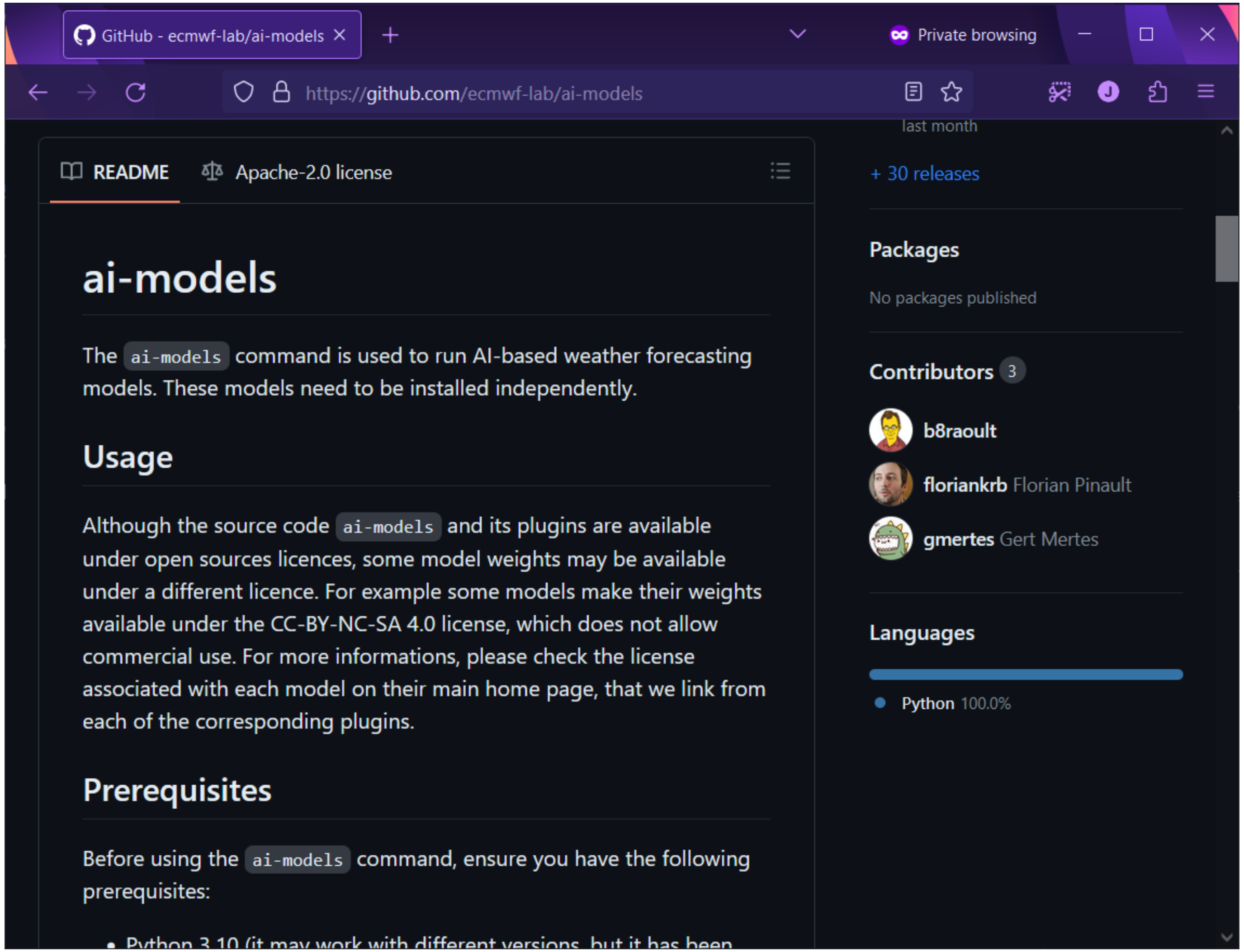
> pip install ai-models-fourcastnet
> ai-models fourcastnet
PyTorch for code and model weights

> pip install ai-models-graphcast (and some)
> 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

