

A Physically Guided CNN Framework for Indian Monsoon rainfall Downscaling

Dr. Subhrajit Rath¹✉; *Presenting author

Dr. Deepeshkumar Jain¹, Dr. M K Phani¹ and Dr. Suryachandra A. Rao¹,

¹Indian Institute of Tropical Meteorology, Pune,411008, India

Correspondence to: subhrajit.rath@tropmet.res.in



Introduction

- Accurate downscaling of monsoon precipitation over India remains a major challenge due to the coupled influence of large-scale moisture transport, low-level jet dynamics, and orographically induced ascent over the North-East Indian plateau and adjoining regions.
- Higher resolution forecasts can be inferred from coarser resolution data by using downscaling techniques. These techniques are critical for delivering precise information on a local or the regional scales.
- There are two approaches to downscaling the coarser resolution forecast data, viz, statistical, and dynamical downscaling. The dynamical downscaling approach employs a high-resolution model (a regional climate model) that integrates in time the initial and boundary conditions from a GCM on a higher resolution grid, however, is computationally expensive.

Motivation

- In the present study, we train and evaluate a deep learning model using ERA5 reanalysis data to analyze the role of multiple physically relevant predictor variables in downscaling the precipitation.
- We conduct a comprehensive sensitivity analysis in which individual predictors and their combinations are systematically excluded (ablation study) to identify the most effective feature set and to establish objective criteria for selecting downscaling inputs.

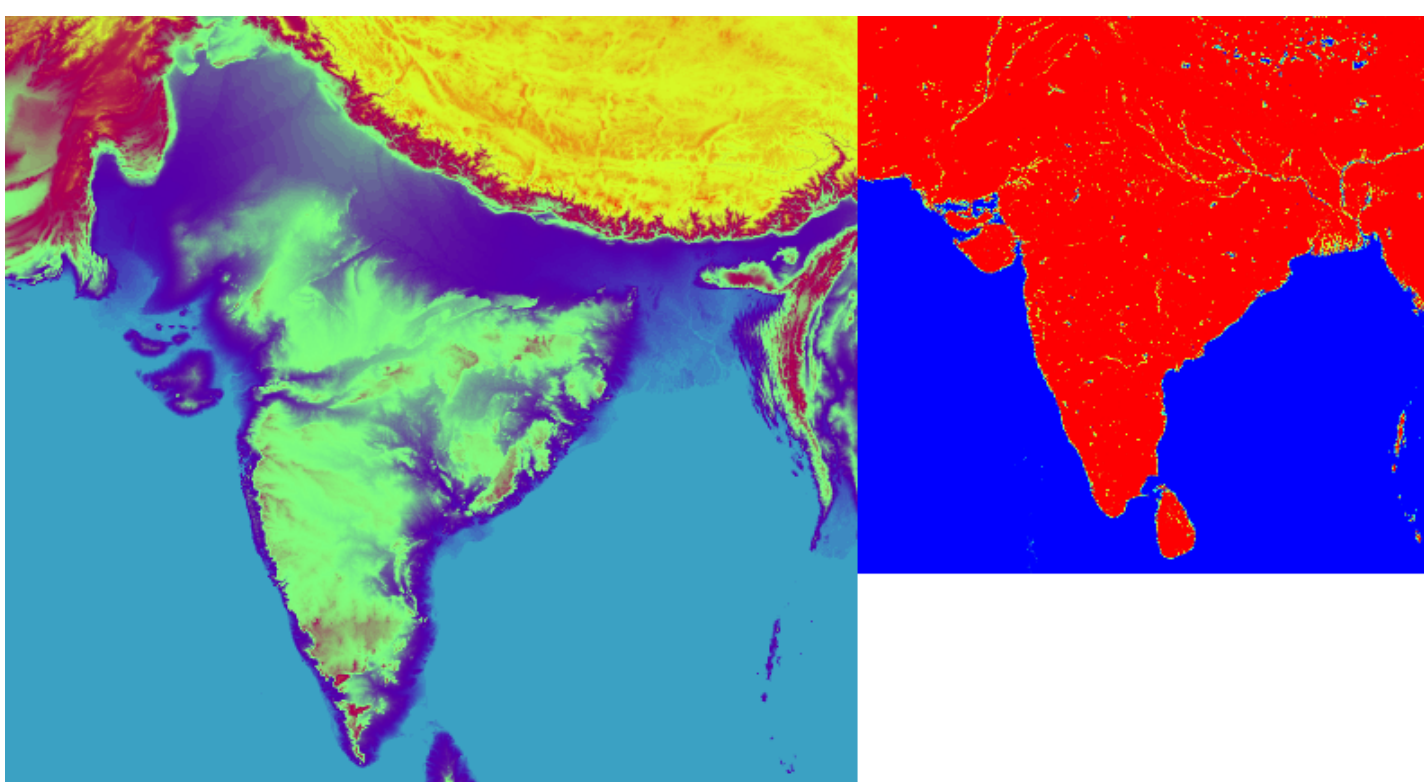
Objective

To use physics-based ML model to downscale precipitation over Indian subcontinent and influence of physical variable.

