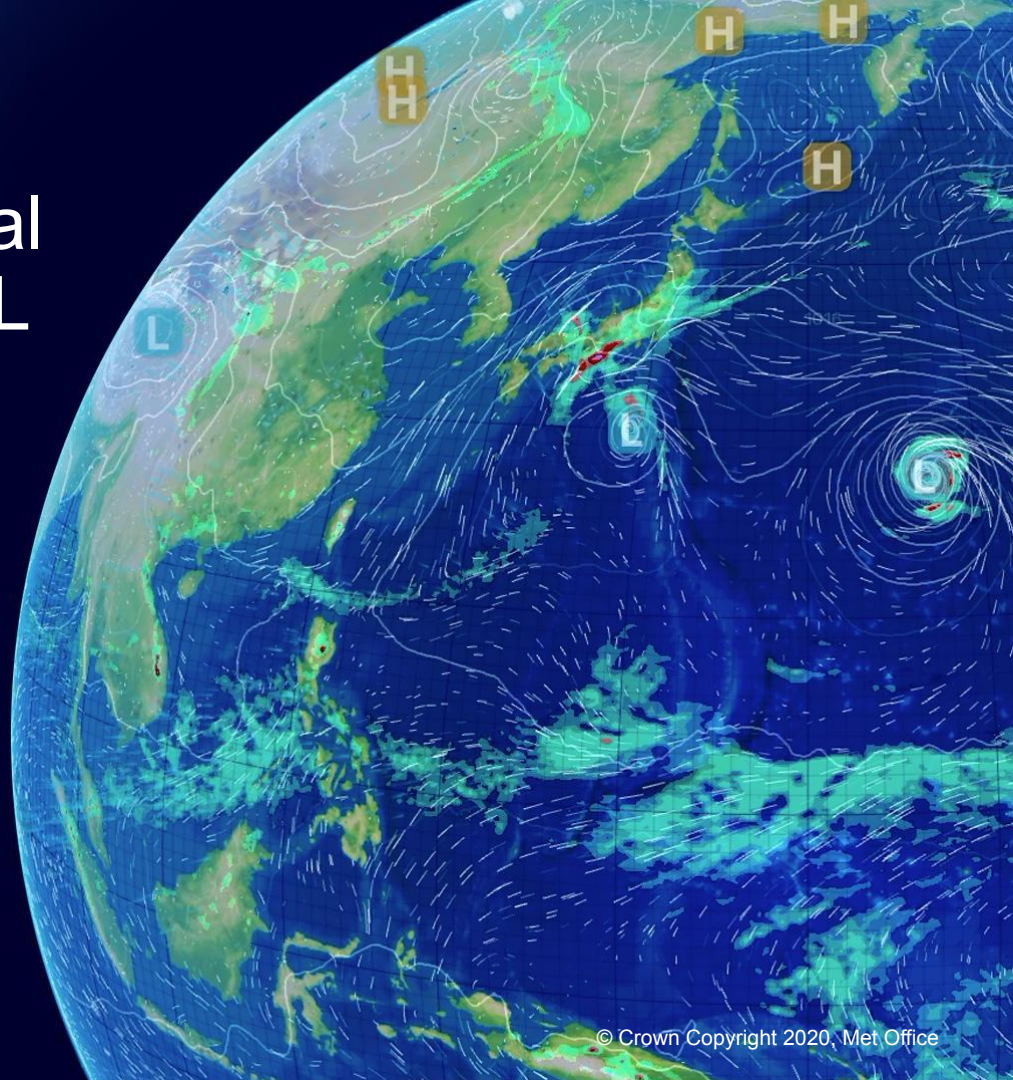


# Systematic errors in global circulation models and ML models for NWP

José M. Rodríguez,  
with thanks to Sean Milton,  
Gill Martin and Mike Cullen

Bristol, 4 June 2025



## **'Dynamical' GCMs**

Despite progress, systematic errors in key Earth system components persist through various model generations.

One of such systems are monsoons: complex patterns and variability on spatial and temporal scales are a significant challenge for modelling systems.

