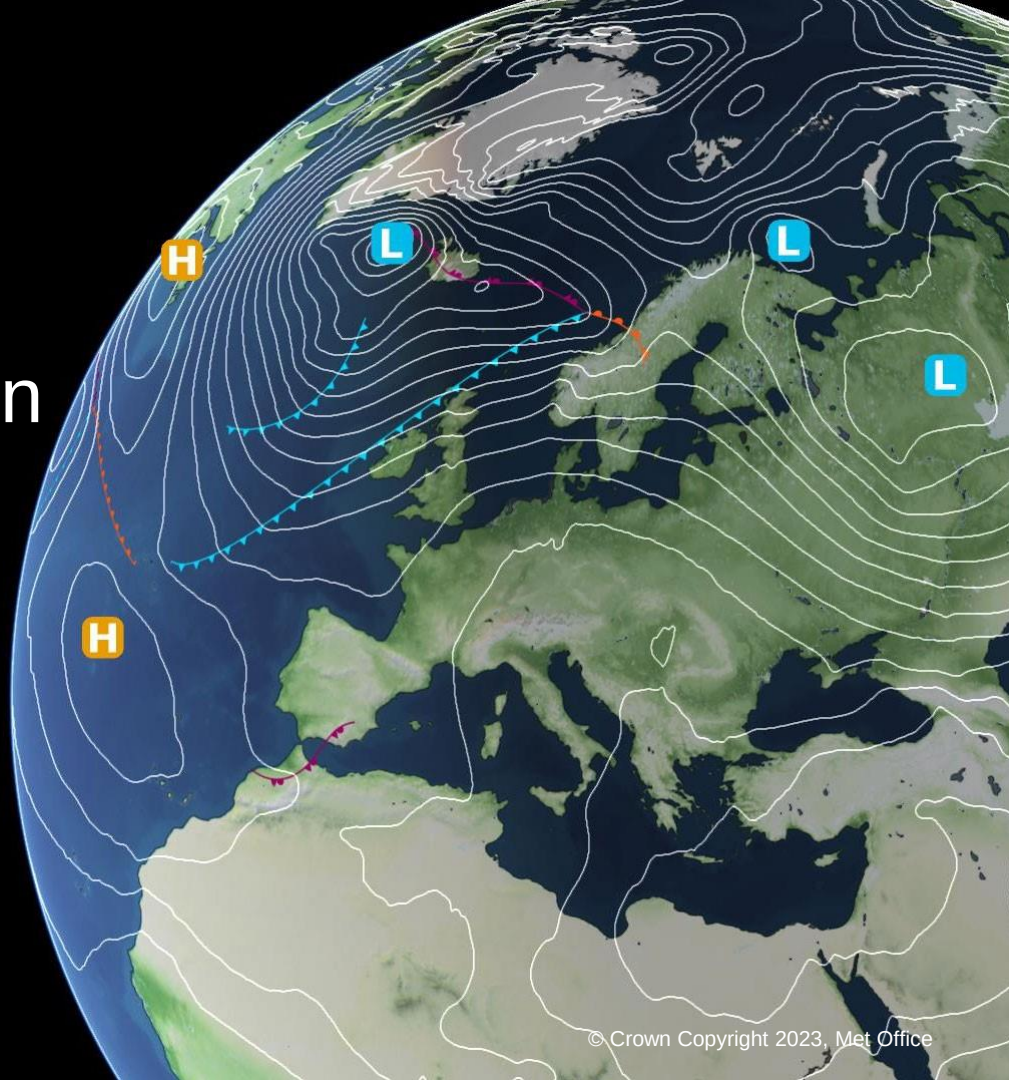


Use of MET/METplus to study and test Machine Learning developments in emulating NWP global deterministic forecasts

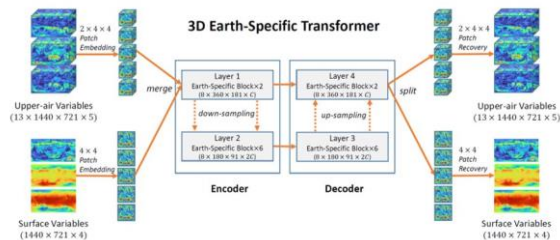
Sebastian Cole & Pangu Weather Testing Team

EWGLAM 2023

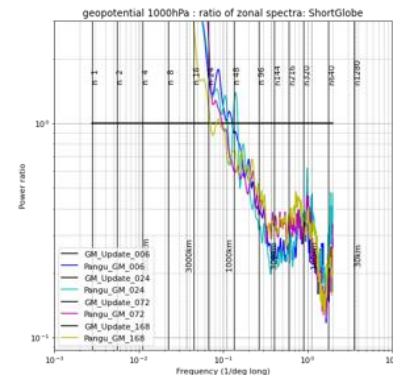


Introduction to Pangu Weather Testing

- Several ML Weather papers/models claiming similar performance to operational NWP systems
- Short project to explore publicly available Pangu ML model
- Key questions for project:
 - How does accuracy of ML models compare to current NWP systems?
 - What infrastructure is required to run ML models in production?
 - Can we run model trained with ERA5 reanalysis from Met Office analysis fields?
 - To what extent are physical processes learnt by ML models? Are fields realistic?

