A high-resolution regional climate model physics ensemble for Northern sub-Saharan Africa

Patrick Laux, Diarra Dieng, Tanja C. Portele, Jianhui Wei, Shasha Shang, Zhenyu Zhang, Joel Arnault, Christof Lorenz, Harald Kunstmann

Angaben zur Veröffentlichung / Publication details:

Laux, Patrick, Diarra Dieng, Tanja C. Portele, Jianhui Wei, Shasha Shang, Zhenyu Zhang, Joel Arnault, Christof Lorenz, and Harald Kunstmann. 2021. "A high-resolution regional climate model physics ensemble for Northern sub-Saharan Africa." *Frontiers in Earth Science* 9 (September): 700249. https://doi.org/10.3389/feart.2021.700249.

Supplementary Material

1 METHODOLOGIES

1.1 Ensemble Structure-Amplitude-Location (eSAL)

The ensemble Structure-Amplitude-Location (eSAL) analysis is conducted for an object-based spatial verification of the simulated precipitation fields, obtained from the 16 WRF physics parameterization runs in Northern sub-Saharan Africa (NSSA). This method compares the simulated ensemble, consisting of ensemble members from different cumulus-, radiation-, and planetary boundary layer scheme runs, with a reference field in terms of the amplitude A (the total precipitation of the domain), the location L (the location of the center of mass of the total domain and the location of the centers of mass of individual precipitation objects), and the structure S (the size, shape, or volume of the precipitation objects). The description of the eSAL method closely follows Portele et al. (2021) in this Special Issue.

First, any contiguous grid cells of precipitation above a given threshold are defined as an object. The threshold is calculated for each ensemble and reference data set separately, as $R^{95} \times f$, with R^{95} being the 95th percentile of all non-zero grid cell values in the domain for the current time step. For the simulations, R^{95}_{sim} is defined by all non-zero grid cell values of all ensemble members. f is a threshold value considered in the following calculations for S and L for daily values, given as:

$$f = \max\left(\frac{1}{15}, \frac{0.01\ \text{mm}}{R^{95}_{\text{sim}}}, \frac{0.01\ \text{mm}}{R^{95}_{\text{ref}}}\right). \tag{S1}$$

f is set to 1/15, except that a threshold value below 0.01 mm is obtained (Radanovics et al., 2018). Note that eSAL is only defined for non-zero precipitation fields and the different components of eSAL are all functions of the time. eSAL is calculated over all (daily) time steps of the study period (2006–2010) and eSAL values only exist if (i) the corresponding reference data is non-zero (i.e. precipitation exists in the field at all) and (ii) at least one ensemble member also produced precipitation. For sake of simplicity the time index is omitted in the following equations.

The ensemble amplitude error is defined as the relative difference of the ensemble mean ($\langle\rangle$) of the domain average precipitation ($\overline{rr}$) in the simulation ensemble and the reference field:

$$eA = \frac{\langle\overline{rr_{\text{sim}}}\rangle - \overline{rr_{\text{ref}}}}{0.5\left(\langle\overline{rr_{\text{sim}}}\rangle + \overline{rr_{\text{ref}}}\right)}. \tag{S2}$$

eA ranges from -2 to 2, with perfect agreement for $eA = 0$, dry biases for $eA < 0$, and wet biases for $eA > 0$. The ensemble structure error (eS) determines the relative difference of the ensemble mean ($\langle\rangle$) of the weighted averaged scaled precipitation volumes (V) in the simulation ensemble and the reference field:

$$eS = \frac{\langle V_{\text{sim}}\rangle - V_{\text{ref}}}{0.5\left(\langle V_{\text{sim}}\rangle + V_{\text{ref}}\right)}, \tag{S3}$$

with

$$V = \frac{\sum_i \left(rr_i \frac{rr_i}{rr_i^{\max}}\right)}{\sum_i rr_i}. \tag{S4}$$

Here, rr_i is the precipitation amount of all contiguous grid cells in object i and $rr_i^{\max}$ the maximum grid cell precipitation of this object. eS indicates if the ensemble is able to simulate the precipitation volumes. Similar to eA, eS ranges from -2 to 2, with equally scaled volumes for $eS = 0$, too small or too peaked simulated objects for $eS < 0$ and too large or too flat simulated objects for $eS > 0$.

