Supplementary Information

Paper Title: Spatial assessment of the reproducibility of Indian summer monsoon rainfall regimes in multiple gridded rainfall products

Authors: Sandipan Paul[1], Priyank J. Sharma[2]*, Ramesh S.V. Teegavarapu[3]

[1]Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai- 400076, Maharashtra, India.
[2]Department of Civil Engineering, Indian Institute of Technology Indore, Madhya Pradesh 453552, India.
[3]Department of Civil, Environmental, and Geomatics Engineering, Florida Atlantic University, Boca Raton 33431, FL, United States.
**Corresponding author, email: priyanksharma@iiti.ac.in*

The supplementary information contains

Two Supplementary Texts (S1 and S2)

Twelve Supplementary Figures (Fig. S1 – Fig. S12)

DOI: 10.1038/s41598-024-75320-5

Supplementary Texts

Text S1: Mathematical Formulation CPM Indices

Composite Performance Measure (CPM) is a variant of Frequency-based Performance Measure (FBPM; defined in Eq. (1)) which incorporates the penalizing indices $\left(\psi_{k,j}\right)$ and is defined in Eq. (2). We used five ($j = 1, 2, \cdots, 5$) CPM indices: (a) variance index (VI) (b) interquartile range index (IQRI), (c) correlation index (CI), (d) bounded relative absolute error index (BRAEI) and (d) Kolmogorov-Smirnov index (KSI), which are defined as the ratio of variance, interquartile range, Pearson's correlation coefficient, absolute differences incorporating the variable value from one step naïve model and bin specific absolute maximum difference between cumulative distribution function (CDF) obtained from time matched observed precipitation in other datasets and IMD for each bin, respectively (cf. Teegavarapu et al., 2022). To readily compare among the indices, indices are normalised by introducing a transformation function $\left(G_q\left(\psi_{k,q}\right)\right)$ in CPM and bounded between 0 and 1. The used CPM indices and normalising procedure are defined in the Eq. (S1.1) to (S1.5).

$$\psi_{k,j} = VI_k = \frac{\left(\sigma_{y^q_{k,t}}\right)^2}{\left(\sigma_{y^{true}_k}\right)^2} \quad \forall k, j = 1\ ; \ G_j(VI_k) = \begin{cases} 1 - VI_k & if\ VI_k \le 1 \\ \left|1 - \dfrac{1}{VI_k}\right| & if\ VI_k > 1 \end{cases} \tag{S1.1}$$

$$\psi_{k,j} = IQRI_k = \frac{\delta_{y^q_{k,t}}}{\delta_{y^{true}_k}} \quad \forall k, j = 2\ ; \ G_j(IQRI_k) = \begin{cases} 1 - IQRI_k & if\ IQRI_k \le 1 \\ \left|1 - \frac{1}{IQRI_k}\right| & if\ IQRI_k > 1 \end{cases} \tag{S1.2}$$

$$\psi_{k,j} = CI_k = \frac{\rho_{y^q_{k,t}}}{\rho_{y^{true}_k}} \quad \forall k, j = 3\ ; \ G_j(CI_k[0,1]) = \frac{1}{2} CI_k[-1,1] + \frac{1}{2} \tag{S1.3}$$

$$\psi_{k,j} = BRAEI_k = \frac{1}{\phi_k} \sum_{p=1}^{\phi_k} \frac{\left|y^q_{p,k,t} - y^{true}_{p,k}\right|}{\left|y^q_{p,k,t} - y^{true}_{p,k}\right| + \left|e^q_{p,k,t} - y^{true}_{p,k}\right|} \quad \forall k, j = 4$$

$$G_j(BRAEI_k) = 1 - BRAEI_k \tag{S1.4}$$

$$\psi_{k,j} = KSI_k = \max_k\left(\left|\mathrm{F}_{Y^{true}_{p,k}}\left(\max\left(y^{true}_{p,k}\right)\right) - \mathrm{F}_{Y^{obs}_{p,k,t}}\left(\max\left(y^q_{p,k,t}\right)\right)\right|\right) \quad \forall k, j = 5$$

$$G_j(KSI_k) = 1 - KSI_k \tag{S1.5}$$

where $G_j\left(\psi_{k,j}\right)$ is the transformation function employed on j^{th} CPM indices in k^{th} bin; σ, δ, and ρ are the standard deviation, interquartile range, and Pearson's correlation coefficient of the corresponding suffix variables, respectively; ϕ_k is the total number of time index-matched observed rainfall in other datasets with respect to base dataset in k^{th} bin; $e^q_{p,k,t}$ is obtained from a one-step naïve model ($X_t = X_{t-1}$); $Y^{true}_{p,k}$ and $Y^{obs}_{p,k,t}$ are the random variables of $y^{true}_{p,k}$ and $y^{obs}_{p,k,t}$, respectively; F is the value CDF of the suffix variable at the maximum observed value of the variable in k^{th} bin.

Text S2: Mathematical Formulation of the Continuous and Categorical Performance Measures

To assess and investigate the ability of precipitation datasets with respect to IMD in capturing precipitation variability, we evaluated their performance using six continuous and two categorical metrics. The adopted six continuous measures are: random RMSE, systematic RMSE, normalised RMSE, percentage bias, correlation coefficient, degree of agreement and two categorical measures are: critical success index and false alarm ratio of q^{th} dataset evaluated with respect to IMD (designated as '*truth*') are defined in Eq. (S1.6)- (S1.13).

$$RMSE_{Rand}^{q} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i^q - \hat{y}_i^q\right)^2} \tag{S1.6}$$

$$RMSE_{Syst}^{q} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i^{true} - \hat{y}_i^q\right)^2} \tag{S1.7}$$

$$NRMSE^{q} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i^{true} - y_i^q\right)^2}}{\bar{y}^{true}} \tag{S1.8}$$

$$P_{Bias}^{q} = \frac{\sum_{i=1}^{n}(y_i^{true} - y_i^q)}{\sum_{i=1}^{n} y_i^{true}} \times 100 \tag{S1.9}$$

$$\tau = \frac{\alpha^q - \beta^q}{\alpha^q + \beta^q} \tag{S1.10}$$

$$D^{q} = 1 - \frac{\sum_{i=1}^{n}\left(y_i^q - y_i^{true}\right)^2}{\sum_{i=1}^{n}\left(\left|y_i^{true} - \bar{y}^{true}\right| + \left|y_i^q - \bar{y}^{true}\right|\right)^2} \tag{S1.11}$$

$$CSI^{q} = \frac{H}{H + M + F} \tag{S1.12}$$

$$FAR^{q} = \frac{F}{H + F} \tag{S1.13}$$

where y_i^{true} and y_i^qrepresents rainfall values (mm/day) in the base and q^{th}dataset at i^{th} time step; $\hat{y}^{obs} = \beta_1 \times y_{true} + \beta_2$; n denotes valid $\left(y_i^{true}, y_i^q\right)$ pair where either one is non-NaN; α^q and β^q indicate the concordant and discordant pairs in q^{th} dataset with respect to IMD; H represents the frequency of precipitation events occurrence in observed both in IMD and q^{th} datasets; M represents the number of precipitation events observed in IMD but absent in q^{th} datasets; F represents the number of precipitation events observed in q^{th} datasets but absent in IMD. Daily precipitation and no-precipitation events are identified by the threshold of 2.5 mm/day instead of 0 mm/day to eradicate the effect of very light drizzle (Bhatla et al., 2016; Sonar, 2014).