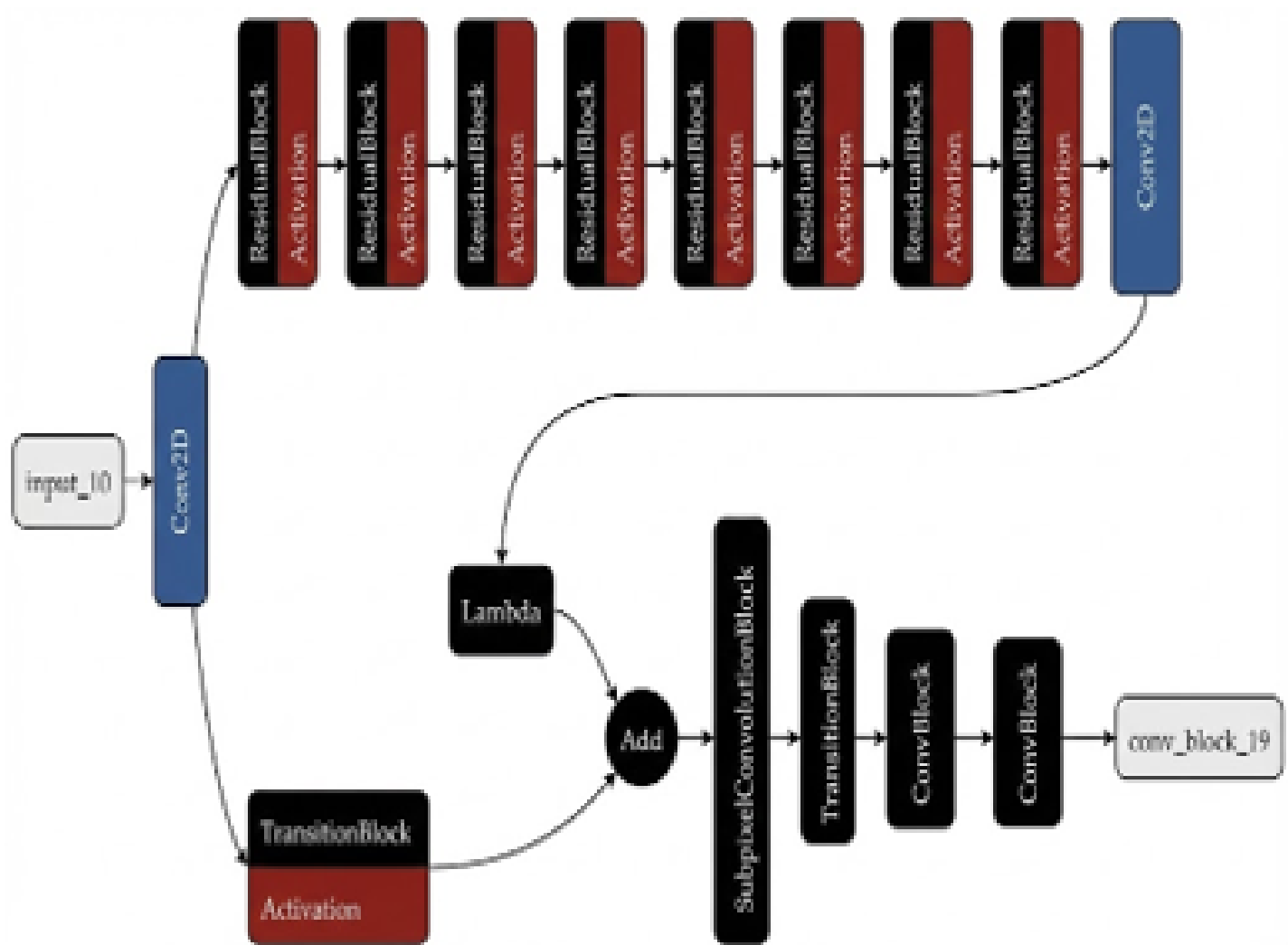
Variable Selection

- The nine variables examined in this study (ERA5), viz D850, D500, D200, GPT500,W500, TCC,VIMD, PWAT as well as T2m.
- These collectively span the four classical ingredients for deep moist convection and stratiform precipitation identified as : (i) a deep, moist layer; (ii) a lifted parcel mechanism; (iii) convective instability; and (iv) a triggering mechanism.
- These collectively span the four classical ingredients for deep moist convection and stratiform precipitation identified as : (i) a deep, moist layer; (ii) a lifted parcel mechanism; (iii) convective instability; and (iv) a triggering mechanism.
- ERA5 at 25 km was selected as the coarse-resolution input because it provides large-scale atmospheric information that is physically relevant to precipitation variability, whereas GPM at 10 km was used as the high-resolution precipitation reference.

