MAUSAM

Measuring **A**I **U**ncertainty during **S**outh **A**sian **M**onsoon

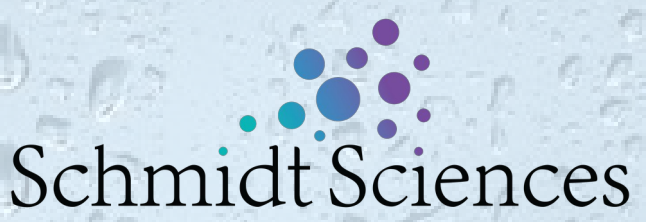
An Observations-focused Assessment of Global AI Weather Prediction Models During the South Asian Monsoon

Aman Gupta, Aditi Sheshadri, and Dhruv Suri

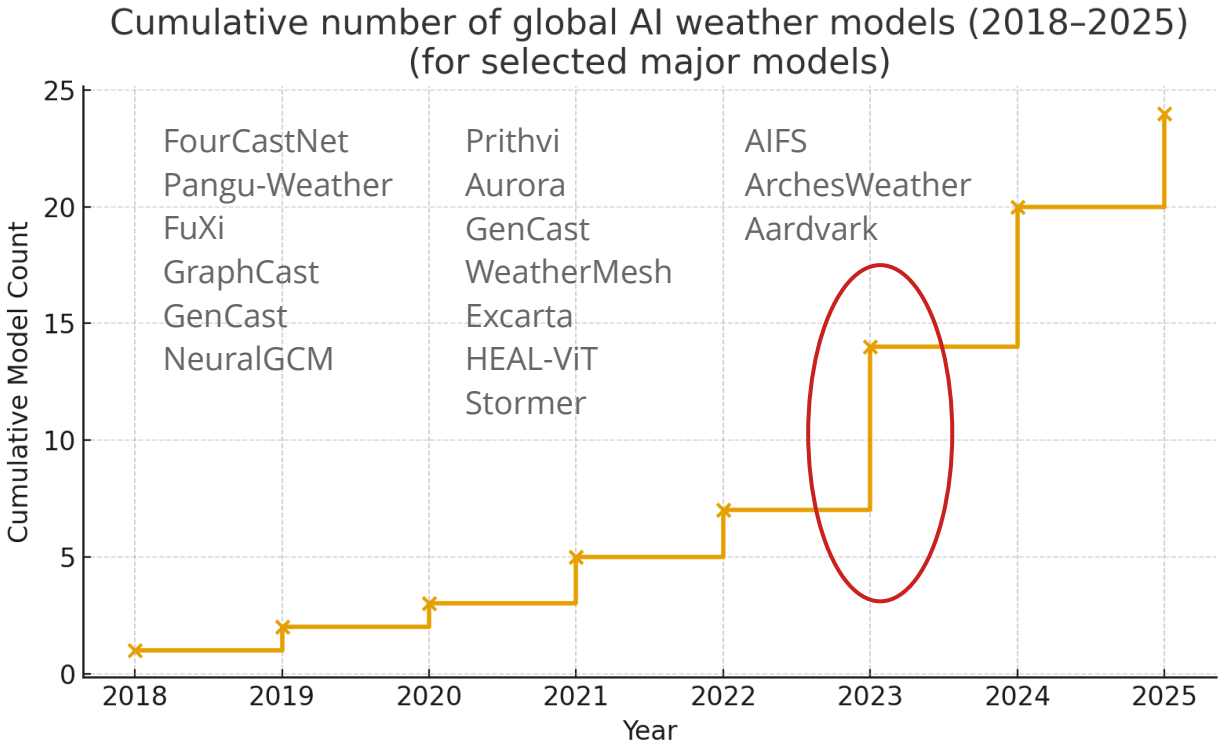
Stanford University

Indian Institute of Technology Bombay

11 December 2025



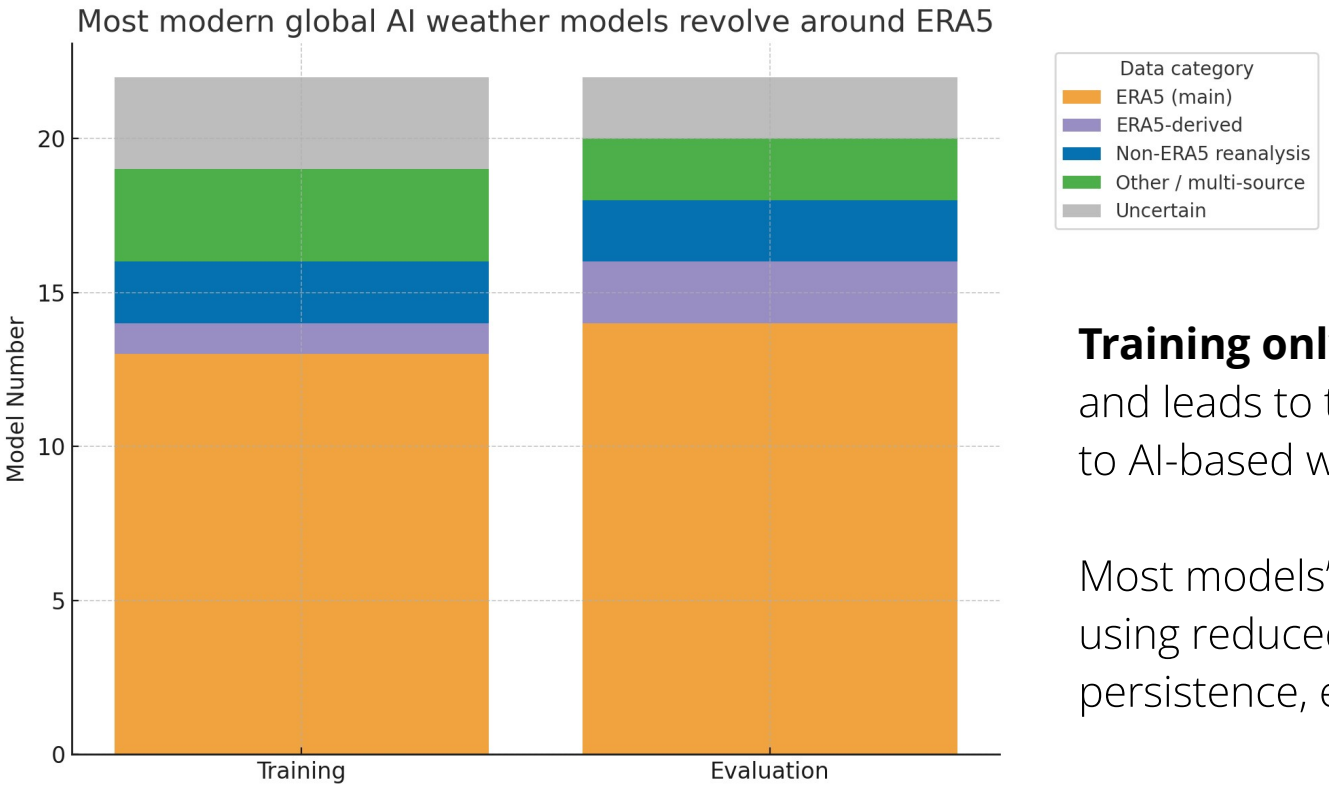
Meteoric Rise of AI Weather Forecasting Models



Over 25 data-driven weather prediction models have been released over the past 8 years.

Most of them released in the past 3 years.

Most AI Weather Models are trained and tested on ERA5 Reanalysis. None on Observations.

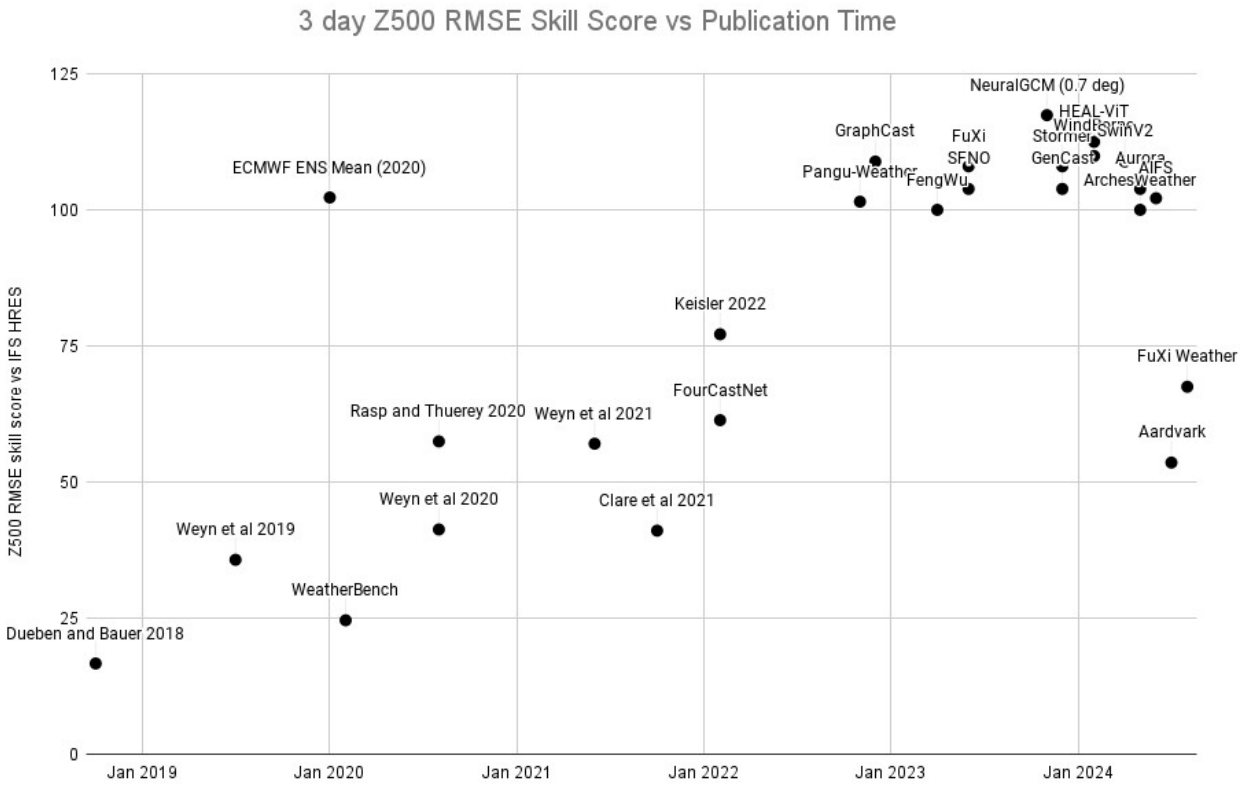


Training only on ERA5 limits learning, and leads to transfer of systemic biases to AI-based weather models.