## **ML models for NWP**

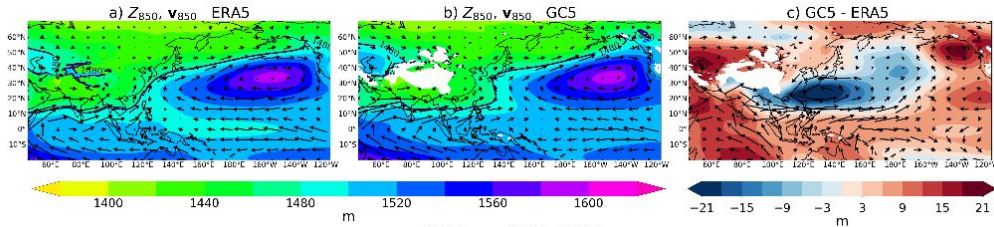
- Recent years have seen marked progress in the development of global medium range weather forecast models based on artificial intelligence techniques.
- A comprehensive assessment of such models is still lacking. Assessment is essential to understand their usability in forecasting systems.

## **Outline**

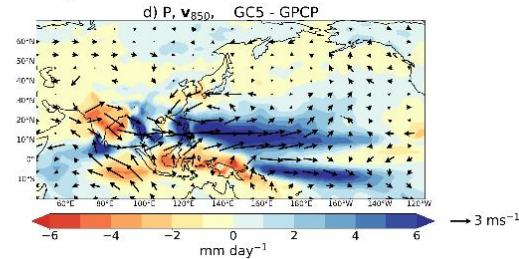
- Systematic errors in GCMs: West Pacific subtropical high.
- Emergence of errors and the seamless approach.
- Error persistence across model configurations.
- Error spin up in ML models for NWP.
- Summary