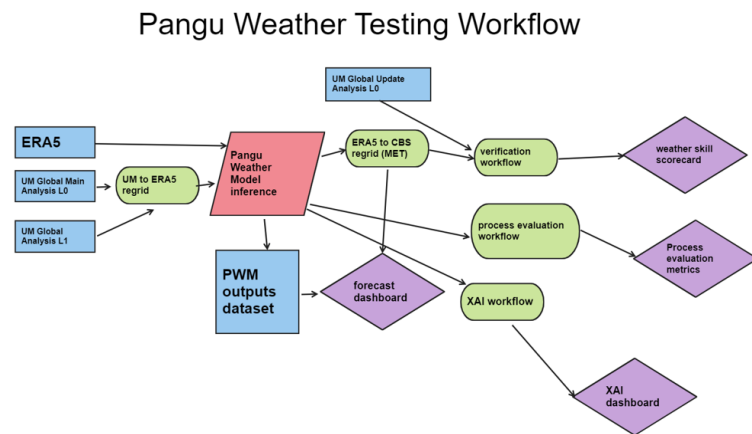


GM initialised Pangu Model 01/07/2018 - 07/07/2018 - T12



Pangu Weather testing

- Multi-disciplinary team to analyse model, resulting in knowledge transfer
- Set up pipeline to run Pangu model with Met Office data and systems
- Produced forecasts and verified for two periods
 - Dec 2017- Feb 2018 and July-Sep 2018
- Initial exploration of learning of physical processes

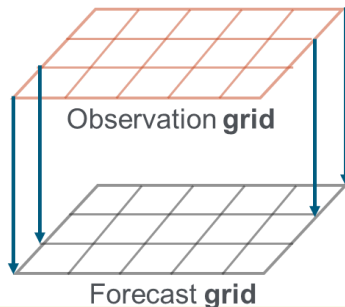


MET Tools

MET is a verification package developed by the National Center of Atmospheric Research (NCAR) Developmental Testbed Center (DTC). METplus provides a suite of python wrappers.

Two forms of gridded verification against analysis were considered in the form of:

- GridStat
- SeriesAnalysis



MET
Model Evaluation Tools

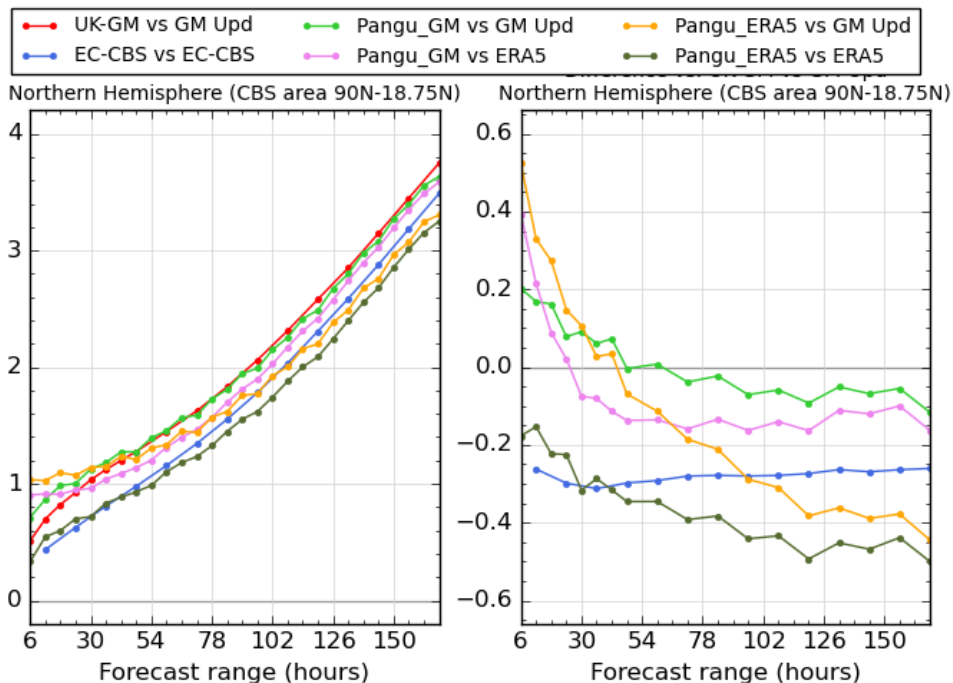
METplus

Verification Strategy

- All forecasts verified 6 hourly
- Area aggregated scores were computed in line with WMO CBS specifications and regridded to 1.5 degree grid using area weighted mean.
- Mapped verification scores were regridded to ERA5 grid using area weighted mean.
- Verification against GM update analysis refers to analysis including observations from outside the operational DA window.
- Results here will focus on 850hPa temperature over the summer period of July to September 2018