Most models' performance assessed using reduced metrics: RMSE, ACC, CRPS, persistence, etc.

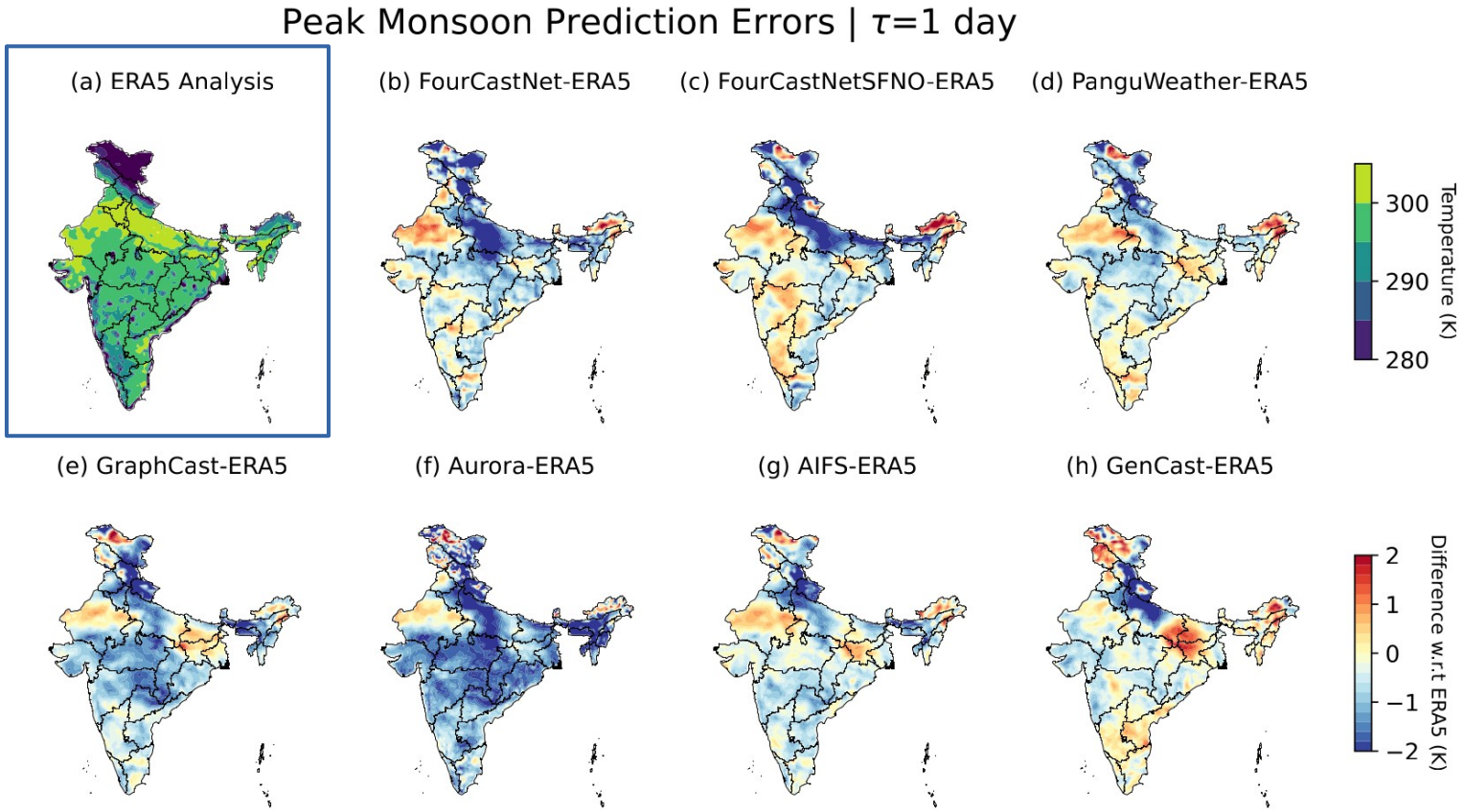
We Assess AI Weather Models During South Asian Monsoon Using *in situ* and Remote Sensing Observations

Improvements in architecture not translating to improved performance anymore



Source: Stephan Rasp's Blog

7 Different AI Models Produce 7 Different Day-ahead Forecasts



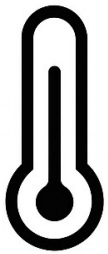
Source: Gupta et al. (submitted), JAMES

Seven AI Models Considered in the Study

AI Model	Developed at	Release Year	Spatial Resolution	Temporal Resolution
FourCastNet	NVIDIA	2022	0.25° x 0.25°	6 hrs
FourCastNet v2	NVIDIA	2024	0.25° x 0.25°	6 hrs
Pangu-Weather	Huawei	2023	0.25° x 0.25°	1hr/6 hrs
GraphCast	Google DeepMind	2023	0.25° x 0.25°	6 hrs
GenCast	Google DeepMind	2025	0.25° x 0.25°	6 hrs
Aurora	Microsoft	2025	0.25° x 0.25°	6 hrs
AIFS	ECMWF	2024	0.25° x 0.25°	6 hrs

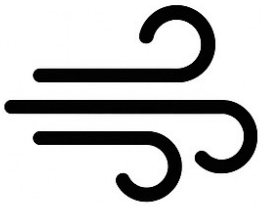
Period of integration: 15 April – 15 October | 2021-2024