# Systematic errors in climate model simulations (GC5): West Pacific subtropical high (WPSH)

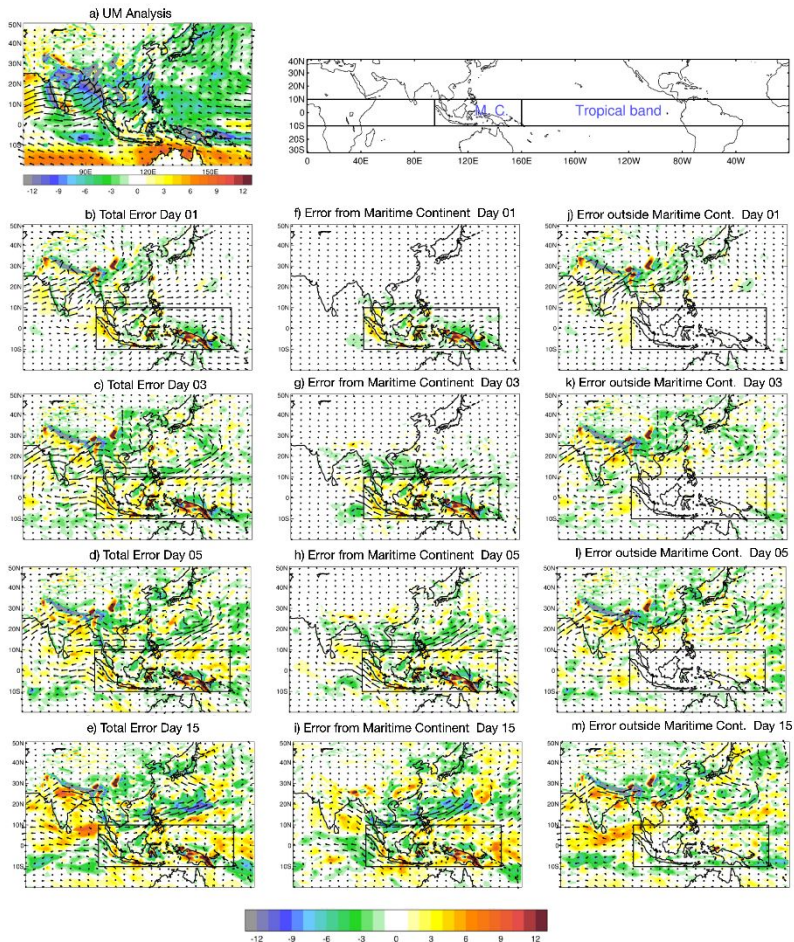
$Z_{850}, v_{850}$



$P, v_{850}$



- JJA mean WPSH is too weak, the north edge is located too far north, and the 'western border' located too far east (especially at tropical West Pacific).
- Biases in the location and strength of the WPSH affect moisture transport to East Asia, from Indian monsoon and western tropical Pacific.
- Additionally, distinct precipitation errors:
  - Dry biases over the Maritime Continent and the Indian peninsula.
  - Excessive rainfall in the tropical W. Pacific and erroneous extension of the Asian Monsoon low level flow.
  - Erroneous cyclonic circulation in the Western Pacific subtropics.
- Robust errors also seen at higher model resolutions and in CMIP6 models.



- Biases in climate models can develop in the first few days of a simulation and then persist to climate.
- Use initialised forecasts to assess physical processes in climate models: large circulation errors are initially minimised  $\Rightarrow$  study error spin-up after initialisation with a small impact of circulation-physics feedbacks and with a separation of physical processes and remote teleconnections: particularly useful in **MetUM seamless approach**.

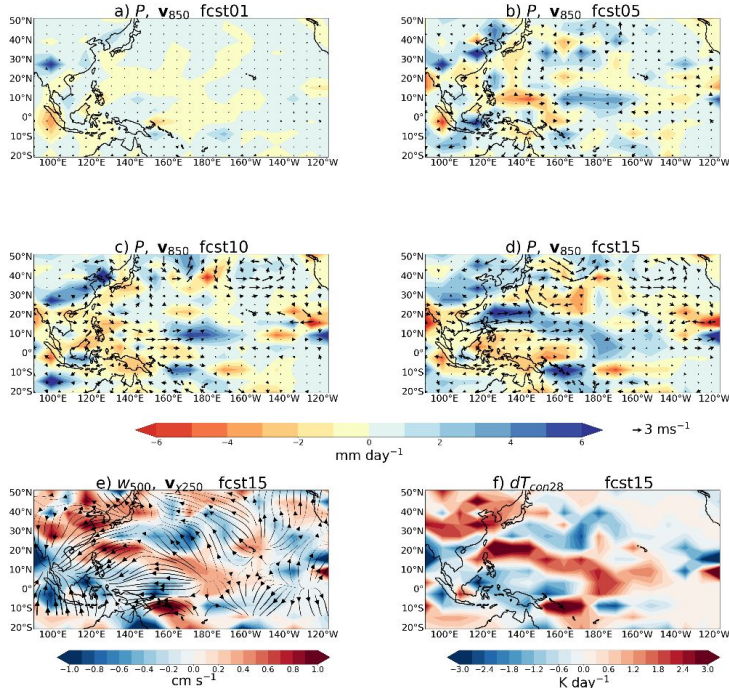
- Use nudging methodology to separate **local vs remote sources of error**:
  - Circulation errors in the Maritime Continent area remotely force excessive westerly monsoon flow into the western Pacific and erroneous cyclonic circulation in the West Pacific.

