Intercomparison of different sources of precipitation data in the Brazilian Legal Amazon

Fabrício Daniel dos Santos Silva [1, *], Cláudia Priscila Wanzeler da Costa [2], Vânia dos Santos Franco [2], Helber Barros Gomes [1], Maria Cristina Lemos da Silva [1], Mário Henrique Guilherme dos Santos Vanderlei [1], Rafaela Lisboa Costa [1], Rodrigo Lins da Rocha Júnior [3], Jório Bezerra Cabral Júnior [4], Jean Souza dos Reis [5], Rosane Barbosa Lopes Cavalcante [2], Renata Gonçalves Tedeschi [2], Naurinete de Jesus da Costa Barreto [2], Antônio Vasconcelos Nogueira Neto [2], Edmir dos Santos Jesus [2], Douglas Batista da Silva Ferreira [2].

[1] Instituto de Ciências Atmosféricas, Universidade Federal de Alagoas, Maceió 57072-900, Brazil; helber.gomes@icat.ufal.br (H.B.G); cristina.lemos@icat.ufal.br (M.C.L.d.S); mario.vanderlei@icat.ufal.br (M.H.G.d.S.V); rafaela.costa@icat.ufal.br (R.L.C.)

[2] Instituto Tecnológico Vale – Desenvolvimento Sustentável, Brazil; Claudia.Costa@itv.org (C.P.W.d.C); vania.franco@pq.itv.org (V.d.S.F); Rosane.Cavalcante@itv.org (R.B.L.C); Renata.Tedeschi@itv.org (R.G.T); naurinete.barreto@pq.itv.org (N.d.J.d.C.B); antonio.nogueira@pq.itv.org (A.V.N.N); edmir.jesus@pq.itv.org (E.d.S.J); douglas.silva.ferreira@itv.org (D.B.d.S.F)

[3] Sistema Meteorológico do Paraná, Curitiba 81530-900, Brazil; rodrigo.lins@simepar.br (R.L.d.R.J)

[4] Instituto de Geografia, Desenvolvimento e Meio Ambiente, Universidade Federal de Alagoas, Maceió 57072-900, Brazil; jorio.cabral@igdema.ufal.br (J.B.C.J)

[5] Departamento de Ciências Climáticas e Atmosféricas, Universidade Federal do Rio Grande do Norte, Natal 59078-970, Brazil; jean.reis@ufrn.br (J.S.d.R)

a- Description of current climatology - seasonal analysis

In the DJF quarter (Figure S1), the accumulated rainfall is well distributed throughout the BLA, with volumes generally higher than 1000 mm, and the two rainiest areas whose values may exceed 2000 mm are between northern Mato Grosso and southern Pará, and in south-central Amazonas. The exception is a dry portion in the extreme north, with an average rainfall of approximately 700 mm. This is a typical climatic response to the South Atlantic Convergence Zone (SACZ) and the Bolivian Highlands, which act intensely in the summer in this region, causing rainfall mainly in the south-central part of the BLA and from the northernmost location of the Intertropical Convergence Zone (ITCZ), causing small, accumulated volumes of rainfall in the state of Roraima [100,40,101]). All data sources underestimate the cited areas with the highest rainfall in the summer of the BLA, especially CPC (Figure S1 (e)) and CRU (Figure S1 (f)). The CMORPH (Figure S1 (d)) extends the dry portion of the Roraima to the north of the Pará and Amapá. PERSIANN-CDR (Figure S1 (b)), CHIRPS (Figure S1 (c), ERA5Land (Figure S1 (g), GPCC (Figure S1 (h)), IMERGE (Figure S1 (i)) and Xavier (Figure S1 (j)) more accurately estimate the spatial distribution of summer rainfall with more intense precipitation nuclei in the rainiest area of the BLA, although lower in accumulated area and volume.

In the MAM quarter (Figure S2), the ITCZ is notable in the north-central part of the BLA (Figure S2 (a)), with the respective weakening of the moisture corridor that provides rainfall in the centre-Amazonian south. These characteristics are better captured by the databases in the MAM quarter than in the DJF quarter, allowing us to infer that the rainfall estimate associated with the ITCZ is efficient in all gridded/reanalysis/remote sensing analyses.

The JJA quarter (Figure S3) is the period of the year when rainfall is restricted to the extreme north of the BLA, especially between Roraima and northwest Amazonas, with a strong gradient

marking a transition zone between this rainy northwest extreme and the dry southeast end. Spring (Figure S4) marks the end of the rains in the northeast Amazon and the beginning of rains in the northwest–southeast diagonal of Amazonas, south of Pará and north of Mato Grosso. For these two areas, which are drier and rainier, the climatology obtained from the data sources is also very similar to that of the observations, with CMORPH overestimating the average volumes of precipitation in the rainiest portion, while the others underestimate.

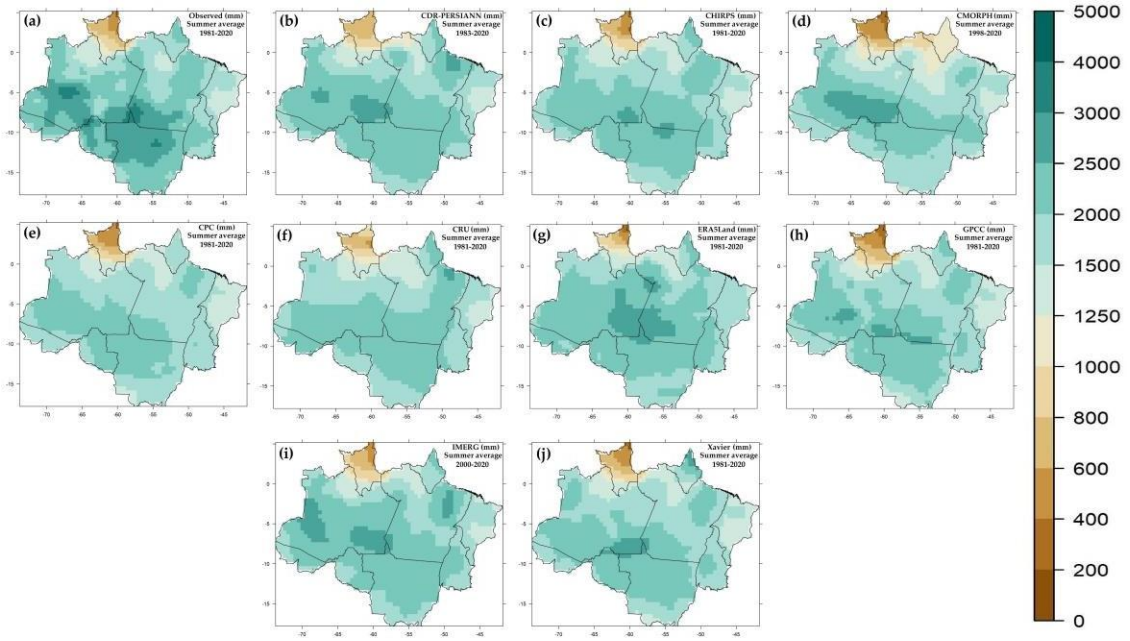

Figure S1. Average accumulated precipitation (mm) in DJF in the BLA: (a) observed data, (b) PER-SIANN-CDR, (c) CHIRPS, (d) CMORPH, (e) CPC, (f) CRU, (g) ERA5Land , (h) GPCC, (i) IMERGE and (j) Xavier.

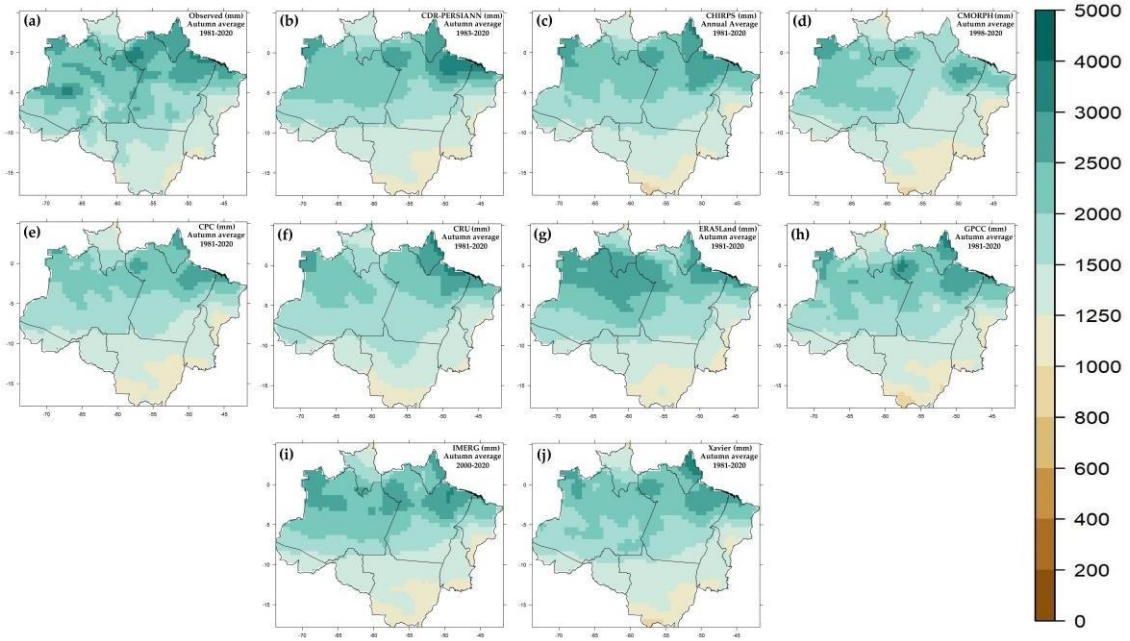

Figure S2. Average accumulated precipitation (mm) in MAM in the BLA: (a) observed data, (b) PER-SIANN-CDR, (c) CHIRPS, (d) CMORPH, (e) CPC, (f) CRU, (g) ERA5Land , (h) GPCC, (i) IMERGE and (j) Xavier.

