A Scalable Framework for Climate-Aware Satellite Precipitation Bias Correction

Yashraj Upase[1], Abhigyan Chakraborty[1], Malarvizhi Arulraj[3], Kishalay Mitra[1], Shruti Upadhyaya[1,2]

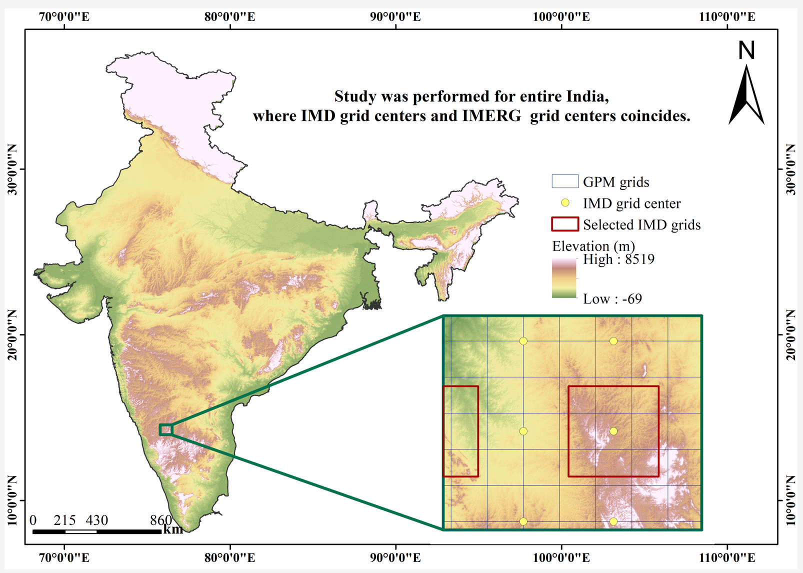
[1] *Indian Institute of Technology, Hyderabad, India* [2] *Advanced Radar Research Center, University of Oklahoma, USA*
[3] *Earth System Science Interdisciplinary Center, University of Maryland, USA*

cc23resch11006@iith.ac.in

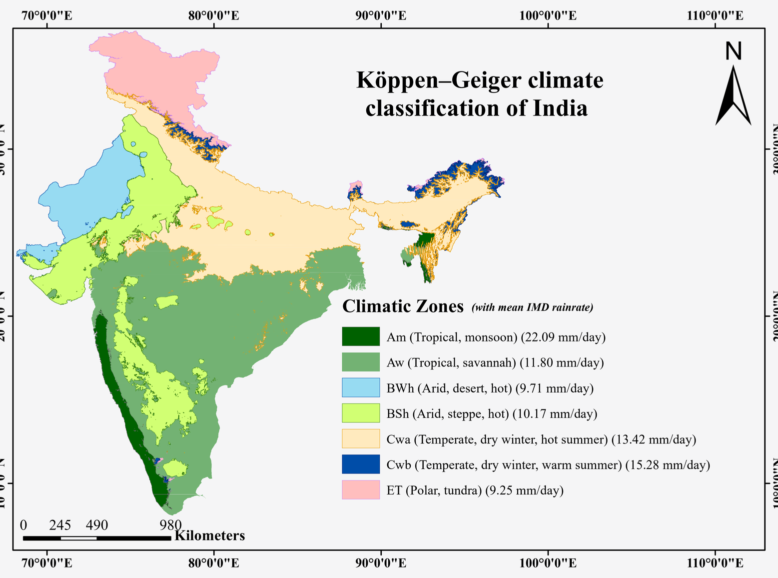
Introduction

- **Challenge**: Global satellite precipitation products like GPM-IMERG are indispensable for climate and hydrological applications, as the ground-based monitoring can be inadequate. However, satellite products exhibit systematic biases whose error structures are dependent on climate regimes (Chen and Wen 2023). A single, uniform correction model may capture only part of this picture, and a model trained for one region's bias pattern may be limited in transferability (Bonaloas et al 2026).
- **Objective**: To introduce a bias correction approach where, rather than training separate models per climate zone (region-specific, accurate but unscalable) or one uniform model ignorant of zone (scalable but bias-blind), we have a learned climate zone embedding. This is a vector representation of the Köppen-Geiger zones, jointly trained with the rest of the network, that allows a single model to internalise zone-specific bias behaviour.
- **Distributional reformulation**: The bias correction is formulated as a distributional regression problem conditioned on terrain, environmental variables, and climate embeddings. By optimising a Gamma-distributed negative log-likelihood (NLL) loss, the network predicts a complete conditional probability distribution rather than a single deterministic value, thereby intending to preserve precipitation variability.
- **Research questions**: Q1. Do the jointly learned climate embeddings organise themselves to reflect physically meaningful differences in how satellite precipitation biases manifest across distinct climates? Q2. Can a single, embedding-augmented network match the skill and regional accuracy of multiple, climate-zone-specific models while remaining a single deployable, scalable entity?