The ensemble location error (eL) consists of two parts, relating to both the entire domain (eL_1) and individual objects (eL_2):

$$eL = eL_1 + eL_2, \tag{S5}$$

whereas eL_1 is the relative distance of the ensemble mean ($\langle\rangle$) centers of mass in the simulation ensemble, and the reference field:

$$eL_1 = \frac{|\langle \boldsymbol{x}\left(rr_{\text{sim}}\right)\rangle - \boldsymbol{x}\left(rr_{\text{ref}}\right)|}{d}, \tag{S6}$$

where d is the largest distance between two domain borders and $\boldsymbol{x}\left(rr\right)$ is the coordinate vector of the center of mass of all precipitation in the domain. eL_2 is finally specified as twice the squared distance between the cumulative distribution functions P, i.e., the *continuous ranked probability score* (CRPS), of the relative weighted average distances between the centers of mass of individual objects and the total center of mass in the simulation ensemble and in the reference field:

$$eL_2 = 2 \times \text{CRPS}\left[P\left(\frac{r_{\text{sim}}}{d}\right), P\left(\frac{r_{\text{ref}}}{d}\right)\right], \tag{S7}$$

with

$$r = \frac{\sum\limits_i rr_i|\boldsymbol{x}_i - \boldsymbol{x}|}{\sum\limits_i rr_i}, \tag{S8}$$

and

$$\text{CRPS}\left(P_{\text{sim}}, P_{\text{ref}}\right) = \int\limits_{-\infty}^{\infty} \left[P_{\text{sim}}(x) - P_{\text{ref}}(x)\right]^2 dx. \tag{S9}$$

Here, $\boldsymbol{x}_i$ is the coordinate vector of the center of mass of precipitation in the object i. For the reference field, the cumulative distribution function P_{ref} is a step-function. Both L_1 and L_2 range between 0 and 1. A value of $L = 0$ defines a perfect ensemble in terms of location. $L = 2$ would be found for total centers of mass located at the opposite domain border (eL_1 close to 1) and for contrarily organized objects, e.g., far from each other in one field and close to each other in the other field (eL_2 close to 1).

1.2 Empirical Copula of T&P pairs

Sklar's theorem (Sklar, 1959) states that any multivariate joint distribution F $(x_1, ..., x_n)$ from independent and identically distributed (iid) random variables $x_1, ..., x_n$ can be written in terms of univariate marginal distribution functions $F_{X_i}(x_i)$ and a Copula function. Thus, the Copula expresses the dependence structure between the variables and can be described with theoretical Copula functions (C_θ):

$$F_{x_1,...,X_n} = C_\theta(F_{X_1}(X_1), ..., F_{X_n}(X_n)); \quad C_\theta : [0,\ 1]^n \rightarrow [0,\ 1] \tag{S10}$$

with the multivariate distribution $F\,(x_1, ..., x_n)$, the theoretical Copula C_θ and the univariate marginals $F_{X_i}(x_i)$.

The multivariate PDF is then given through

$$f\,(x_1, ..., x_n) = c\,(f_{X_1}\,(x_1)\,, ..., f_{X_n}\,(x_n)) \cdot f_{X_1}\,(x_1) \cdot ... \cdot f_{X_n}\,(x_n) \tag{S11}$$

The dependency between two or more random variables is fully described by the Copula-PDF c and independent from the univariate marginal distributions f_X. In this study, however, we focus on the bivariate case of P&T data only. We compare the dependence structure of the modelled data obtained by the 16 parameterization runs which is obtained by using the observation stations data across entire NSSA. It is assumed that the observation stations represent well the entire modelling domain. The approach has already been applied in a similar manner to evaluate the performance of parameterization runs for Kenya (Laux et al., 2019). More detailed descriptions of the empirical Copula, also known as Deheuvels Copula, can be found e.g. in Deheuvels (1983); Mao et al. (2015).

