Supplementary Information for:

Potential yield simulated by Global Gridded Crop Models: a process-based emulator to explain their differences

B.Ringeval*, C.Müller, T.A.M.Pugh, N.D.Mueller, P.Ciais, C.Folberth, W.Liu, P.Debaeke, S.Pellerin

* corresponding author: Bruno Ringeval, bruno.ringeval@inrae.fr

Supplementary Text

Text S1: Growing season as input/output of GGCMs

SMM also needs the begin/end of the growing season and we used respectively the planting day (t_p) and the timing of maturity (t_m), both being provided in the output of each GGCM. t_p is called *plant-day* in the GGCMI nomenclature and t_m is approached by *plant-day* + *maty-day* where *maty-day* is the number of days from planting to maturity in the GGCMI protocol. In *harmnon*, all GGCMs are forced by the same t_p and t_m (derived from a combination between MIRCA and SAGE). However, some GGCMs allow flexibility in regards to t_p and t_m prescribed as input (Müller et al., 2019), as suggested by the GGCMI protocol: *"crop variety parameters (e.g., required growing degree days to reach maturity, vernalization requirements, photoperiodic sensitivity) should be adjusted as much as possible to roughly match reported maturity dates"*. The comparison of t_p and t_m between GGCM input and GGCM output for some GGCMs (Fig.S4) suggests that we cannot use t_p and t_m of GGCM input to approach t_p and t_m of GGCM output when these latter are not available, i.e. for EPIC-BOKU, PEPIC, EPIC-TAMU and PEGASUS. Thus, these GGCMs have been excluded from our analysis.

Text S2: Implementation of heat stress

The effect of a heat stress on crop development is assessed by replacing the Eq.6 of the Main Text:

$$NPP_{biom}(d) = RUE * APAR(d) \quad \text{(Eq.6)}$$

with the following equation:

$$NPP_{biom}(d) = RUE * f_{heat}(d) * APAR(d) \quad \text{(Eq.6bis)}$$

with $f_{heat}(d)$=1 if $tas(d)$<30°C, $f_{heat}(d)$=0 if $tas(d)$>42°C and $f_{heat}(d)$ decreases linearly with $tas(d)$ from 1 to 0 for tas in [30, 42°C]. Eq.6bis is similar to the parametrization of heat stress in the EPIC model (Fig.S9).

References

van Bussel, L. G. J., Stehfest, E., Siebert, S., Müller, C. and Ewert, F.: Simulation of the phenological development of wheat and maize at the global scale: Simulation of crop phenology at global scale, Global Ecology and Biogeography, 24(9), 1018–1029, doi:10.1111/geb.12351, 2015.

Flénet, F., Kiniry, J. R., Board, J. E., Westgate, M. E. and Reicosky, D. C.: Row Spacing Effects on Light Extinction Coefficients of Corn, Sorghum, Soybean, and Sunflower, Agronomy Journal, 88(2), 185, doi:10.2134/agronj1996.00021962008800020011x, 1996.

Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L. B., Obersteiner, M. and van der Velde, M.: Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations, Nature Communications, 7(1), doi:10.1038/ncomms11872, 2016.

Folberth, C., Elliott, J., Müller, C., Balkovič, J., Chryssanthacopoulos, J., Izaurralde, R. C., Jones, C. D., Khabarov, N., Liu, W., Reddy, A., Schmid, E., Skalský, R., Yang, H., Arneth, A., Ciais, P., Deryng, D., Lawrence, P. J., Olin, S., Pugh, T. A. M., Ruane, A. C. and Wang, X.: Parameterization-induced uncertainties and impacts of crop management harmonization in a global gridded crop model ensemble, edited by J. M. Martínez-Paz, PLoS ONE, 14(9), e0221862, doi:10.1371/journal.pone.0221862, 2019.

Kiniry, J. R., Jones, C. A., O'toole, J. C., Blanchet, R., Cabelguenne, M. and Spanel, D. A.: Radiation-use efficiency in biomass accumulation prior to grain-filling for five grain-crop species, Field Crops Research, 20(1), 51–64, 1989.

Muchow, R. C.: Comparative productivity of maize, sorghum and pearl millet in a semi-arid tropical environment I. Yield potential, Field Crops Research, 20(3), 191–205, 1989.

Müller, C., Elliott, J., Kelly, D., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Hoek, S., Izaurralde, R. C., Jones, C. D., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Wang, X., de Wit, A. and Yang, H.: The Global Gridded Crop Model Intercomparison phase 1 simulation dataset, Scientific Data, 6(1), doi:10.1038/s41597-019-0023-8, 2019.

Plénet, D., Etchebest, S., Mollier, A. and Pellerin, S.: Growth analysis of maize field crops under phosphorus deficiency, Plant and Soil, 223(1–2), 119–132, 2000a.

Plénet, D., Mollier, A. and Pellerin, S.: Growth analysis of maize field crops under phosphorus deficiency. II. Radiation-use efficiency, biomass accumulation and yield components, Plant and Soil, 224(2), 259–272, 2000b.