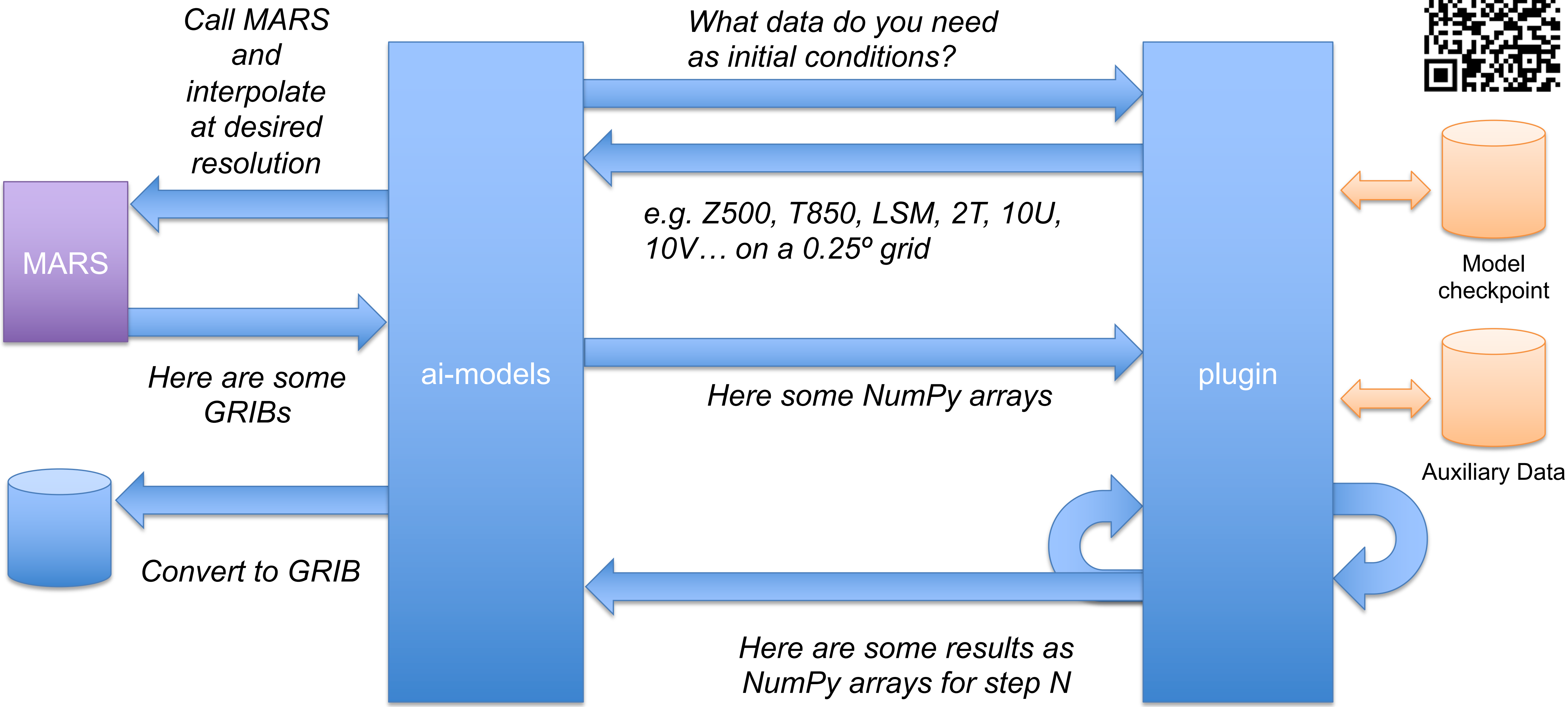


Running 10-day forecasts in Minutes on ATOS

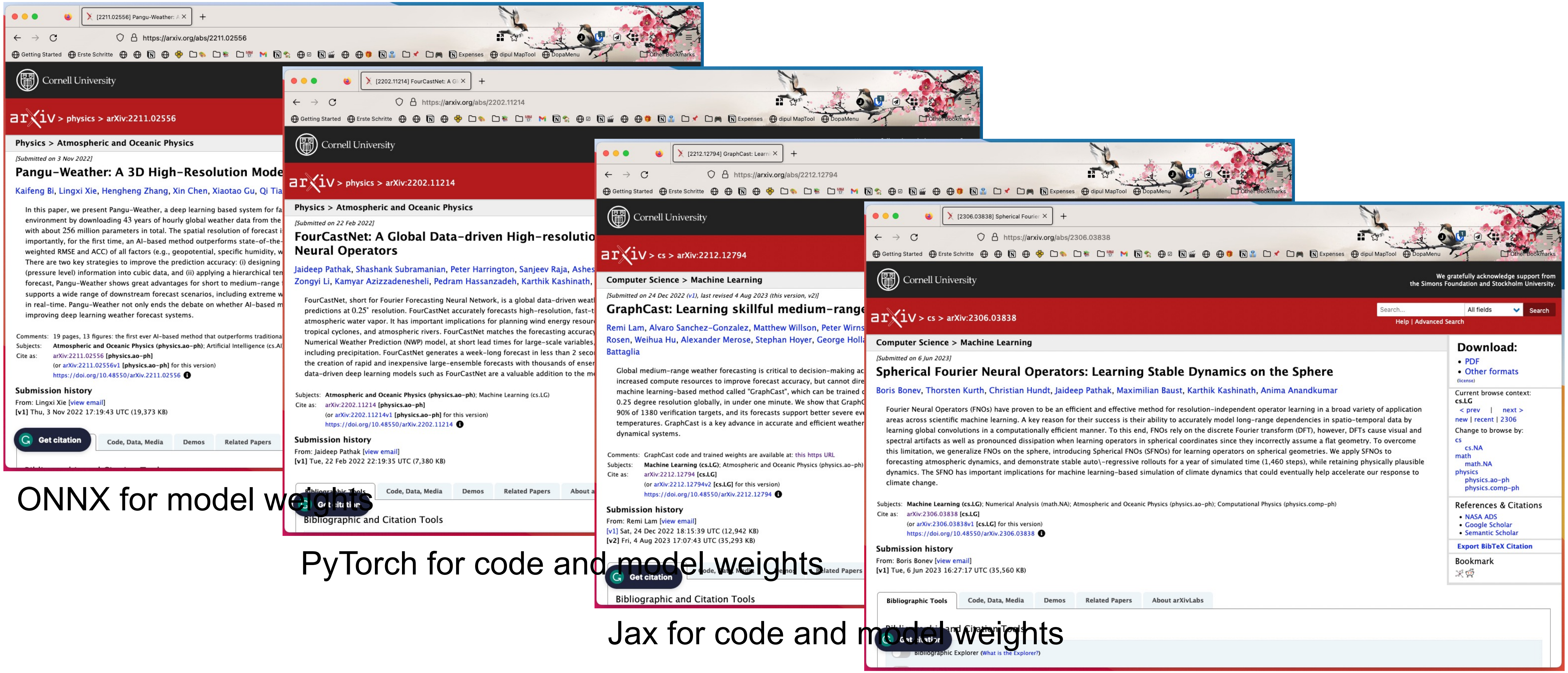
```
2023-09-03 13:25:00,810 INFO Writing results to panguweather.grib.
2023-09-03 13:25:00,810 INFO Loading pressure fields from MARS
2023-09-03 13:25:02,350 INFO Loading surface fields from MARS
2023-09-03 13:25:02,476 INFO ONNXRuntime providers: ['CUDAExecutionProvider', 'CPUExecutionProvider']
2023-09-03 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.
```


Call MARS and interpolate at desired resolution
What data do you need as initial conditions?
MARS
e.g. Z500, T850, LSM, 2T, 10U, 10V… on a 0.25º grid
Model checkpoint
ai-models
plugin
Here are some GRIBs
Here some NumPy arrays
Auxiliary Data
Convert to GRIB
Here are some results as NumPy arrays for step N