Study Area

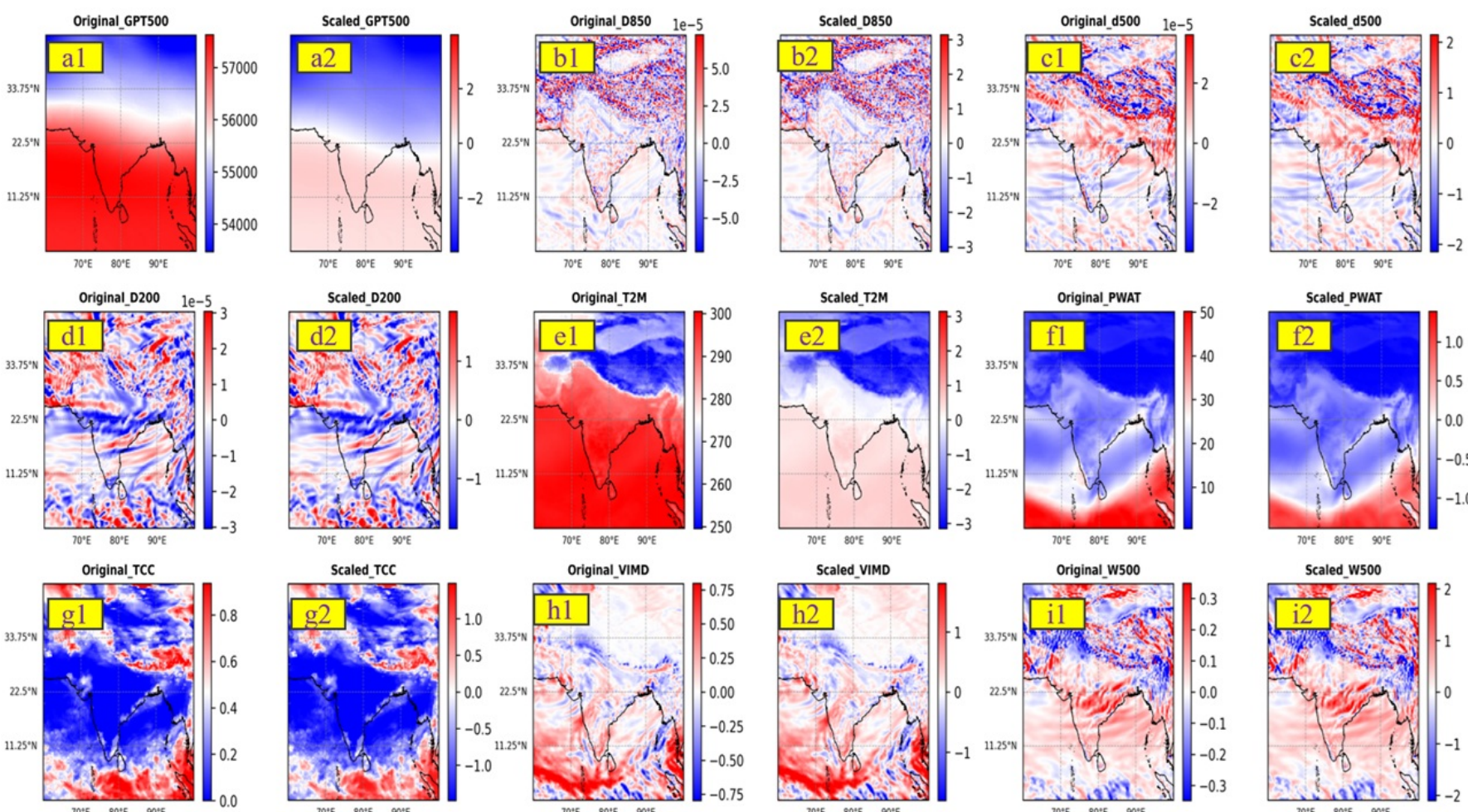


CNN architecture and Feature ablation



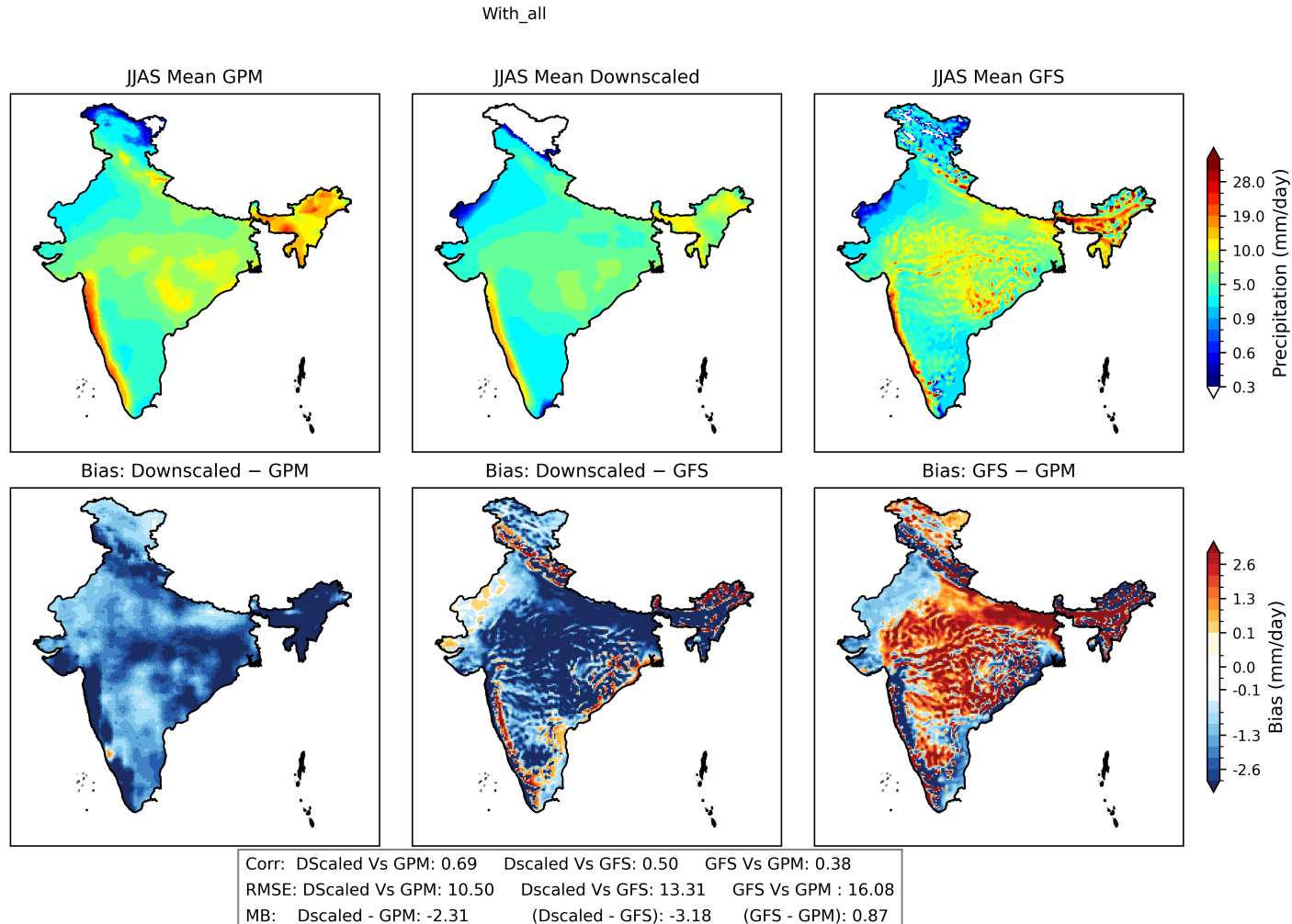
- The baseline model used all input variables ("With all").
- Subsequent experiments systematically excluded one predictor at a time (denoted "X") or multiple predictors together (denoted "X_Y") to evaluate their individual and joint influence on model performance.
- The entire time series was divided into training, validation and testing segments respectively (Segment 1: January 1, 2000 - December 31, 2015, Segment 2: January 1, 2016 - December 31, 2016, and Segment 3: January 1, 2017 - December 31, 2019).