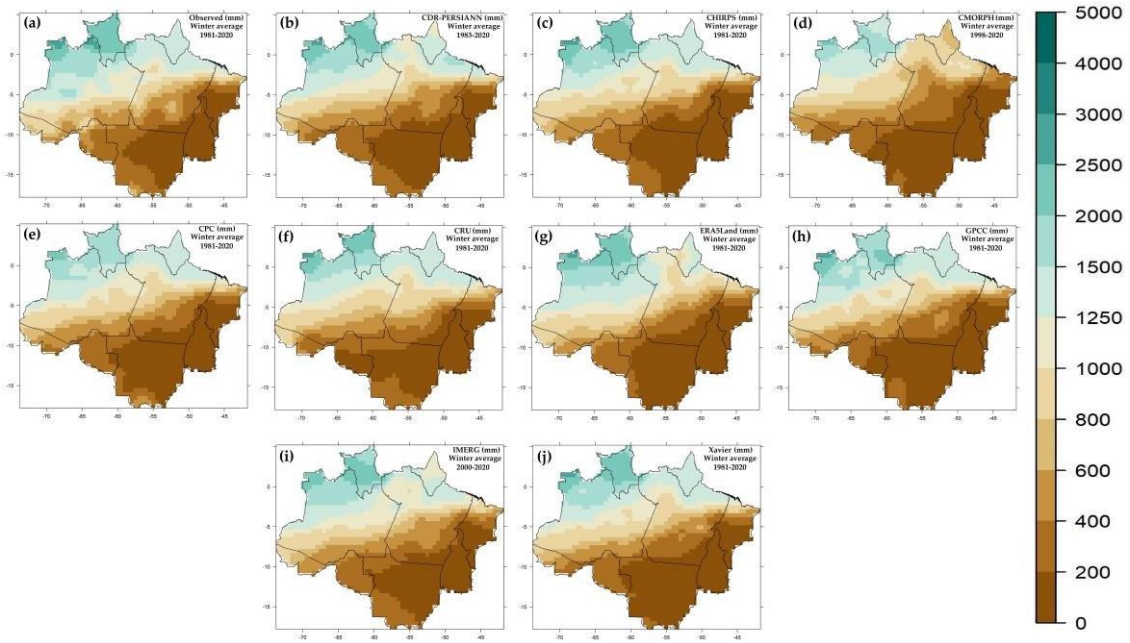

Figure S3. Average accumulated precipitation (mm) in JJA in the BLA: (a) observed data, (b) PER-SIANN-CDR, (c) CHIRPS, (d) CMORPH, (e) CPC, (f) CRU, (g) ERA5Land , (h) GPCC, (i) IMERGE and (j) Xavier.

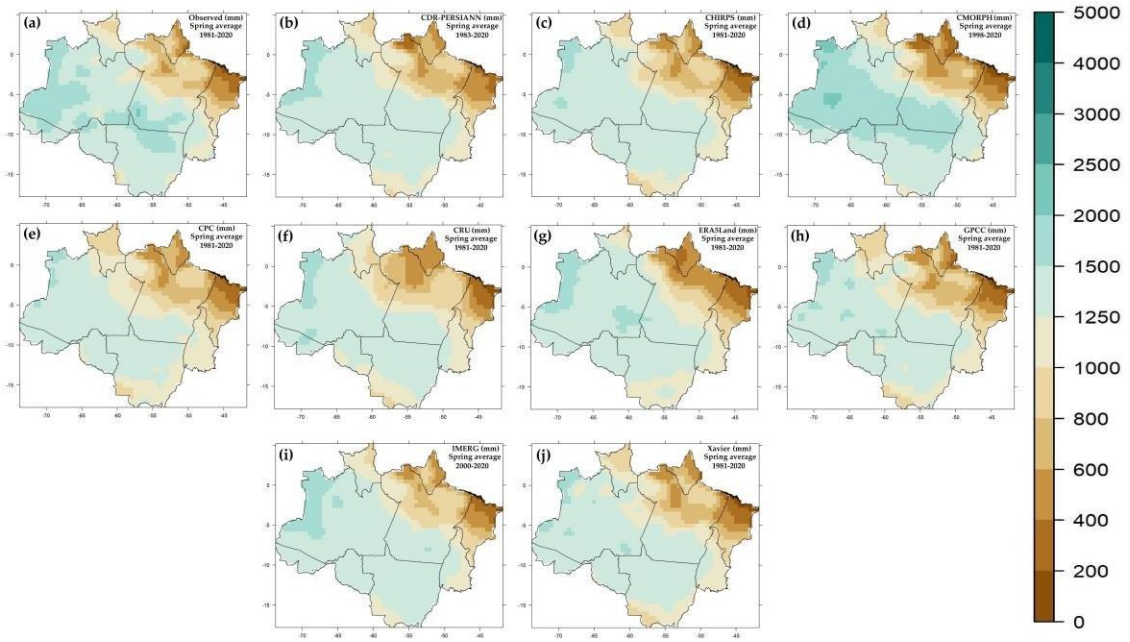

Figure S4. Average accumulated precipitation (mm) in SON in the BLA: (a) observed data, (b) PER-SIANN-CDR, (c) CHIRPS, (d) CMORPH, (e) CPC, (f) CRU, (g) ERA5Land , (h) GPCC, (i) IMERGE and (j) Xavier.

b- Skill assessment using bias – seasonal analysis

Figure S5 shows the *bias* for the DJF quarter, predominantly negative for the CHIRPS, CPC, GPCC and Xavier estimates (Figures S5 (b), S5 (d), S5 (g) and S5 (i)) and with a negative predominance, but with positive nuclei to the north and centre of the BLA for PERSIANN-CDR, CMORPH, CRU, ERA5Land and IMERG (Figures S5 (a), S5 (c), S5 (e), S5 (f)) and S5 (h)), respectively. In the MAM quarter (Figure S6), negative *biases* also predominated in most of the BLA for all data sources, with positive core *biases* observed in the north and centre of the BLA with PERSIANN-CDR and CRU (Figure S6 (a) and S6 (e)), further south of the Maranhão in CMORPH (Figure S6 (c)), and

more located in the northwest portion of the BLA with ERA5Land and IMERG (Figure S6 (f) and S6 (h)). In the JJA quarter, the *biases* are lighter, with a predominance of more areas with underestimation of precipitation than with overestimation, with more significant negative *biases* observed in the north/northwest sectors of the BLA, especially the CMORPH (Figure S7). Figure S8 shows the *bias* in the SON quarter, with the databases tending to underestimate precipitation in most of the BLA, with the exception of CMORPH (Figure S8 (c)), which tends to overestimate precipitation in most of the BLA, and ERA5Land (Figure S8 (f)) and IMERG (Figure S8 (h)), which also have positive nuclei scattered in the BLA.

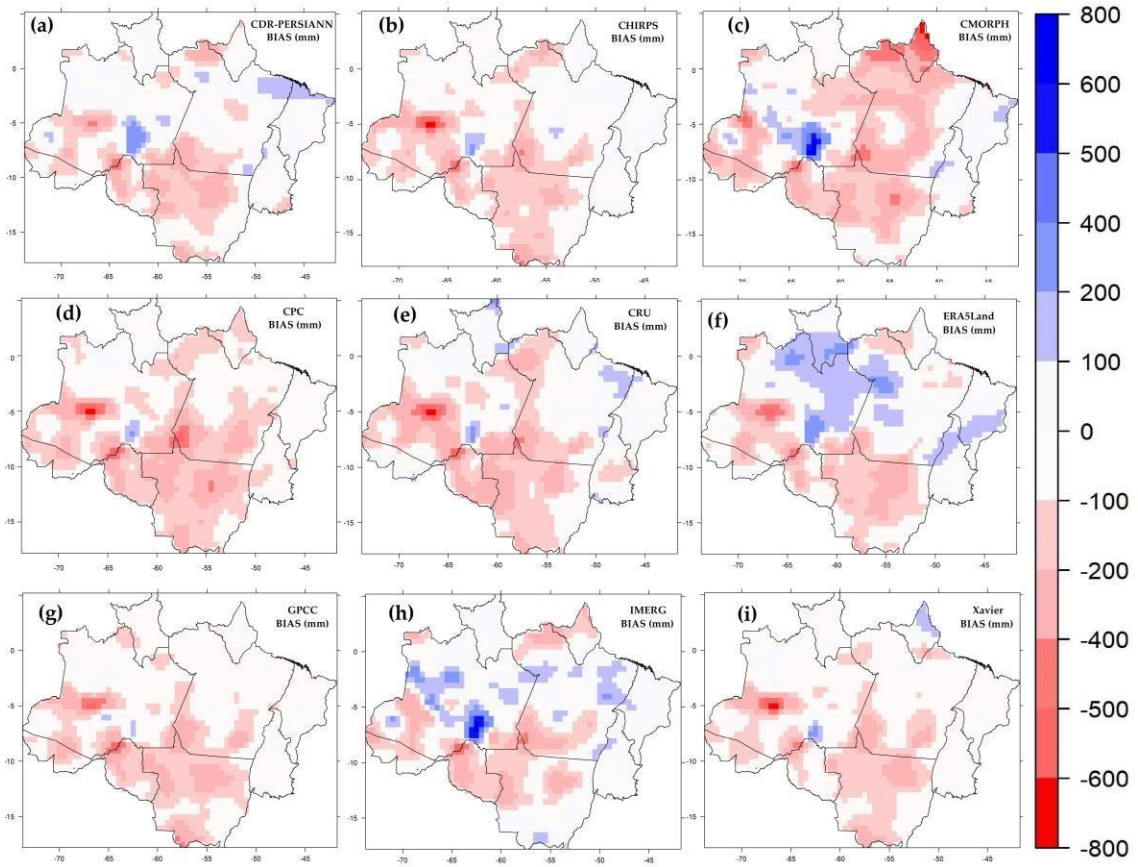

Figure S5. Precipitation *bias* (mm) estimated by each database compared to the DJF observations: (a) PERSIANN-CDR, (b) CHIRPS, (c) CMORPH, (d) CPC, (e) CRU, (f) ERA5Land, (g) GPCC, (h) IMERGE and (i) Xavier.