Supplementary Figures

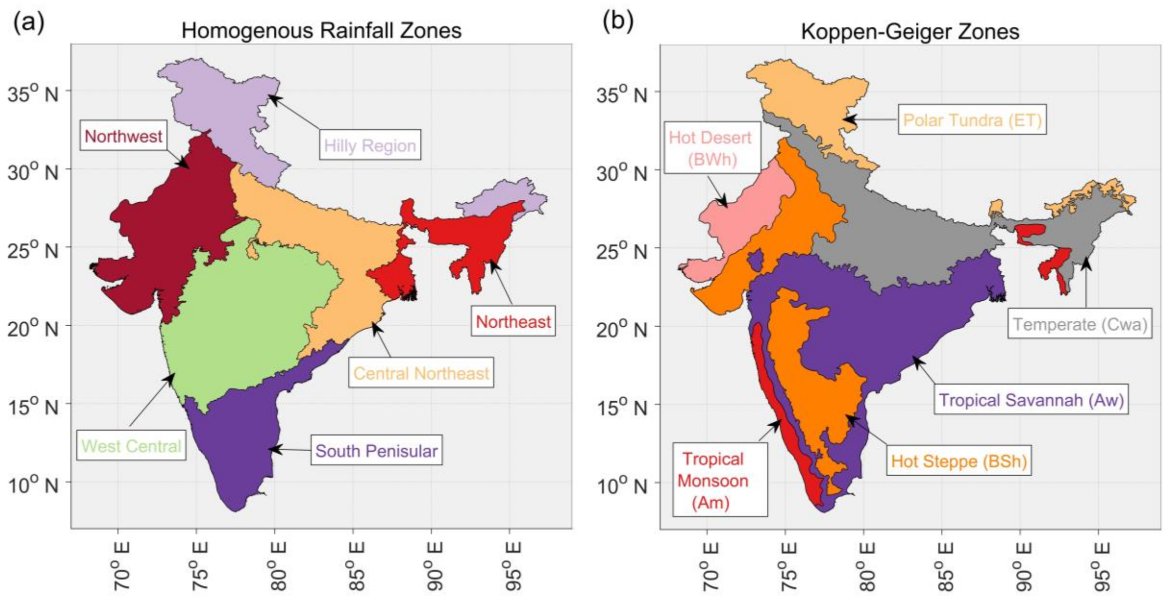

***Fig. S1**: Spatial location of homogenous rainfall regions (a) (Parthasarathy, 1995) and Koppen-Gieger climate zones (b) (Beck et al., 2018).*