Sangoi, L., Gracietti, M. A., Rampazzo, C. and Bianchetti, P.: Response of Brazilian maize hybrids from different eras to changes in plant density, Field Crops Research, 79(1), 39–51, 2002.

Sinclair, T. R. and Horie, T.: Leaf Nitrogen, Photosynthesis, and Crop Radiation Use Efficiency: A Review, Crop Science, 90–98, 1989.

Sinclair, T. R. and Muchow, R. C.: Radiation Use Efficiency, Advances in Agronomy, 215–265, 1999.

Testa, G., Reyneri, A. and Blandino, M.: Maize grain yield enhancement through high plant density cultivation with different inter-row and intra-row spacings, European Journal of Agronomy, 72, 28–37, doi:10.1016/j.eja.2015.09.006, 2016.

Zhu, P., Zhuang, Q., Archontoulis, S. V., Bernacchi, C. and Müller, C.: Dissecting the nonlinear response of maize yield to high temperature stress with model data integration, Global Change Biology, doi:10.1111/gcb.14632, 2019.

Supplementary Tables

Table S1: Parameter values during the calibration procedure

SMM parameter	Unit	Initial estimate	Range of values tested (in percent of the initial estimate)*	References used to constrain the initial estimates
T_0	°C	8	[50 - 150]	(van Bussel et al., 2015; Zhu et al., 2019)
GDD_{1leaf}	°C	43	[50 - 150]	Averaged value between leaf 4 and 18 for P1.5 treatment in (Plénet et al., 2000a)
max_{nleaf}	-	19	[50 - 150]	19-24 (Sangoi et al., 2002), 19 (Plénet et al., 2000b)
f	-	0.48	No variation	(Sinclair and Muchow, 1999)
C	-	0.12 ≈ 0.45*0.038 * 7	[50 - 150]	$C=k*S_{leaf}*d_{plant}$ with k: coefficient of extinction of radiation in canopy, S_{leaf}: individual leaf area and d_{plant}: plant density. k=0.45 (mean values over different row spacings: 0.4; mean values over different times of the day: 0.5) (Flénet et al., 1996) S_{leaf}=0.038 m² leaf^{-1} (average over the growing season for treatment P1.5 in (Plénet et al., 2000a)). d_{plant}=7 plants m^{-2} (derived from common values in USA (8) and in Europe (6-8) and from values tested in a field trial in Italia (7.5-12) and in Brazilia (2.5-10) (Sangoi et al., 2002; Testa et al., 2016)
RUE	g DM (MJ of absorbed PAR)$^{-1}$	2.0	[50 – 150]	3.1-4.0 (values representative to the whole growing season) (Sinclair and Muchow, 1999) ** 3.5 (value representative to the flowering) (Kiniry et al., 1989) 4.0 (Sinclair and Horie, 1989) ** 3.2 (values representative to the whole growing season) (Muchow, 1989) **
n_{thresh}	-	8	[0 – 200] ***	This study
$frac$	-	0.5	[80 – 160] ***	This study

* for a given parameter, 5 values are tested in the range provided in this column. These values are regularly distributed in this range, e.g. [50-150] corresponds to 50%, 75%, 100%, 125% and 150% of the initial guess.

** the values here given in [g DM (MJ of absorbed PAR)$^{-1}$] are derived from values in [g DM (MJ of intercepted solar radiation)$^{-1}$] given in the references and dividing by 0.425 following (Sinclair and Muchow, 1999)

*** n_{thersh} and $frac$ are correlated and the ranges used for their variation during the calibration aim to mimic all strategies of GGCMs: from i) GGCMs that start grain filling from emergence with very low fraction of NPP towards the grains to ii) GGCMs that start grain filling late in the growing season with large fraction of NPP towards the grains.