Data normalisation

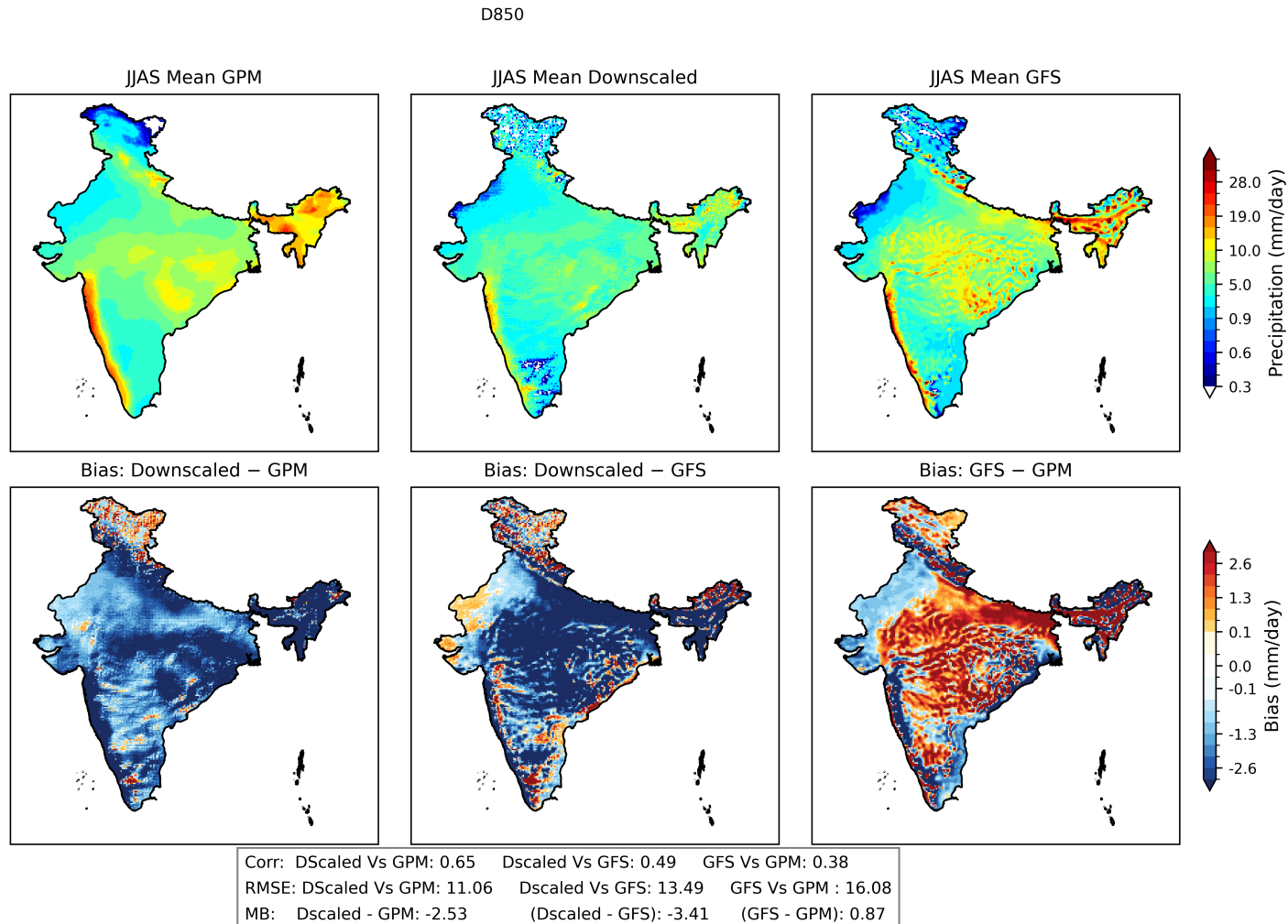


Data-Normalisation. Spatial distribution of the predictor variables before (left) and after (right) feature scaling for a representative test sample. Panels (a1–i1) show the original predictor fields, while panels (a2–i2) show their normalized counterparts used as inputs to the neural network. The predictor set comprises geopotential height, divergence at 850, 500, and 200 hPa, 2-m air temperature, total column water vapour, total cloud cover, moisture divergence, and vertical velocity. For each predictor, identical contour levels based on the 0.5th and 99.5th percentile values of the combined original and scaled datasets are used to facilitate a direct comparison of the effects of feature scaling