We Test a Wide Range of Atmospheric Features

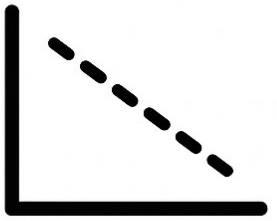


Temperature

Geopotential

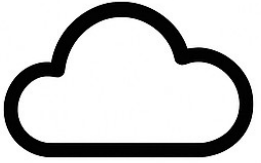


Winds



Kinetic Energy Spectrum

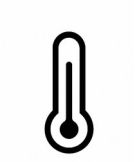
Precipitation



Cloud Cover

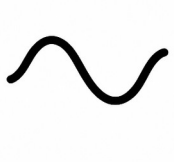
Cyclone Trajectory

Wide Range of Multi-modal Observations Used for Testing



450+ IMD Weather Stations

IMD Rain Gauge Data / ERA5 / IMERG



ERA5 / IFS / HRES model output

INSAT 3D-R/S / ERA5

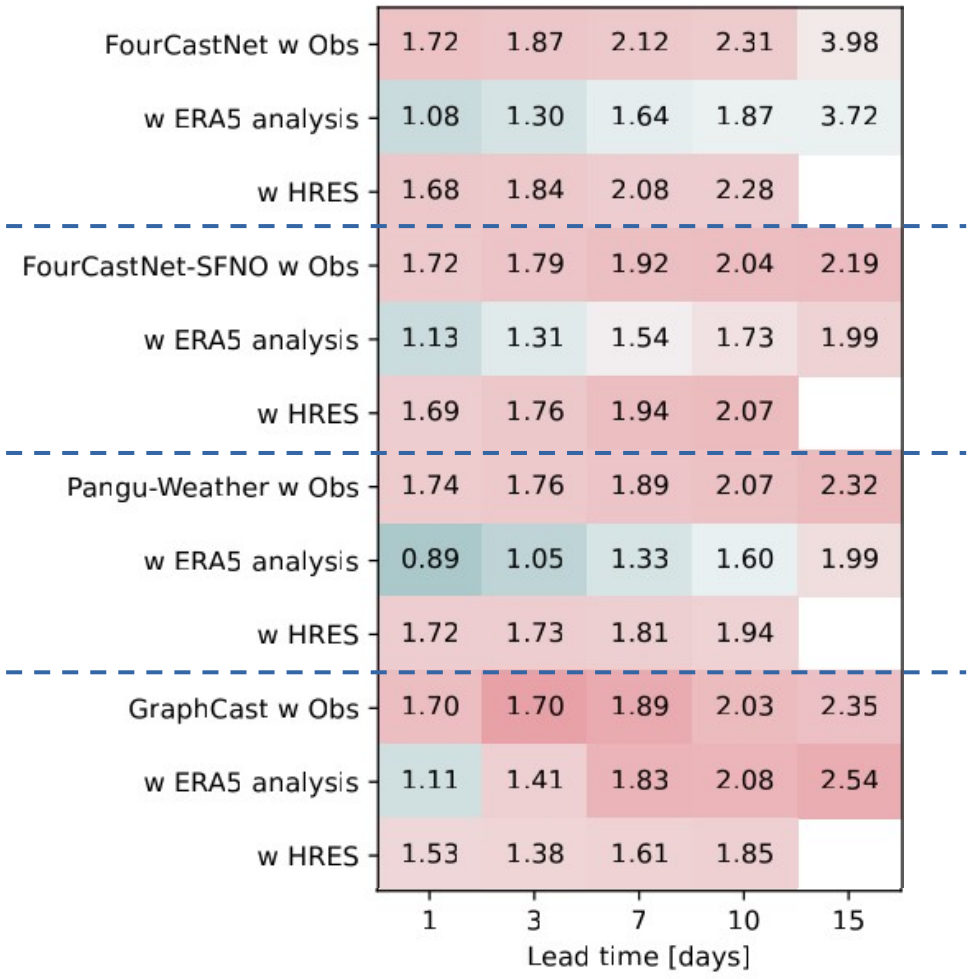
450+ IMD Weather Stations / ERA5

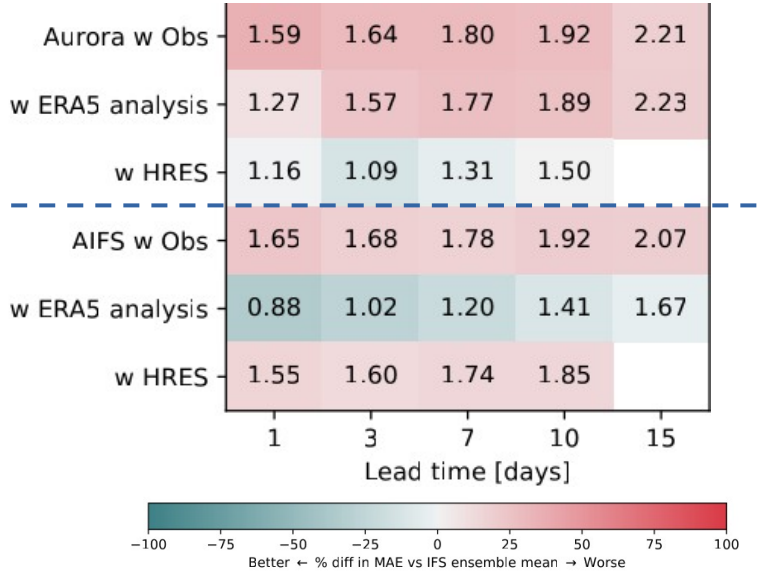
IBTRaCS Best Guess Data

ERA5 Reanalysis

Results

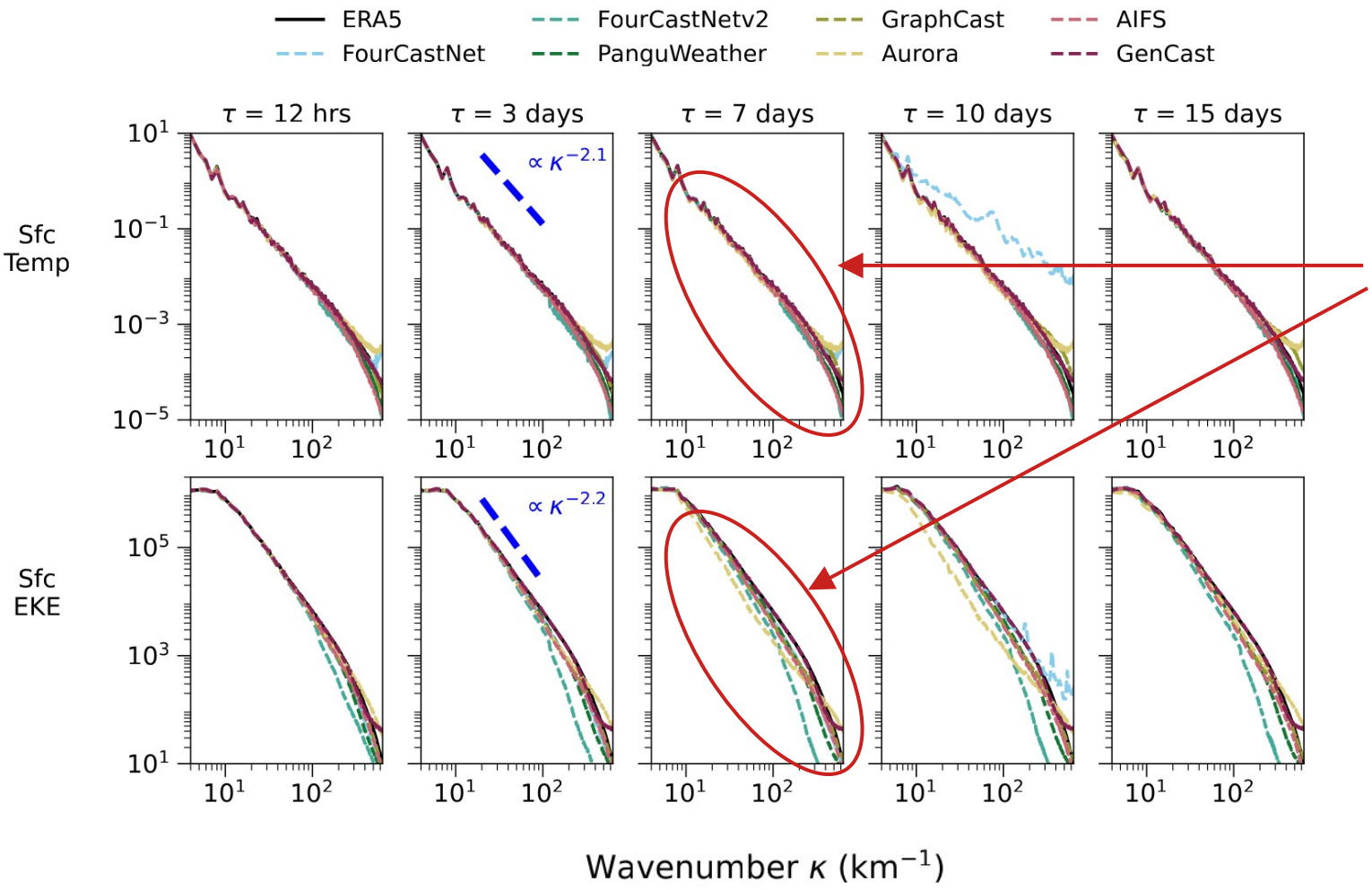
Weather Bench-like Error Analysis: **Surface Temperature**





- All models **overfit** to ERA5 and perform poorly against 9 km ECMWF HRES
- All MAEs 50-70% higher against ground observations Errors higher for surface winds speeds.
- Similar pattern for other variables at various altitudes (500 hPa, 750 hPa)

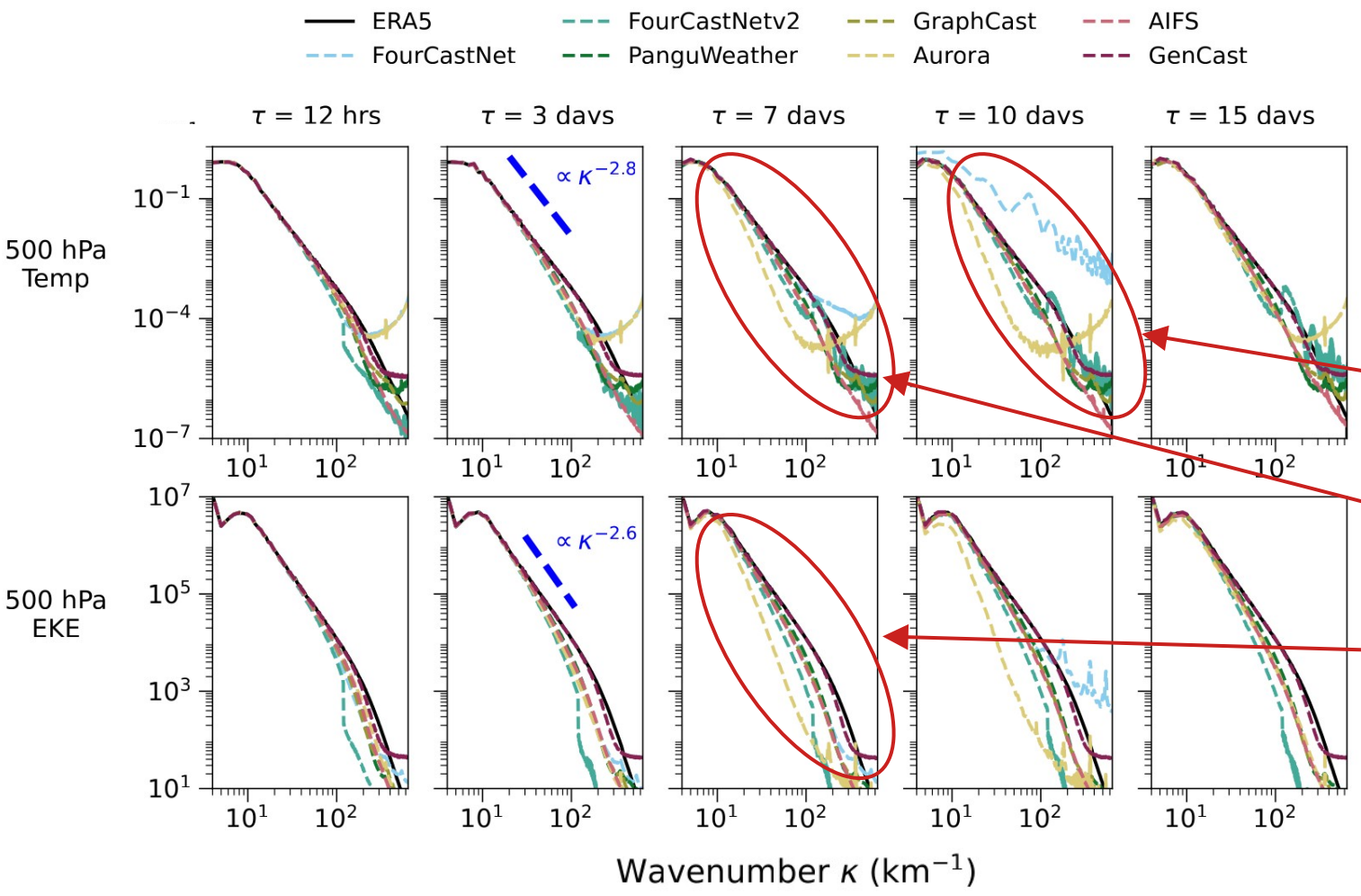
EKE Spectrum
Surface



Strong agreement in surface temperature power spectrum than EKE spectrum

Strong agreement in surface EKE spectrum over planetary and synoptic scales

EKE Spectrum
500 hPa



Disagreements much higher in the upper atmosphere.