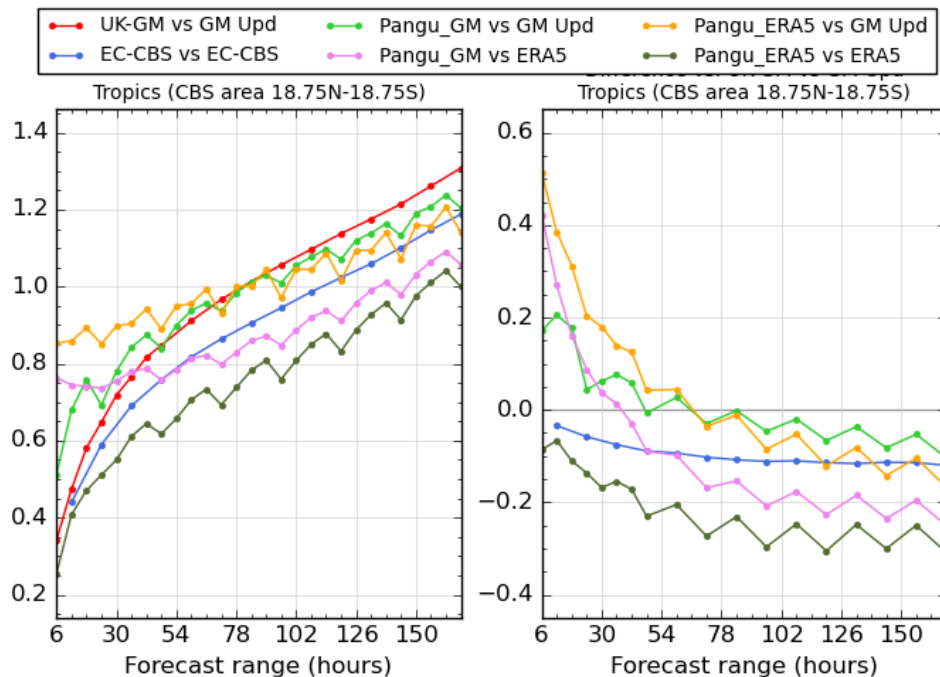
Initial Results – Northern Hemisphere

Temperature (K) @ 850hPa,
Root Mean Square Error (Forecast - Analysis), 1.5deg grid



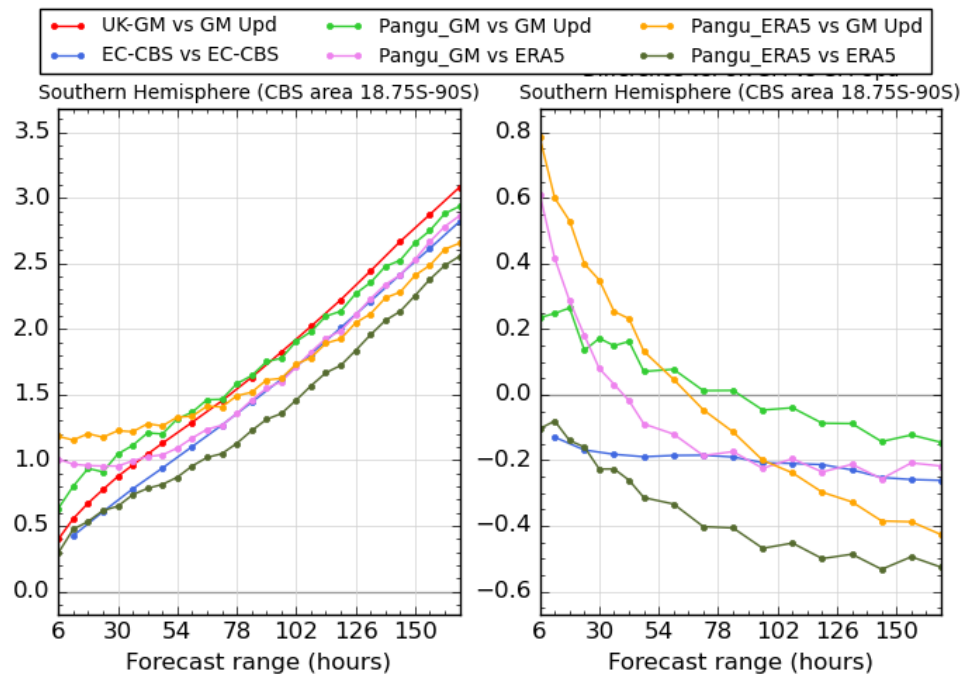
Initial Results – Tropics

Temperature (K) @ 850hPa,
Root Mean Square Error (Forecast - Analysis), 1.5deg grid