Supplementary Figures

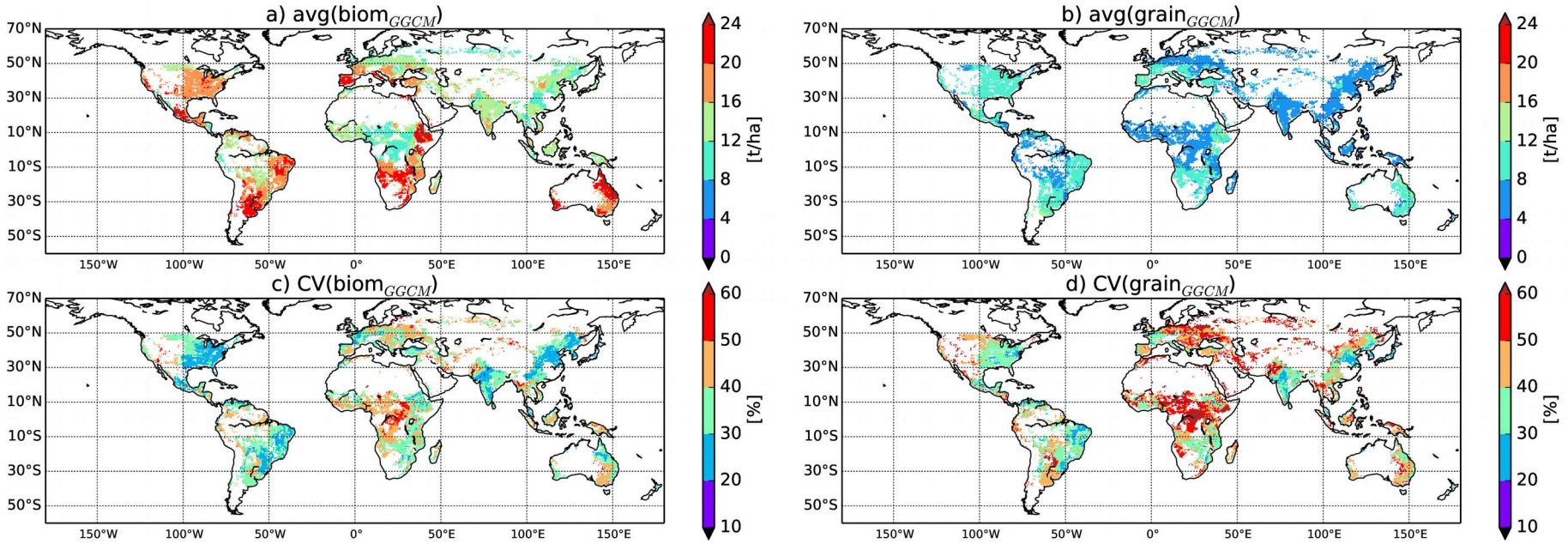

Fig.S1: Average (avg) and coefficient of variation (CV) for both aboveground biomass (*biom*) and yield (*grain*) of 11 GGCMs for simulations approaching potential yield in GGCMI (i.e. *harmnon* x *irrigated* for: LPJ-GUESS, LPJmL, CLM-crop, pDSSAT, pAPSIM, GEPIC, EPIC-IIASA, EPIC-TAMU, PEPIC, PEGASUS and *default* x *irrigated* for CGMS-WOFOST). For models of the EPIC family, the variable *biom* has been corrected (see Sect.2.2.1). EPIC-TAMU, ORCHIDEE-crop and PRYSBI2 participated to GGCMI but are not considered in this figure as: ORCHIDEE-crop did not provide *biom* in the GGCMI data archive, PRYSBI2 did not perform *harmnon* simulations and we did not succeed in getting the cultivar map for EPIC-TAMU required to correct *biom*.