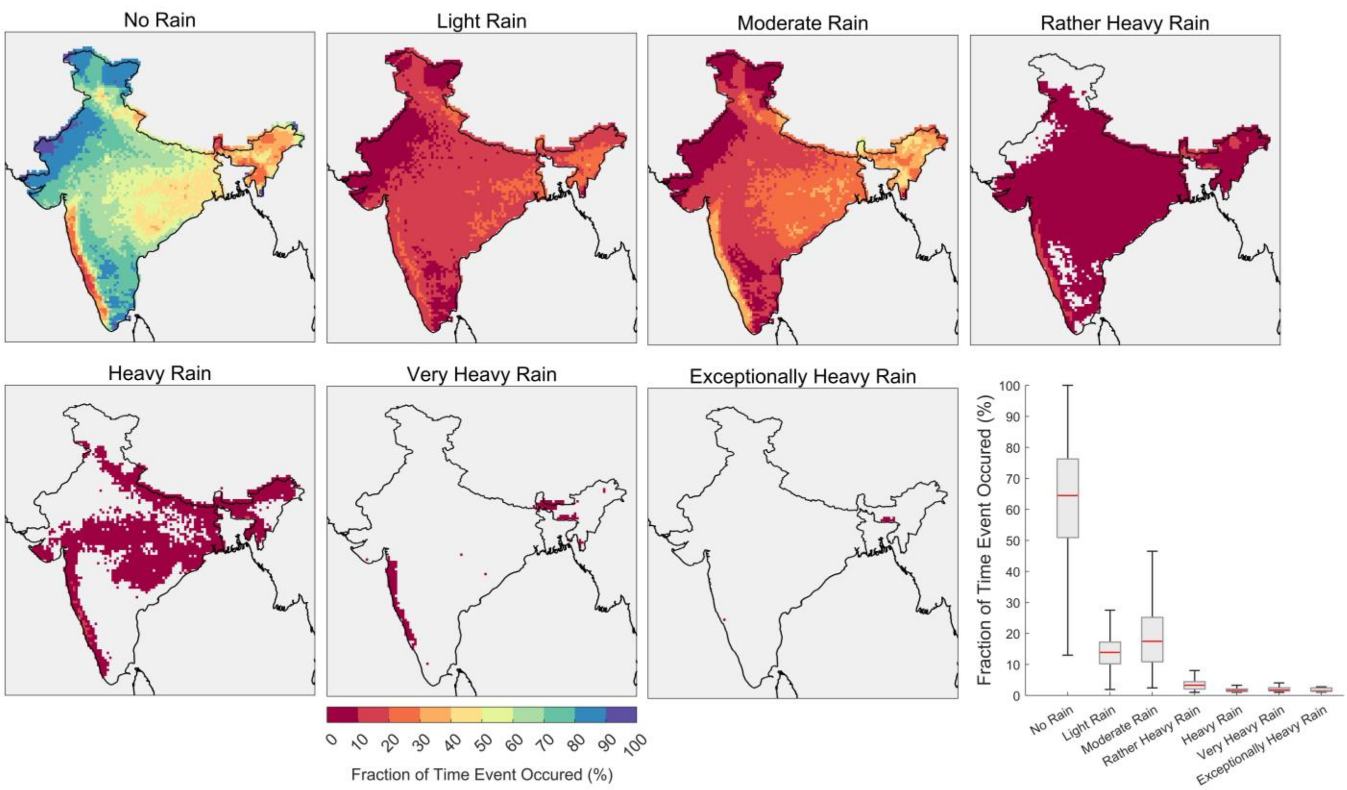

***Fig. S2**: Spatial distribution of temporal frequency of seven characteristics rainfall events, which are (1) no-rain or drizzle (< 2.5 mm/day), (2) light rain (2.5-7.5 mm/day), (3) moderate rain (7.5-36 mm/day), (4) rather heavy rain (36-65 mm/day), (5) heavy rain (65-125 mm/day), (6) very heavy rain (125-245 mm/day), (7) exceptionally heavy rain (≥ 245 mm/day) (Bhatla et al., 2016; Sonar, 2014) derived from IMD data. Event frequency indicates the ratio of number of times a rainfall event occurred in daily scale to the total observation* *days used in this study (i.e., 32 years × 122 days in each year during JJAS months). Saptial variability of the respective seven rainfall events (bottom right pannel).*

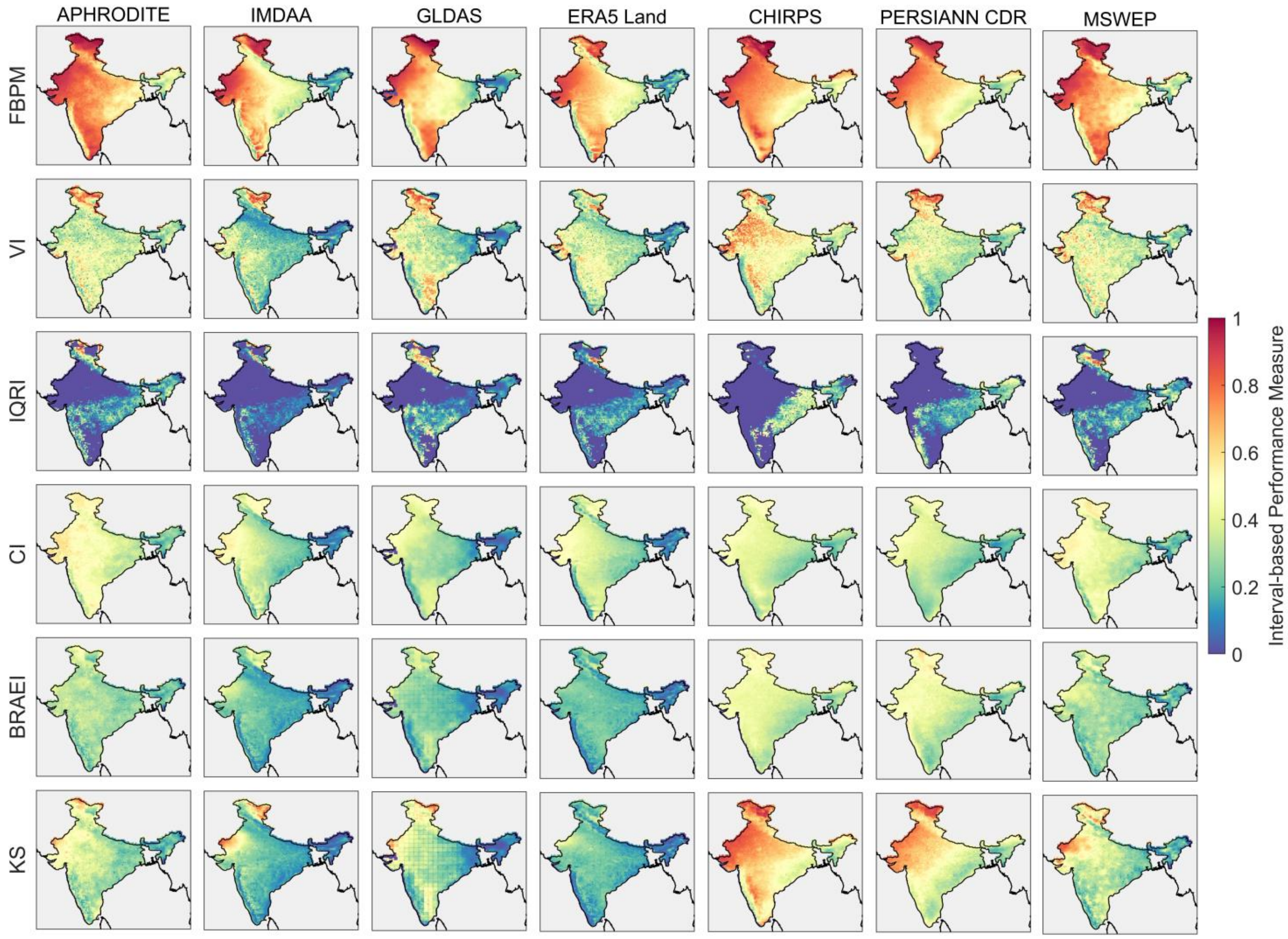

***Fig. S3**: Spatial distribution of six (frequency-based performance measure (FBPM), variance index (VI), interquartile range index (IQRI), correlation index (CI), (d) bounded relative absolute error index (BRAEI) and Kolmogorov-Smirnov index (KSI)) interval based performance measures (IBPMs) of seven examined dataset (indicated by columns) with respect to IMD (indicated in each row) for no -rain or drizzle event.*