Rodríguez, J. M. and Milton, S. F. (2019), *Atmosphere*

Martin, G. M. et al, (2021), *Geoscientific Model Development*

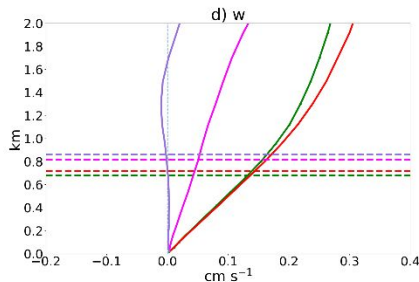
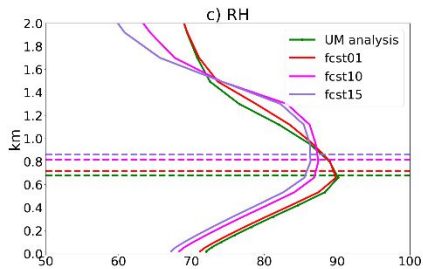
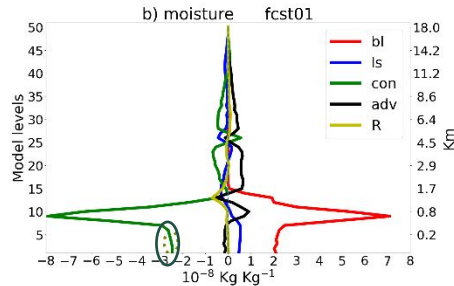
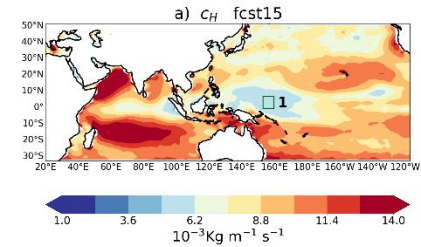
Martin, G. M. and Rodríguez, J. M., (2024), *Weather and Climate Dynamics*

# Emergence of circulation errors



- We use an ensemble of NWP-N216 hindcasts, ran for 15 days → composite analysis at various lead times during the July 2016 period (31 cases).
- Gradual emergence of P and  $v_{850}$  errors consistent with AMIP-N96 simulations: by day 15 the distinct pattern is fully formed.
- Diabatic heating errors: convection tendency errors in the tropics from a pattern in agreement with P errors.
- Vertical motion errors in agreement with P and convection tendency errors: deficient upward motion in the MC circumscribed by an area of excessive lift
- Confluence of errors highlights **key role played by tropical convection** in the development of circulation errors in the western flank of WPSH.

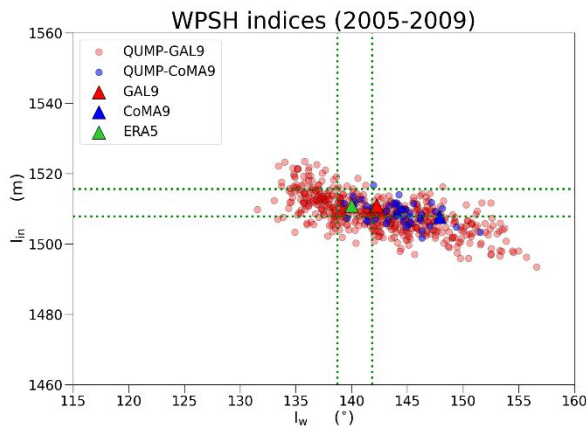
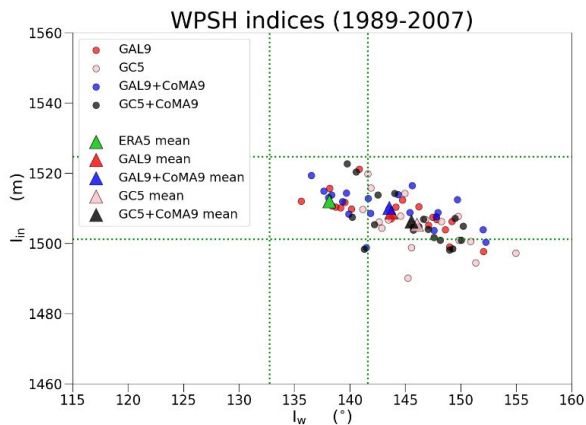
## Emergence of convection bias: deficit case



- Model convection is inhibited on places of light surface winds (small exchange coefficient  $c_H$ ).
- Moisture budget shows a large convection increment in lower boundary layer (bl), which is balanced mainly by large bl increment, **mainly surface fluxes**.
- This is not sustainable in places with low  $c_H$  and RH decreases.

$$Q = c_H(q^*(SST) - q_0),$$

- Lifting condensation level linearly related to RH in convection scheme. As RH decreases, lcl. increases  $\rightarrow$  reducing ability to generate deep convection.
- As the lead time increases convection becomes weaker and weaker and the deficit bias is established.
- We need to **device a convection scheme that does not produce such a large drying in the lower boundary layer**. A revised convection parameterization scheme, CoMorph, is under development.