Let $\{r_1(1), \ldots, r_1(n)\}$ and $\{r_2(1), \ldots, r_2(n)\}$ denote the rank space values that are derived from the fitted theoretical marginal distributions. Then the empirical Copula C_n is a rank based estimator of the theoretical Copula C_θ:

$$C_n(u, v) = 1/n \sum_{t=1}^{n} \mathbf{1}\left(\frac{r_1(t)}{n} \leqslant u, \frac{r_2(t)}{n} \leqslant v\right) \tag{S12}$$

with $u = F_X(x)$, $v = F_Y(y)$ and $\mathbf{1}(\ldots)$ denoting the indicator function and n the sample size. C_n is a discontinuous approximation of C_θ, to which it converges uniformly. Since it is completely non-parametric, C_n can be considered to be the most objective approximation of the underlying true Copula C_θ (e.g. Genest et al., 2009), which makes it a suitable candidate for GoF test statistics.

2 DATA AVAILABILITY

The 16 WRF parameterization runs (2006–2010) are made available at the *World Data Centre for Climate* (WDCC) of *German Climate Computing Center* (DKRZ) under the SaWaM project `https://cera-www.dkrz.de/WDCC/ui/cerasearch/q?query=SaWaM` to facilitate own evaluation studies.

Table S1. Full factorial WRF parameterization combinations, consisting of two cumulus parameterization (CU), two microphysic parameterization (MP), two planatery boundary layer parameterization (PBL), and two radiation parameterization (RA) schemes.

WRF ensemble member	CU	MP	PBL	RA	Reference
run1	G3D + shallow	WSM3	ACM2	RRTMG	Laux (2021a)
run2	G3D + shallow	WSM3	ACM2	RRTM + Dudhia	Laux (2021b)
run3	G3D + shallow	WSM3	YSU	RRTMG	Laux (2021c)
run4	G3D + shallow	WSM3	YSU	RRTM + Dudhia	Laux (2021d)
run5	G3D + shallow	WSM6	ACM2	RRTMG	Laux (2021e)
run6	G3D + shallow	WSM6	ACM2	RRTM + Dudhia	Laux (2021f)
run7	G3D + shallow	WSM6	YSU	RRTMG	Laux (2021g)
run8	G3D + shallow	WSM6	YSU	RRTM + Dudhia	Laux (2021h)
run9	Tiedtke	WSM3	ACM2	RRTMG	Laux (2021i)
run10	Tiedtke	WSM3	ACM2	RRTM + Dudhia	Laux (2021j)
run11	Tiedtke	WSM3	YSU	RRTMG	Laux (2021k)
run12	Tiedtke	WSM3	YSU	RRTM + Dudhia	Laux (2021l)
run13	Tiedtke	WSM6	ACM2	RRTMG	Laux (2021m)
run14	Tiedtke	WSM6	ACM2	RRTM + Dudhia	Laux (2021n)
run15	Tiedtke	WSM6	YSU	RRTMG	Laux (2021o)
run16	Tiedtke	WSM6	YSU	RRTM + Dudhia	Laux (2021p)