Study Region

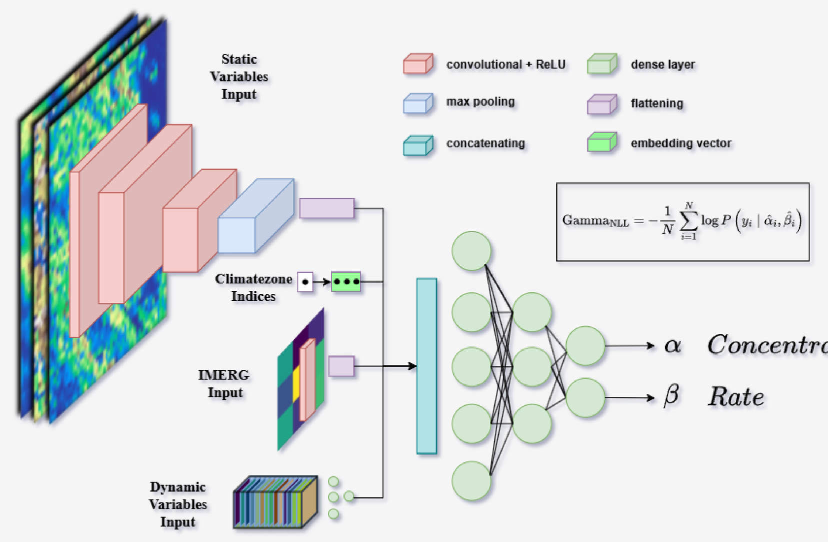


1. Study area



2. Climate zones for study

CNN Model with Embedding

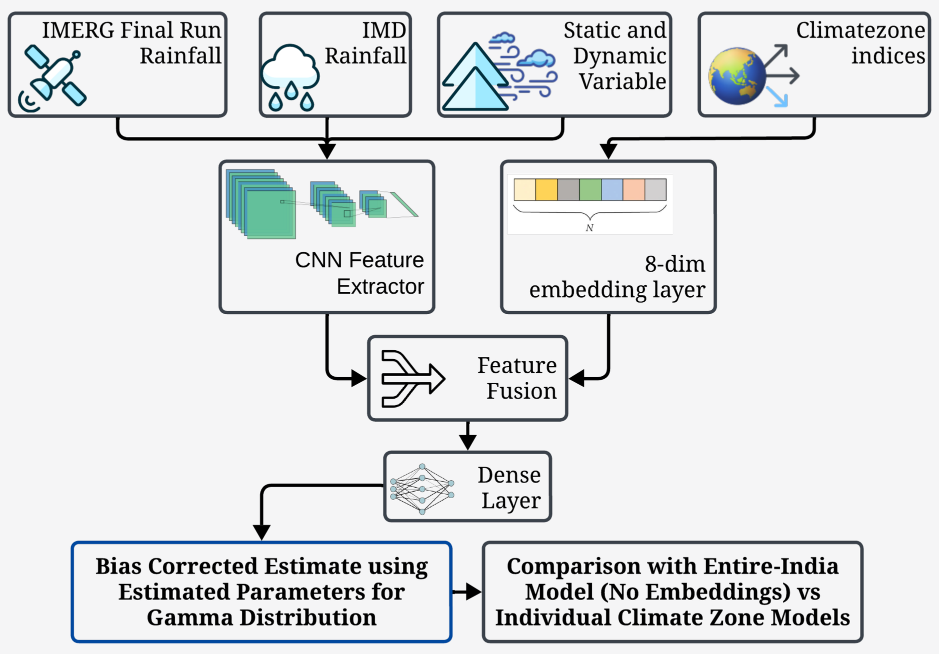


3. Model architecture

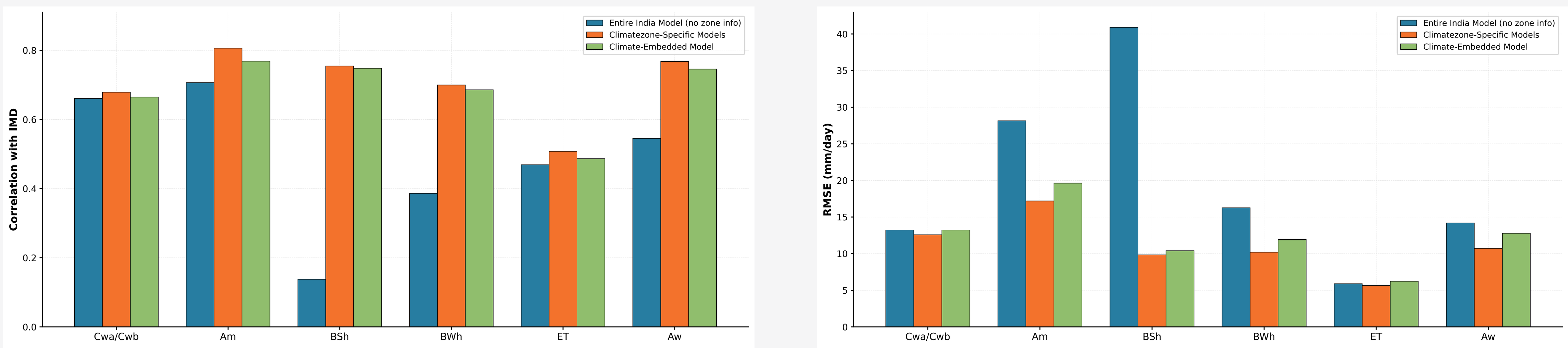
Dataset Used

Data	Resolution
Rainfall 2015-23, JJAS months, Entire India, Koppen-Geiger Zones	
IMERG Final Run V7	0.1° x 0.1°, 30 min data
IMD Gridded Rainfall	0.25° x 0.25°, Daily data
Static Variables	
Elevation	500m (NASA SRTM+)
Aspect	500m
Slope	500m
Dynamic Variables	
Precipitable water	0.25° x 0.25°, Daily data
Convective Inhibition (CIN)	(NCEP GDAS/FNL
Relative humidity	Tropospheric Analysis Grids)
Potential temperature	
Surface temperature	
Wind shear	
Level of the 0° isotherm	
Vertical motion	
Convective Available Potential Energy (CAPE)	
Moisture flux convergence	
Wet bulb temperature	

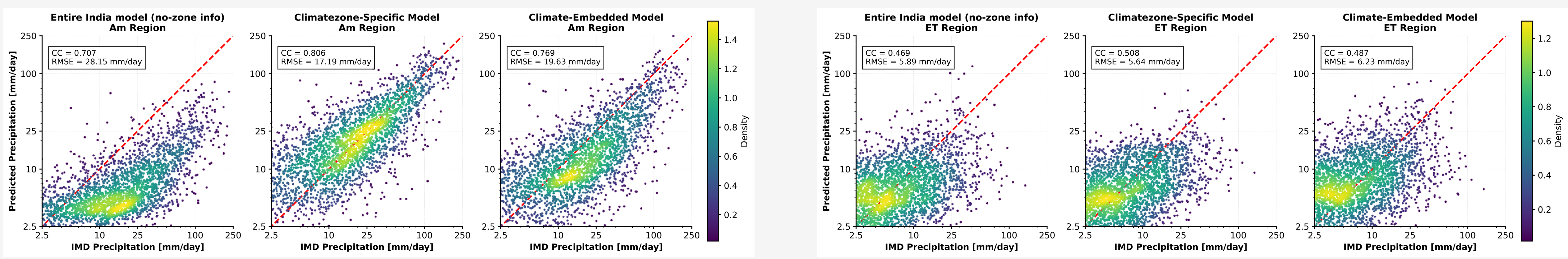
Methodology



Results and Discussion

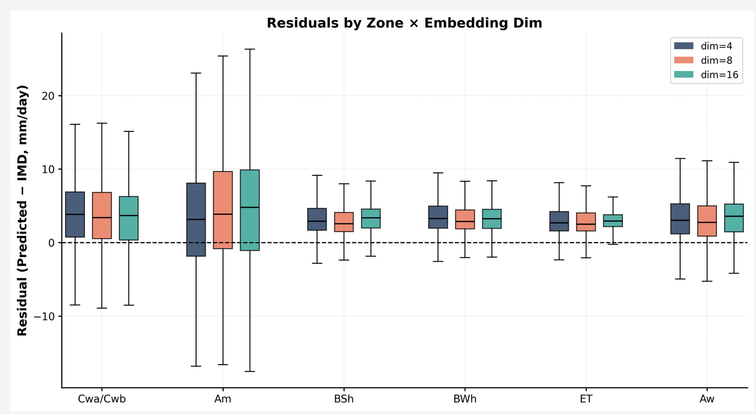


4. Correlation(left) and RMSE(right) of bias-corrected estimation by all models across climate zones



5. Bias-corrected estimations for climate zone of Am(right) and ET(left) by all models

- The climate zone specific models performs best as compared to both, entire-India model(without embedding) validated for that climate zone, and entire-India model with climate aware embeddings (Fig 4 and 5)
- Increasing embedding dimension from 4 to 16 yielded no notable performance change (Fig 6)



6. Performance of different embedding dimensions

Conclusion and Future Study

- A single climate-zone-aware model can substantially close the accuracy gap between a computationally-cheap-but-less-accurate global bias correction model and expensive-but-accurate region-specific models.
- Evaluated across six Köppen-Geiger climate zones in India, the climate-embedded model came close to matching climate-zone-specific-model accuracy; achieves this with one trained model instead of six region-specific models
- Offers a practical, scalable path to deploy climate-aware bias correction operationally, especially relevant for global products where maintaining a per-zone bias correction method can be infeasible
- Future work: investigate how dynamical variables help differentiate climate zones through learned embeddings, and performance will be tested for the climate zone in different regions like US.

Acknowledgements

I would like to thank the India Meteorological Department, NCEP and NASA for their open-source data products. This research is supported by the Greenko Fellowship at IIT Hyderabad. I would like to thank IPWG organisers for the opportunity and the support to attend this workshop.