

Ensemble Error Growth Dynamics in ML-Based Models for the Prediction of Tropical Cyclone Nargis (2008)



Apoorva K¹ and Govindan Kutty^{1*}

¹Department of Earth and Space Sciences, Indian Institute of Space Science and Technology, Valiamala – 695 547, India

*Corresponding author: Dr. Govindan Kutty · govind@iist.ac.in

APARC General Assembly 2026 · IITM Pune | Theme 4

BACKGROUND

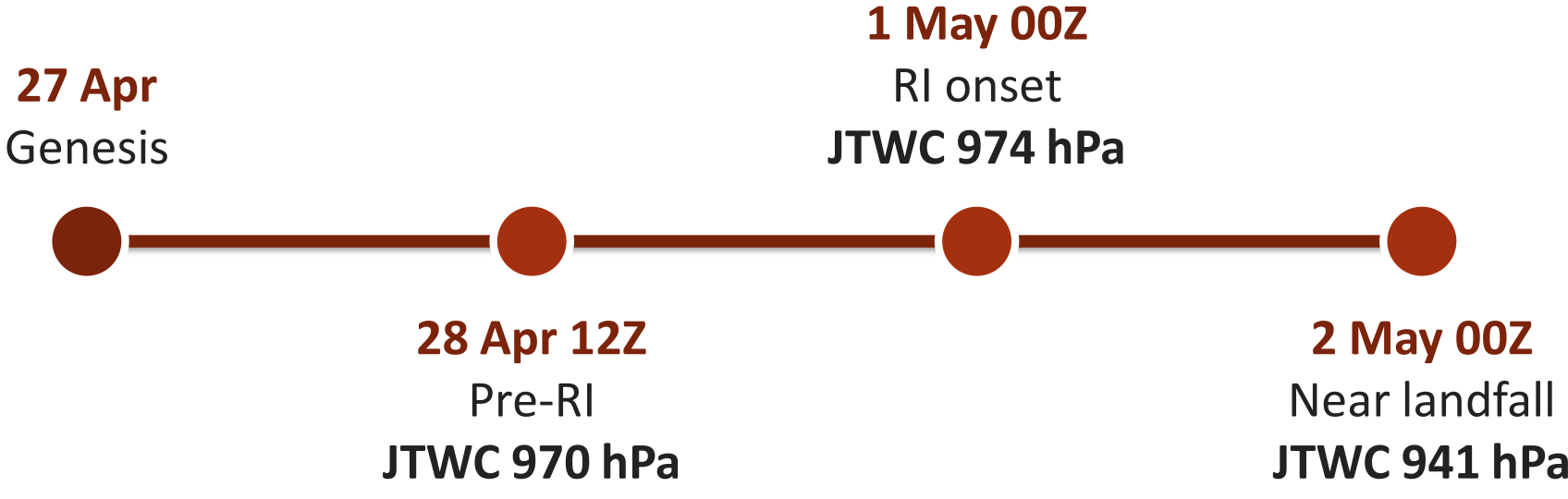
- ML weather models (Pangu-Weather, FourCastNetv2, GenCast) now reach skill comparable to operational NWP on global verification metrics.
- Skill alone is not physical realism. ML forecasts need cautious interpretation (Bonavita 2024) and may grow small perturbations too slowly, overestimating predictability (Selz & Craig 2023).
- Gap: these analyses used synoptic and global scales. Error growth in tropical cyclones (TCs), especially during rapid intensification (RI), is not established for ML ensembles.

OBJECTIVES

- Do ML ensembles show NWP-like uncertainty evolution, error growth and predictability for a TC with recurvature and RI?
- How well are track and intensity forecast?
- Is ensemble spread calibrated against error?
- How fast do initial perturbations grow (DTE)?

CASE OVERVIEW

- TC Nargis (2008): formed over the central Bay of Bengal on 27 April, moved northwest, then recurved northeast.
- Rapid intensification over warm water; JTWC MSLP fell from about 970 to 937 hPa.
- Landfall 2 May, Ayeyarwady Delta, Myanmar: more than 138,000 fatalities (Fritz et al. 2009).