3 SUPPLEMENTARY FIGURES

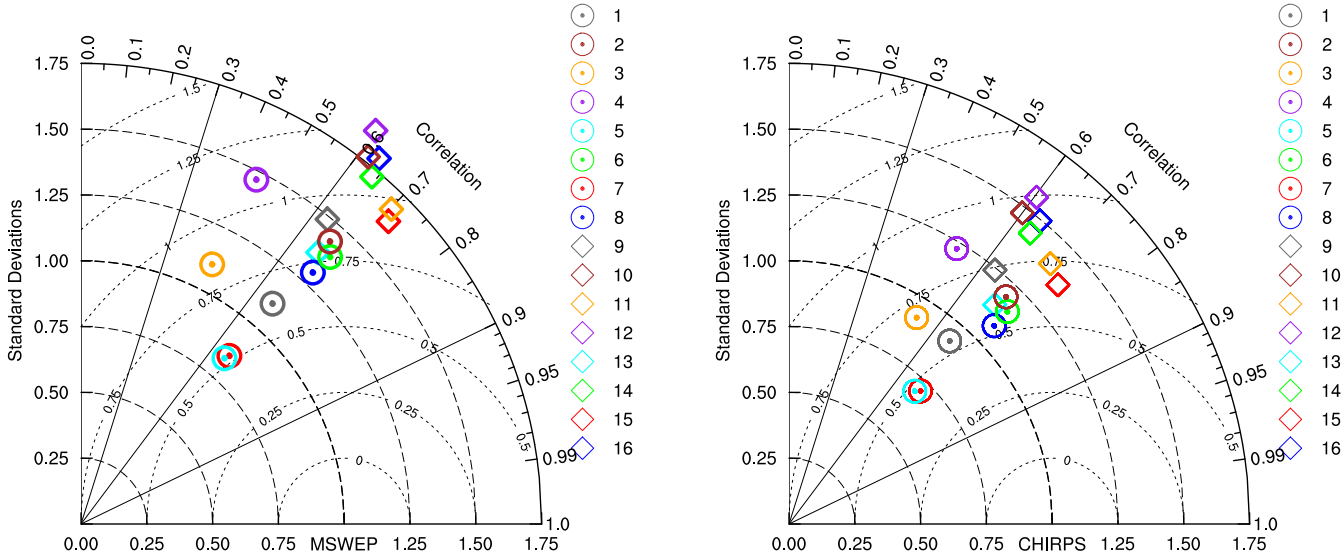

Figure S1: Taylor diagrams, depicting the Pearson correlation coefficient (straight lines), the normalized standard deviation (dashed lines) and the root-mean-square error (RMSE) (dotted lines) between simulated JJA precipitation amounts of the 16 ensemble members and MSWEP (top) and CHIRPS (bottom) reference data for the period 2006–2010.

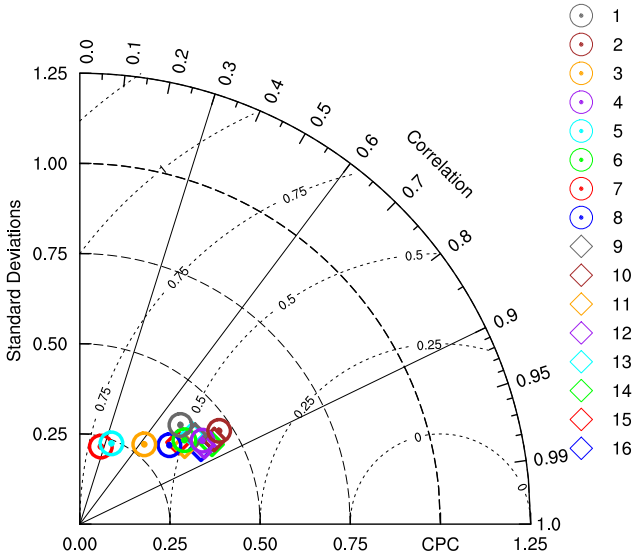

Figure S2: Taylor diagrams, depicting the Pearson correlation coefficient (straight lines), the normalized standard deviation (dashed lines) and the root-mean-square error (RMSE) (dotted lines) between simulated JJA temperature mean values of the 16 ensemble members CPC reference data for the period 2006–2010.