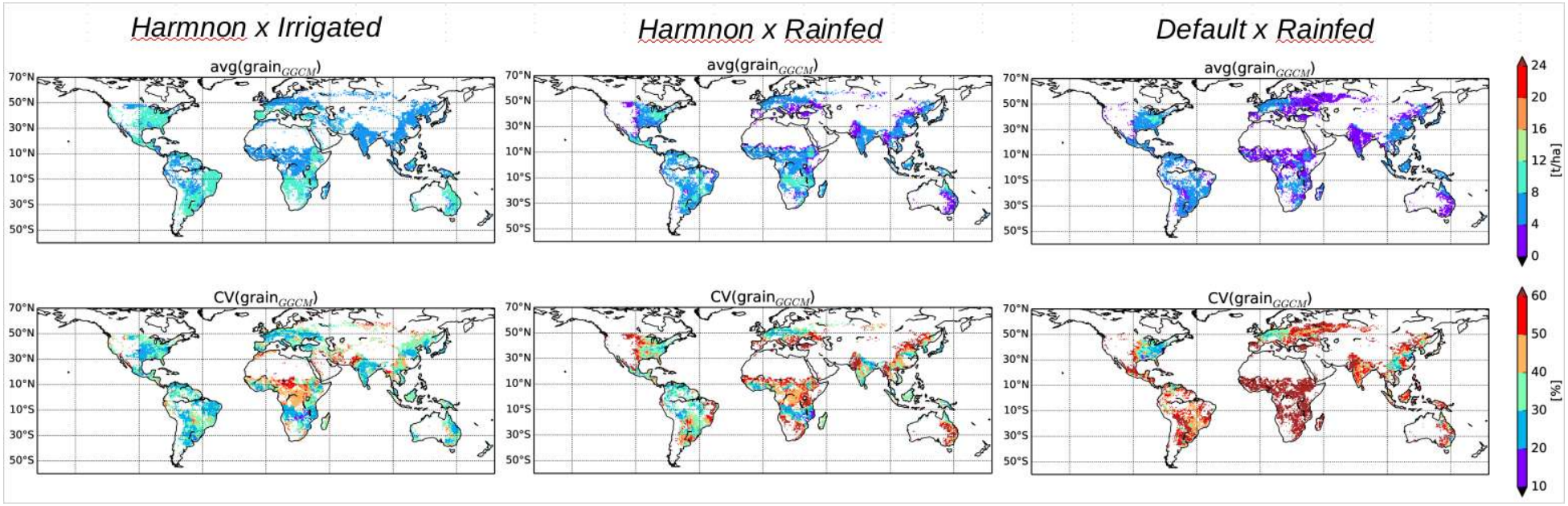

Fig.S2: GGCM divergence in yield simulated for different GGCMI simulations: *harmnon* and *irrigated* (left), *harmnon* and *rainfed* (middle) and *default* and *rainfed* (right column). The average (avg) and coefficient of variation (CV) of yield (*grain*) are computed among 8 GGCMs used in the current analysis (LPJ-GUESS, LPJmL, CLM-crop, pDSSAT, pAPSIM, CGMS-WOFOST, GEPIC and EPIC-IIASA). Note that *default* replace the *harmnon* configuration for CGMS-WOFOST (see Sect.2.2.1). Only grid-cells common to the 8 GGCMs are considered for the figure. The variable *biom* is not displayed as it is difficult to correct GEPIC and EPIC-IIASA in simulations where stresses (water, nutrient) occur (see Sect.2.2.1). Left column is similar to panels *b* and *d* of Fig.1.

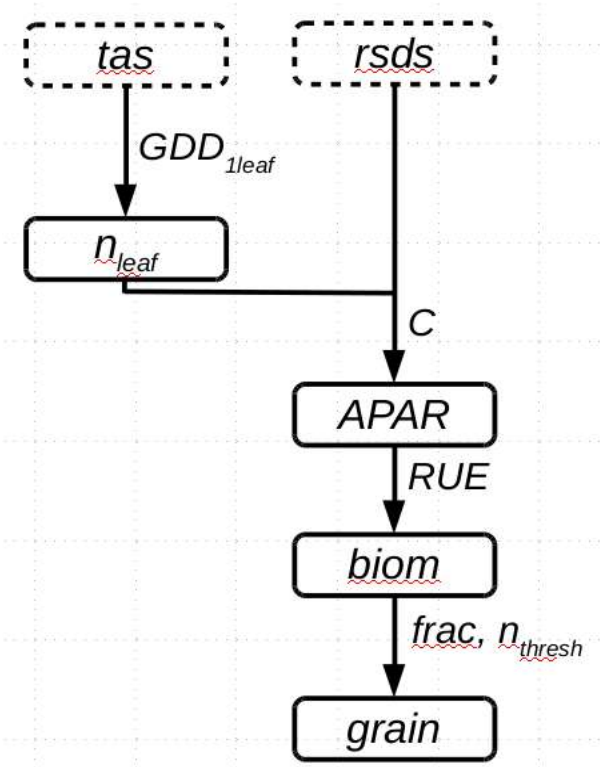

Fig.S3: Simplified flow chart of SMM. SMM variables are in solid boxes, input variables are in dashed boxes. Arrows represent relationship between variables. Parameters involved in the relationship between some variables are written on the right of the corresponding arrow. Only key variables are plotted. Only SMM parameters subject to the calibration procedure are given. The meaning of variables and parameters are given in Table 1.

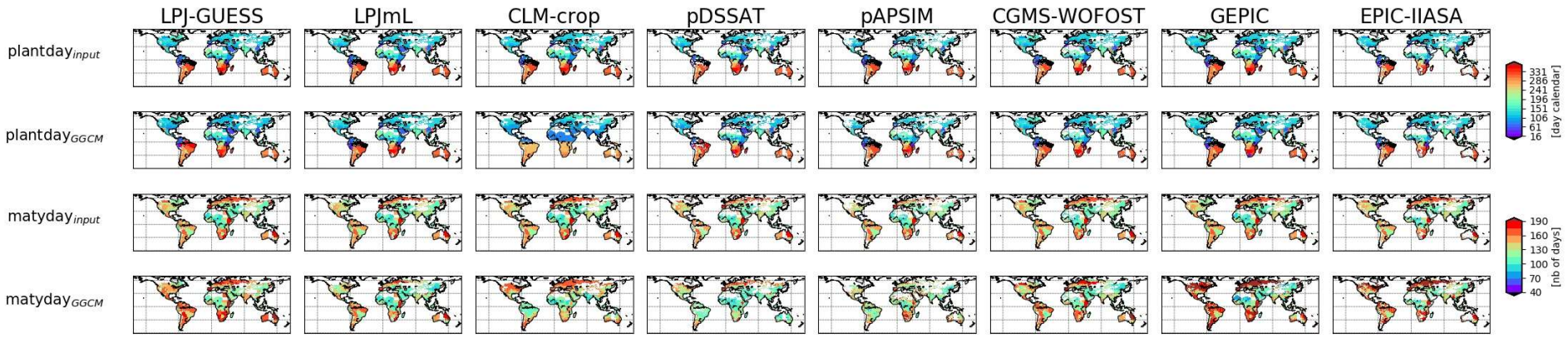

Fig.S4: Comparison of growing season between GGCM input and GGCM output. Variables plotted are planting day (*plant-day*, in [calendar day]; first two rows) and the length of the growing season (*matyday* [in days]; last two rows). GGCM input (1st and 3rd lines) are variables provided to GGCM modellers in the GGCMI protocol. Except some differences in the grid-cells considered, GGCM input are similar for all GGCMs. GGCM output (2nd and 4th lines) are variables provided in each GGCM output.