EXPERIMENTAL DESIGN

1 DATA

- ECMWF IFS ensemble: TIGGE archive (Bougeault et al. 2010)
- ERA5 ensemble, 10 members: initial conditions for Pangu-Weather and FourCastNetv2 (Hersbach et al. 2020)
- Verification: JTWC best track

2 MODELS & EXPERIMENT

ECMWF IFS · physics NWP · SV + SPPT
GenCast · diffusion · 50 stochastic samples
Pangu-Weather · transformer · lagged IC
FourCastNetv2 · SFNO · lagged IC
50 members each · init 12 UTC 26 Apr 2008 · 144 h

3 PROCESSING

- Lagged ensembles (Hoffman & Kalnay 1983): 10 ERA5 members × 5 start times, 25–26 Apr = 50 members
- All systems on a common 0.25° grid
- ML models run with ECMWF ai-models

4 DIAGNOSTICS

- Track error and MSLP bias vs JTWC
- Spread/error ratio (1 = reliable)
- DTE = $\frac{1}{2} \Sigma (\delta u^2 + \delta v^2 + \kappa \delta T^2)$, $\pm 10^\circ$ storm box at 850, 500, 200 hPa; doubling time τ_2 from fit over days 0.5–3

RESULTS

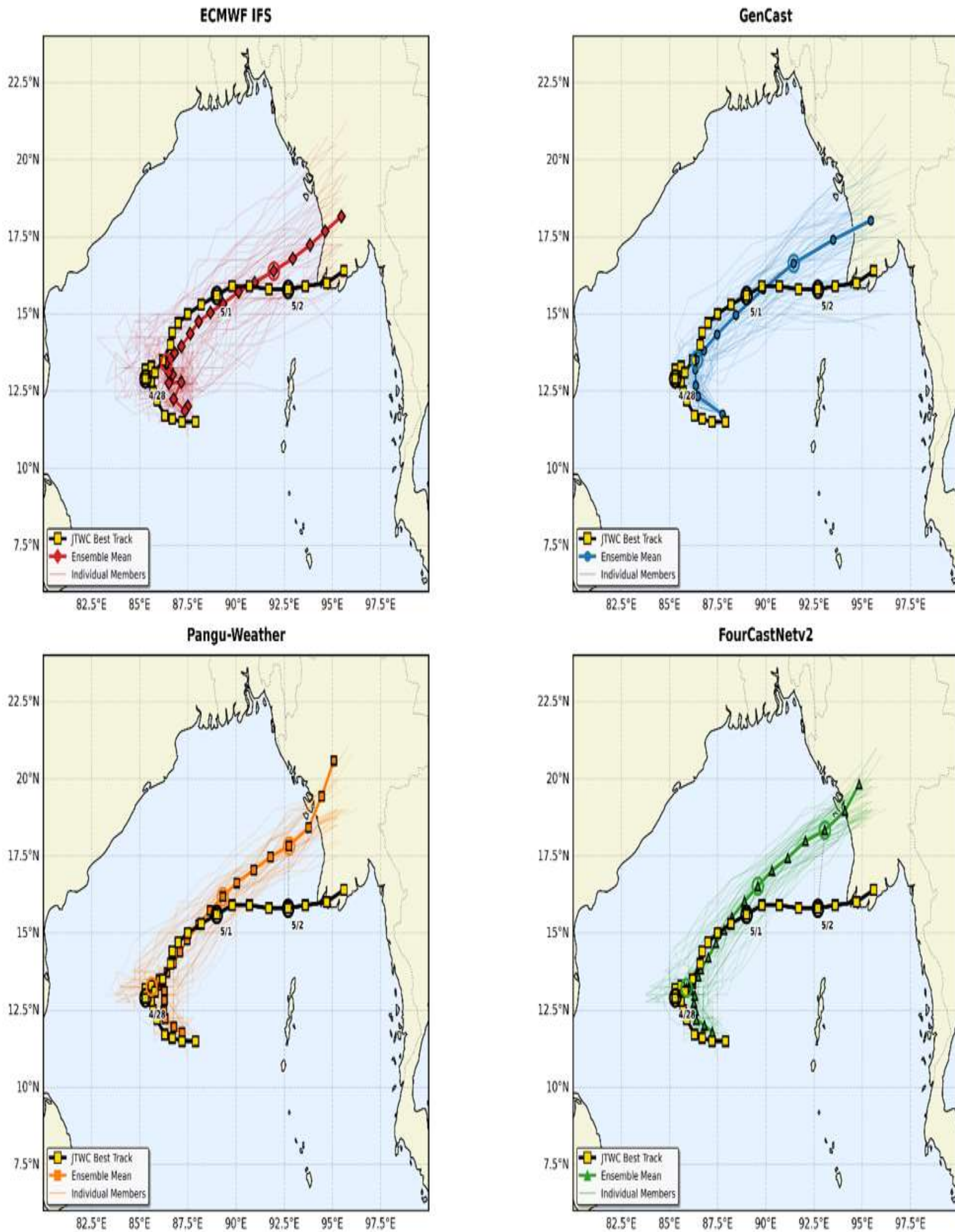


Fig. 1 All four systems capture recurvature; Pangu-Weather and FourCastNetv2 are tightly clustered, with a northward landfall bias.