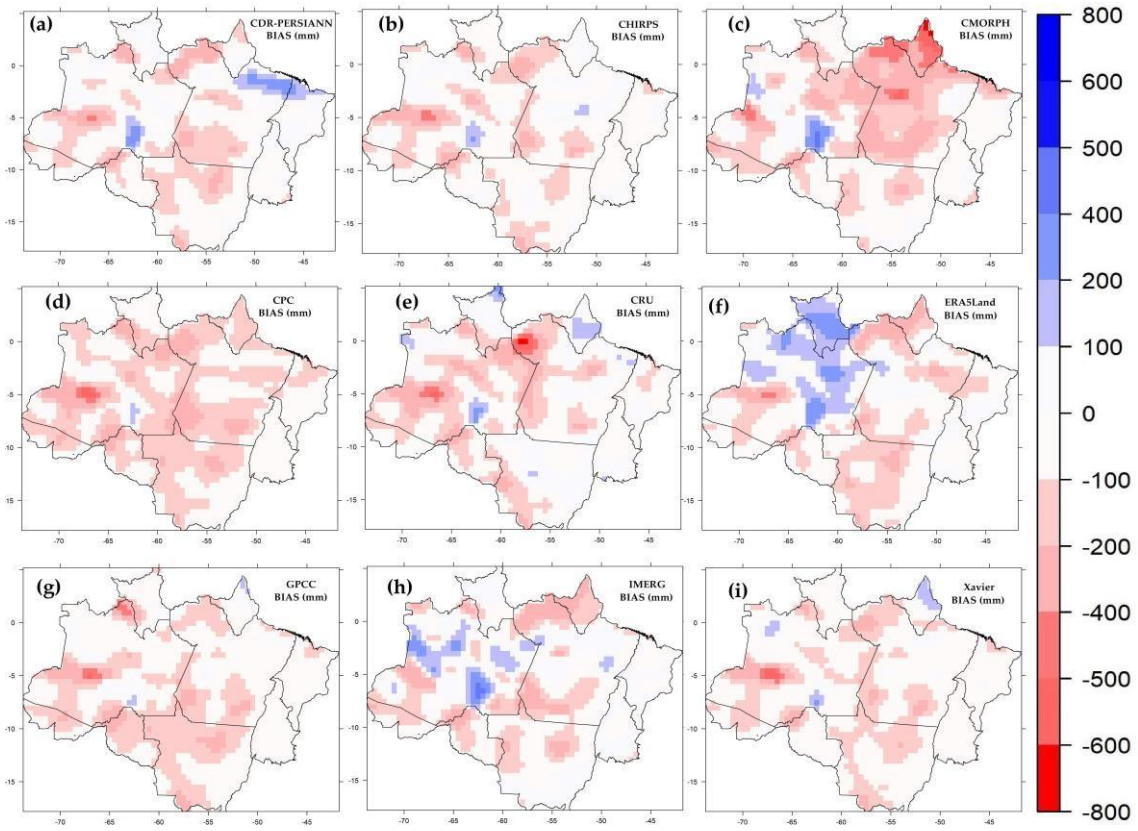

Figure S6. Precipitation *bias* (mm) estimated by each database compared to the MAM observations: (a) PERSIANN-CDR, (b) CHIRPS, (c) CMORPH, (d) CPC, (e) CRU, (f) ERA5Land, (g) GPCC, (h) IMERGE and (i) Xavier.

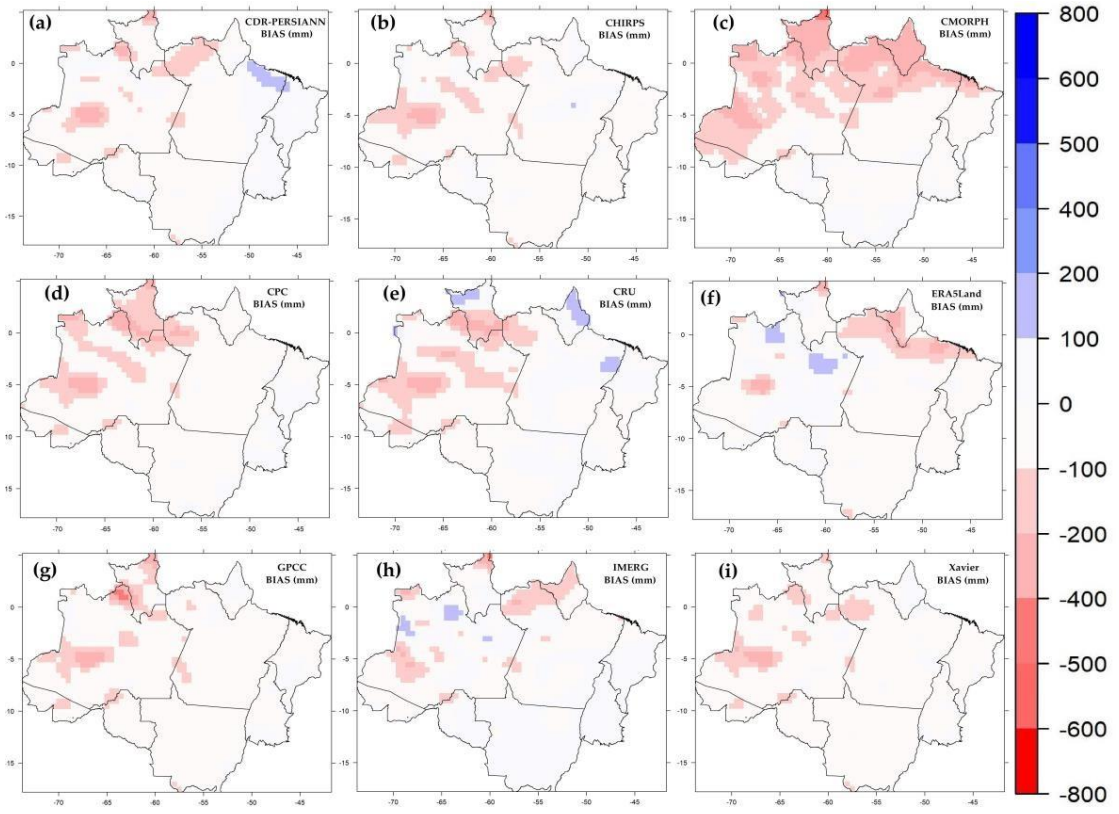

Figure S7. Precipitation *bias* (mm) estimated by each database compared to the JJA observations: (a) PERSIANN-CDR, (b) CHIRPS, (c) CMORPH, (d) CPC, (e) CRU, (f) ERA5Land, (g) GPCC, (h) IMERGE and (i) Xavier.

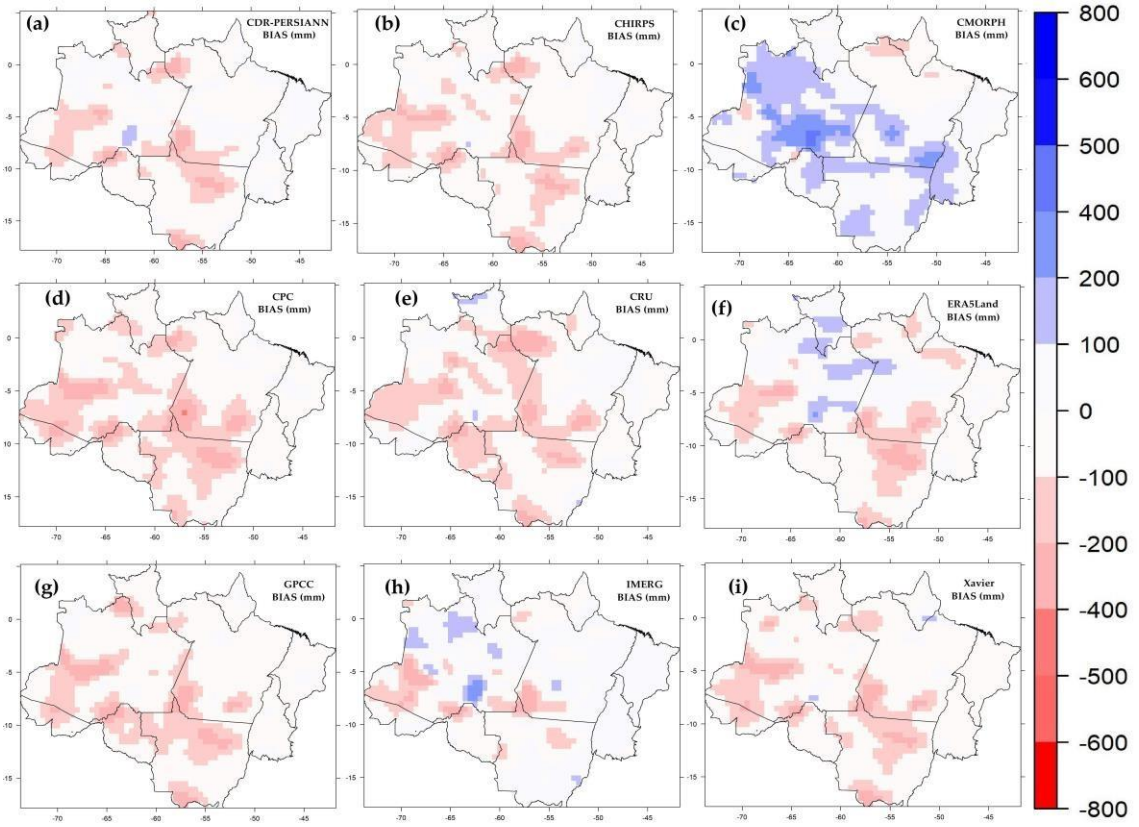

Figure S8. Precipitation *bias* (mm) estimated by each database compared to the SON observations: (a) PERSIANN-CDR, (b) CHIRPS, (c) CMORPH, (d) CPC, (e) CRU, (f) ERA5Land, (g) GPCC, (h) IMERGE and (i) Xavier.

c- Skill assessment using RMSE – seasonal analysis

Figure S9 shows the spatialization of the *RMSE* in the BLA for the DJF quarter, with the smallest errors in brown and the largest errors in shades of green. Although it is not difficult to see that some data sources have a greater predominance of areas in stronger shades of brown than others, indicating larger areas with lower errors, such as Xavier (Figure S9 (i)), and others are the opposite, such as PERSIANN-CDR (Figure S9 (a)), the spatialization of this index shows similarities between the sources that can hinder purely visual evaluations. To remedy this, we calculated the spatial average value of the *RMSE* across the entire BLA area. The two sources with the lowest spatial *RMSE* were Xavier with 251 mm, followed by IMERG with 254 mm (Figure S9 (h)), and the two sources with the highest spatial *RMSE* values were PERSIANN-CDR (Figure S9 (a)) with 353 mm, followed by CRU with 321 mm (Figure S9 (e)). In the MAM quarter (Figure S10), IMERG (Figure S10 (h)) and Xavier (Figure S10 (i)) had the lowest values in the area with 227 mm and 230 mm, respectively, while ERA5Land and CRU had the highest mean *RMSE* values in the area, with 287 mm and 292 mm (Figures S10 (f) and S10 (e)), respectively. In the JJA quarter, IMERG and Xavier also presented the lowest *RMSE* values (Figures S11(h) and S11 (i)), at 94 and 98 mm, with ERA5Land and CRU (Figures S11 (e) and S11 (f)) with the highest values, 124 and 125 mm. In the SON quarter, IMERG and CMORPH presented the lowest mean *RMSE* values of 166 mm and 176 mm (Figures S12 (h) and S12 (c)), respectively, while CRU and ERA5Land

(Figures S12 (e) and S12 (f)) showed the highest mean *RMSE* values of 216 mm and 219 mm, respectively.