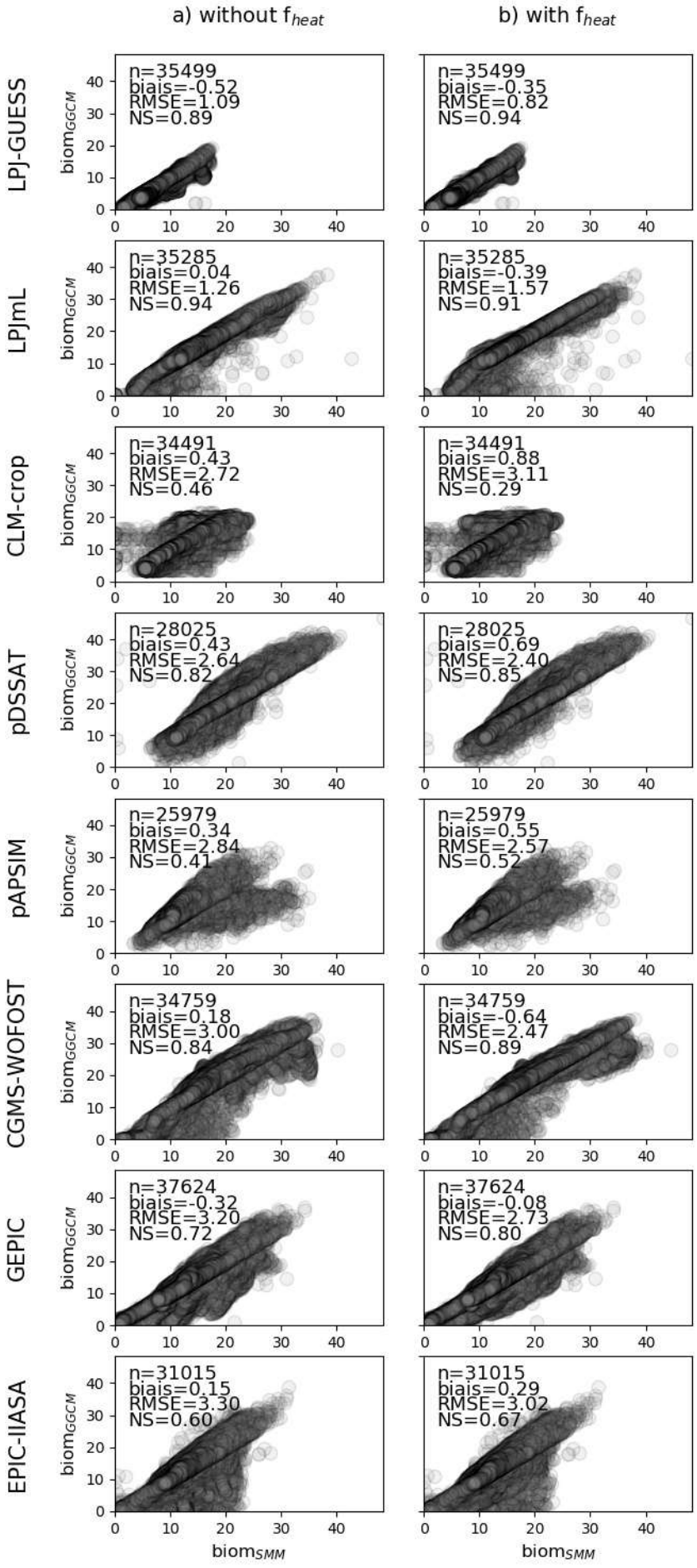

Fig.S5: $biom_{GGCM}$ vs. $biom_{SMM}$ and effect of the implementation of a heat stress. As in Fig.5, the figure displays scatter-plots of $biom_{GGCM}$ (y-axis) vs $biom_{SMM}$ (x-axis) for SMM simulations after calibration (i.e. calibration of global GGCM-dependent *C* and *RUE* and spatial varying GDD_{1leaf}). Each dot corresponds to one grid-cell. Scatter-plots are given without (panels a, similar to Fig.5a) and with implementation of a heat stress (panels b, see Text S1).