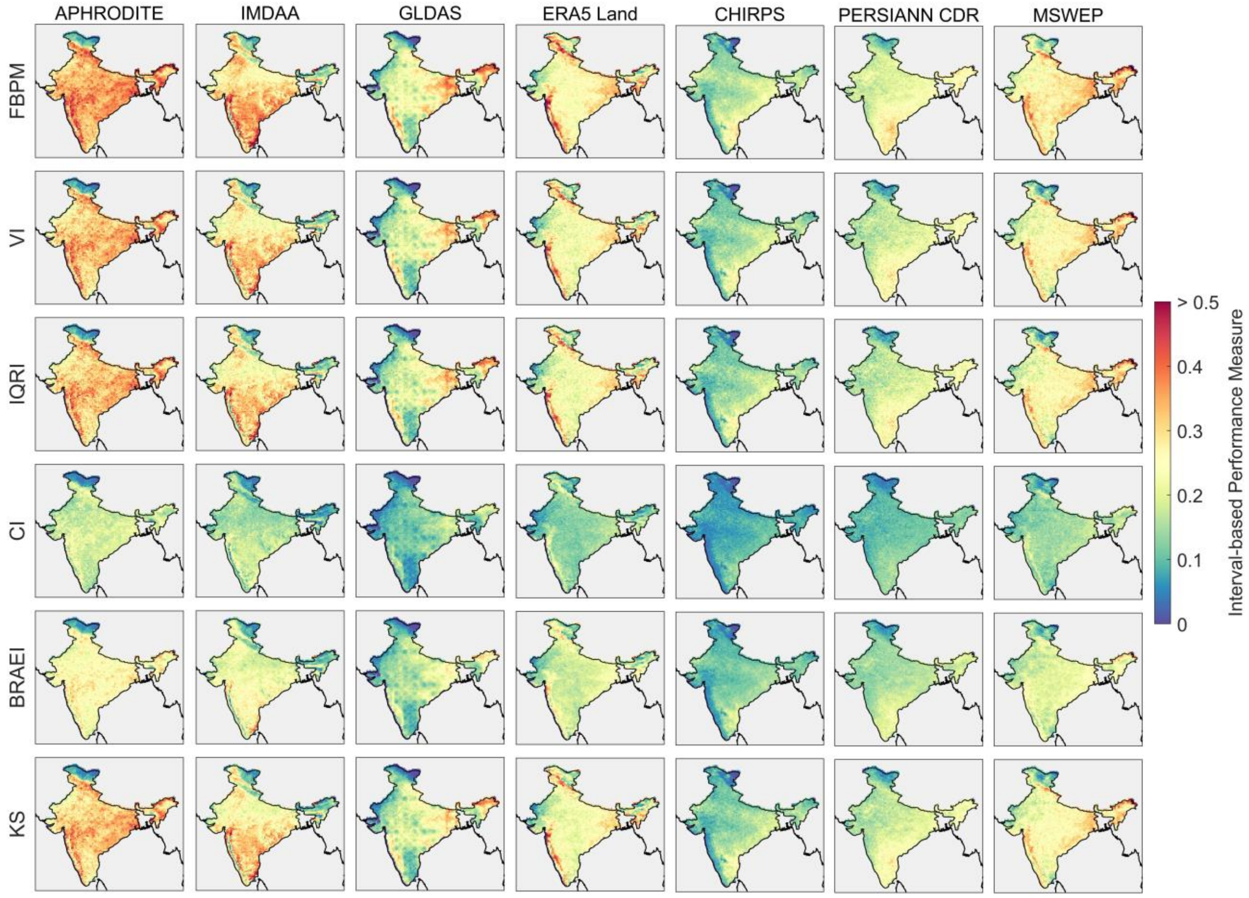

Fig. S4: *Same as Fig. S3, but for little rain events*

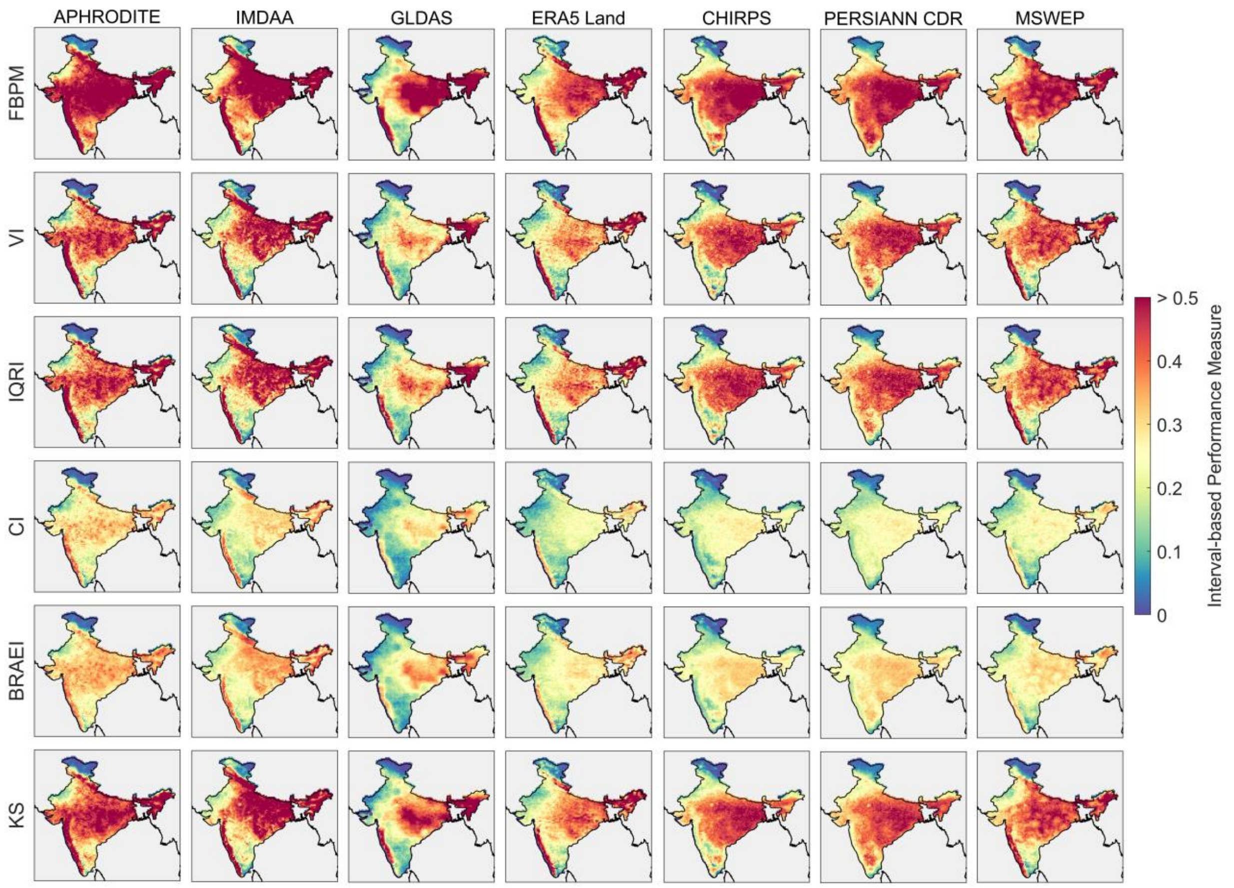

Fig. S5: Same as Fig. S3, but for rather moderate rain events

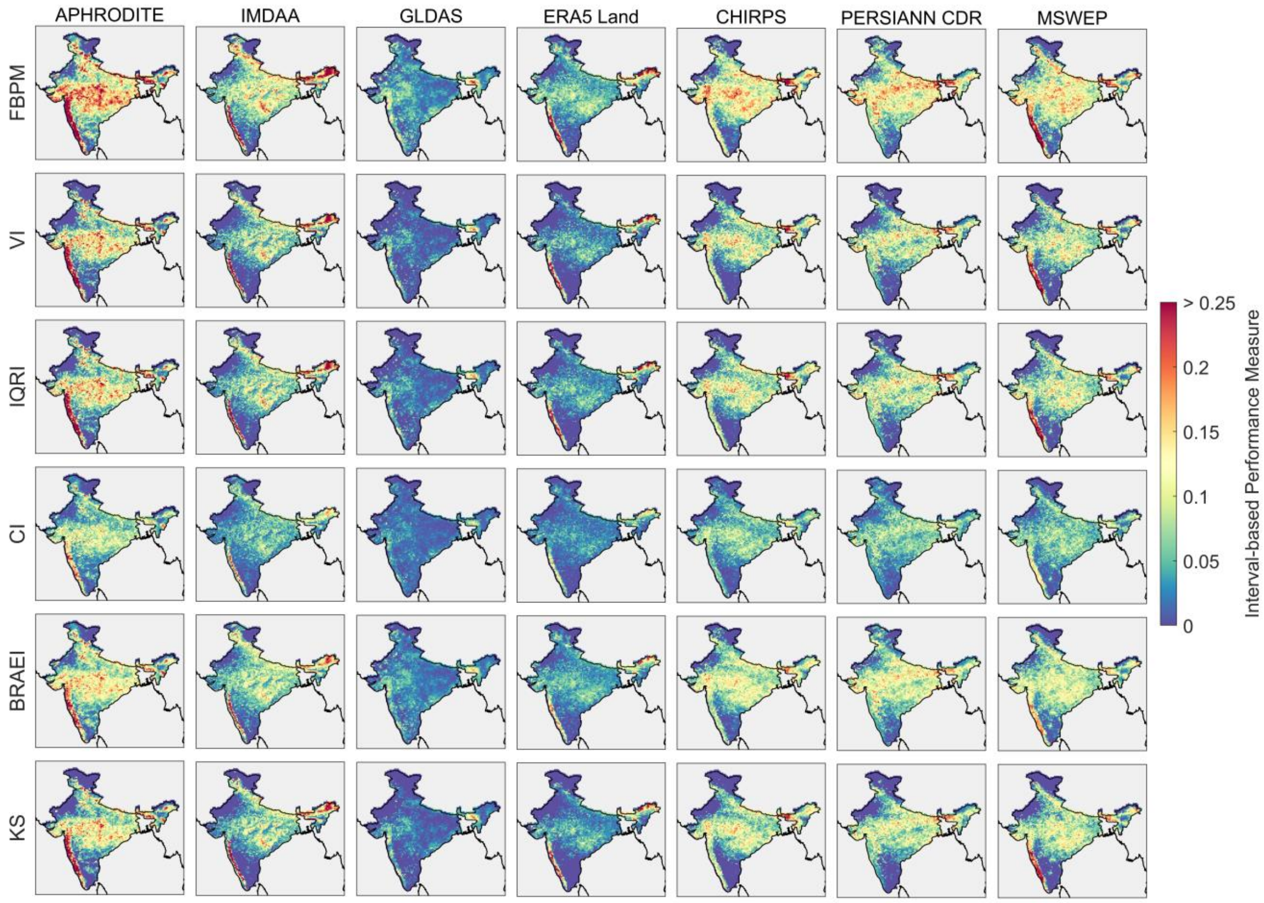

Fig. S6: Same as Fig. S3, but for rather heavy rain events

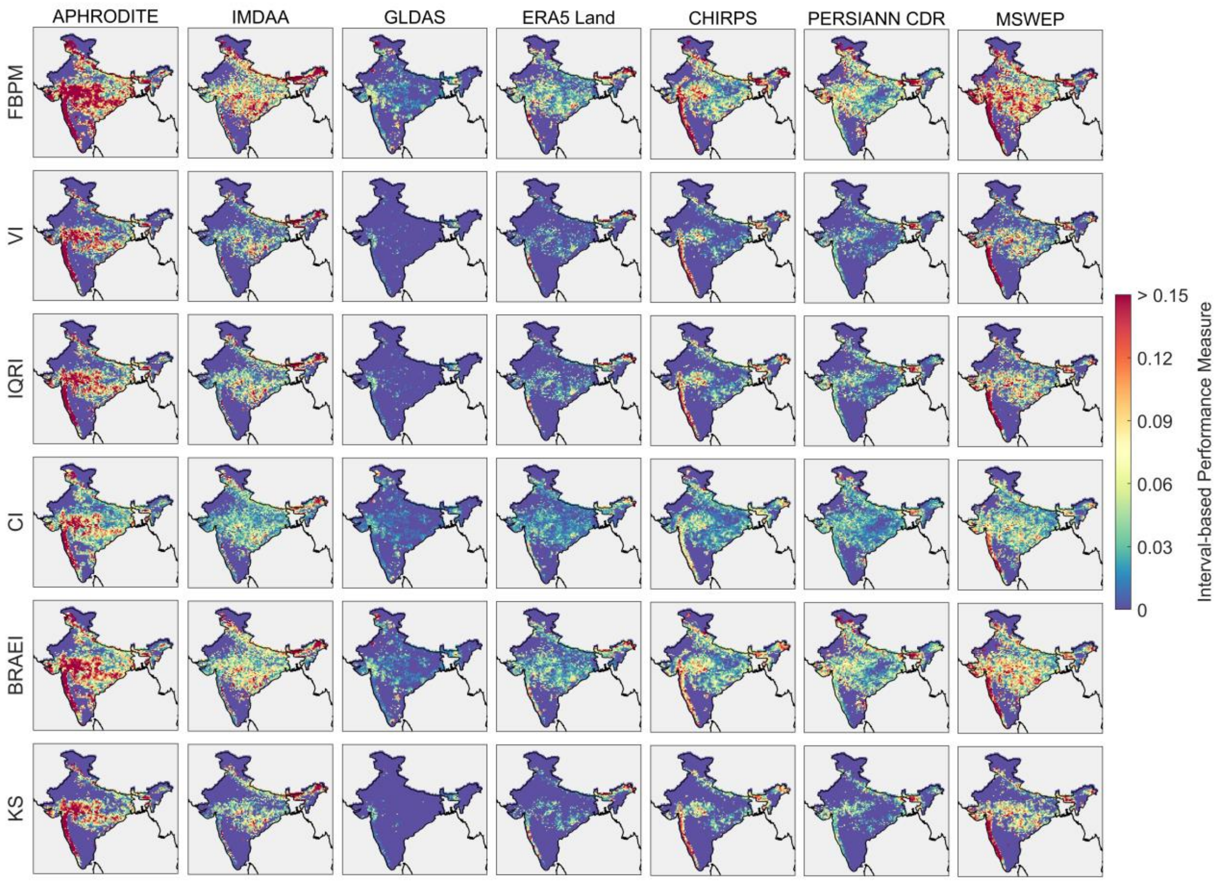

Fig. S7: Same as Fig. S3, but for heavy rain events

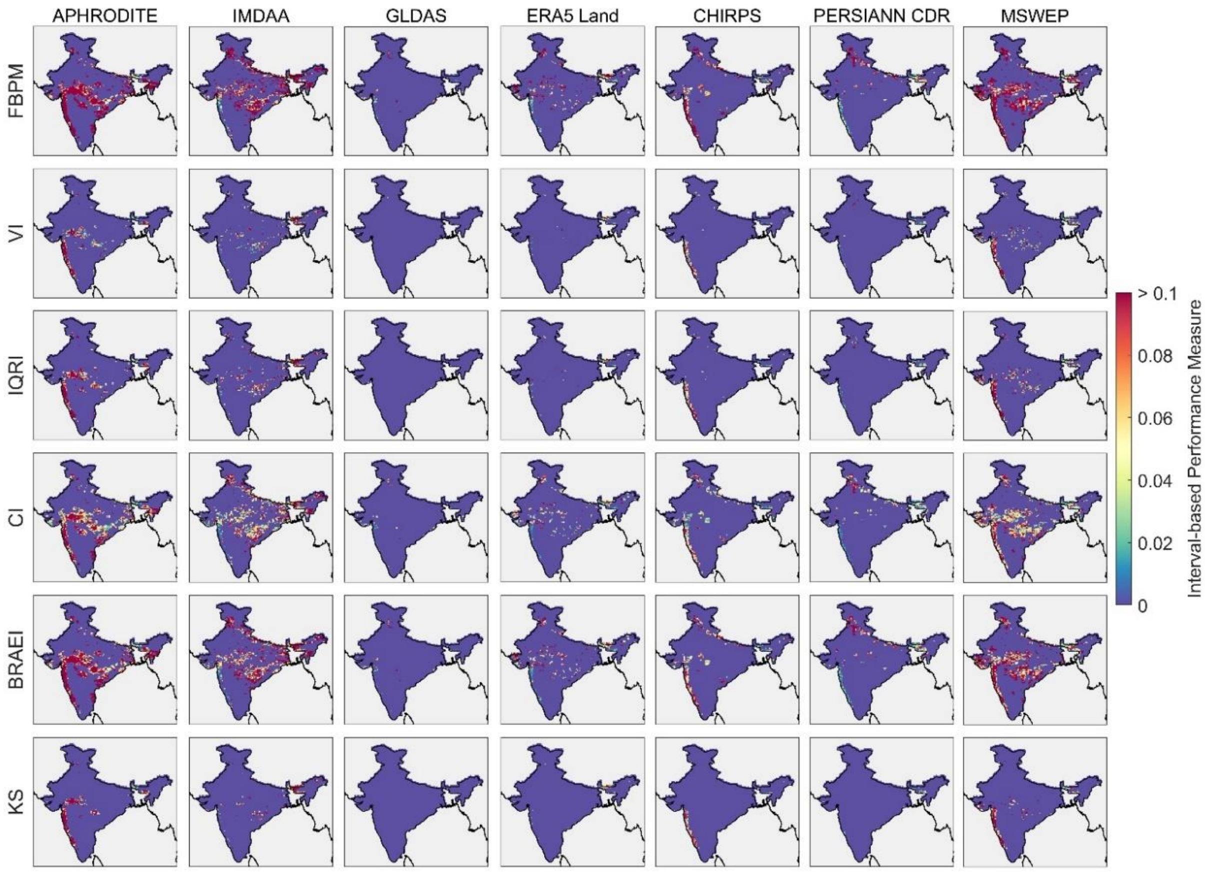