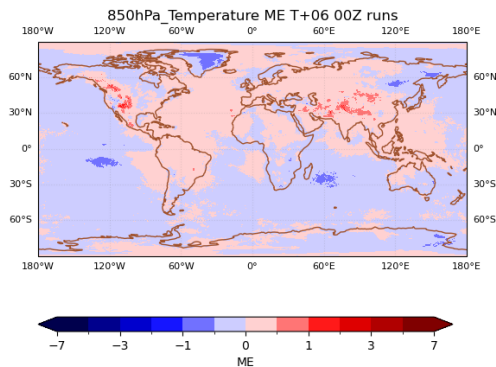
Initial Results – Southern Hemisphere

Temperature (K) @ 850hPa,
Root Mean Square Error (Forecast - Analysis), 1.5deg grid

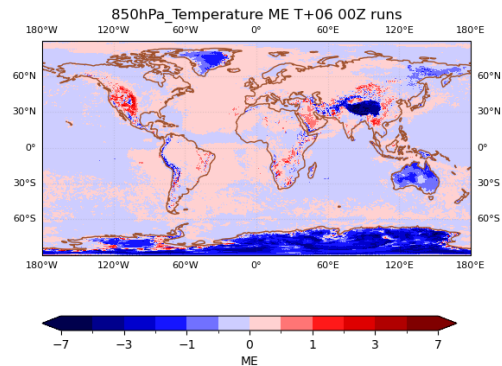


ME Comparison T+6

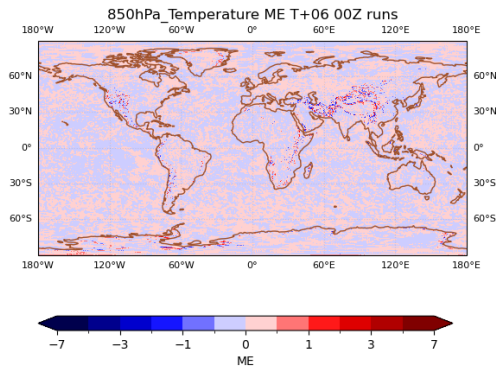
GM vs GM update



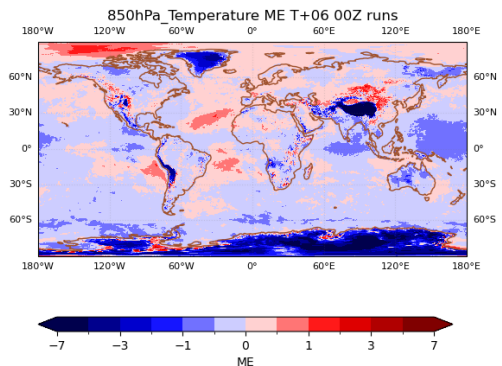
GM Pangu vs GM update



ERA5 Pangu vs ERA5

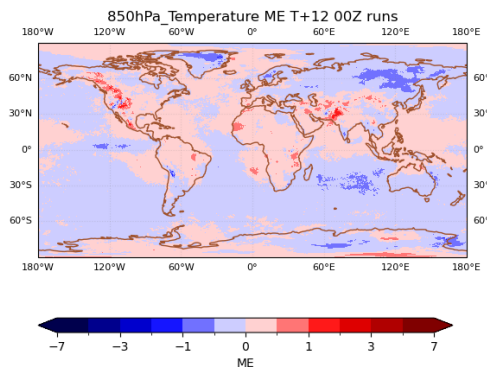


ERA5 Pangu vs GM update

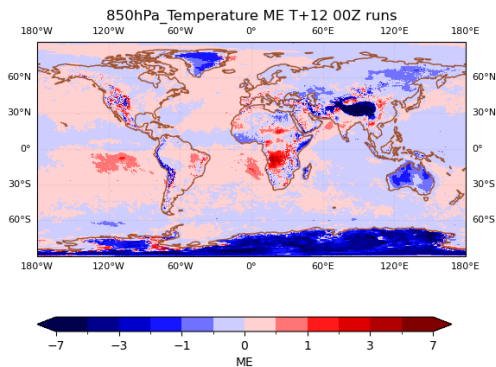


ME Comparison T+12

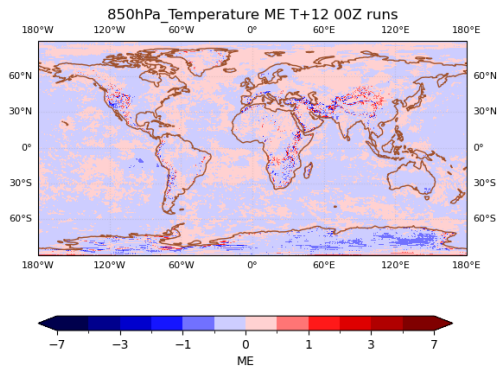
GM vs GM update



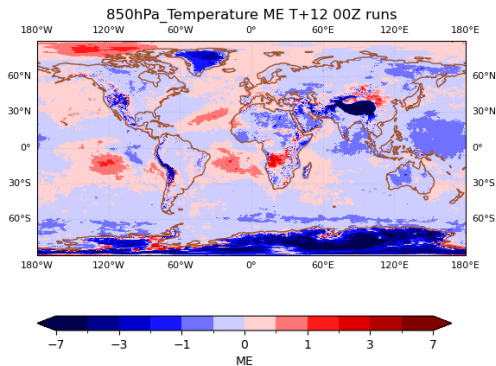
GM Pangu vs GM update



ERA5 Pangu vs ERA5

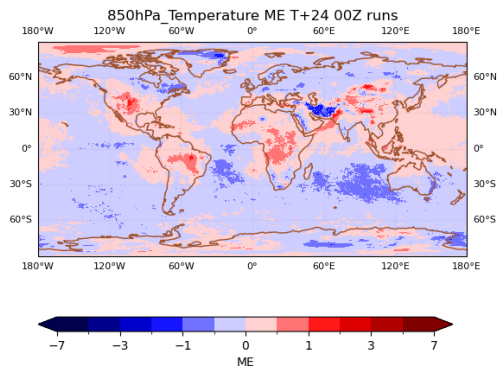


ERA5 Pangu vs GM update

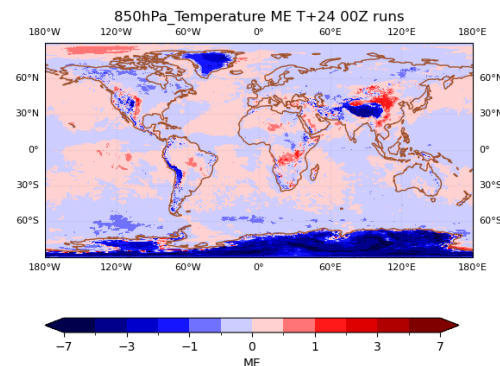


ME Comparison T+24

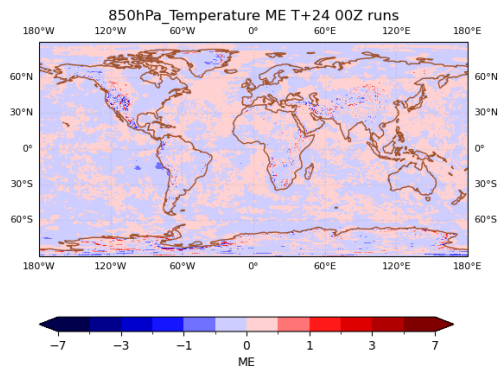
GM vs GM update



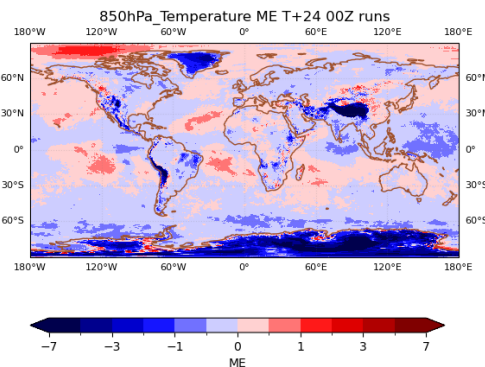
GM Pangu vs GM update



ERA5 Pangu vs ERA5



ERA5 Pangu vs GM update



Summary of initial results

- Reproduced the results initially claimed by Pangu developers.
- Initialising Pangu from Met Office analysis hinders model performance, particularly at short leadtimes.
- Different relationship between error and leadtime in Pangu
- A caveat being the use of CBS scores based on old versions of the GM.

Next steps – ML Model Intercomparison

- Follow-up through *AI4NWP* programme
- Run and evaluate latest ML models – Fourcastnet, Graphcast
 - Compare results to ML Weather Model being developed by Met Office
- Create leaderboards of current ML models
- Explore further verification metrics (e.g. anomaly correlation) and scenarios (extremes, tropical cyclones)
 - Inform future ML model development
- Use process evaluation and Explainable AI to understand how ML models produce forecasts

Thank you for your attention

Pangu Weather Testing Team & Contributors

Sebastian Cole, Nathan Creaser, Anna- Louise Ellis, Phil Gill, Stephen Haddad, Roger Harbord, Richard Hattersley, Rachel North, Kelly O'Meara, Aurore Porson, Matthew Shin, Chris Steele, David Walters, Martin Willet & more!