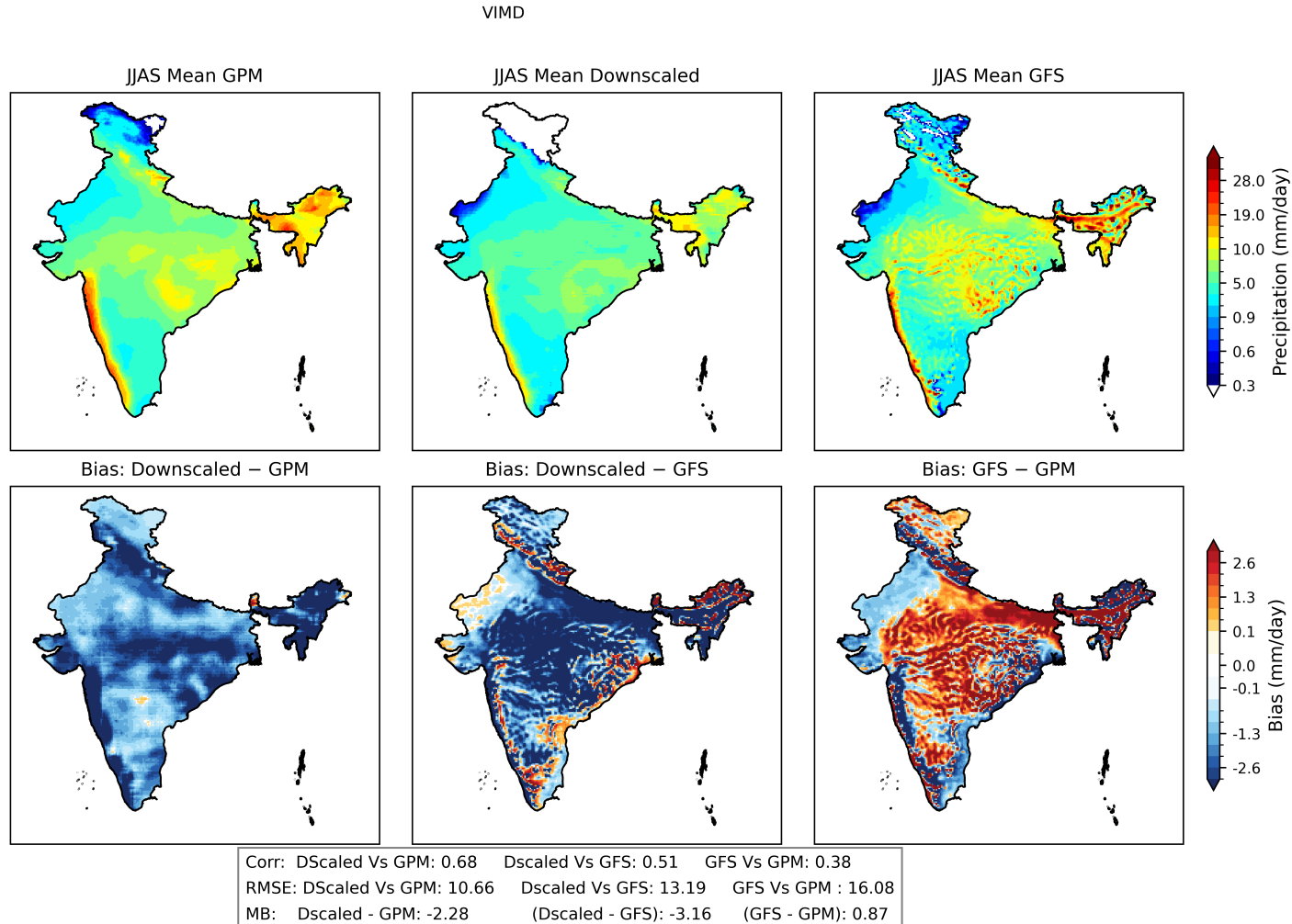
With ALL variable



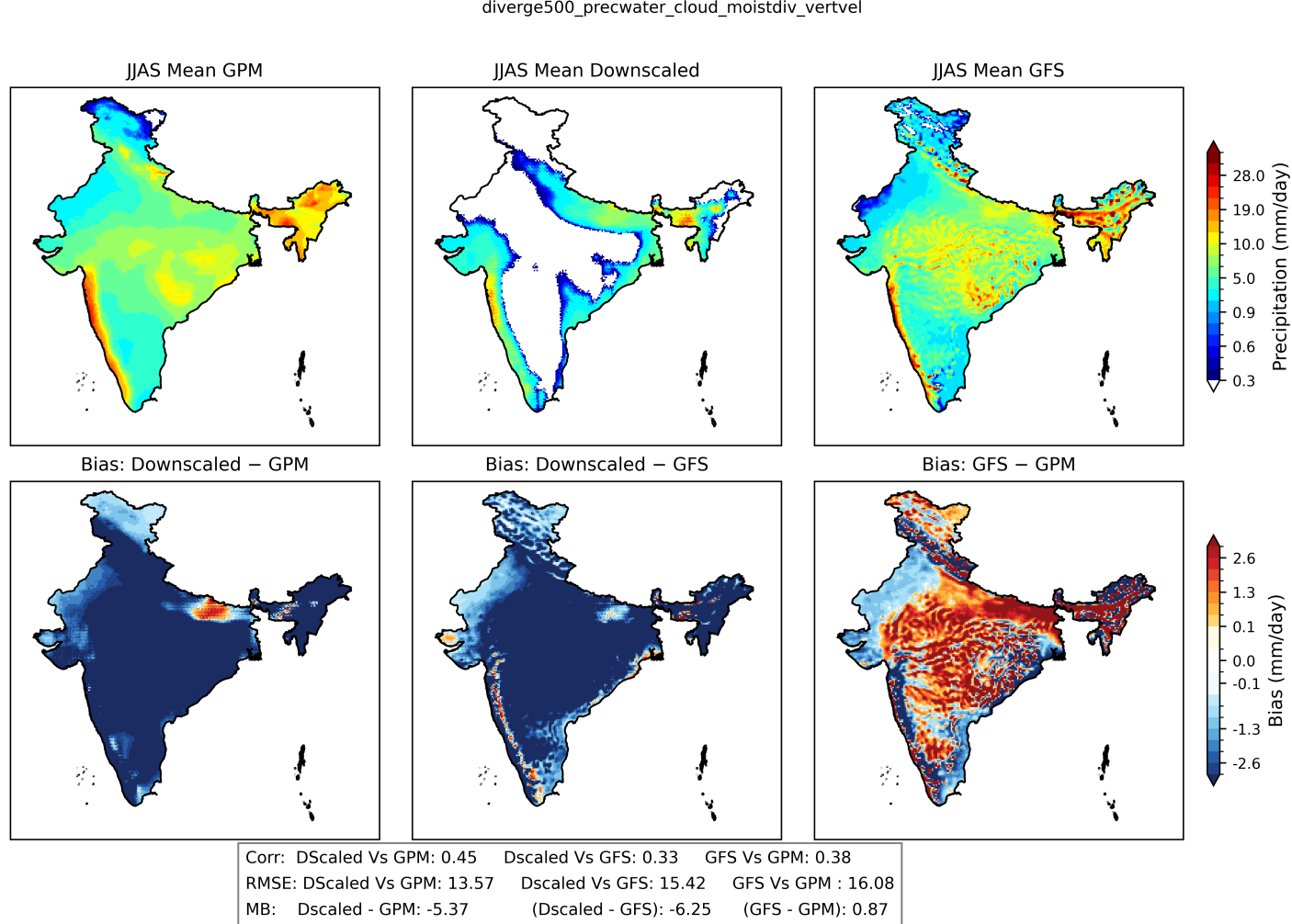
Without D850 variable