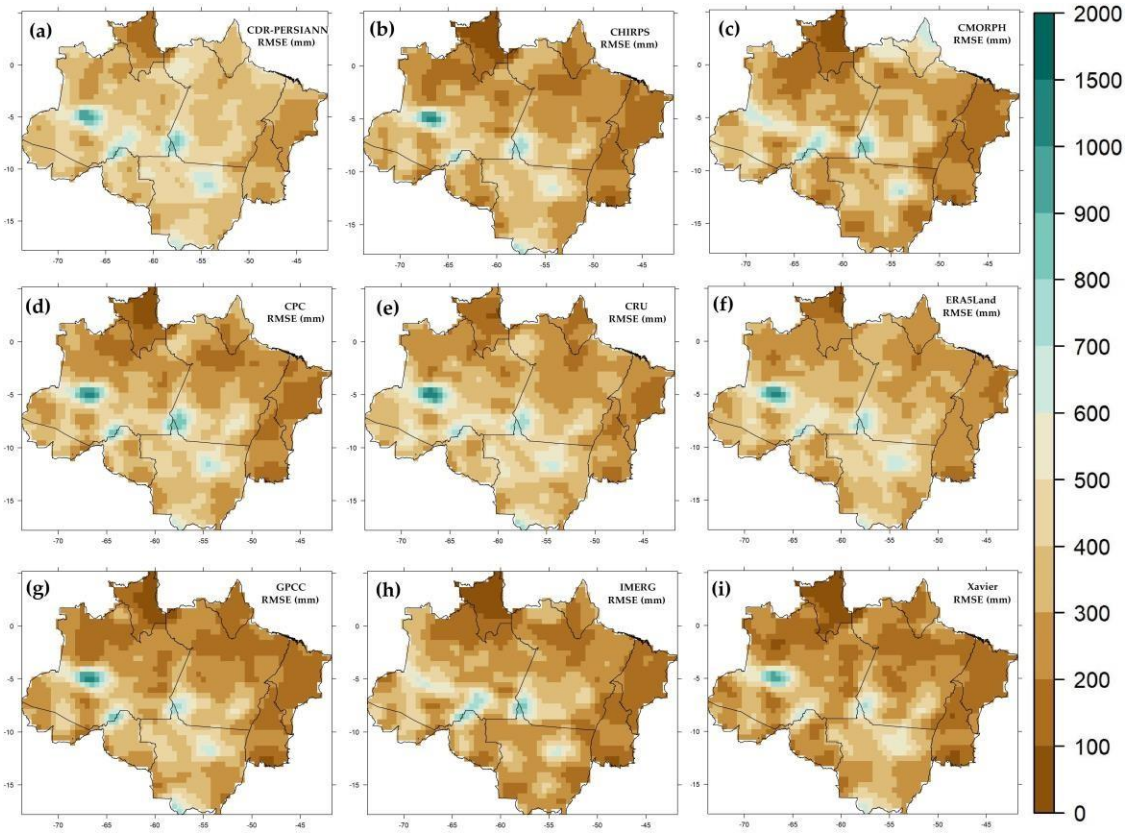

Figure S9. DJF RMSE (mm) of each database compared to the observations: (a) PERSIANN-CDR, (b) CHIRPS, (c) CMORPH, (d) CPC, (e) CRU, (f) ERA5Land, (g) GPCC, (h) IMERGE and (i) Xavier.

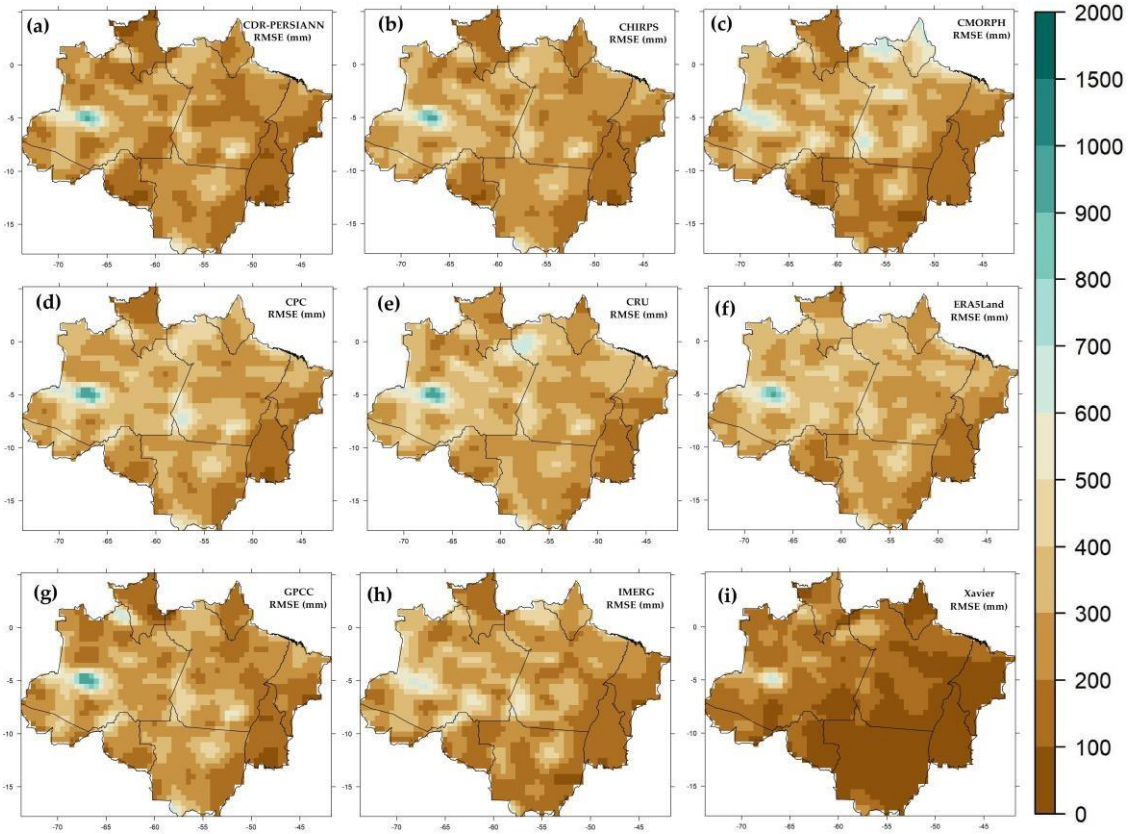

Figure S10. MAM RMSE (mm) of each database compared to the observations: (a) PERSIANN-CDR, (b) CHIRPS, (c) CMORPH, (d) CPC, (e) CRU, (f) ERA5Land, (g) GPCC, (h) IMERGE and (i) Xavier.

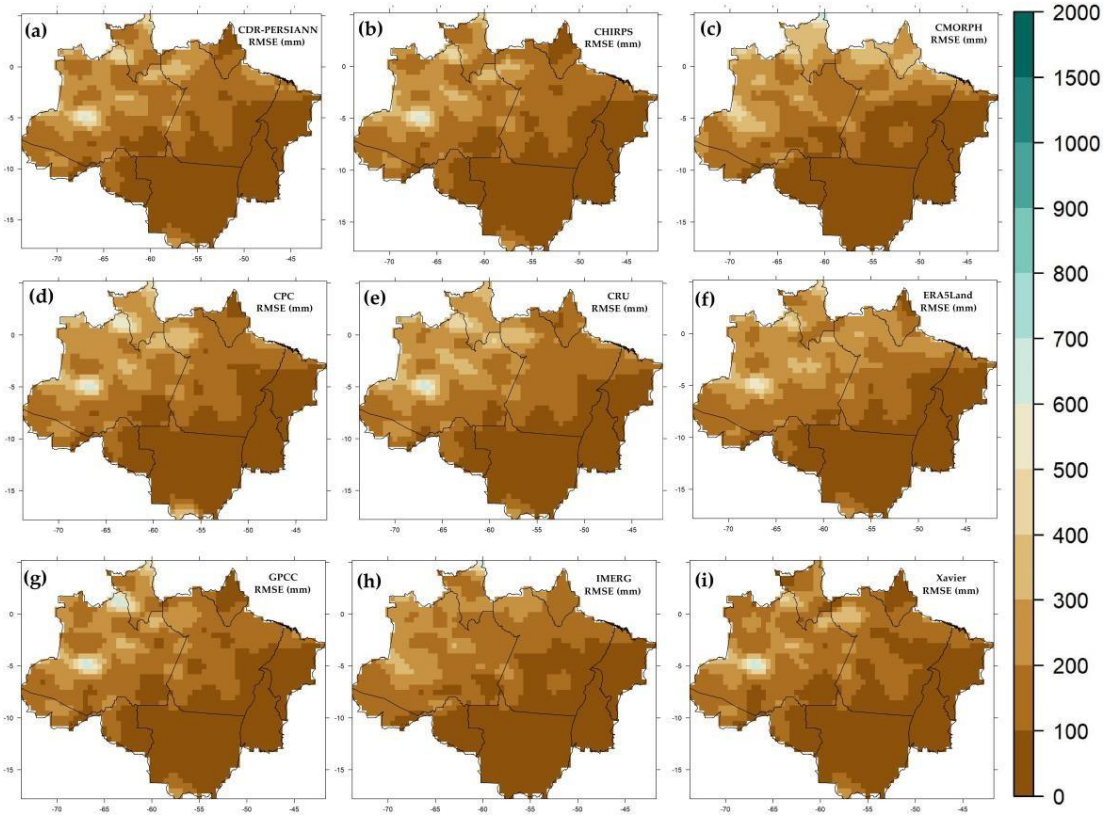

Figure S11. JJA RMSE (mm) of each database compared to the observations: (a) PERSIANN-CDR, (b) CHIRPS, (c) CMORPH, (d) CPC, (e) CRU, (f) ERA5Land, (g) GPCC, (h) IMERGE and (i) Xavier.