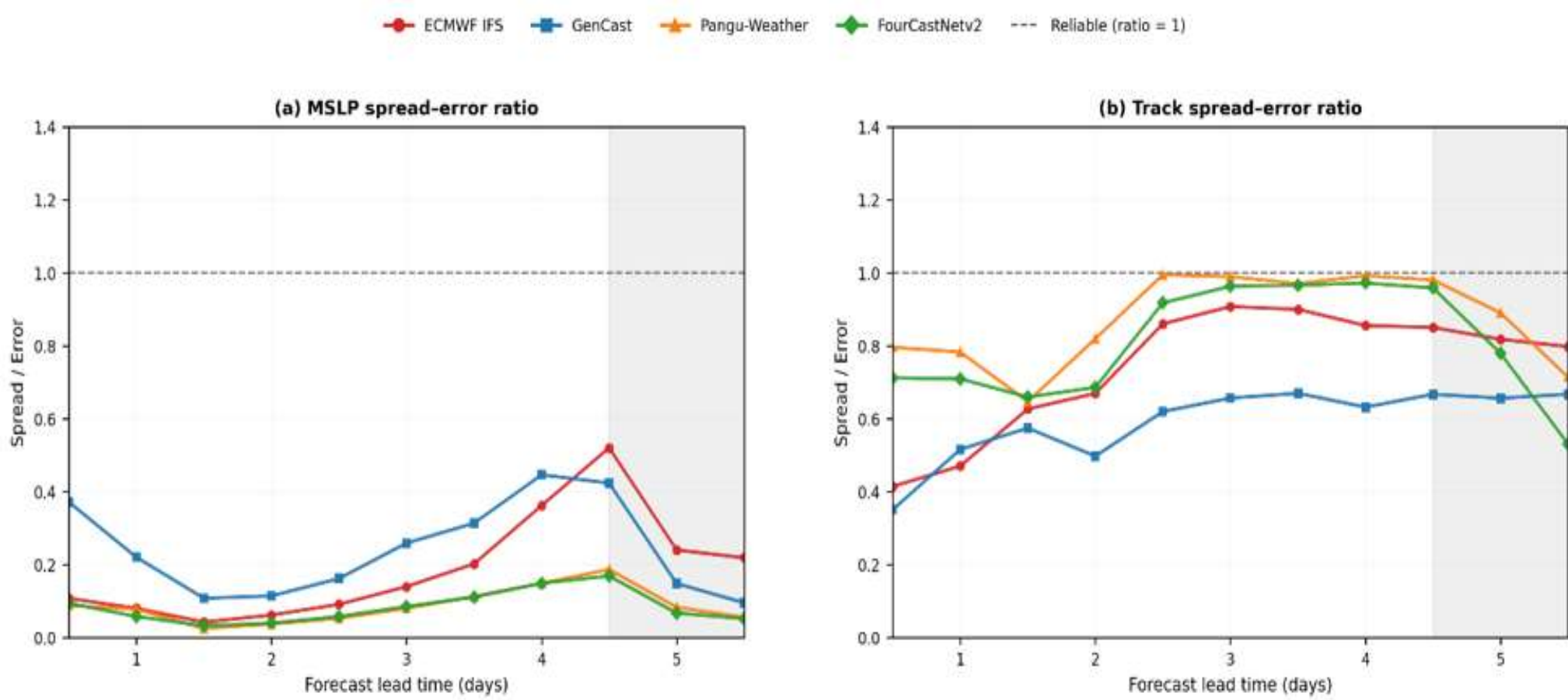


Fig. 3 Intensity ensembles are under-dispersive; track spread is better calibrated.

Models	850 hPa	500 hPa	200 hPa
ECMWF IFS	0.69	0.71	1.64
GenCast	0.78	0.82	1.69
Pangu-Weather	0.86	0.74	1.93
FourCastNetv2	1.05	0.81	1.79