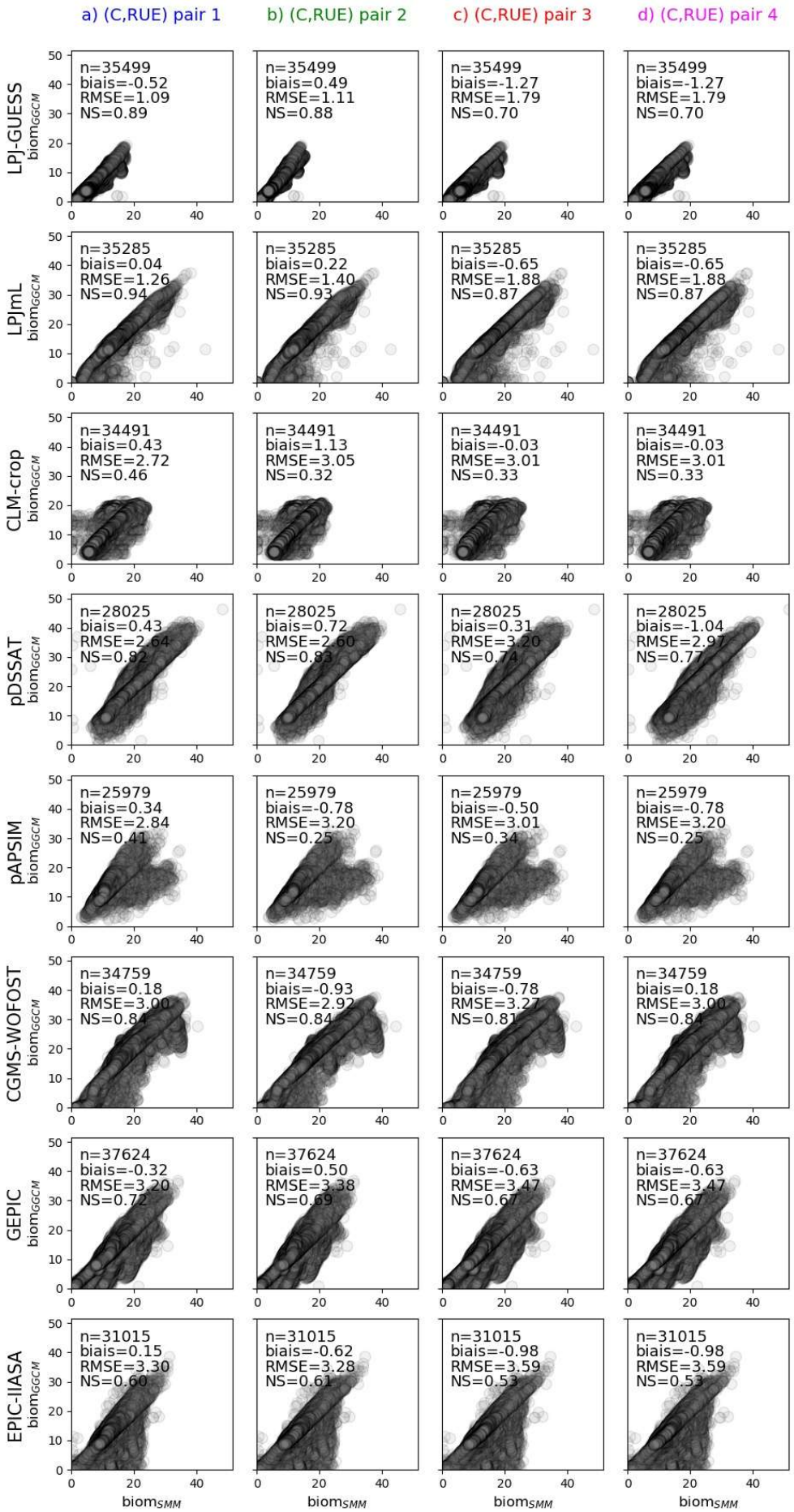

Fig.S6: $biom_{GGCM}$ vs. $biom_{SMM}$ and sensitivity to the chosen (*C*, *RUE*) pair. As in Fig.5, the figure displays scatter-plots of $biom_{GGCM}$ (y-axis) vs $biom_{SMM}$ (x-axis) for SMM simulations after calibration (i.e. calibration of global GGCM-dependent *C* and *RUE* and spatial varying GDD_{1leaf}). Each dot corresponds to one grid-cell. Scatter-plots are given for each (*C*, *RUE*) pair (corresponding to the different columns). Pair n°4 corresponds to the (*C*, *RUE*) minimizing the RMSE with *C* equal to its initial estimate. Pair n°1 (1st column) and pair n°4 (last column) correspond to panels a and b of Fig.5, respectively.