Fig. S8: Same as Fig. S3, but for very heavy rain events

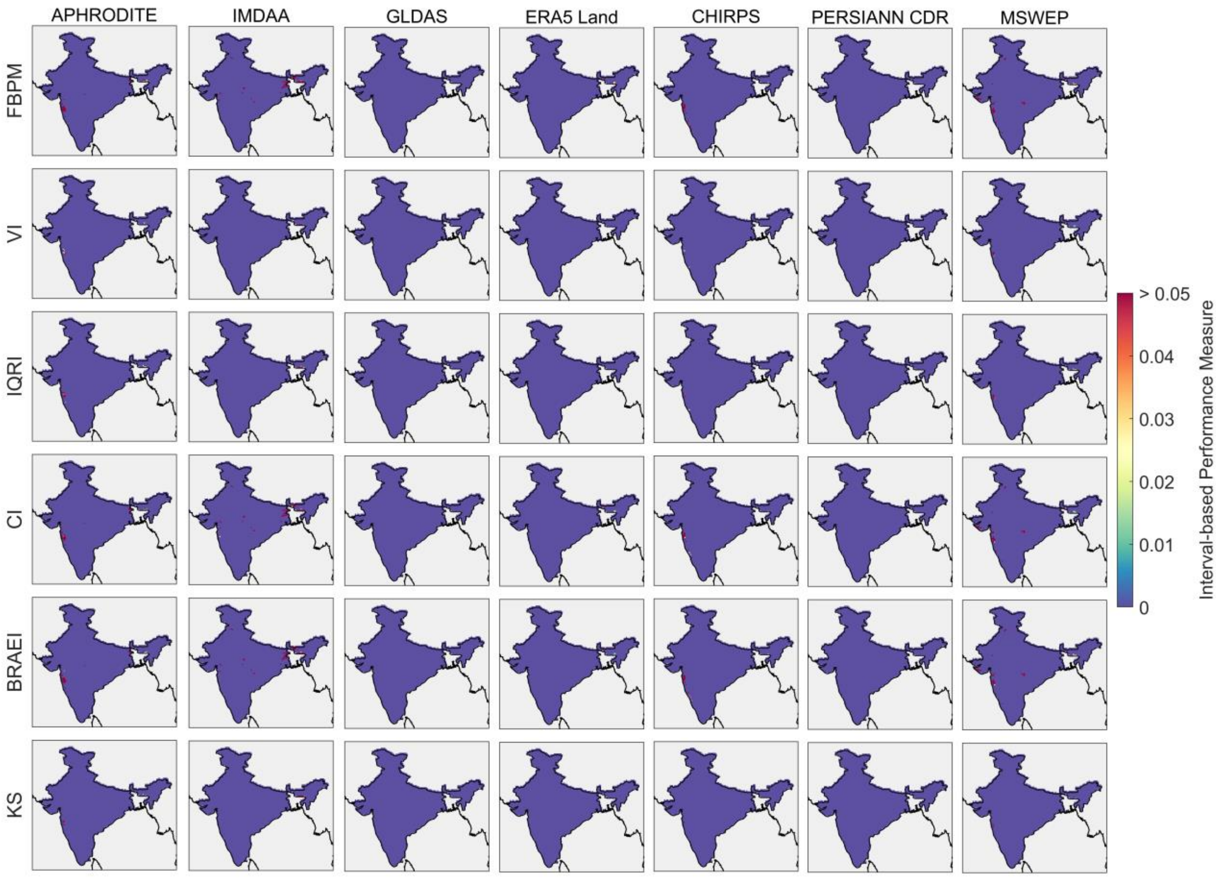

Fig. S9: Same as Fig. S3, but for exceptionally very heavy rain events

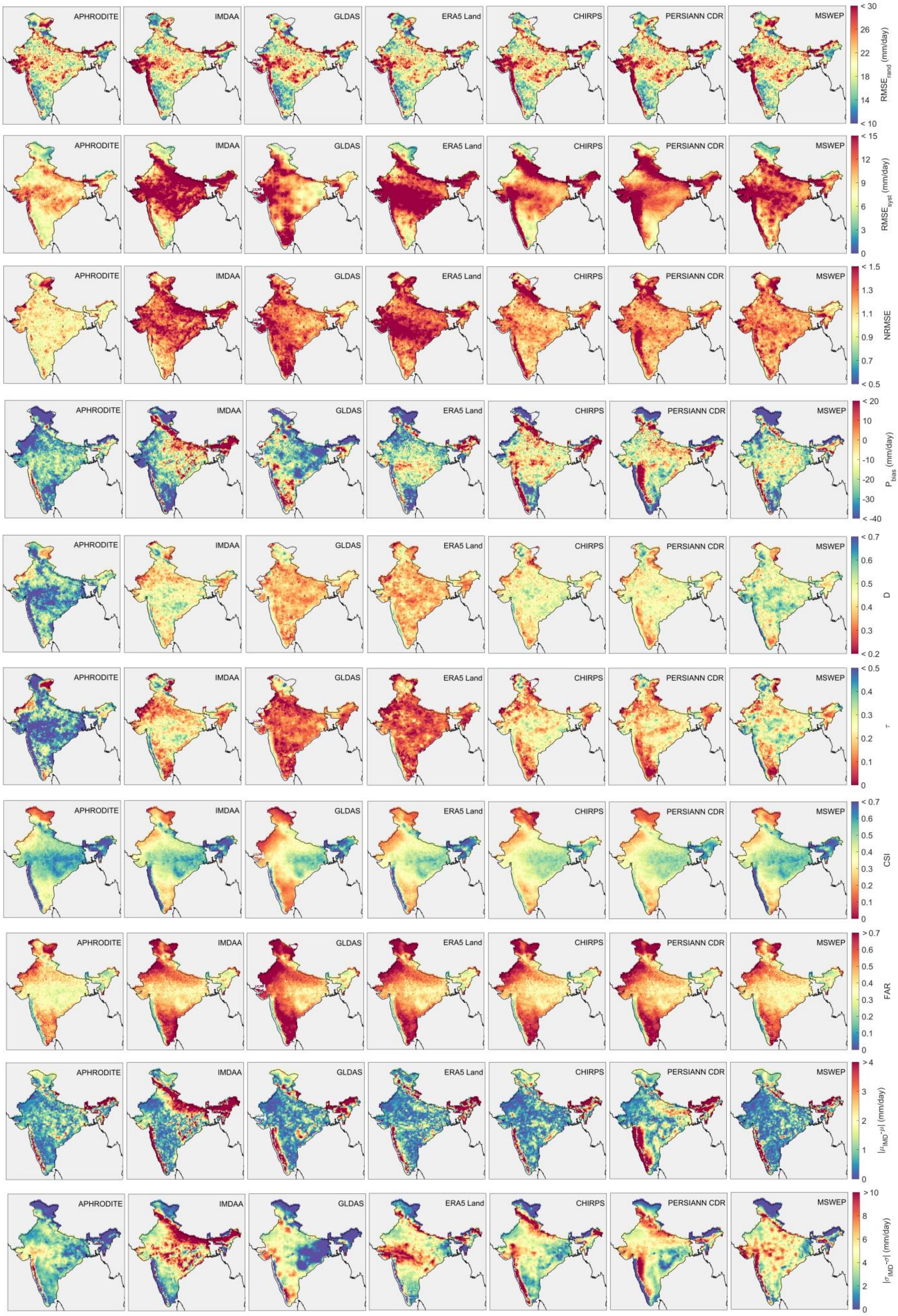
APHRODITE
IMDAA
GLDAS
ERA5 Land
CHIRPS
PERSIANN CDR
MSWEP
< 30
26
22
18
14
< 10
RMSE$_{rand}$ (mm/day)
< 15
12
9
6
3
0
RMSE$_{syst}$ (mm/day)
< 1.5
1.3
1.1
0.9
0.7
< 0.5
NRMSE
< 20
10
0
-10
-20
-30
< -40
P$_{bias}$ (mm/day)
< 0.7
0.6
0.5
0.4
0.3
< 0.2
D
< 0.5
0.4
0.3
0.2
0.1
0
τ
< 0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
CSI
> 0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FAR
> 4
3
2
1
0
$|\mu_{IMD}-\mu|$ (mm/day)
> 10
8
6
4
2
0
$|\sigma_{IMD}-\sigma|$ (mm/day)

***Fig. S10**: Spatial distribution of six continuous and two categorical perfromacne measures and magnitude of mean and standard deviation difference (each row indicates the one metric) of the seven datasets (indicated in each column) with respect to IMD. Six continous measures used here are: random root mean square* $(RMSE_{rand})$, *systematic RMSE* $(RMSE_{syst})$, *normalised RMSE* $(NRMSE)$, *percentage bias* (P_{bias}), *Kendall's correlation coefficient* (τ), *degree of agreement* (D), *critical success index* (CSI) *and false alarm ratio* (FAR)