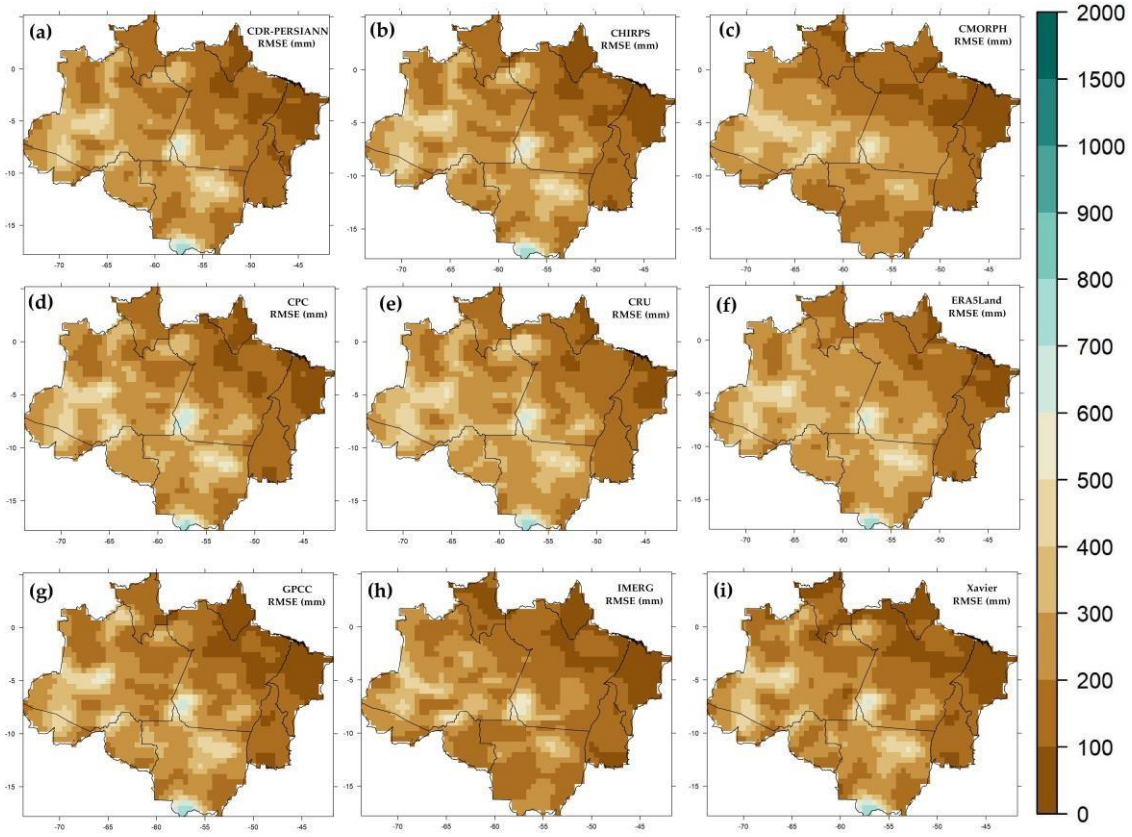

Figure S12. SON RMSE (mm) of each database compared to the observations: (a) PERSIANN-CDR, (b) CHIRPS, (c) CMORPH, (d) CPC, (e) CRU, (f) ERA5Land, (g) GPCC, (h) IMERGE and (i) Xavier.

d- Comparison using Pearson's correlation – seasonal analysis

Figures S13 to S16 show the correlations between each database and the observations on a seasonal scale. In the representative quarters of the year, it is evident that three data sources stand out with higher and statistically significant correlations based on the Student's t test, in shades of blue (above 0.4 on the scale): Xavier with an average correlation in the BLA of 0.63 in the DJF quarter (Figure S13 (i)), 0.63 in the MAM quarter (Figure S14 (i)), 0.65 in the JJA quarter (Figure S15 (i)) and 0.58 in the SON quarter (Figure S16 (i)). Next, there is the GPCC with a correlation of 0.57 in DJF (Figure S13 (g)), 0.54 in MAM (Figure S14 (g)), 0.56 in JJA (Figure S15 (g)) and 0.50 in SON (Figure S16 (g)). The third database with the highest correlations was CHIRPS, 0.52 in DJF (Figure S13 (b)), 0.52 in MAM (Figure S14 (b)), 0.52 in JJA (Figure S15 (b)) and 0.46 in SON (Figure S16 (b)).

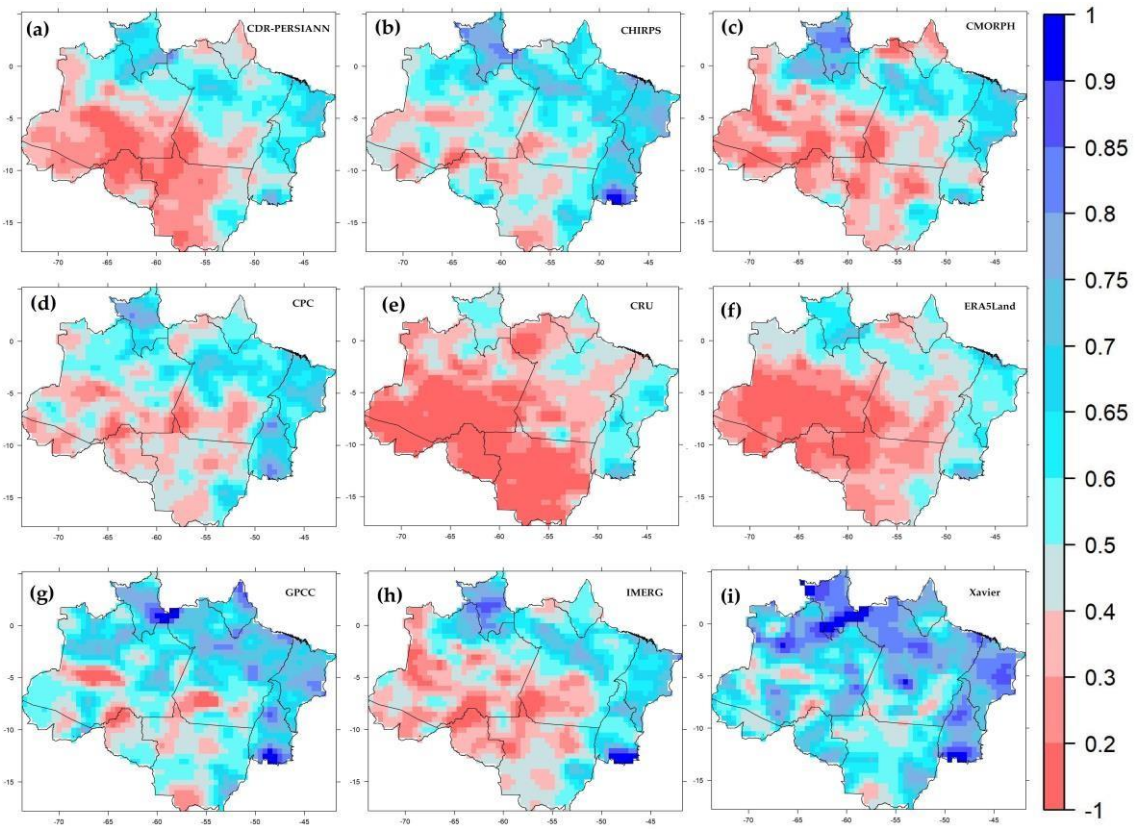

Figure S13. DJF correlation between each database and observations: (a) PERSIANN-CDR, (b) CHIRPS, (c) CMORPH, (d) CPC, (e) CRU, (f) ERA5Land, (g) GPCC, (h) IMERGE and (i) Xavier.