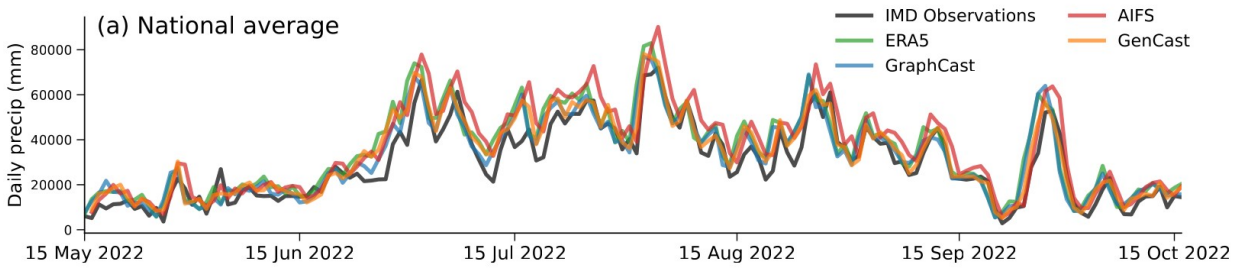
FourCastNet blows up after 10 days

Aurora produces aphysical spectrum

All moders underestimate mesoscale power

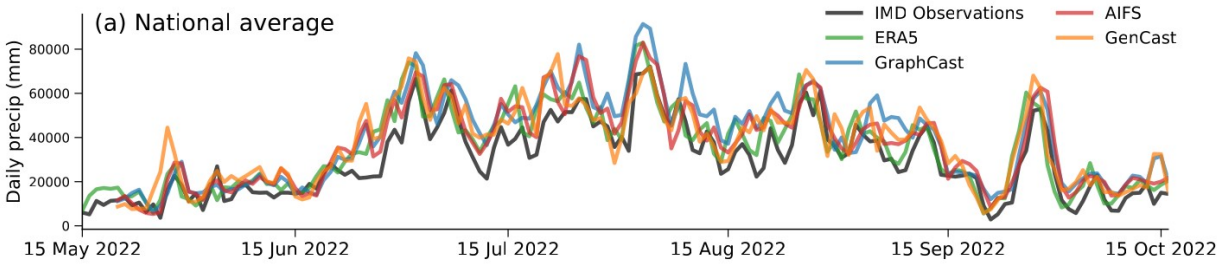
Nationally-Averaged Precipitation for Monsoon 2022

1-day ahead forecasts



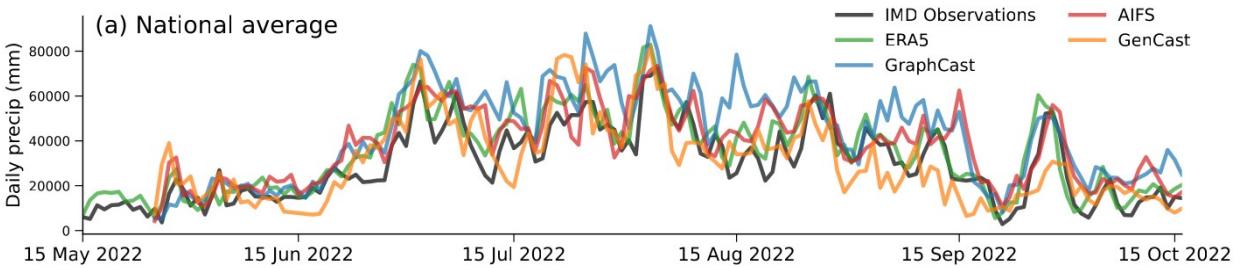
Only three out of the seven models output precipitation

5-day ahead forecasts



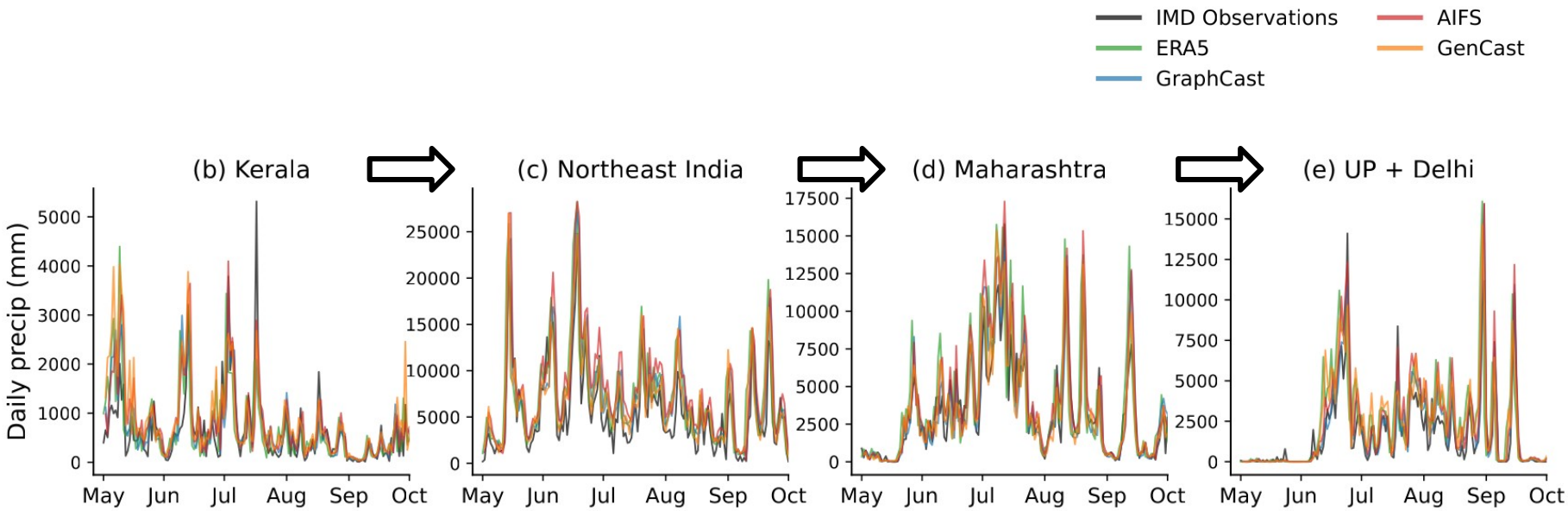
Strong agreement in national avg precip between IMD and ERA5

10-day ahead forecasts



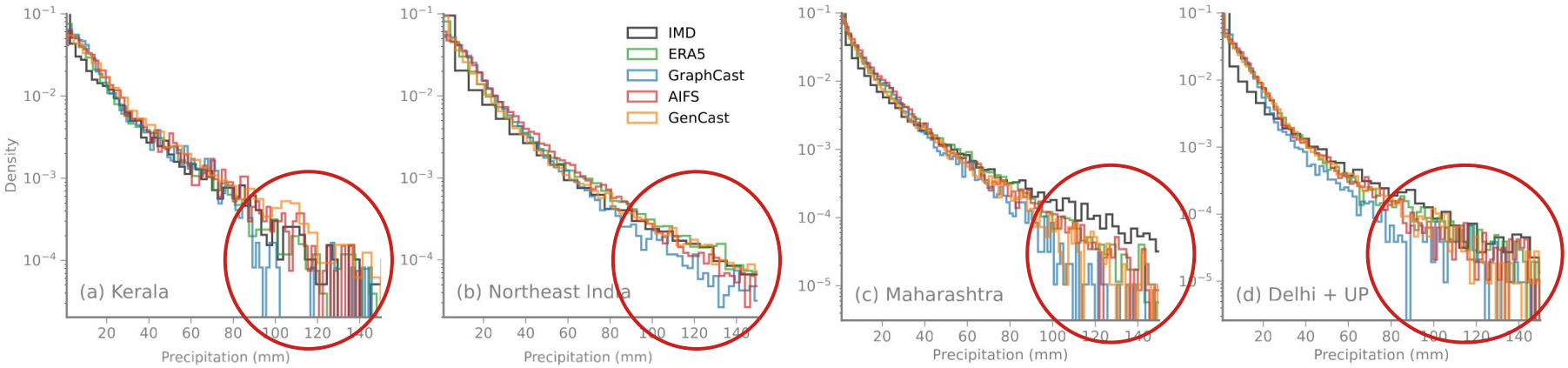
Model forecasts diverge notably after 10 days

State-wise Cumulative Precipitation for Monsoon 2022



1-5 day-ahead state-wise precipitation forecasts are well replicated by the models as well. Most episodes in precipitation intensification are identified, even as the full intensity may not be captured.