Extra slides

Inference workflows: Algorithm

- For each cycle, produce hourly forecasts from T+1 to T+168.
 - Pangu trained for 1-, 3-, 6- and 24-hour forecasts.
 - E.g., for T+10, we'd run a 6-hour forecast, followed by 3-hour and then 1-hour.
 - To obtain the forecasts for the full range of lead times, we need to run the forecasts in dependency orders – as an acyclic graph.
 - Application manages the graph with Python's [graphlib.TopologicalSorter](#) and [multiprocessing.Pool](#).

(1,)	(24, 24, 1, 1)
(1, 1)	(24, 24, 3)
(3,)	(24, 24, 3, 1)
(3, 1)	(24, 24, 3, 1, 1)
(3, 1, 1)	(24, 24, 6)
(6,)	(24, 24, 6, 1)
(6, 1)	(24, 24, 6, 1, 1)
(6, 1, 1)	(24, 24, 6, 3)
(6, 3)	(24, 24, 6, 3, 1)
(6, 3, 1)	(24, 24, 6, 3, 1, 1)
(6, 3, 1, 1)	(24, 24, 6, 6)
(6, 6)	(24, 24, 6, 6, 1)
(6, 6, 1)	(24, 24, 6, 6, 1, 1)
(6, 6, 1, 1)	(24, 24, 6, 6, 3)
(6, 6, 3)	(24, 24, 6, 6, 3, 1)
(6, 6, 3, 1)	(24, 24, 6, 6, 3, 1, 1)
(6, 6, 3, 1, 1)	(24, 24, 6, 6, 6)
(6, 6, 6)	(24, 24, 6, 6, 6, 1)
(6, 6, 6, 1)	(24, 24, 6, 6, 6, 1, 1)
(6, 6, 6, 1, 1)	(24, 24, 6, 6, 6, 3)
(6, 6, 6, 3)	(24, 24, 6, 6, 6, 3, 1)
(6, 6, 6, 3, 1)	(24, 24, 6, 6, 6, 3, 1, 1)
(6, 6, 6, 3, 1, 1)	(24, 24, 24)
(24,)	(24, 24, 24, 1)
(24, 1)	(24, 24, 24, 1, 1)
(24, 1, 1)	(24, 24, 24, 3)
(24, 3)	(24, 24, 24, 3, 1)
(24, 3, 1)	(24, 24, 24, 3, 1, 1)
(24, 3, 1, 1)	(24, 24, 24, 6)
(24, 6)	(24, 24, 24, 6, 1)
(24, 6, 1)	(24, 24, 24, 6, 1, 1)
(24, 6, 1, 1)	(24, 24, 24, 6, 3)
(24, 6, 3)	(24, 24, 24, 6, 3, 1)
(24, 6, 3, 1)	(24, 24, 24, 6, 3, 1, 1)
(24, 6, 3, 1, 1)	(24, 24, 24, 6, 6)
(24, 6, 6)	(24, 24, 24, 6, 6, 1)
(24, 6, 6, 1)	(24, 24, 24, 6, 6, 1, 1)
(24, 6, 6, 1, 1)	(24, 24, 24, 6, 6, 3)
(24, 6, 6, 3)	(24, 24, 24, 6, 6, 3, 1)
(24, 6, 6, 3, 1)	(24, 24, 24, 6, 6, 3, 1, 1)
(24, 6, 6, 3, 1, 1)	(24, 24, 24, 6, 6, 6)
(24, 6, 6, 6)	(24, 24, 24, 6, 6, 6, 1)
(24, 6, 6, 6, 1)	(24, 24, 24, 6, 6, 6, 1, 1)
(24, 6, 6, 6, 1, 1)	(24, 24, 24, 6, 6, 6, 3)
(24, 6, 6, 6, 3)	(24, 24, 24, 6, 6, 6, 3, 1)
(24, 6, 6, 6, 3, 1)	(24, 24, 24, 24)
(24, 6, 6, 6, 3, 1, 1)	(24, 24, 24, 24, 1)
(24, 24)	(24, 24, 24, 24, 1, 1)
(24, 24, 1)	