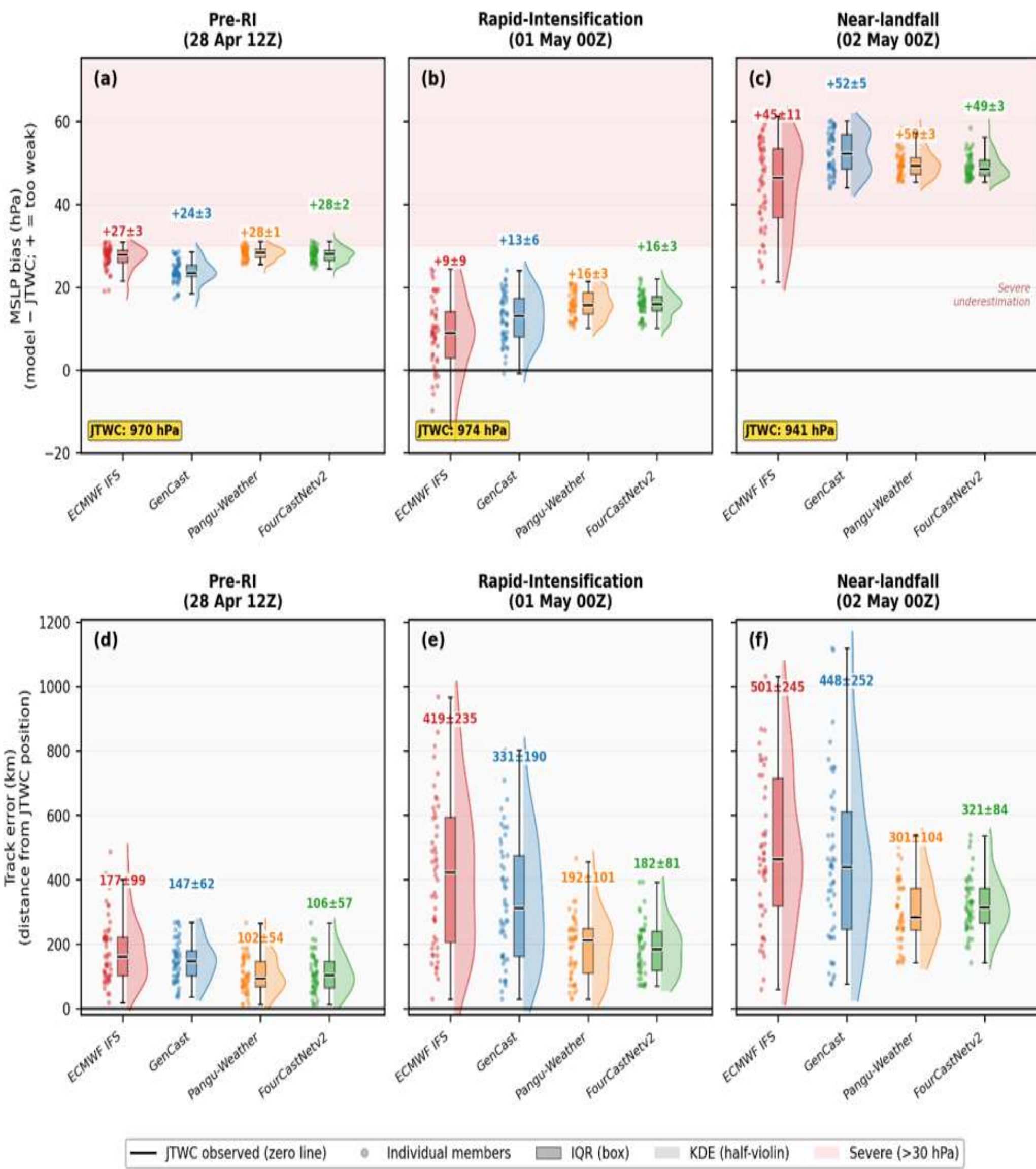


Fig. 2 MSLP bias grows from 24–28 hPa before RI to 45–52 hPa near landfall.