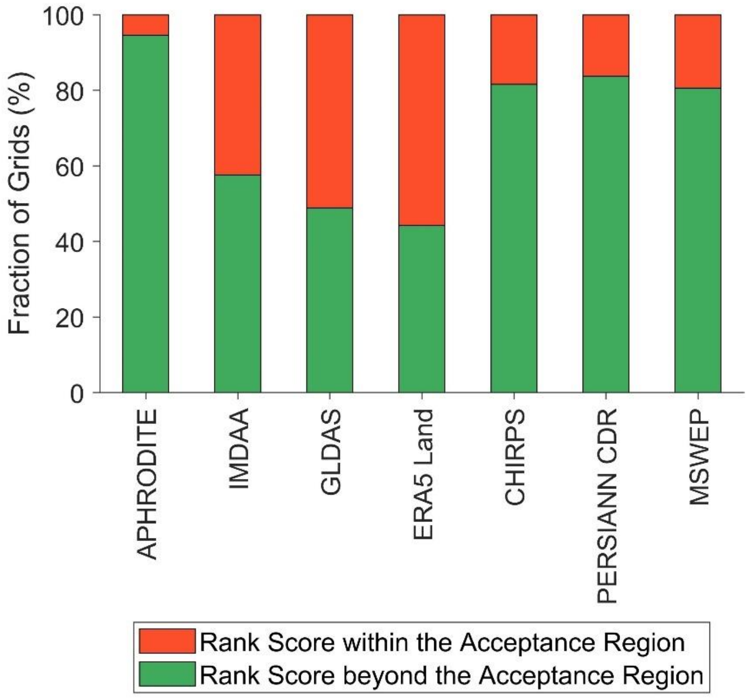

***Fig. S11**: Fraction of grid points represent the significant rank score in seven datasets. Percentage of fraction of grids is calculated using the number of grid points to total number of valid grids observed within Indian domain. Significance of the rank score derived from the bootstrap samples are derived within the 95% confidence interval.*

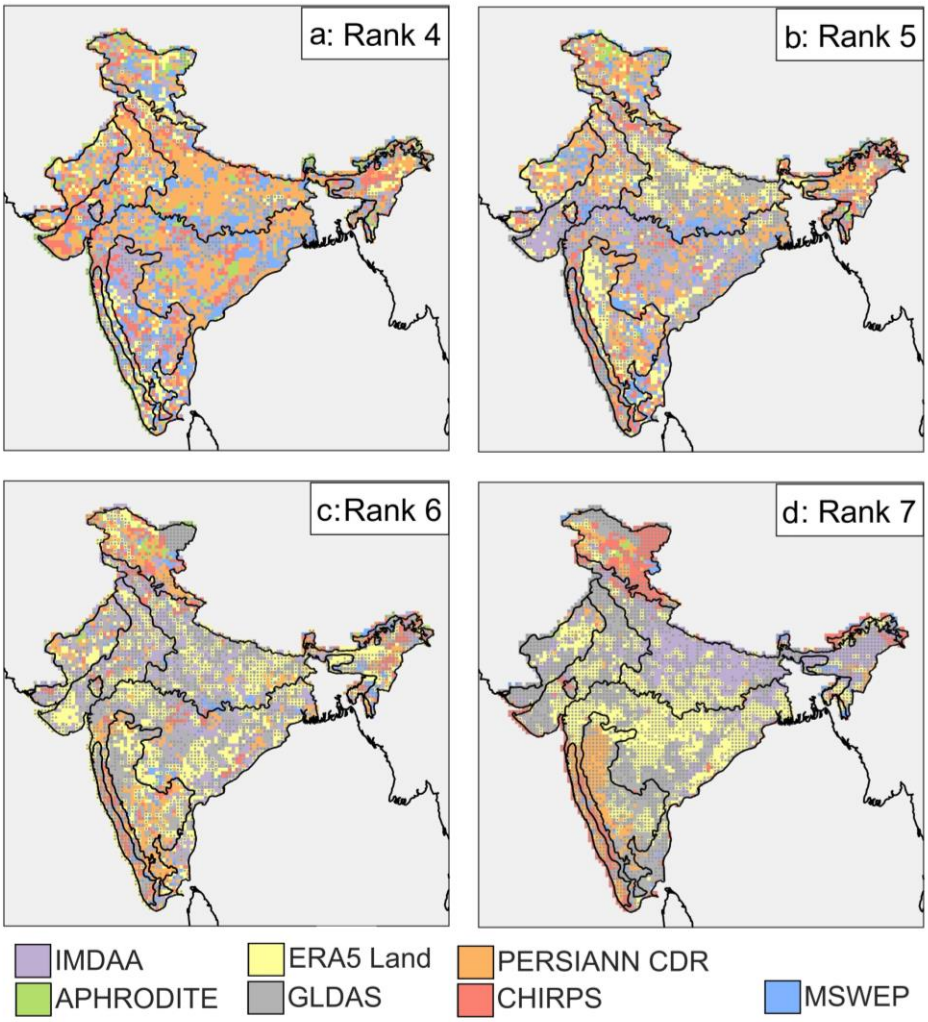

***Fig S12**: Same as Fig. 5, but for rank four (a), five (b), six (c) and seven (d).*

References

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), Article 1. https://doi.org/10.1038/sdata.2018.214

Bhatla, R., Tripathi, A., & Singh, R. S. (2016). Analysis of rainfall pattern and extreme events during southwest monsoon season over Varanasi during 1971-2010. *MAUSAM*, *67*(4), Article 4. https://doi.org/10.54302/mausam.v67i4.1418

Parthasarathy, B. (1995). Monthly and seasonal rainfall series for all India homogeneous regions and meteorological subdivisions: 1871-1994. *Ndian Institute of Tropical Meteorology Research Report*, *RR-65*.

Sonar, R. (2014). Observed trends and variations in rainfall events over Ratnagiri (Maharashtra) during southwest monsoon season. *MAUSAM*, *65*(2), Article 2. https://doi.org/10.54302/mausam.v65i2.965

Teegavarapu, R. S. V., Sharma, P. J., & Lal Patel, P. (2022). Frequency-based performance measure for hydrologic model evaluation. *Journal of Hydrology*, *608*, 127583. https://doi.org/10.1016/j.jhydrol.2022.127583