State-wise Precipitation Distribution Reveals Biases in Extreme Precip

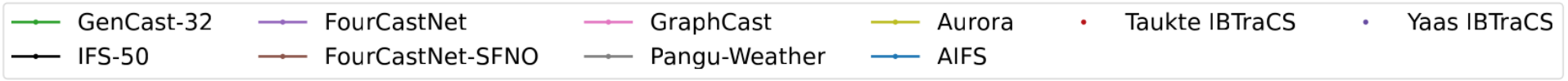


ERA5 and IMD Observations largely agree on precipitation distribution.

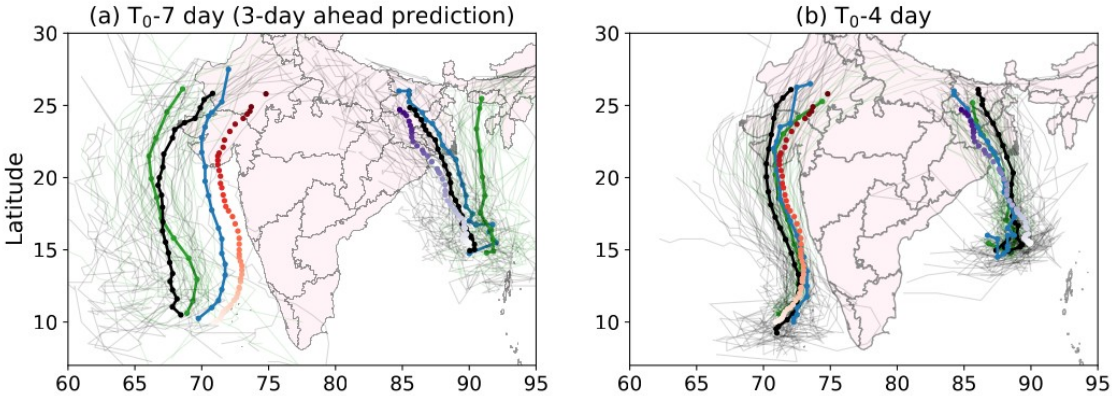
Weak (drizzle) bias in most AI models.

GraphCast fails to capture precipitation extremes. **AIFS produces the most reliable distribution.**

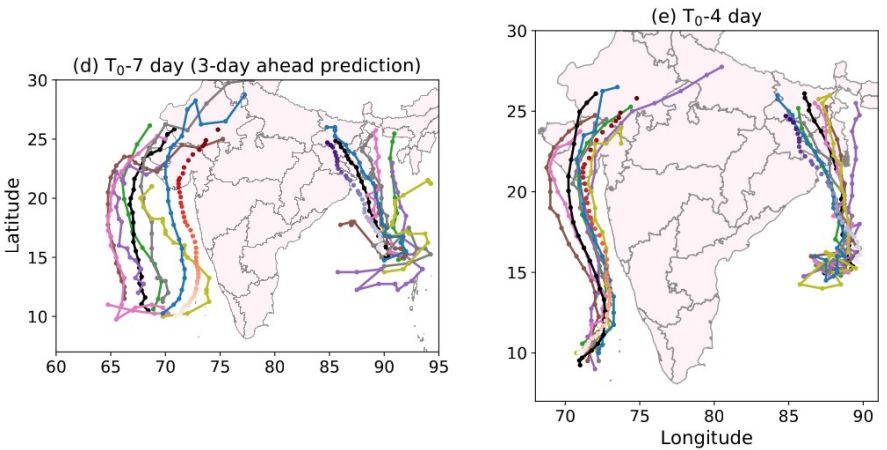
Monsoon-time Cyclones Trajectories: **Cyclones Tauktae and Yaas**



Ensembles -->



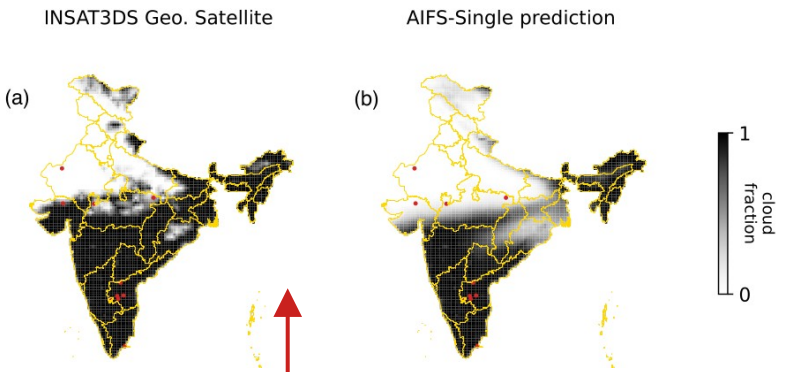
Deterministic -->



Deterministic AIFS trajectory (blue) consistently better than both ensemble means (top) and other deterministic models (bottom)