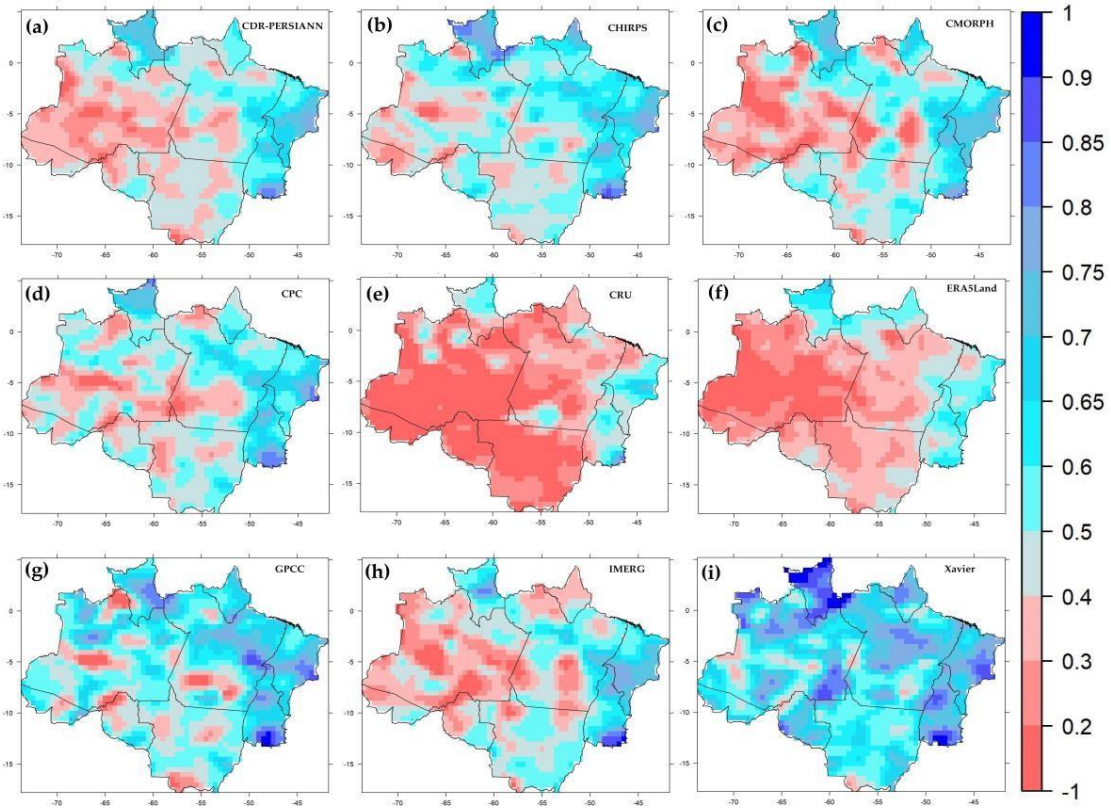

Figure S14. MAM correlation between each database and observations: (a) PERSIANN-CDR, (b) CHIRPS, (c) CMORPH, (d) CPC, (e) CRU, (f) ERA5Land, (g) GPCC, (h) IMERGE and (i) Xavier.

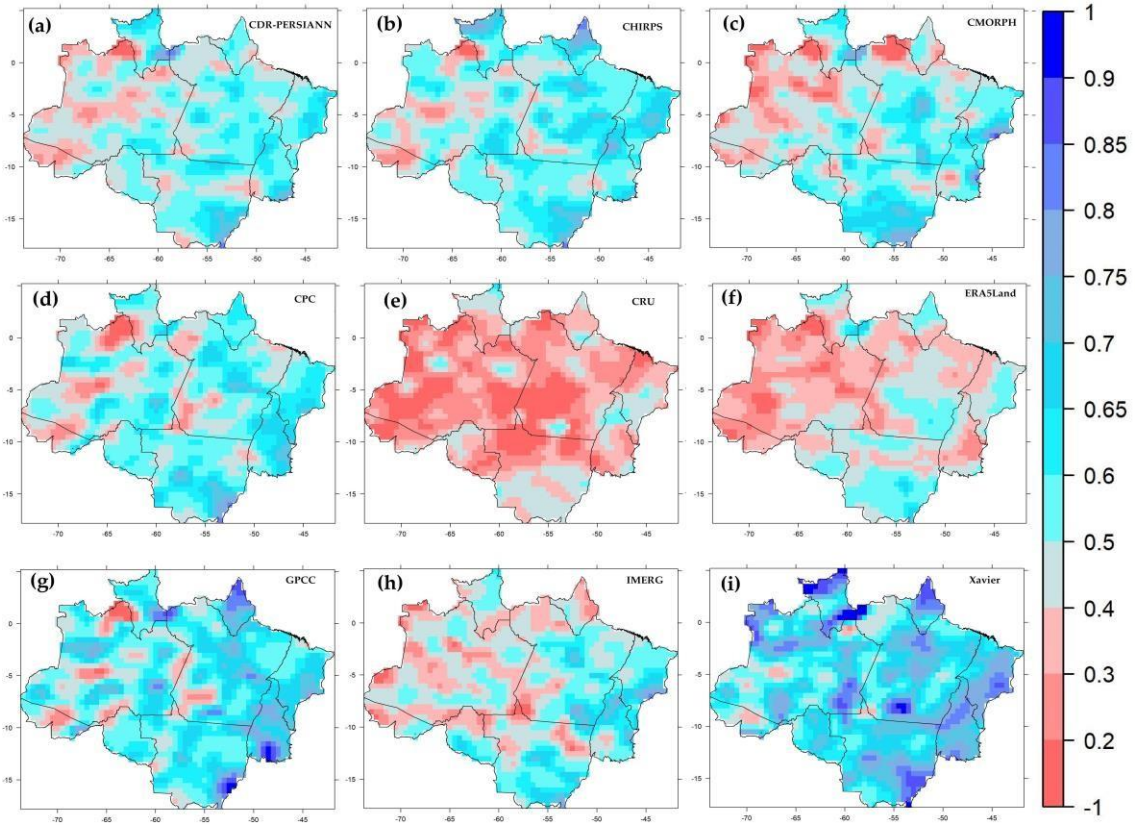

Figure S15. JJA correlation between each database and observations: (a) PERSIANN-CDR, (b) CHIRPS, (c) CMORPH, (d) CPC, (e) CRU, (f) ERA5Land, (g) GPCC, (h) IMERGE and (i) Xavier.

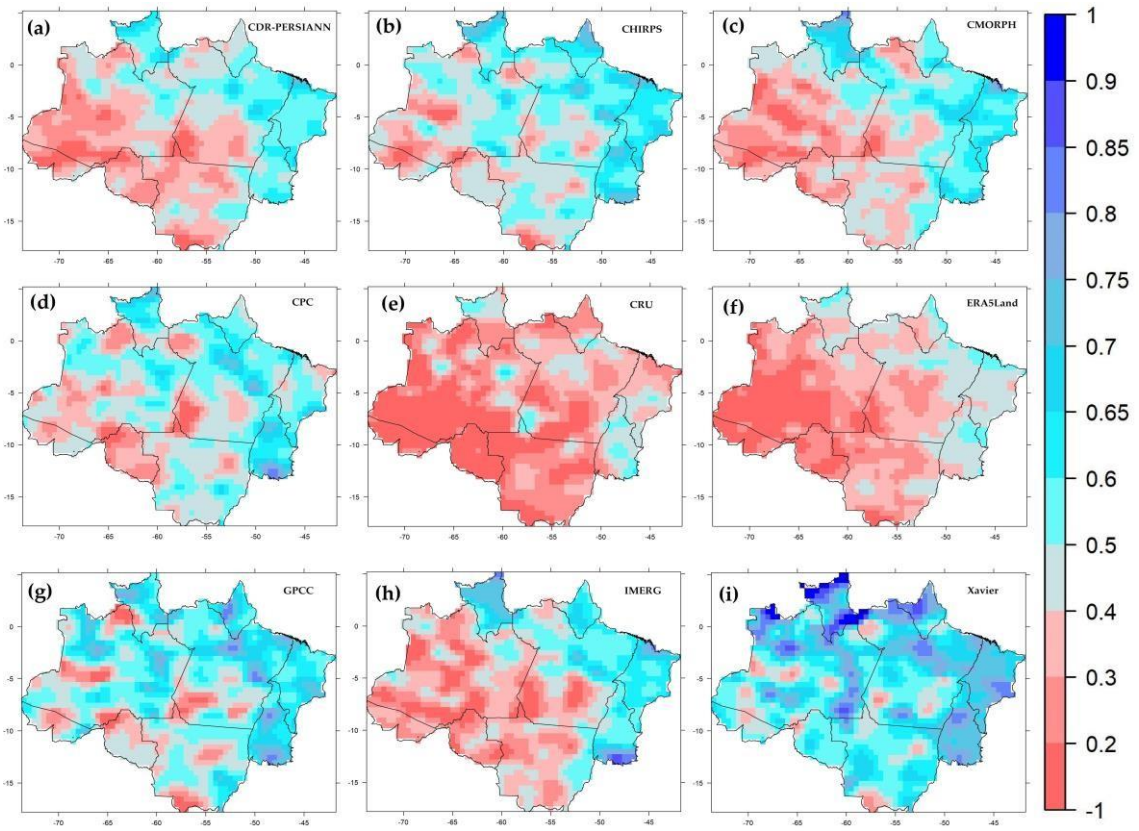

Figure S16. SON correlation between each database and observations: (a) PERSIANN-CDR, (b) CHIRPS, (c) CMORPH, (d) CPC, (e) CRU, (f) ERA5Land, (g) GPCC, (h) IMERGE and (i) Xavier.

e- PDF analysis

In Group 2 (Figure S17), the *PDFs* have a normal distribution from April to November, i.e., from the end of the rainiest period passing through the less rainy period until the beginning of the new rainy period, which is from December to May. This characteristic is captured by the data sources, but with clear underestimations of the values and some peculiarities in the *PDFs* of some sources, such as the flatter May distribution observed in CHIRPS, the bimodal June distribution observed in the CPC, the bimodal May distribution observed in the ERA5Land, GPCC and CRU, and trimodal in May from IMERG. From the observation, the *PDFs* of December and January have a trimodal distribution, a characteristic that was captured by Xavier and CHIRPS, while the CPC and IMERG showed a normal distribution in December, and the other sources showed a normal or bimodal distribution.

In Group 3 (Figure S18), the *PDFs* in the less rainy period from May to December were better captured by all sources, while for the wetter period from January to April, shape differences are more common, as the normal distribution observed in ERA5Land, CRU and IMERG for January compared to a flatter and almost trimodal distribution observed. For March, a normal distribution is observed, while most sources show a bimodal or normal distribution more pronounced than the observation.

For Group 4 (Figure S19), the characteristics of the *PDFs*, both observed and of the other sources, were similar to those of Group 2, with greater difficulty of the sources in the reliable representation of the *PDFs* of the wettest period from November to April, while for Group 5 (Figure S20), the analyses are similar to those performed for Group 3. Finally, for Group 6 (Figure

S21), with the rainy season concentrated in the May-June-July quarter, there are contractions in the *PDFs* estimated for May, with a more normalized distribution than that observed in CHIRPS, CRU and IMERGE and with bimodal or trimodal variation. In the CPC, GPCC, PERSIANN-CDR and CMORPH, ERA5Land maintains a *PDF* similar to that observed, although it overestimates the May rainfall. These results demonstrate a greater difficulty of the data sources in representing the first month of the rainiest period of Group 6. Again, the driest period of Group 6 shows the greatest similarities between the estimated and observed *PDFs*.