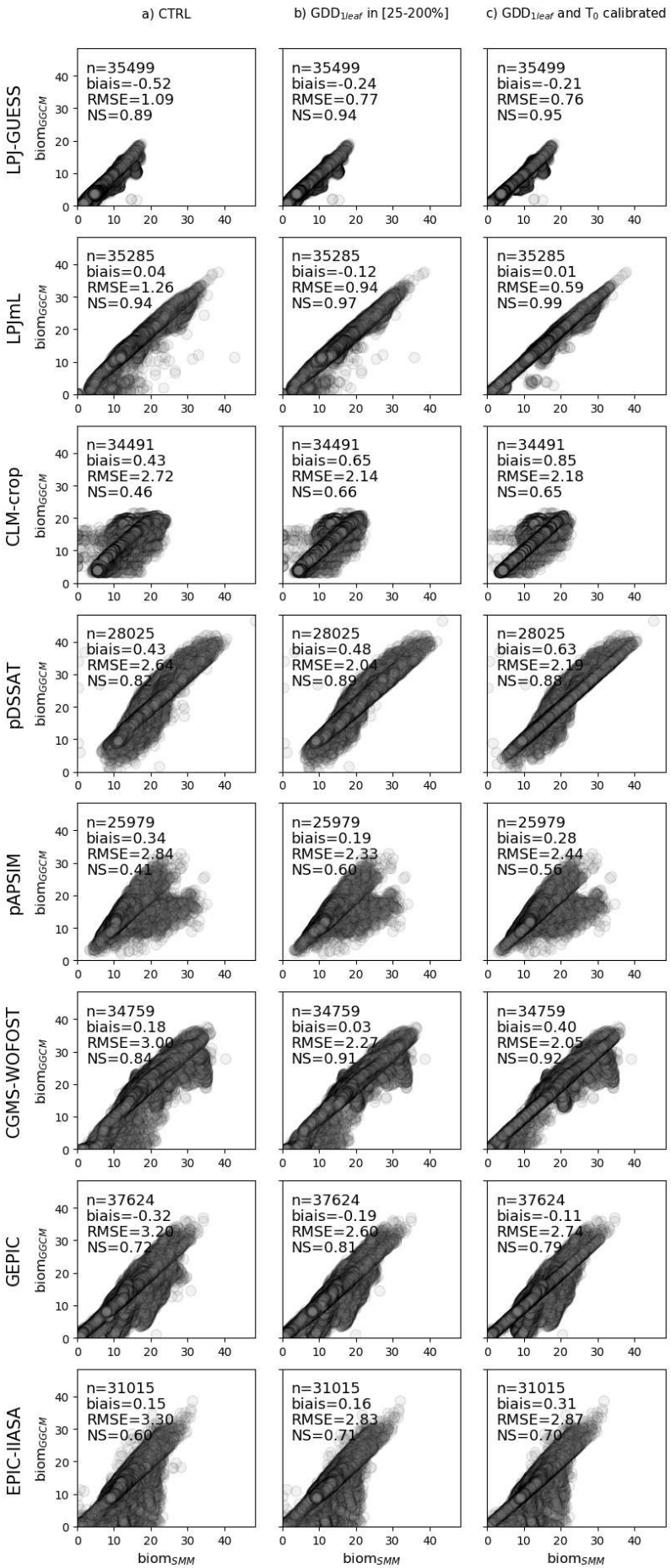

Fig.S7: $biom_{GGCM}$ vs. $biom_{SMM}$ for different calibrations. In 1st column (CTRL), GDD_{1leaf} is calibrated and the range of variation allowed during the calibration is [50-150%] of its initial estimate. This 1st column is similar to Fig.5a. In the 2nd column, the range of variation allowed for GDD_{1leaf} is increased ([25-200%] of its initial estimate). In the 3rd column, both GDD_{1leaf} and T_0 are allowed to vary ([50-150%] of their initial estimate) at the same time. In all cases, globally constant (*C*, *RUE*) have been first calibrated for each GGCM. During the calibration, the five values allowed for a given parameter are uniformly distributed within the range of variation (e.g. 50, 75, 100, 125, 150 for [50-150%]).