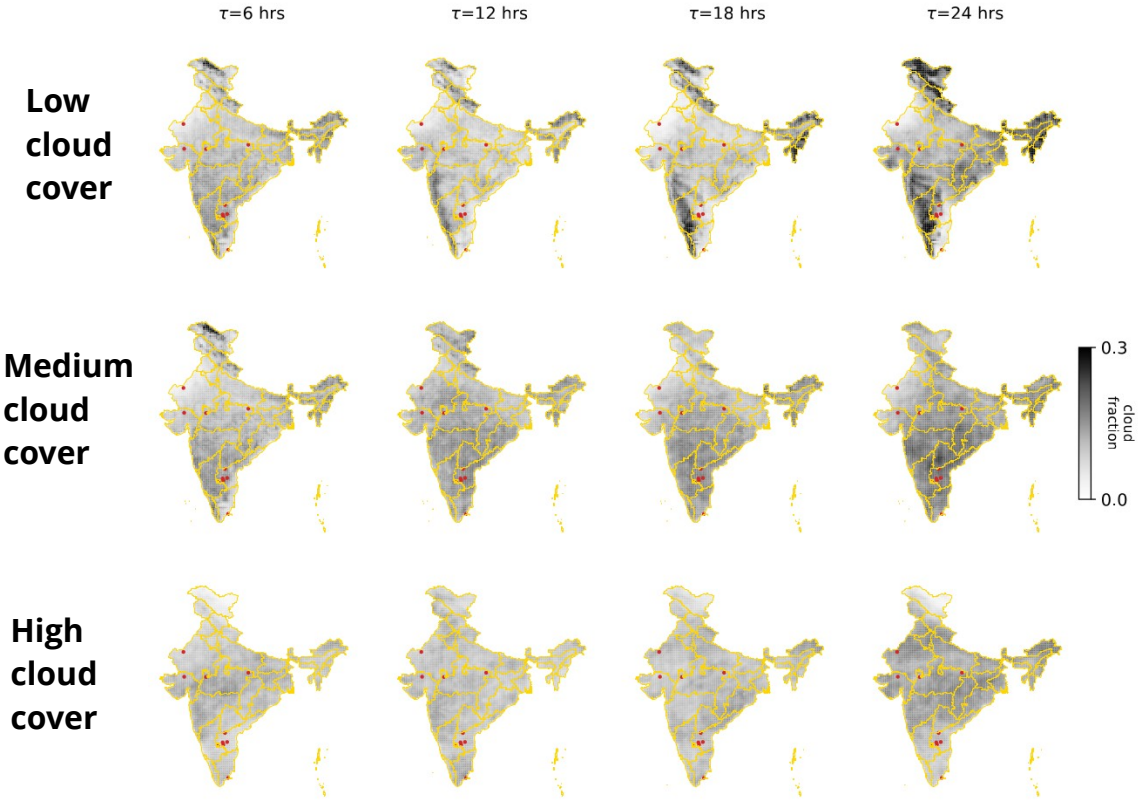
Cloud Cover

20 May 2025 1200 UTC

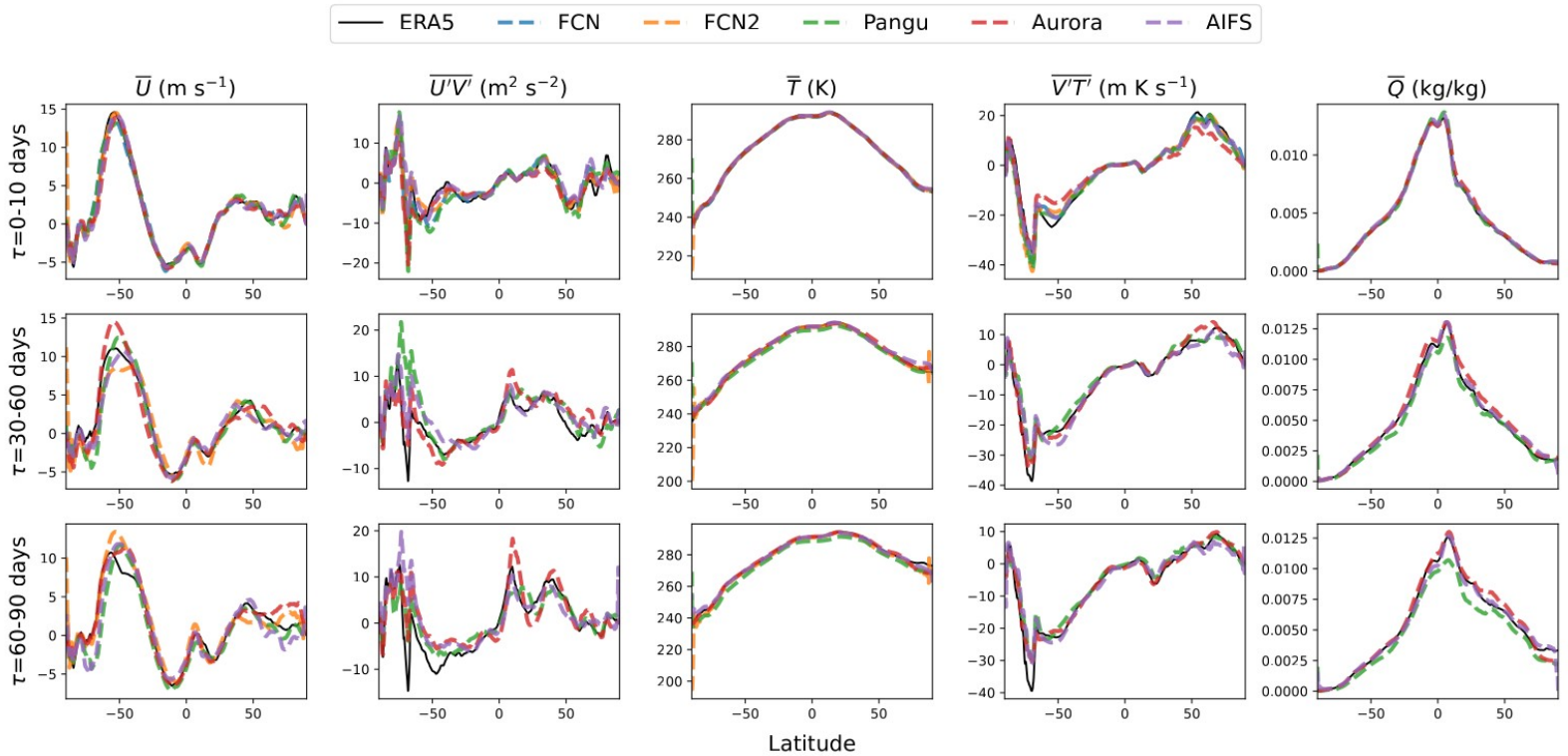


Excellent agreement of TCC with INSAT-3DS Satellite in some instances, but …

Prominent biases in low cloud cover nowcasting over the whole monsoon period (even when compared against ERA5 reanalysis)



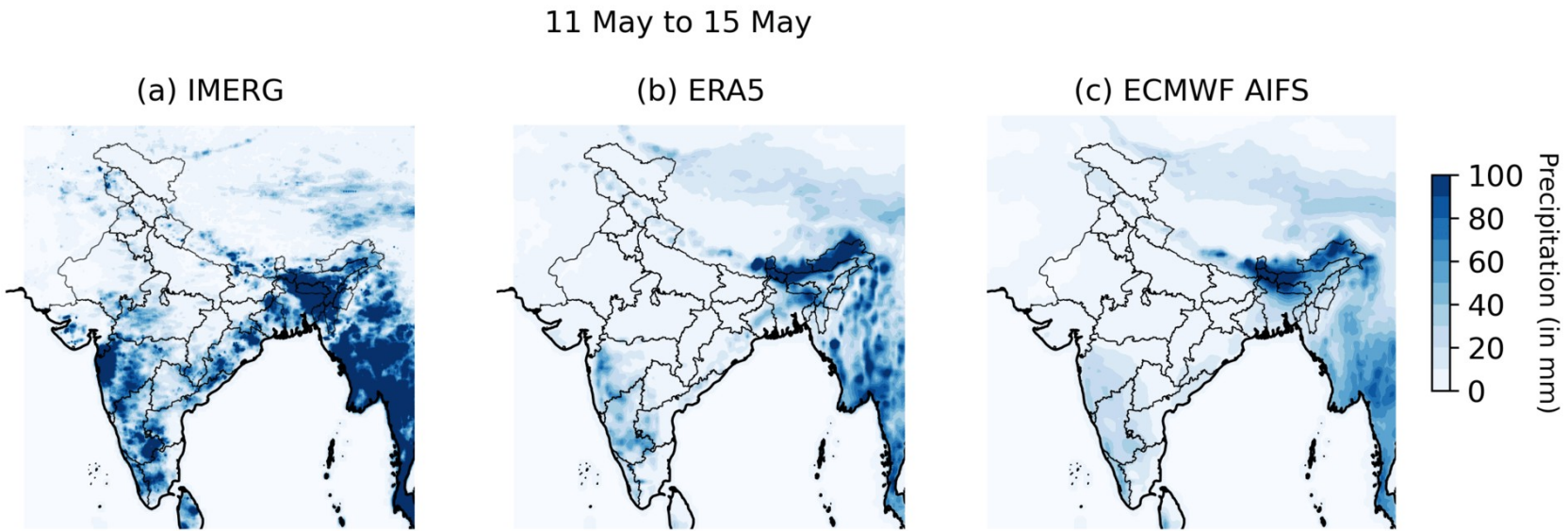
Models Produce Reliable Subseasonal-to-seasonal (S2S) Statistics



Longer-term auto regressive rollouts stable for most models.
Models generate realistic jet structure and momentum and heat flux even on S2S timescales, even though they were trained to produce 15-day forecasts only.

AI Model Captures Aspect of Abrupt Monsoon Pause of 2025

This means the monsoon usually covers the entire country in about 38 days. However, the monsoon never maintains a uniform pace throughout its journey and experiences several bouts of acceleration and deceleration. Despite a long pause of about 20 days between 26th May and 15th June 2025, the monsoon is set to cover the entire country within the stipulated timeline. A highlight of this season is that the entire country will be covered within the month of June itself.



AIFS has initial weak bias during mid-May Monsoon onset, but captures the abrupt pause and subsequent intensification of the 2025 Indian Monsoon.

Key Takeaways

Feature	Status
Errors against ground observations	Higher in all models. Lowest in AIFS.
EKE Spectrum	1st Gen models misrepresent it. 2nd Gen models underestimate it. GenCast and AIFS most accurate.
National and Regional Precipitation	Drizzle bias and underestimated extremes. Distribution closer to ERA5 than IMD rain gauge data. AIFS most accurate.
Cloud Cover	Strong bias in low cloud cover around steep topography. Key disagreements with satellite imagery.
Cyclone Trajectories	Deterministic and Ensemble AIFS outperform traditional ensembles from ECMWF.
S2S Statistics	Stable long-term rollouts and consistent jet structure, humidity profiles, and eddy momentum and heat fluxes. Biases strongest over poles.

Strengths of current AI models

AI driven weather forecasting is a rapidly evolving field:

1) 2nd generation models generate realistic medium-range forecasts and are even **stable on longer-term (S2S) rollouts**. A big improvement on 1st generation weather models.

2) Models produce **realistic EKE spectrum** (proxy for energy transfers),
and national and regional precipitation

3) Based on accuracy and applicability, **AIFS offers the best monsoon-time prediction** providing reasonable accurate cloud cover and providing a more accurate deterministic cyclone trajectory than ECMWF's IFS Ensemble.

Limitations and Scope for Improvement of AI models

1) **Limited spatio-temporal resolution and output set:** 25 km 6-hourly insufficient for operational applications. Where does higher-res training data come from? Only one model outputs cloud cover.

2) **Finetuning of AI models** and foundation models (like Prithvi WxC) on critical regional variables:
    - precip, cloud cover (in progress)

3) **Towards limited area models:** Optimizing for global metrics does little to ensure regional accuracy.

4) **Beyond ERA5:** urgent need to find training alternatives beyond ERA5 reanalysis.
    - One idea is to use the limited are India reanalysis (IMDAA).
    - Another is to rethink the data assimilation stage.