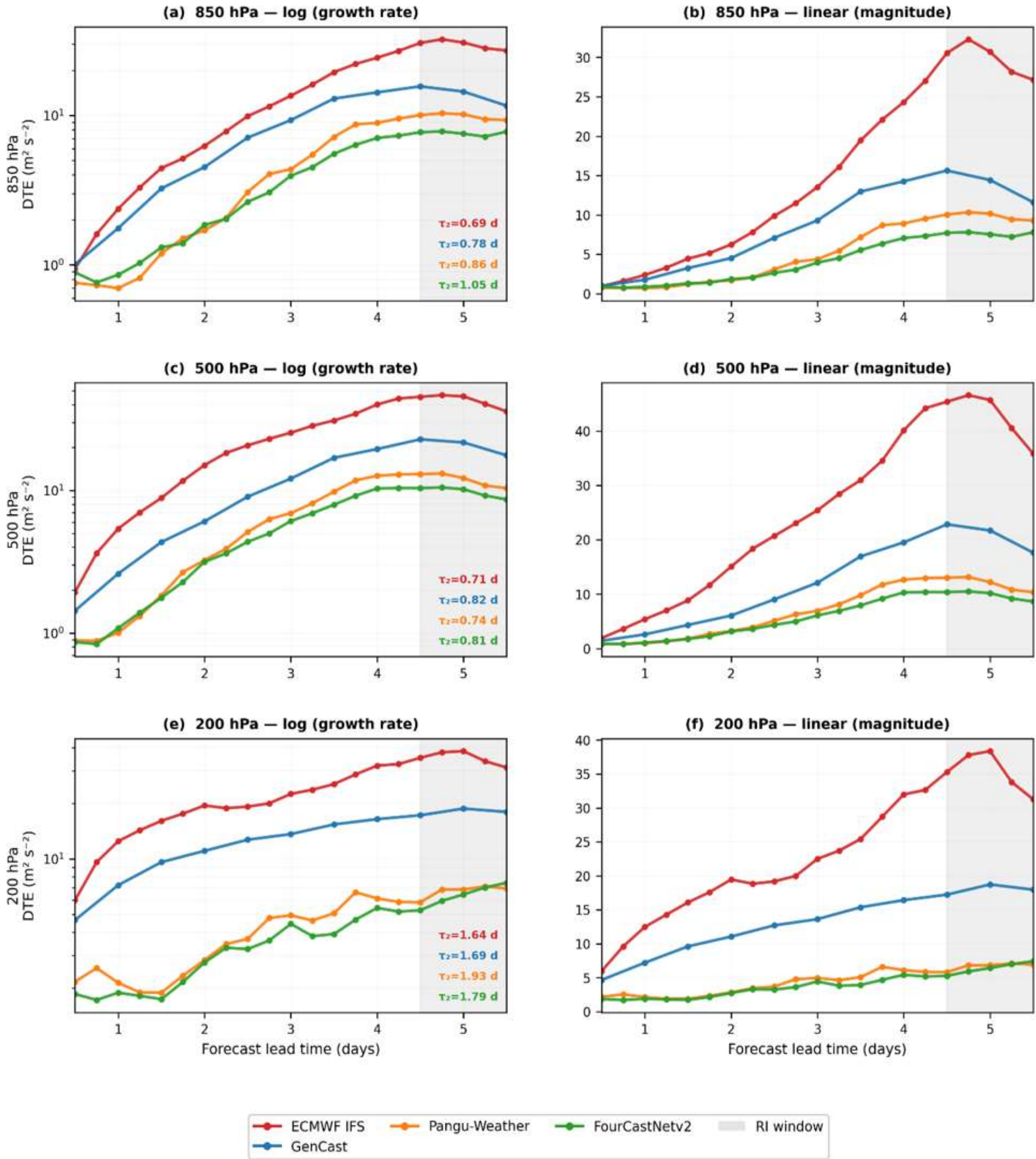


Fig. 4 ECMWF reaches the largest DTE at all levels, followed by GenCast. DTE (m² s⁻²); storm-centred $\pm 10^\circ$ box. Grey: RI window.

KEY FINDINGS & CONCLUSIONS

- Track:** all four systems capture recurvature; landfall track errors are 301–501 km (mean).
- Intensity:** every system underestimates it; bias grows from 24–28 to 45–52 hPa.
- Calibration:** intensity ensembles are under-dispersive (spread/error ≤ 0.52); track is better calibrated.
- Error growth:** ECMWF and GenCast show the largest DTE; GenCast's spread and growth are closest to ECMWF among the ML models.

INTERPRETATION

Stochastic sampling (GenCast) represents uncertainty better than lagged-IC deterministic models, but intensity and its uncertainty needs calibration.

FUTURE WORK

- More tropical cyclones across basins and intensity regimes
- How ML models capture nonlinear error growth during intensification
- Hybrid ML–NWP ensembles and multi-model super-ensembles

Limitations

Single case; ensemble-generation methods differ between systems; 2008 lies within the ML training periods.

REFERENCES

- Bi et al., Nature 619, 533–538 (2023)
- Bonavita, GRL 51, e2023GL107377 (2024)
- Bonev et al., ICML 202 (2023)
- Bougeault et al., BAMS 91, 1059–1072 (2010)
- Fritz et al., Nat. Geosci. 2, 448–449 (2009)
- Hersbach et al., QJRM 146, 1999–2049 (2020)
- Hoffman & Kalnay, Tellus A 35, 100–118 (1983)
- Price et al., Nature 637, 84–90 (2025)
- Selz & Craig, GRL 50, e2023GL105747 (2023)