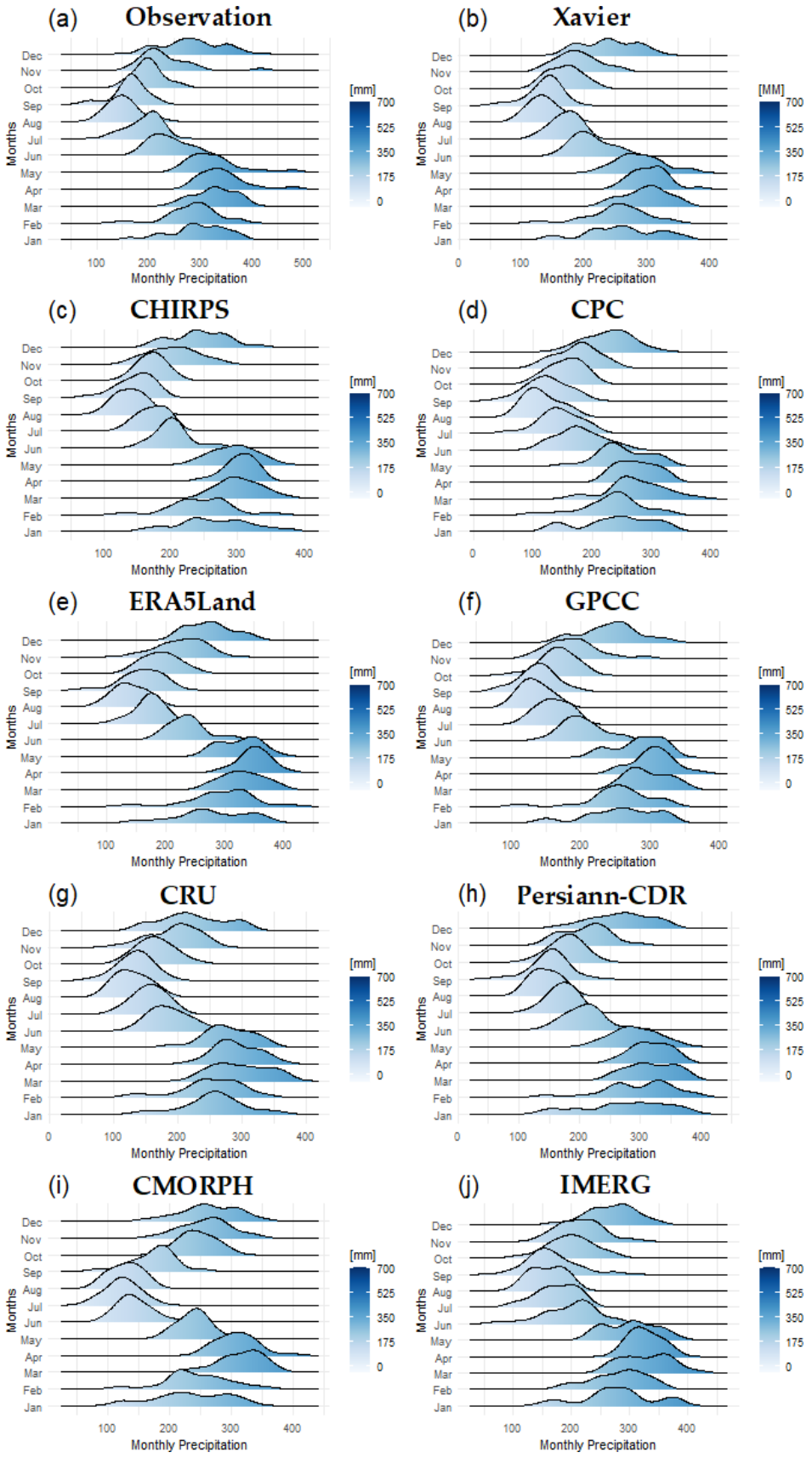

Figure S17. For Group 2, probability density of monthly rainfall (mm) for (a) Observations, (b) Xavier, (c) CHIRPS, (d) CPC, (e) ERA5Land, (f) GPCC, (g) CRU, (h) PERSIANN-CDR, (i) CMORPH and (j) IMERG.

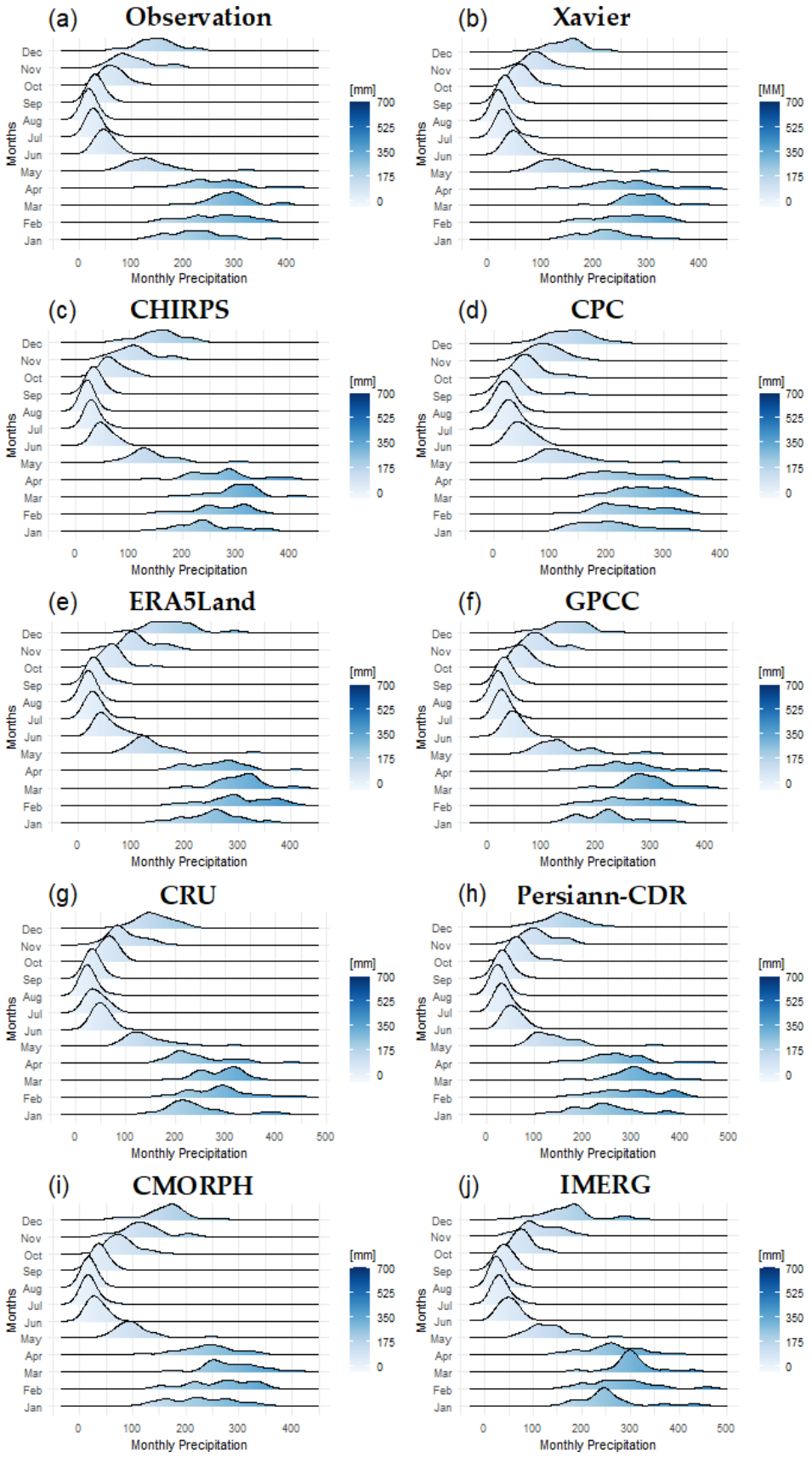

Figure S18. For Group 3, probability density of monthly rainfall (mm) for (a) Observations, (b) Xavier, (c) CHIRPS, (d) CPC, (e) ERA5Land, (f) GPCC, (g) CRU, (h) PERSIANN-CDR, (i) CMORPH and (j) IMERG.