Supplementary Plots

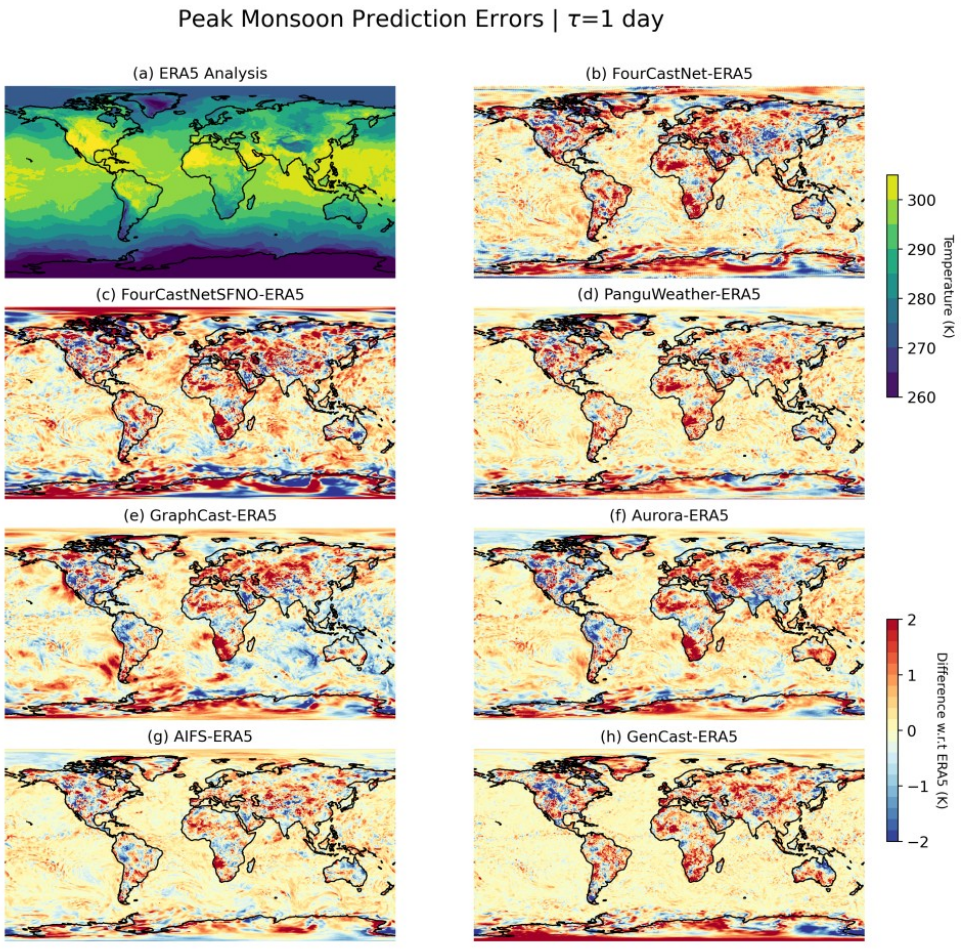
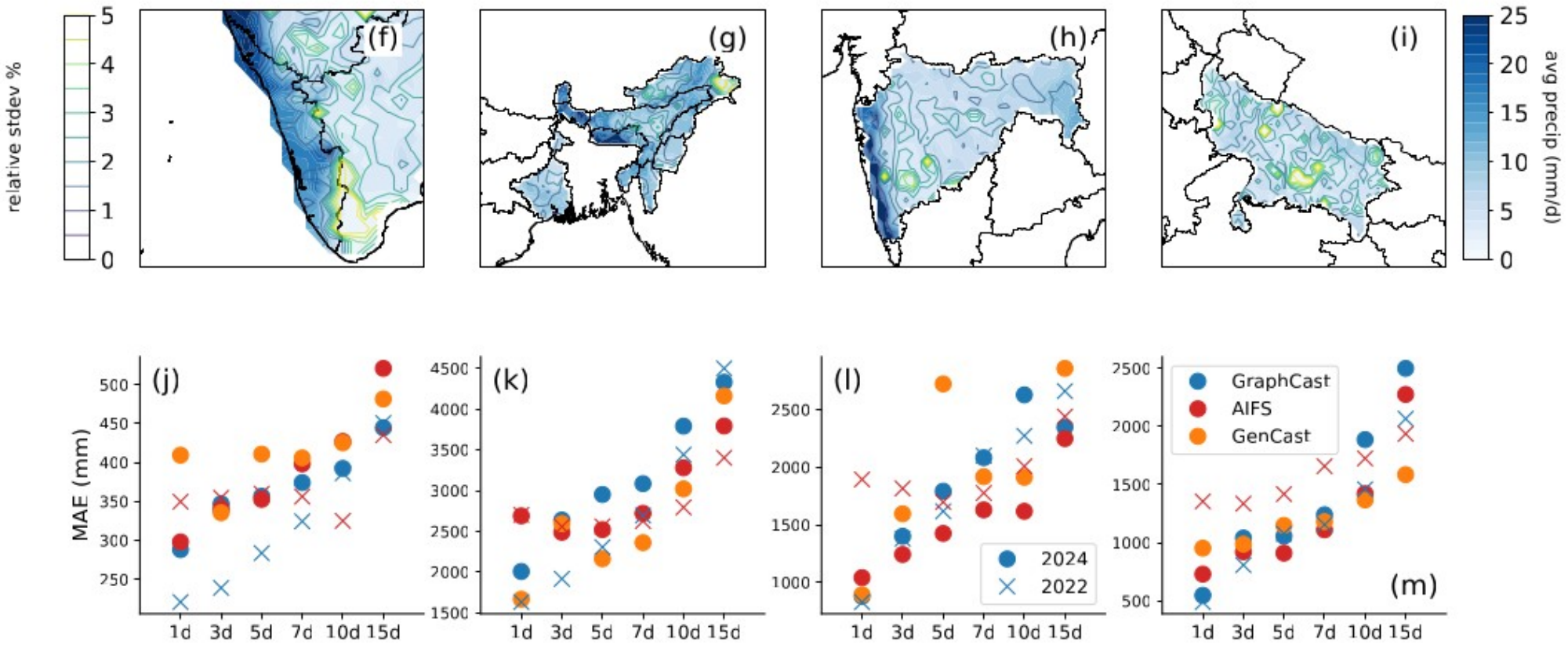
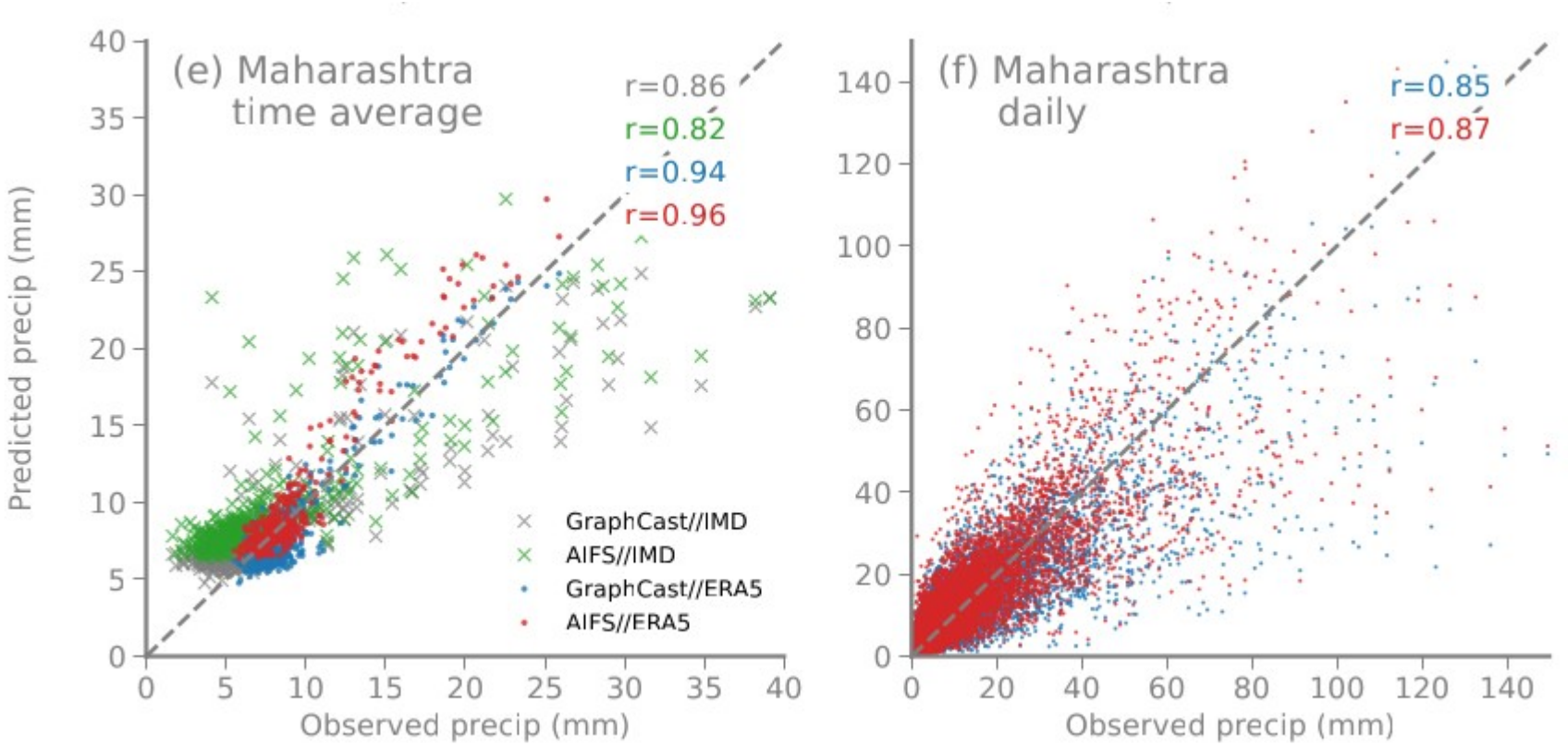
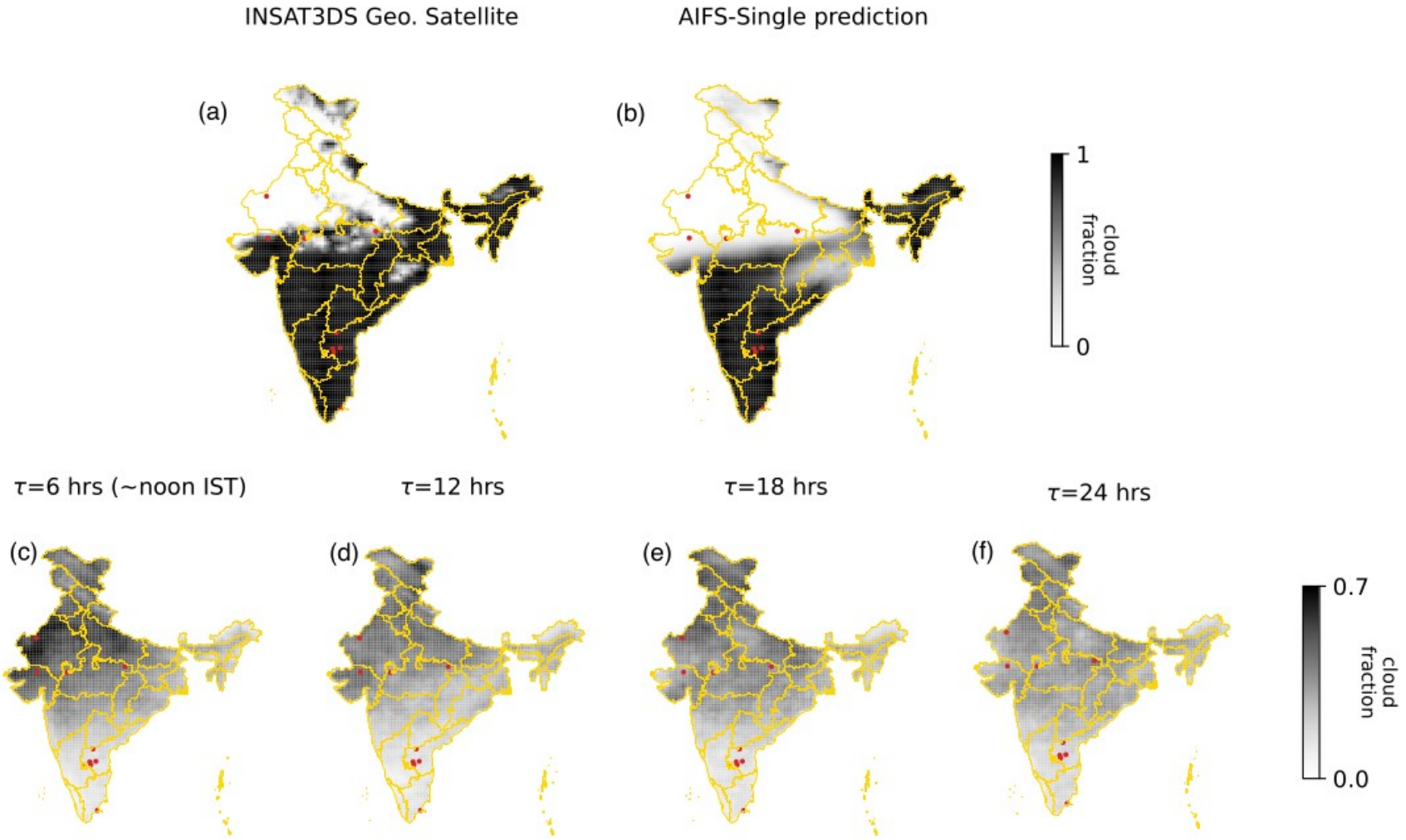