GM vs GM update on ERA 5 grid

06Z

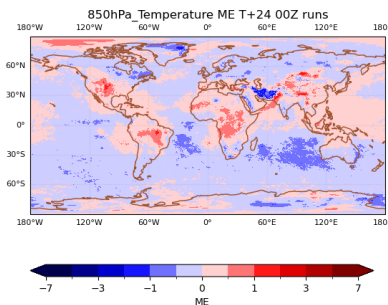
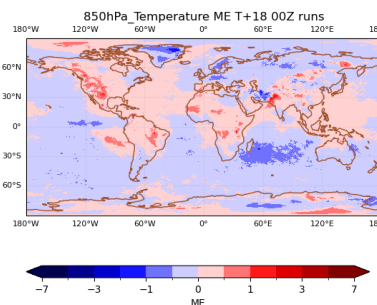
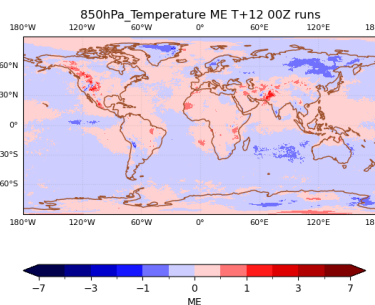
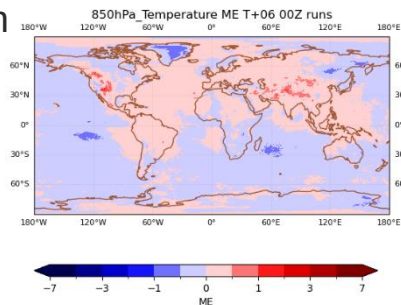
12Z

18Z

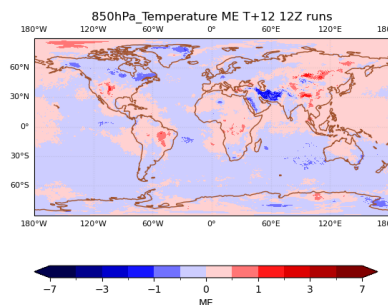
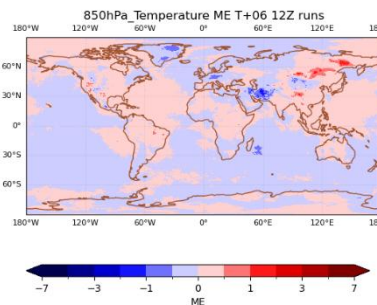
00Z

Model Run

0Z



12Z



GM vs GM update on ERA 5 grid Day 2

06Z

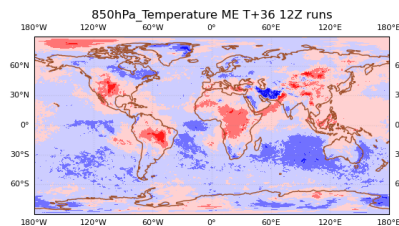
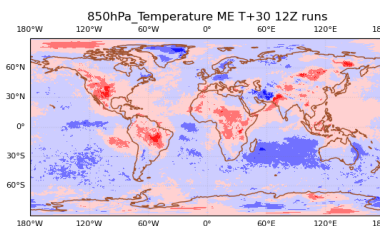
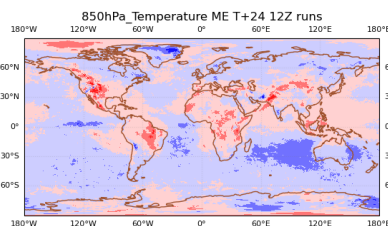
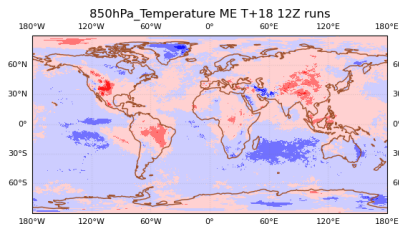
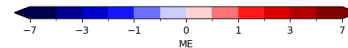
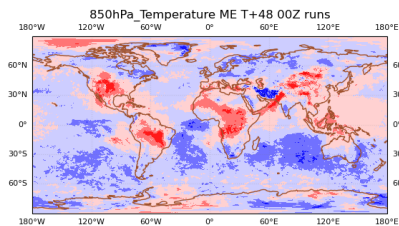
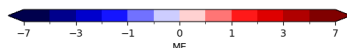
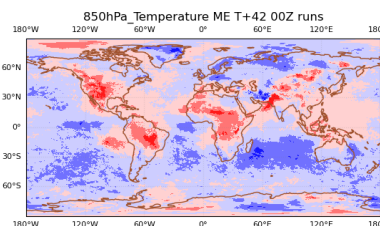
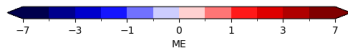
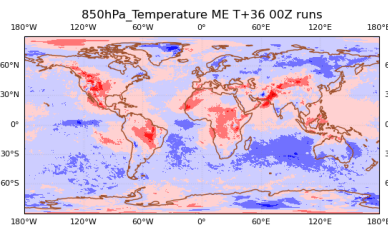
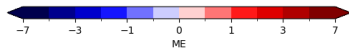
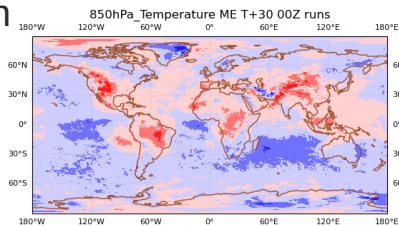
12Z