Different Frameworks and Checkpoint-files



ONNX for model weights

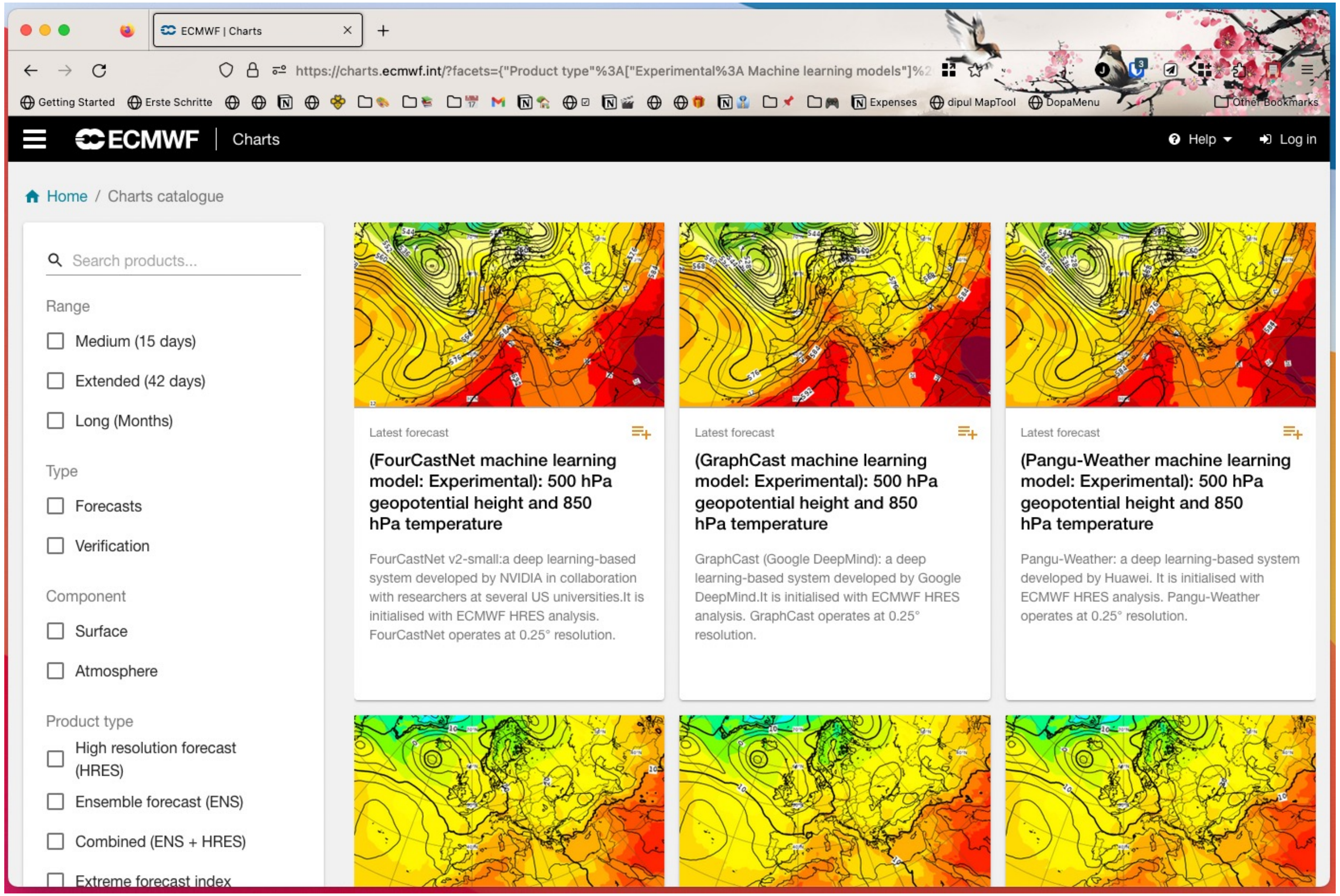
PyTorch for code and model weights

Jax for code and model weights

PyTorch for code and model weights

Operational Data-driven NWP: Enabling Evaluation & Verification

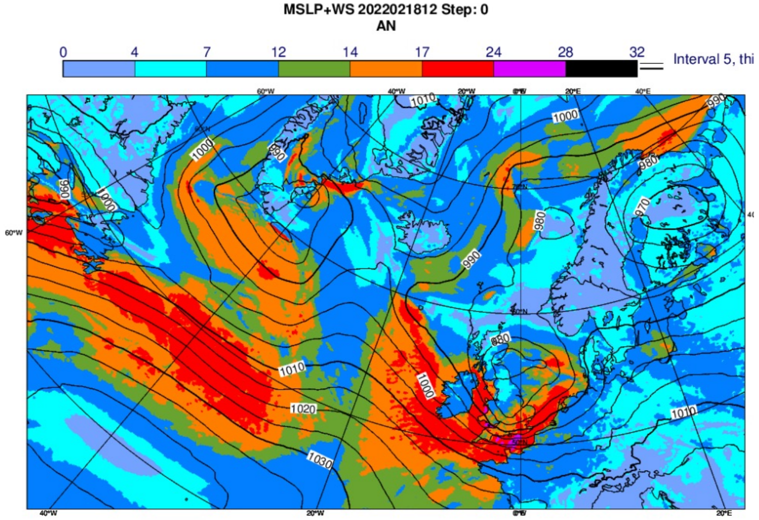
Operational Public Models



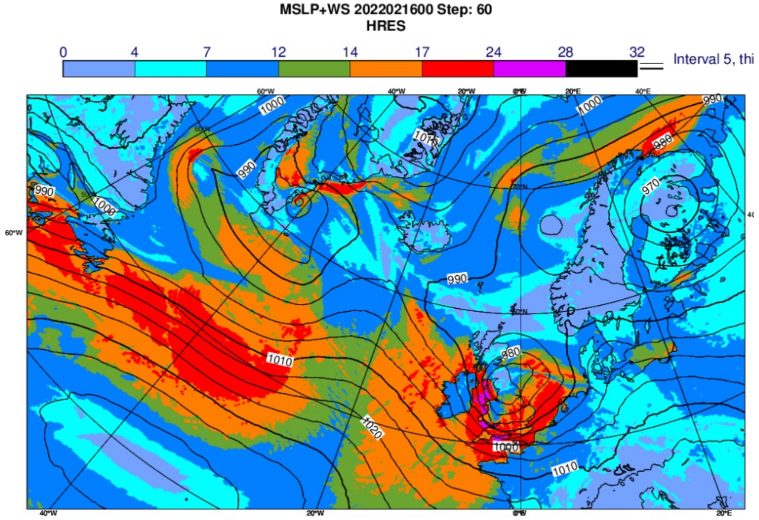
Evaluation Storm Eunice over UK 2022-02-16 00z + 60h