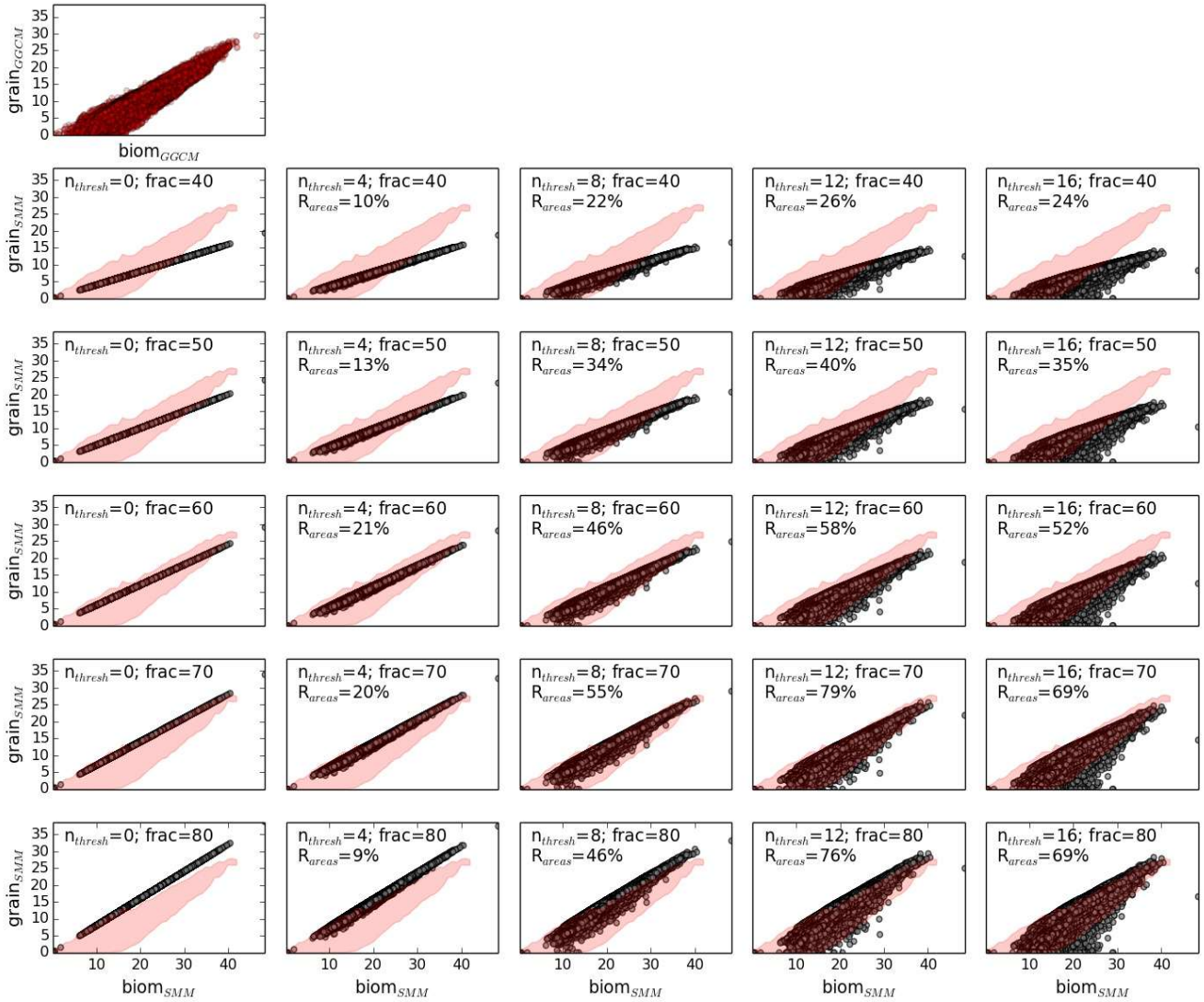

Fig.S8: Relationship $grain_{GGCM}$ vs $biom_{GGCM}$ (top panel) and comparison to $grain_{SMM}$ vs $biom_{SMM}$ for different (n_{thresh}, *frac*) combinations (*frac* varies with rows while n_{thresh} varies with columns). pDSSAT is here chosen as example. Pink area in panels related to SMM corresponds to A_{GGCM} (see Methods). The (n_{thresh}, *frac*) combination and R_{areas} criteria is given in top of each panel. The resolution along x-axis and y-axis in the computation of A_{GGCM} and A_{SMM} is 1 t DM ha^{-1}.

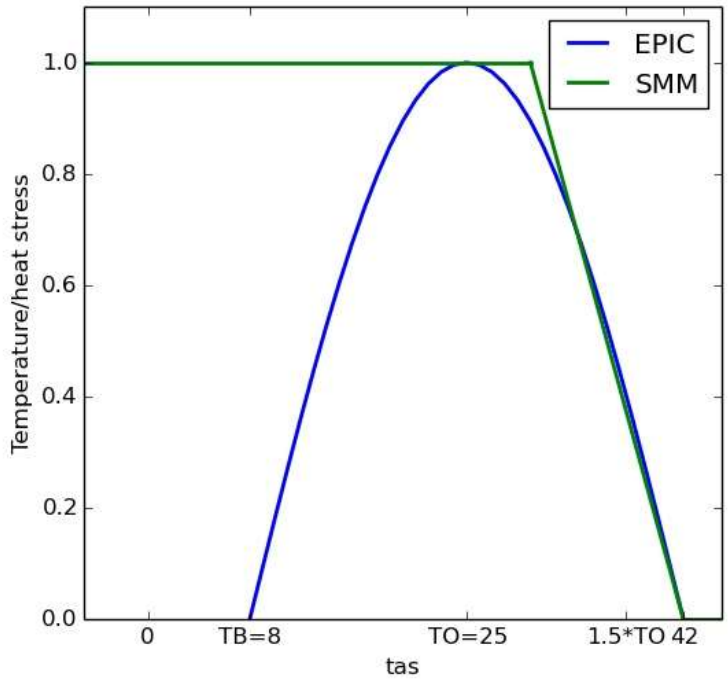

Fig.S9: Parametrization of temperature stress in EPIC models and of heat stress in SMM. The plot displays the temperature stress for EPIC models as function of the daily average temperature (*tas*) (blue). In EPIC models, temperature stress follows a sinus function and parameters are related to base temperature (TB) and optimum temperature (TO) (Eq.14 in (Folberth et al., 2016)). TB and TO vary as function of the cultivar and values used here corresponds to the default cultivar (cultivar 1 or "high-yielding variety", see Table D of (Folberth et al., 2019)). Parametrization used in SMM (Eq.6bis, green curve) mimics EPIC temperature stress for high temperatures (so called heat stress) and does not vary with cultivar.