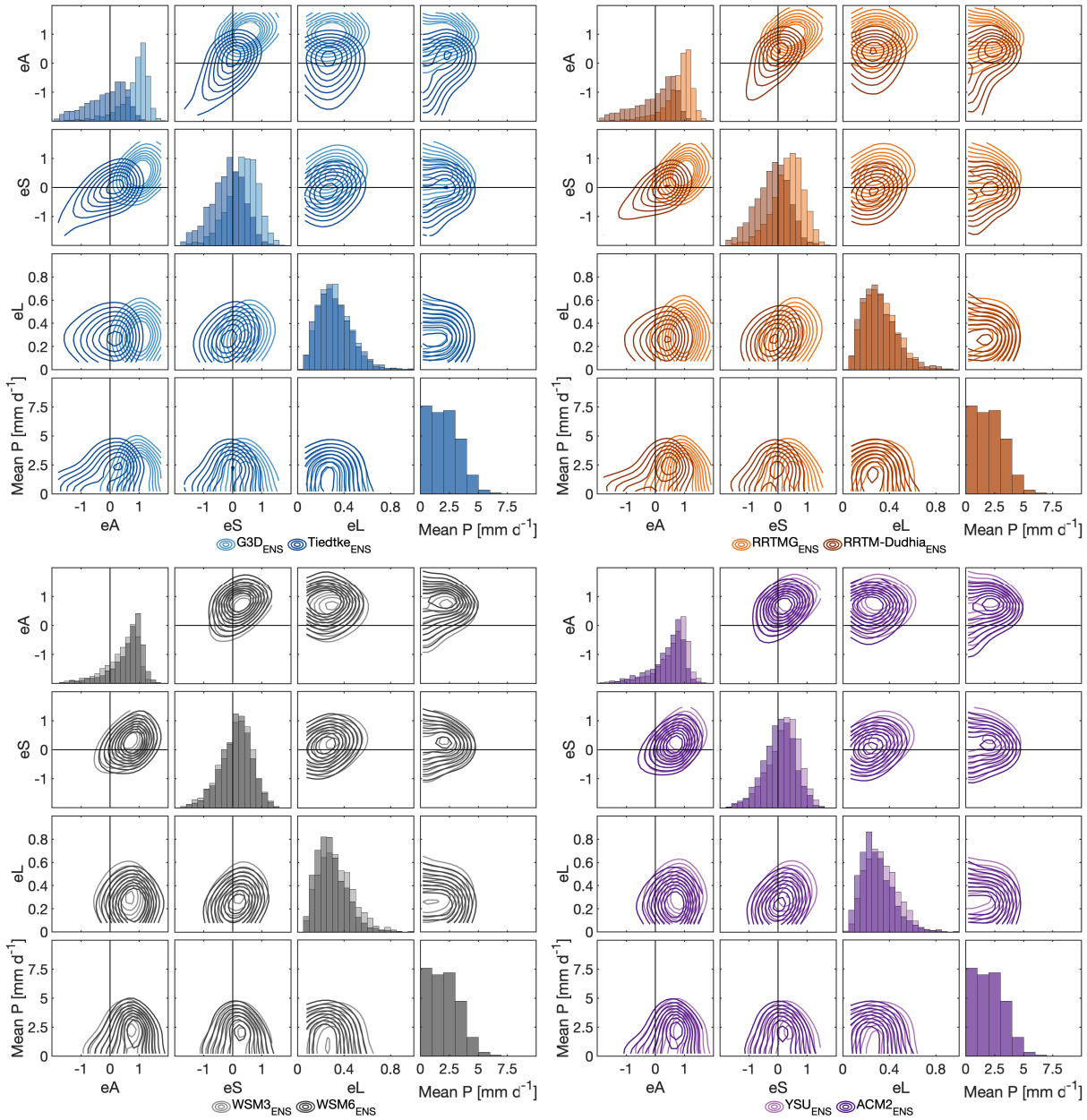

Figure S3: Validation of the domain-averaged precipitation (amplitude, A), the structure (volume, S), and the location (L) of precipitation objects for the ensembles (e), consisting of eight cumulus- (blue), radiation- (orange), microphysics- (grey), and planetary boundary layer (magenta) parameterization runs, respectively, for the period 2006–2010. The reference data is the MSWEP.