See ECMWF Newsletter 176

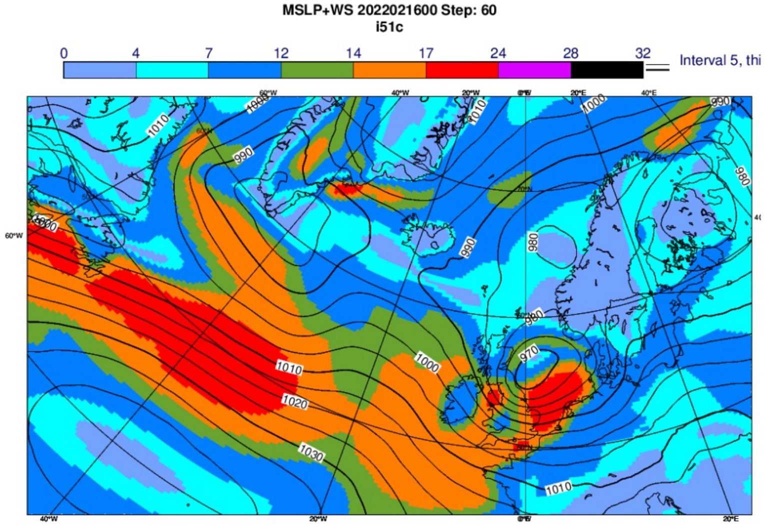
Analysis



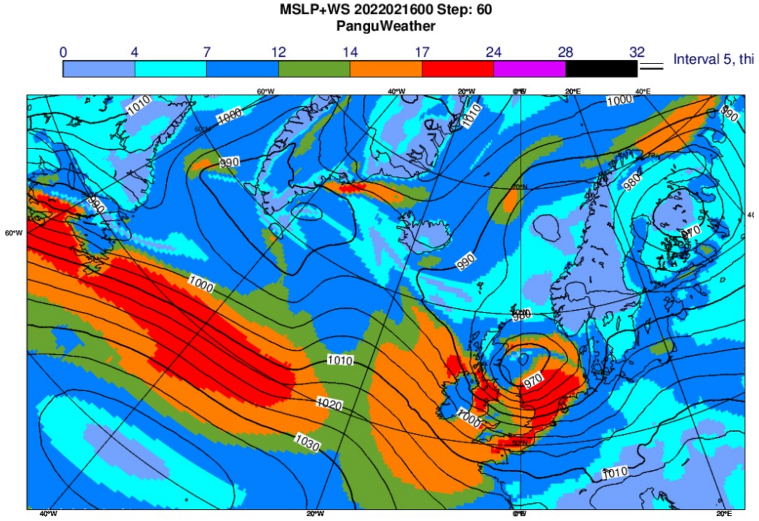
HRES



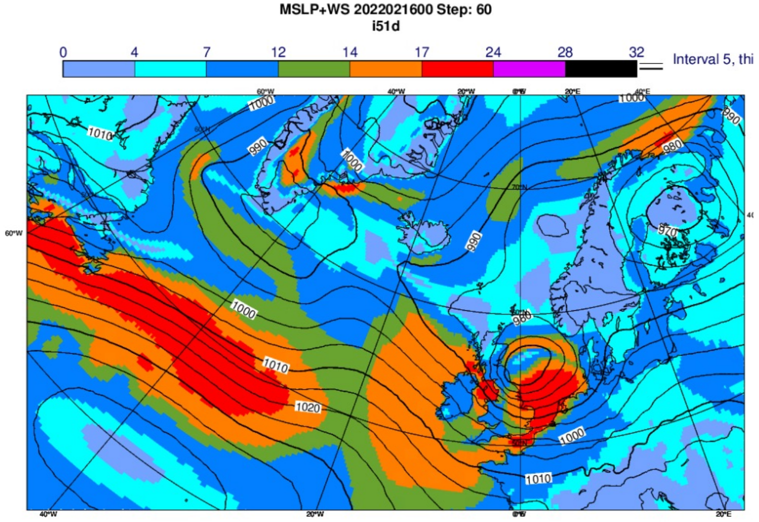
Fourcastnet



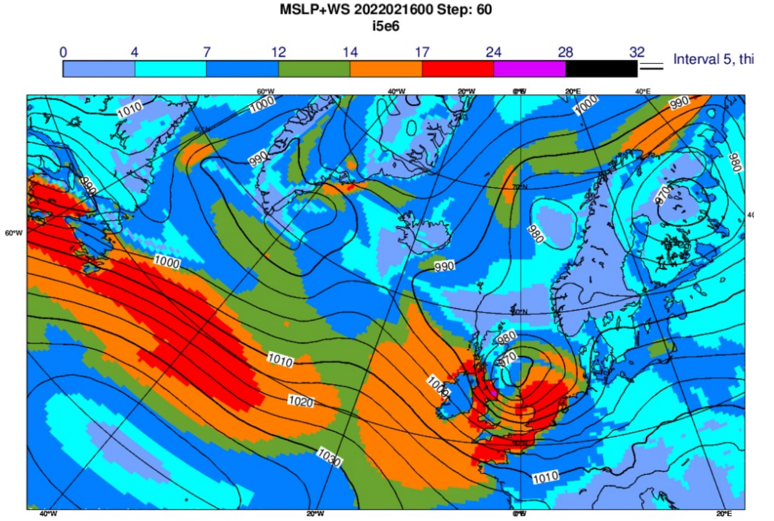
PanguWeather



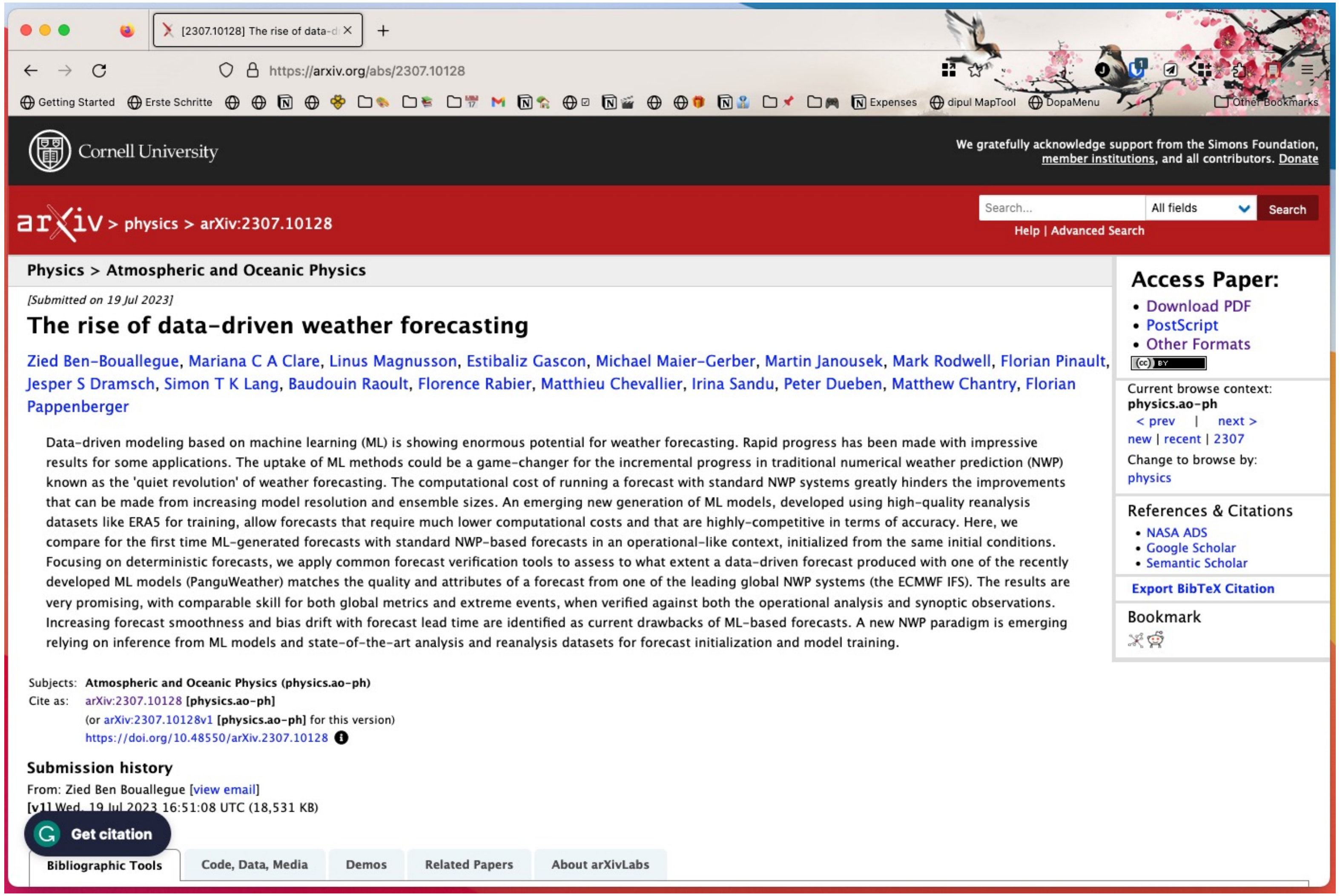
Graphcast



AIFS n320



Enables Research like this:



Cornell University

We gratefully acknowledge support from the Simons Foundation, member institutions, and all contributors. Donate

arXiv > physics > arXiv:2307.10128

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)