Figure A1. Global anomalies in day-ahead weather prediction across seven state-of-the-art AIWP models for models initialized on 15 July 2022 0000UTC.


relative stdev %
0
1
2
3
4
5
(f)
(g)
(h)
(i)
avg precip (mm/d)
0
5
10
15
20
25
(j)
(k)
(l)
(m)
MAE (mm)
250
300
350
400
450
500
1500
2000
2500
3000
3500
4000
4500
1000
1500
2000
2500
500
1000
1500
2000
2500
1d
3d
5d
7d
10d
15d
2024
2022
GraphCast
AIFS
GenCast


(e) Maharashtra
time average
r=0.86
r=0.82
r=0.94
r=0.96
GraphCast//IMD
AIFS//IMD
GraphCast//ERA5
AIFS//ERA5
Predicted precip (mm)
Observed precip (mm)
0
5
10
15
20
25
30
35
40
(f) Maharashtra
daily
r=0.85
r=0.87
Observed precip (mm)
0
20
40
60
80
100
120
140


INSAT3DS Geo. Satellite
AIFS-Single prediction
(a)
(b)
1
cloud fraction
0
τ=6 hrs (~noon IST)
τ=12 hrs
τ=18 hrs
τ=24 hrs
(c)
(d)
(e)
(f)
0.7
cloud fraction
0.0

Precipitation Predictions

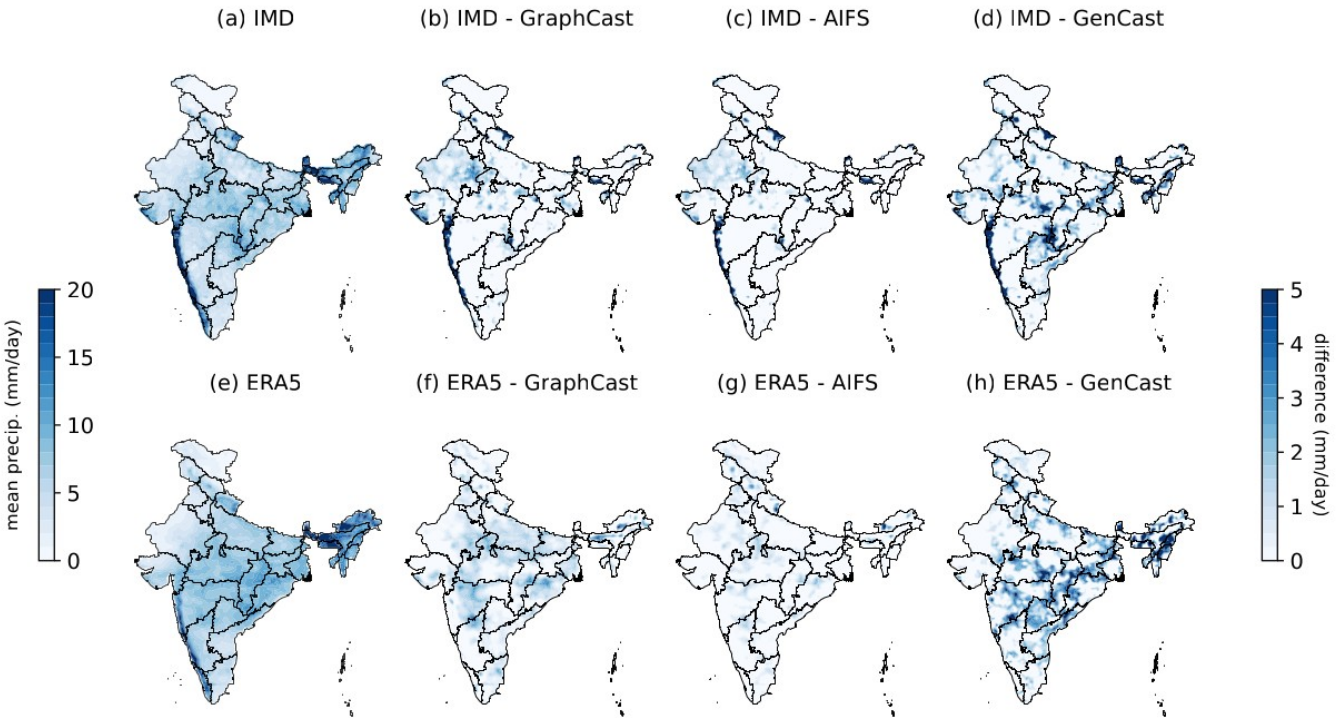
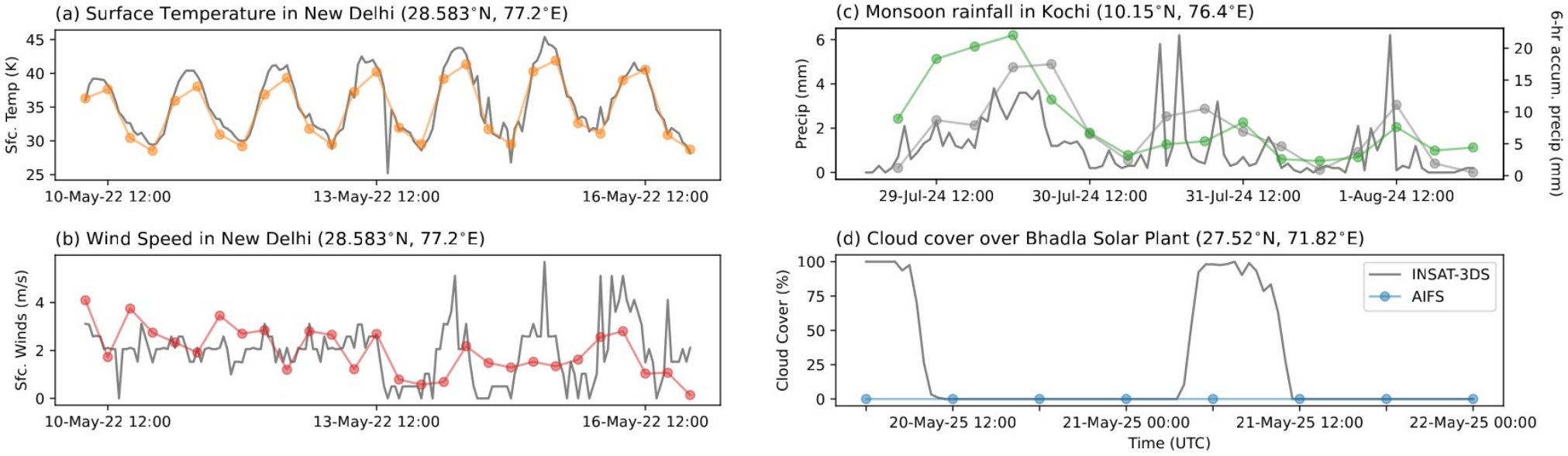


Figure A5. Systematic spatial precipitation biases persist across contrasting monsoon conditions. Mean precipitation fields and model-observation differences for the 2024


(a) Surface Temperature in New Delhi (28.583°N, 77.2°E)
Sfc. Temp (K)
45
40
35
30
25
10-May-22 12:00
13-May-22 12:00
16-May-22 12:00
(b) Wind Speed in New Delhi (28.583°N, 77.2°E)
Sfc. Winds (m/s)
4
2
0
10-May-22 12:00
13-May-22 12:00
16-May-22 12:00
(c) Monsoon rainfall in Kochi (10.15°N, 76.4°E)
Precip (mm)
6
4
2
0
6-hr accum. precip (mm)
20
15
10
5
0
29-Jul-24 12:00
30-Jul-24 12:00
31-Jul-24 12:00
1-Aug-24 12:00
(d) Cloud cover over Bhadla Solar Plant (27.52°N, 71.82°E)
Cloud Cover (%)
100
75
50
25
0
INSAT-3DS
AIFS
20-May-25 12:00
21-May-25 00:00
21-May-25 12:00
22-May-25 00:00
Time (UTC)

Ensemble Dispersion: GenCast vs. IFS Ensemble