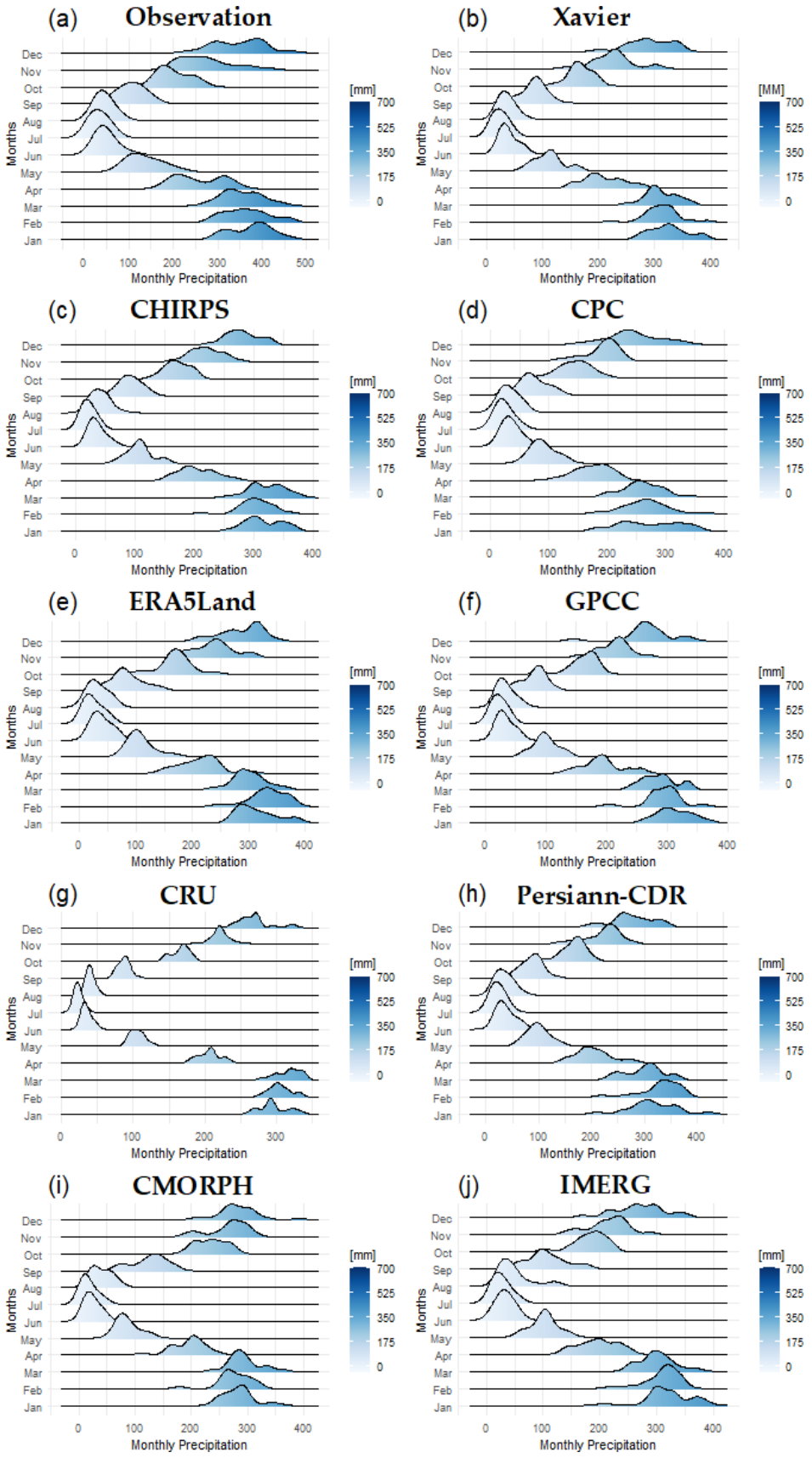

Figure S19. For Group 4, probability density of monthly rainfall (mm) for (a) Observations, (b) Xavier, (c) CHIRPS, (d) CPC, (e) ERA5Land, (f) GPCC, (g) CRU, (h) PERSIANN-CDR, (i) CMORPH and (j) IMERG.

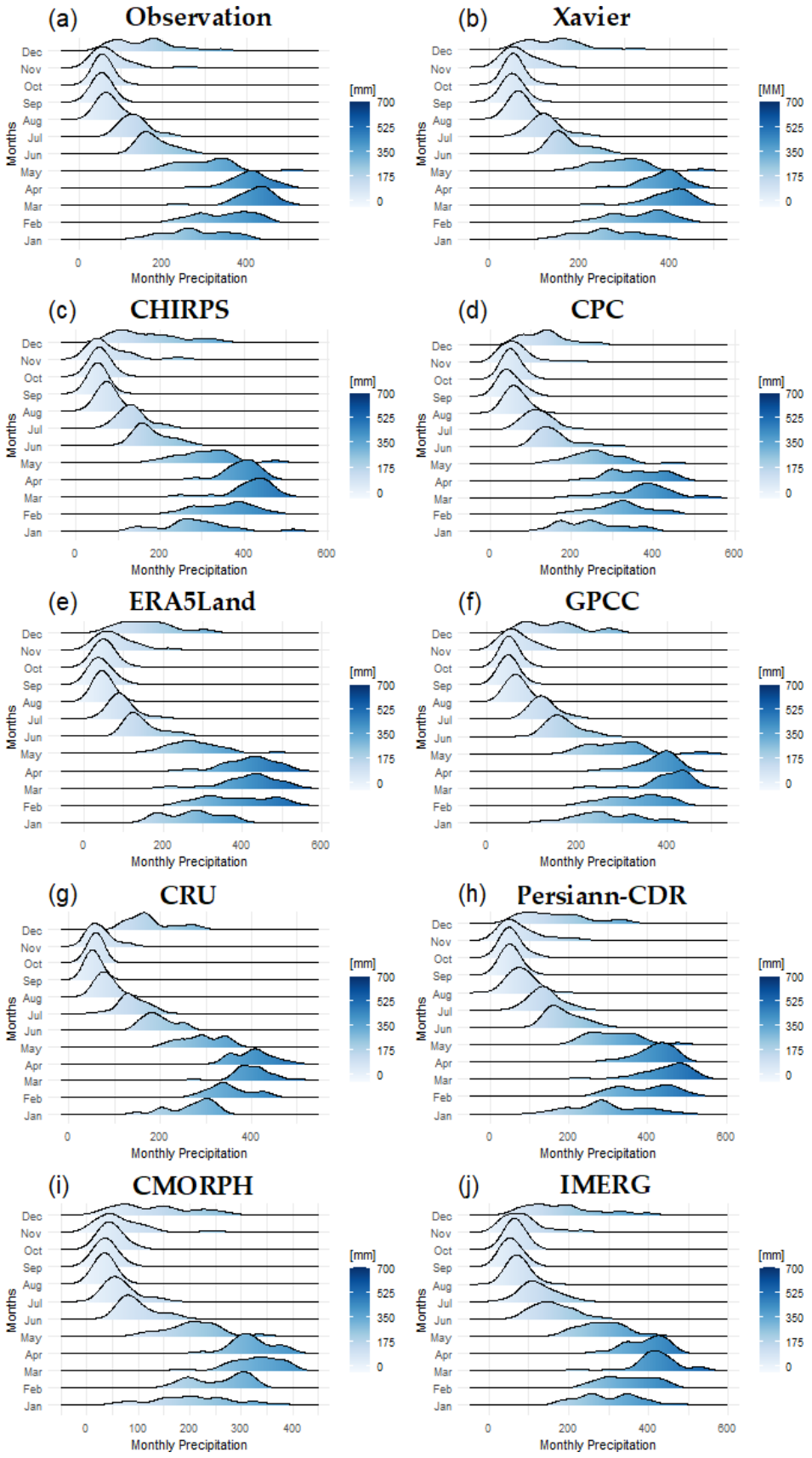

Figure S20. For Group 5, probability density of monthly rainfall (mm) for (a) Observations, (b) Xavier, (c) CHIRPS, (d) CPC, (e) ERA5Land, (f) GPCC, (g) CRU, (h) PERSIANN-CDR, (i) CMORPH and (j) IMERG.

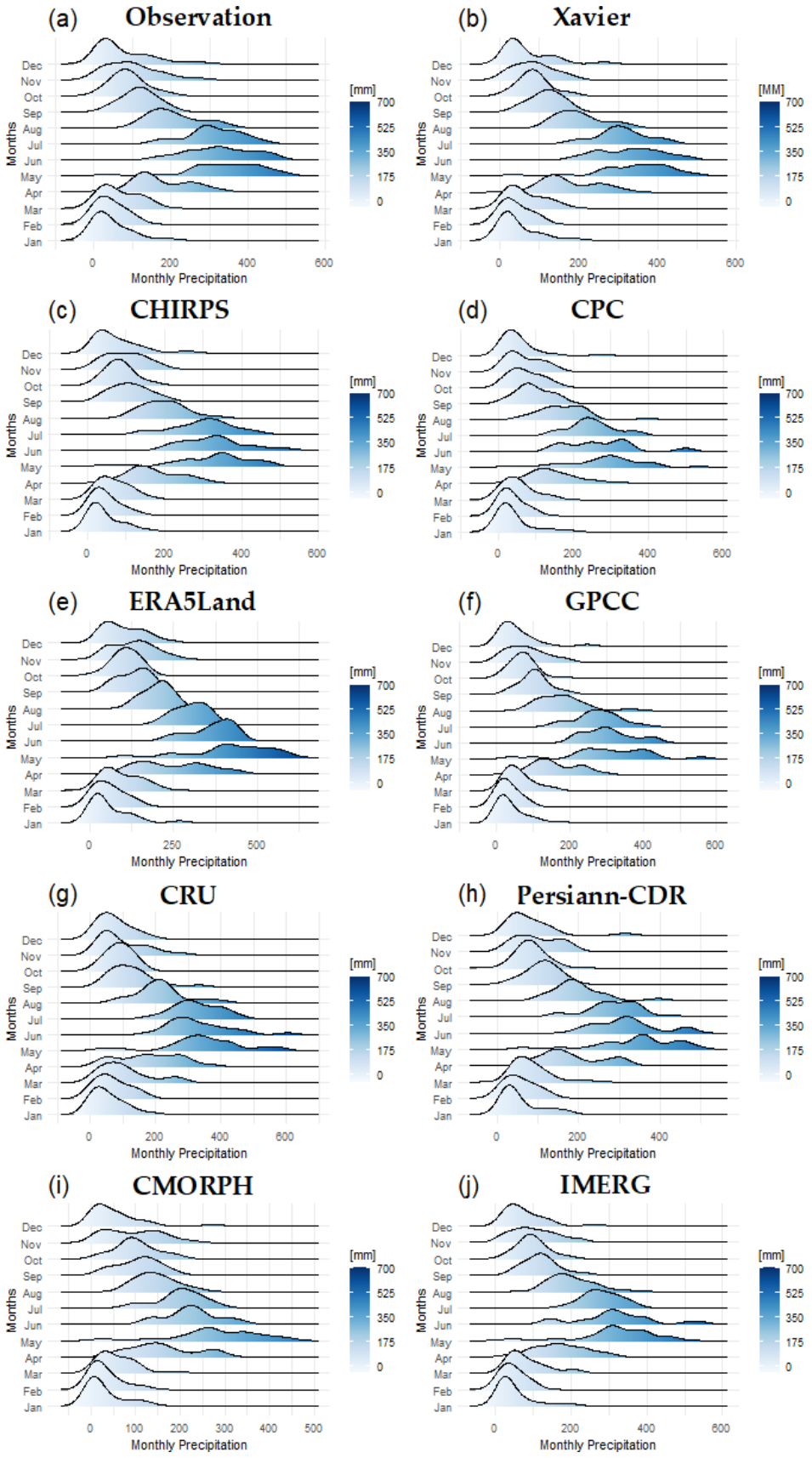

Figure S21. For Group 6, probability density of monthly rainfall (mm) for (a) Observations, (b) Xavier, (c) CHIRPS, (d) CPC, (e) ERA5Land, (f) GPCC, (g) CRU, (h) PERSIANN-CDR, (i) CMORPH and (j) IMERG.