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

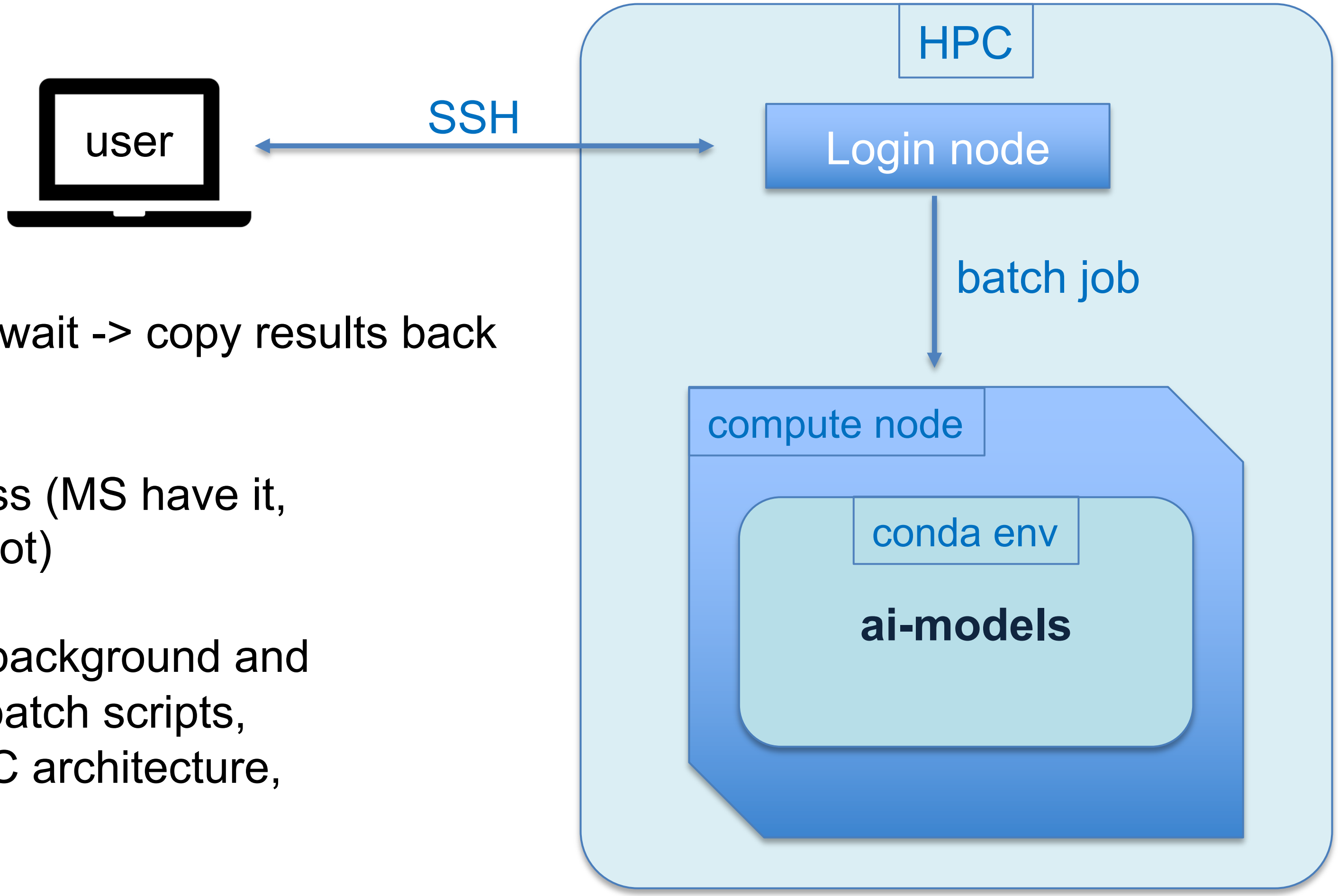
Bookmark

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

Running ai-models anywhere

ai-models in production