18Z

00Z

Model Run

0Z



12Z

ERA5 Pangu vs GM update

06Z

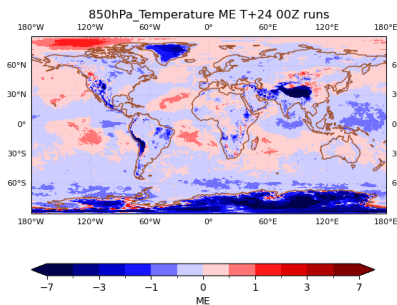
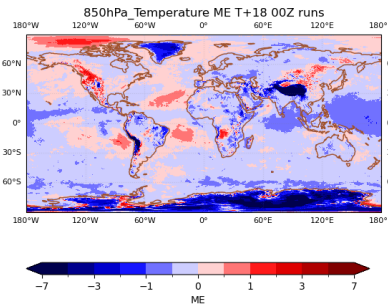
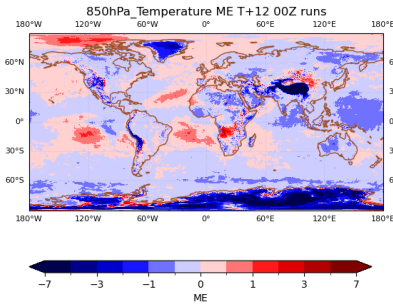
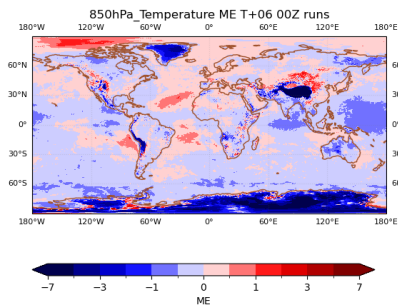
12Z

18Z

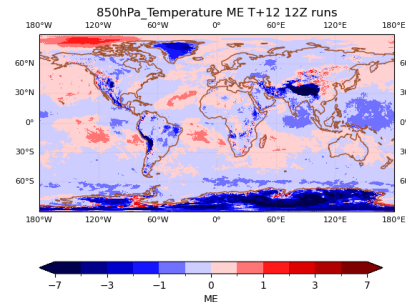
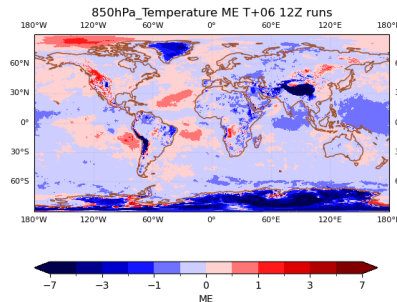
00Z

Model Run

0Z



12Z



ERA5 Pangu vs GM update D2

06Z

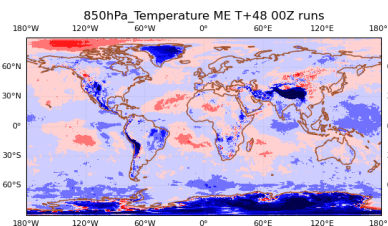
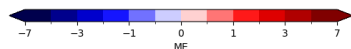
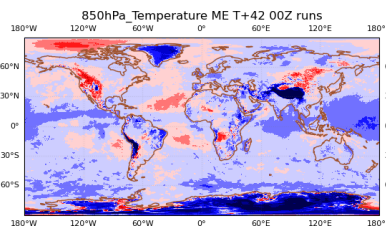
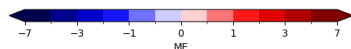
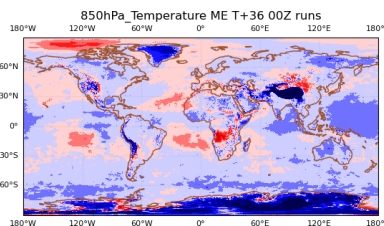
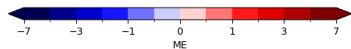
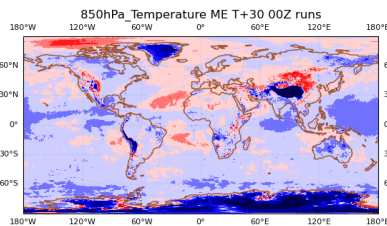
12Z

18Z

00Z

Model Run

0Z



12Z