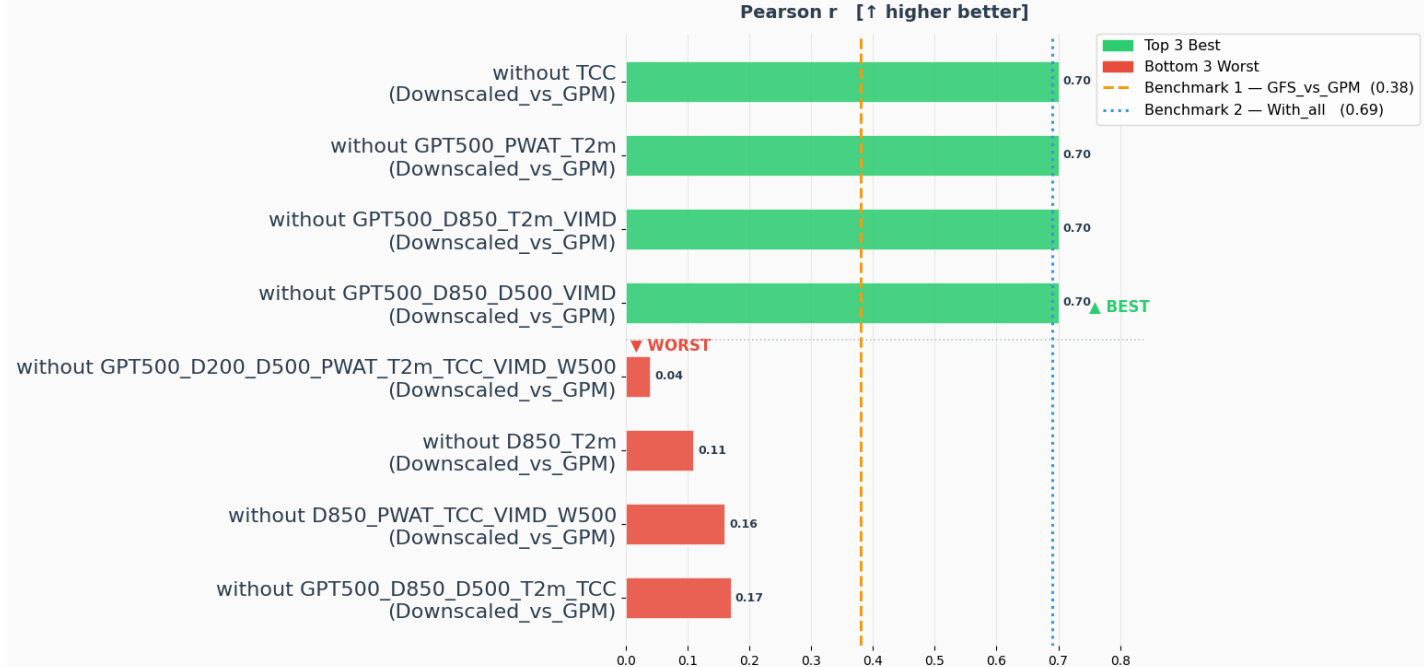
Without VIMD variable



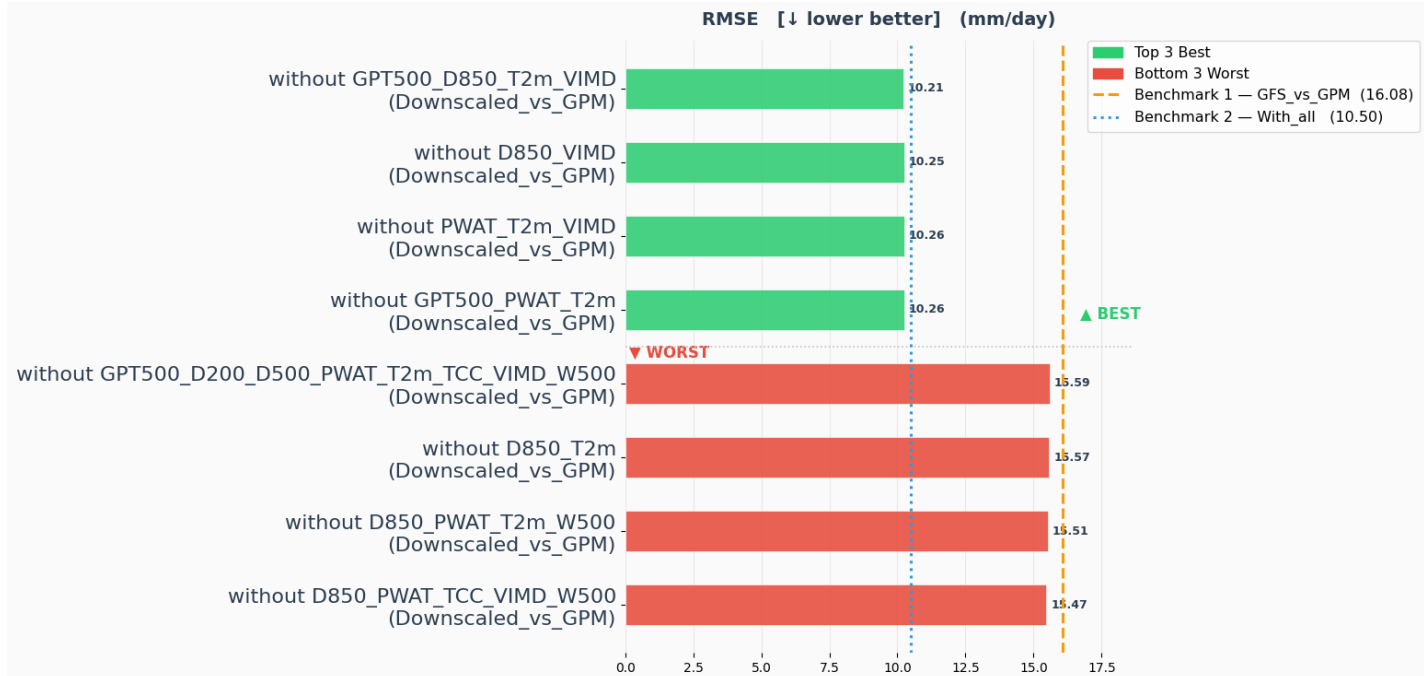
Without D500 PWAT TCC VIMD W500 variable