Typical workflow:

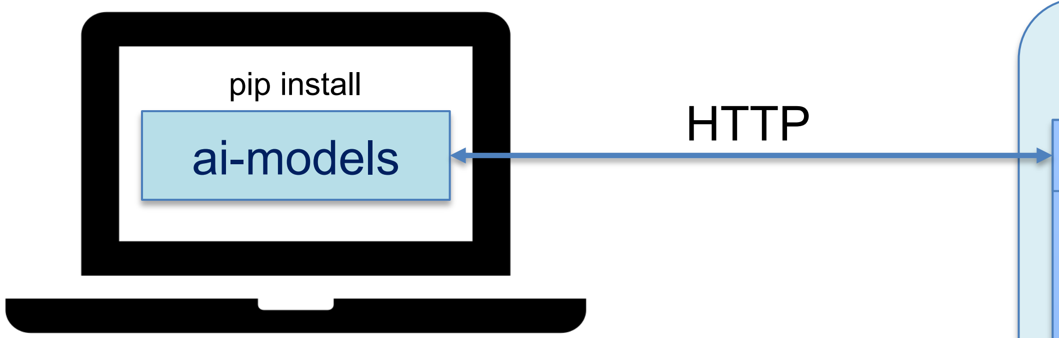


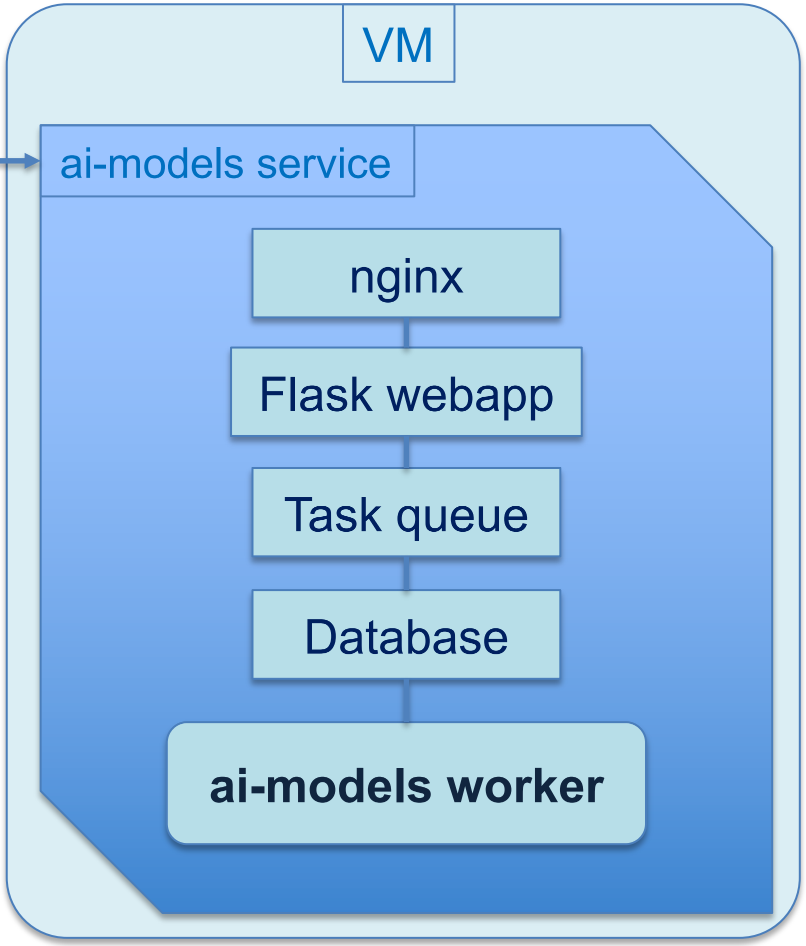
SSH to HPC -> submit job -> wait -> copy results back

- 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!