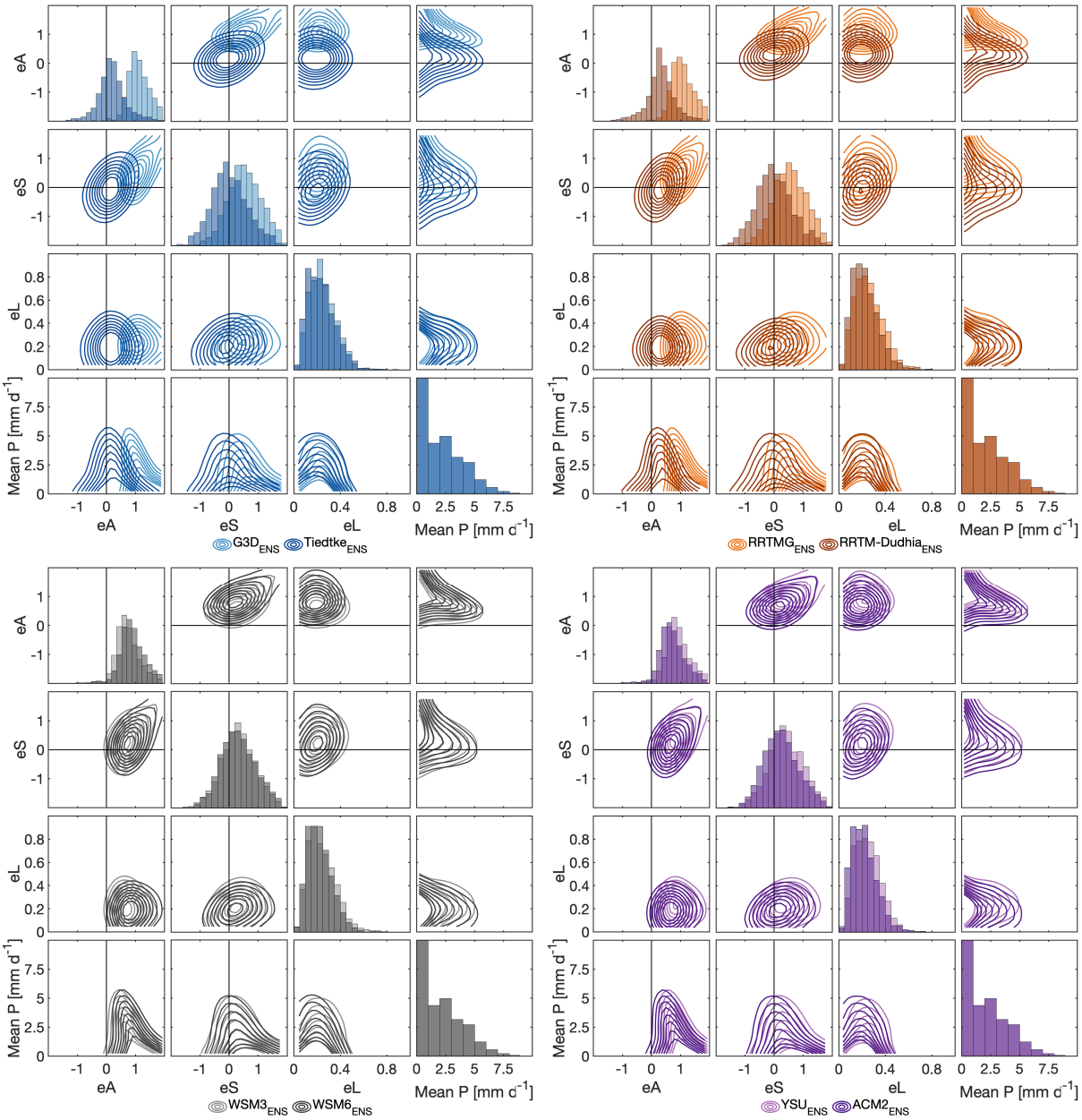

Figure S4: Same as for Figure S3, but using CHIRPS as reference data.