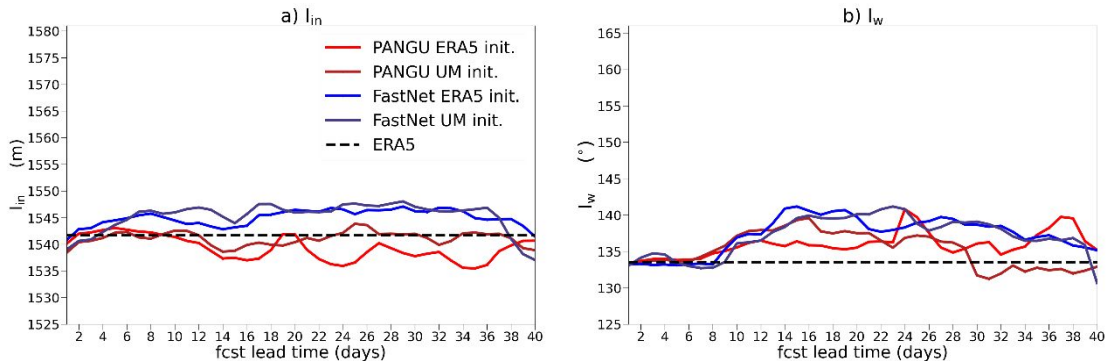


## Error in GC5 and GC5 with new convection scheme

- Use two indices to examine year to year variability: Intensity and western extent.
- Western extent ( $I_W$ ): Values located **too far east** and variability in all climate simulations larger than in ERA5, **independently of the convection scheme used** or if it is AMIP or coupled configuration.
- Intensity ( $I_{IN}$ ): AMIP simulation within limits of ERA5 variability, improving error common in previous model generations. Coupled: larger spread, with a tendency to lie in the weaker side.

## Alternative behaviours: error in QUMP PPE members

- PPEs: ensembles where each member uses a different set of values for key parameters in the model's physics parameterisation schemes.
- Control simulations: red and blue triangles. Circles and triangles represent 5-year mean indices.
- In various PPE members the WPSH western boundary is not only within the variability of ERA5, but some of them even have it located too-far west. **Study some of those members to understand drivers of different behaviour.**
- $I_W$  distinctly sensitive to the formulation of convection parameterisation.



Change with lead time of the intensity and westward extension indices in Pangu and FastNet

- Met Office currently developing a machine learning evaluation intercomparison project, with the goal to compare and evaluate various ML weather models for different uses.