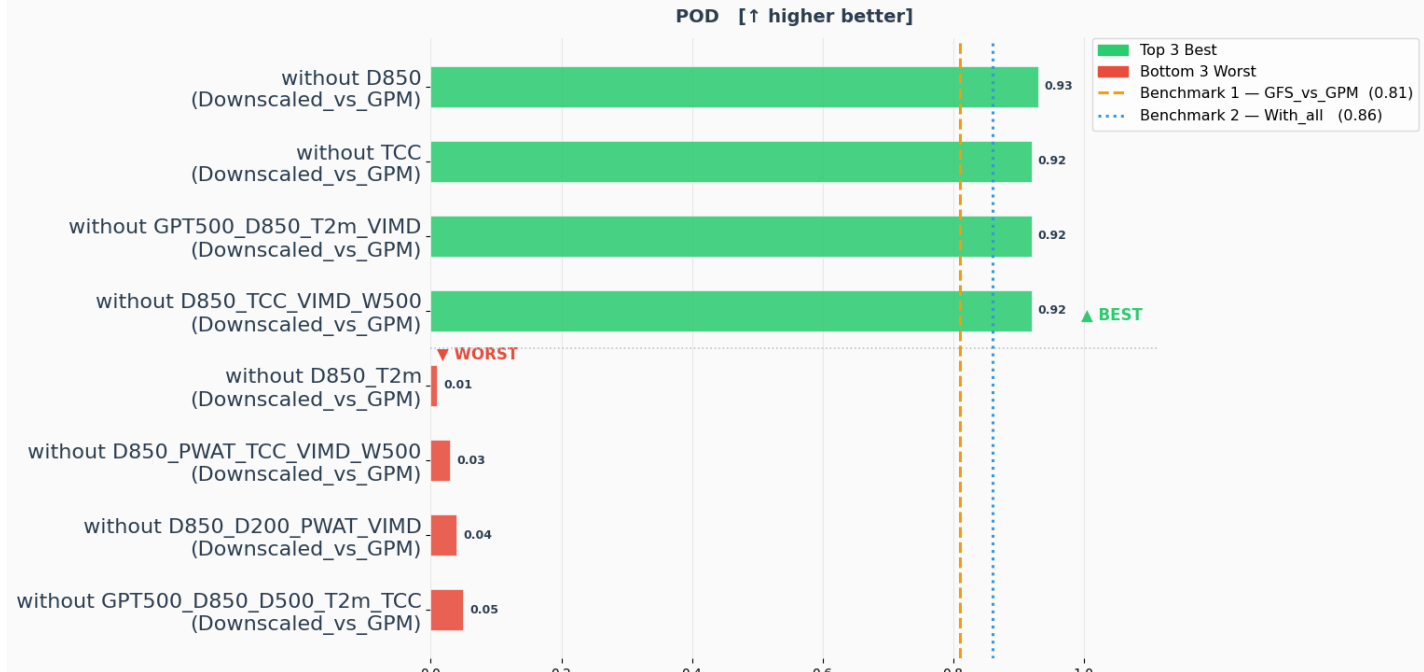
Pearson's r



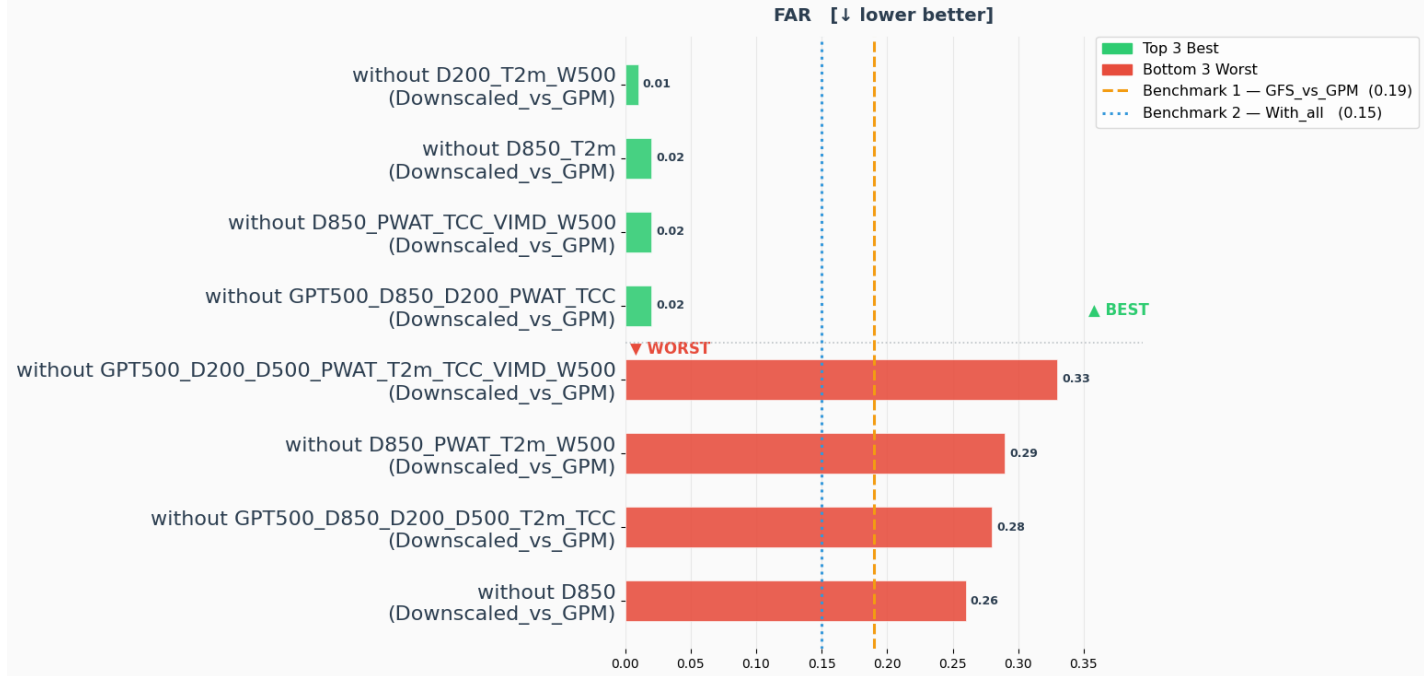
RMSE



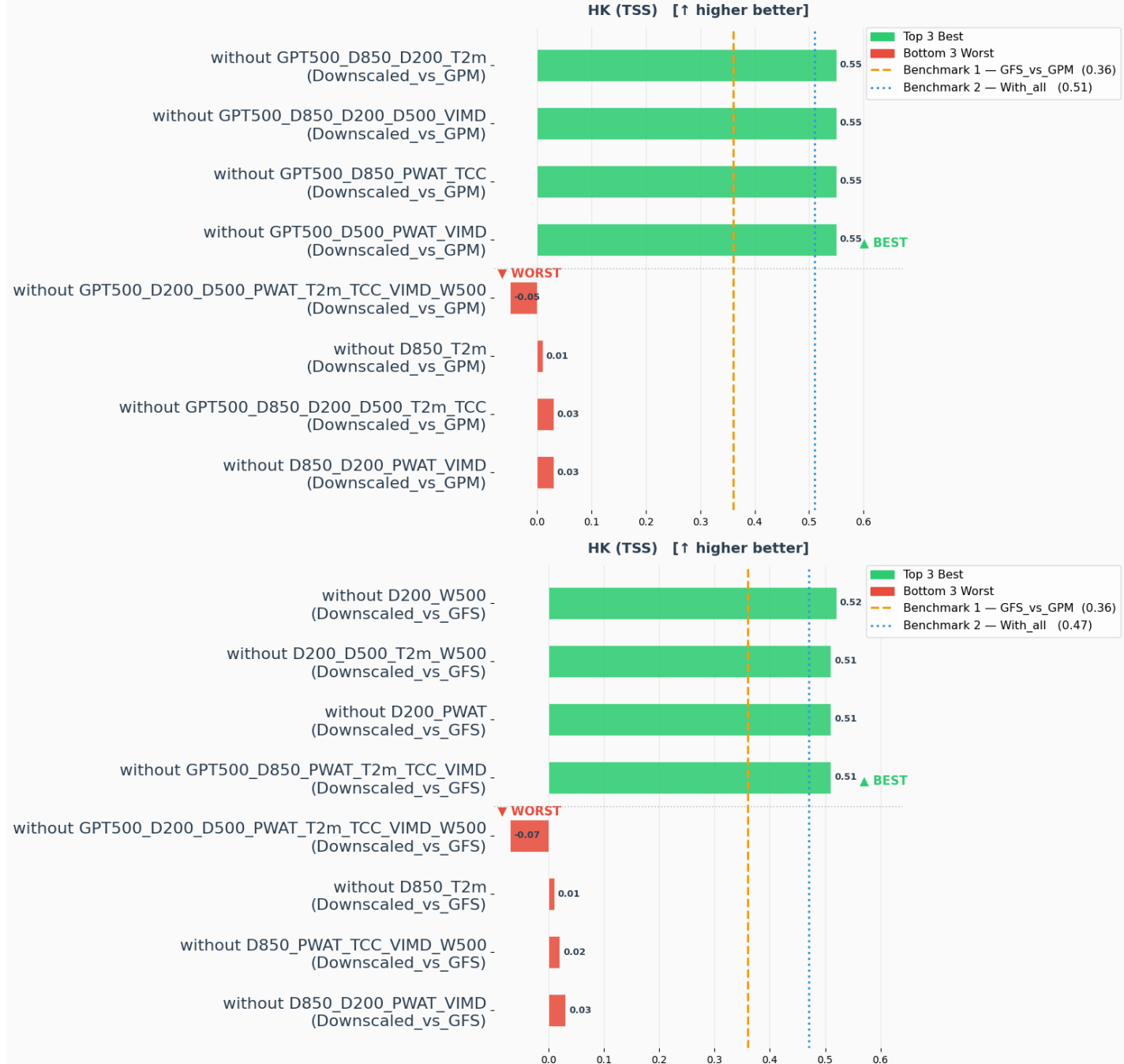
POD



FAR



TSS



Conclusion

- Across both continuous and categorical analyses, a consistent physical interpretation emerges.
- D850 and T2m form the primary dynamical and thermodynamic controls on precipitation.
- Their combined presence allows the model to represent the coupling between moisture convergence and boundary-layer processes that governs rainfall occurrence and intensity.
- W500 contributes to vertical motion and supports the representation of deep convection, while VIMD and PWAT provide additional information on moisture availability.

Acknowledgement

The authors acknowledge Dr. Suryachandra A. Rao, Director, IITM for their support and cooperation. The ERA5 dataset provided by ECMWF (<https://cds.climate.copernicus.eu/cdsapp#!dataset/>) was used to initialize thermodynamical parameters. The authors also thank IITM for providing Peta-scale Computing Resources and simulation support.