REFERENCES

Deheuvels, P. (1983). Point processes and multivariate extreme values. *Journal of Multivariate Analysis* 13, 257–272. doi:10.1016/0047-259X(83)90025-8

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics* 44, 199–213. doi:10.1016/j.insmatheco.2007.10.005

[Dataset] Laux, P. (2021a). SaWaM WRF physics parameterization scheme combination #1 (RUN1). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r1

[Dataset] Laux, P. (2021b). SaWaM WRF physics parameterization scheme combination #10 (RUN10). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r10

[Dataset] Laux, P. (2021c). SaWaM WRF physics parameterization scheme combination #11 (RUN11). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r11

[Dataset] Laux, P. (2021d). SaWaM WRF physics parameterization scheme combination #12 (RUN12). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r12

[Dataset] Laux, P. (2021e). SaWaM WRF physics parameterization scheme combination #13 (RUN13). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r13

[Dataset] Laux, P. (2021f). SaWaM WRF physics parameterization scheme combination #14 (RUN14). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r14

[Dataset] Laux, P. (2021g). SaWaM WRF physics parameterization scheme combination #15 (RUN15). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r15

[Dataset] Laux, P. (2021h). SaWaM WRF physics parameterization scheme combination #16 (RUN16). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r16

[Dataset] Laux, P. (2021i). SaWaM WRF physics parameterization scheme combination #2 (RUN2). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r2

[Dataset] Laux, P. (2021j). SaWaM WRF physics parameterization scheme combination #3 (RUN3). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r3

[Dataset] Laux, P. (2021k). SaWaM WRF physics parameterization scheme combination #4 (RUN4). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r4

[Dataset] Laux, P. (2021l). SaWaM WRF physics parameterization scheme combination #5 (RUN5). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r5

[Dataset] Laux, P. (2021m). SaWaM WRF physics parameterization scheme combination #6 (RUN6). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r6

[Dataset] Laux, P. (2021n). SaWaM WRF physics parameterization scheme combination #7 (RUN7). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r7

[Dataset] Laux, P. (2021o). SaWaM WRF physics parameterization scheme combination #8 (RUN8). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r8

[Dataset] Laux, P. (2021p). SaWaM WRF physics parameterization scheme combination #9 (RUN9). doi:10.26050/WDCC/SaWaM_WRF_phys_par_comb_r9

Laux, P., Kerandi, N., and Kunstmann, H. (2019). Physics parameterization selection in RCM and ESM simulations revisited: New supporting approach based on empirical copulas. *Atmosphere* 10. doi:10.3390/atmos10030150

Mao, G., Vogl, S., Laux, P., Wagner, S., and Kunstmann, H. (2015). Stochastic bias correction of dynamically downscaled precipitation fields for Germany through copula-based integration of gridded observation data. *Hydrology and Earth System Sciences* 11, 7189–7227. doi:10.5194/hessd-11-7189-2014

Portele, T. C., Laux, P., Lorenz, C., Janner, A., Horna, N., Fersch, B., et al. (2021). Ensemble-Tailored Pattern Analysis of High-Resolution Dynamically Downscaled Precipitation Fields: Example for Climate Sensitive Regions of South America. *Frontiers in Earth Science* 9, 1–23. doi:10.3389/feart.2021.669427

Radanovics, S., Vidal, J.-P., and Sauquet, E. (2018). Spatial Verification of Ensemble Precipitation: An Ensemble Version of SAL. *Weather and Forecasting* 33, 1001–1020. doi:10.1175/waf-d-17-0162.1

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Universite de Paris* 8, 229–231