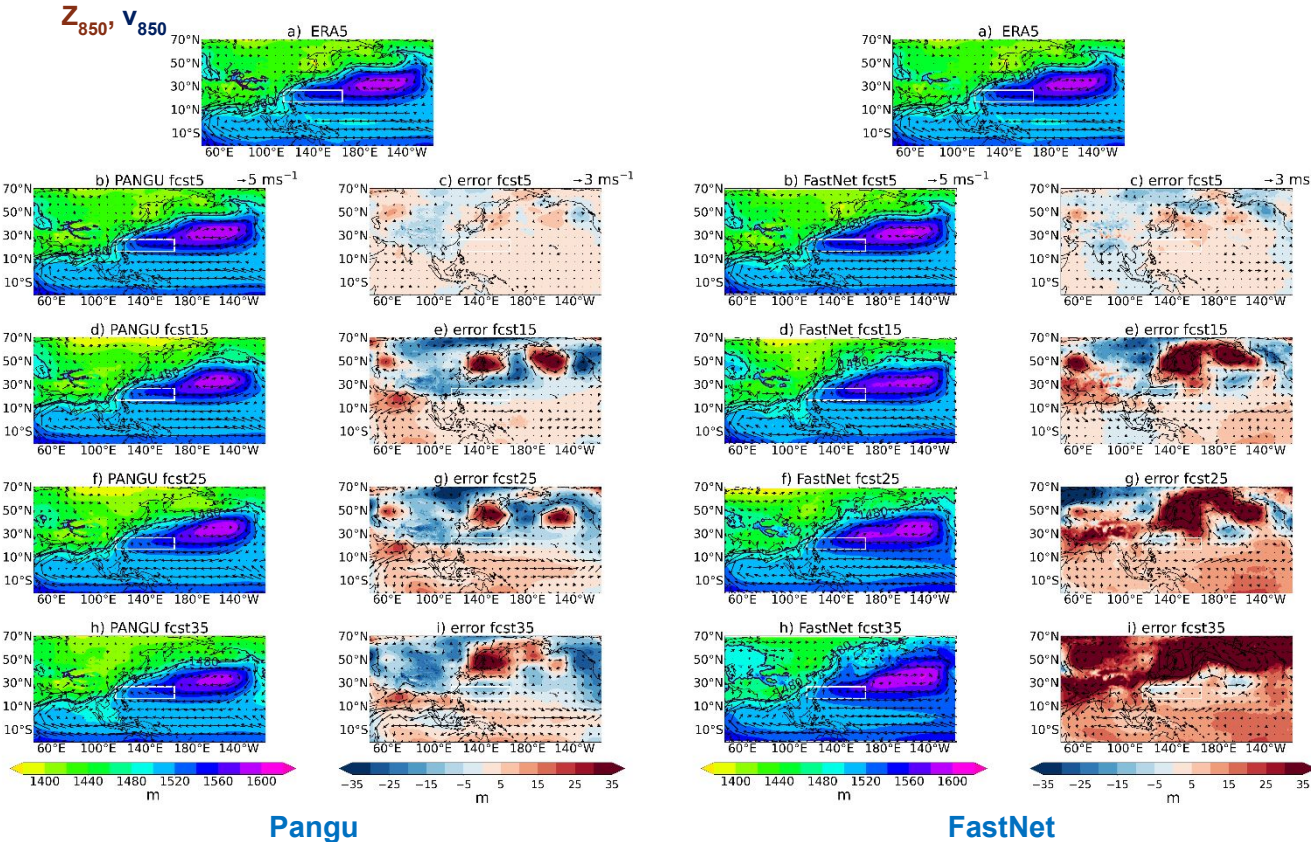
- As part of this effort, test physical realism of simulations, understand systematic errors in the models, and find potential limits of usability

- $I_w$ : time limit in the accuracy of forecasts?  
First 7 days: index in both models is almost identical with ERA5 and UM analysis (133°E).

After 8-10 days: systematic bias, with a large spread in the location of the western boundary ( $\approx 10^\circ$ ).

- $I_{IN}$ : Different errors for different architectures?  
Pangu: good agreement with analysis values for approximately 9 days, after which it becomes consistently negative (systematic weak bias).

FastNet: positive (stronger) bias.



**Pangu**

- Negligible bias during first 4 days.
- From day 5: wavy pattern of troughs and ridges at mid latitudes, and weak intensity error in the zone of maximum variability in the tropical west Pacific (white box).
- Day 15 onwards: erroneous cyclonic circulation starts to develop in tropical W Pacific that weakens the anticyclone and similar to climatological error in GC5.

**FastNet**

- Different error growth.
- As time evolves, the error becomes an extensive overly strong geopotential, saturating almost the entire region, except parts of the WPSH western extent, where it is slightly weaker.

- We have reported progress in understanding systematic circulation errors in the East Asian summer monsoon in GC5, as an example of systematic errors common in GCMs.
- Circulation errors in the area include weakening of the anticyclone and a location too far east, which leads to an underestimation of the south-westerly monsoon flow over East Asia. and contribute to seasonal precipitation errors in the area.
- There is a strong connection between these errors and diabatic heating errors associated with the tropical convection.
- Biases in the representation of the East Asian summer monsoon have persisted through various generations. To look for potential configurations on which the WPSH is better represented, we have started to explore QUMP PPE. We will continue this work with an ensemble produced with perturbations to the latest convection scheme, CoMorph B.
- We are starting to evaluate ML models for NWP to test physical realism of simulations, understand systematic errors in the models, and find potential limits of usability. We will continue this work analysing various aspects of Asian monsoon and comparing with more architectures and ML models proposed for seasonal forecasting.

For more information please contact



[jose.rodriguez@metoffice.gov.uk](mailto:jose.rodriguez@metoffice.gov.uk)

