Potential yield simulated by global gridded crop models: using a process-based emulator to explain their differences

Bruno Ringeval[1], **Christoph Müller**[2], **Thomas A. M. Pugh**[3], **Nathaniel D. Mueller**[4], **Philippe Ciais**[5], **Christian Folberth**[6], **Wenfeng Liu**[7], **Philippe Debaeke**[8], **and Sylvain Pellerin**[1]

[1]ISPA, Bordeaux Sciences Agro, INRAE, 33140, Villenave d'Ornon, France
[2]Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Potsdam, Germany
[3]School of Geography, Earth & Environmental Science and Birmingham Institute of Forest Research,
University of Birmingham, Birmingham, UK
[4]Department of Earth System Science, University of California, Irvine, CA, USA
[5]Laboratoire de Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ,
Université Paris-Saclay, 91191, Gif-sur-Yvette, France
[6]Ecosystem Services and Management Program, International Institute for Applied Systems Analysis,
2361 Laxenburg, Austria
[7]College of Water Resources and Civil Engineering, China Agricultural University, Beijing 100083, China
[8]AGIR, University of Toulouse, INRAE, 31326, Castanet-Tolosan, France

Correspondence: Bruno Ringeval (bruno.ringeval@inrae.fr)

Abstract. How global gridded crop models (GGCMs) differ in their simulation of potential yield and reasons for those differences have never been assessed. The GGCM Intercomparison (GGCMI) offers a good framework for this assessment. Here, we built an emulator (called SMM for simple mechanistic model) of GGCMs based on generic and simplified formalism. The SMM equations describe crop phenology by a sum of growing degree days, canopy radiation absorption by the Beer–Lambert law, and its conversion into aboveground biomass by a radiation use efficiency (RUE). We fitted the parameters of this emulator against gridded aboveground maize biomass at the end of the growing season simulated by eight different GGCMs in a given year (2000). Our assumption is that the simple set of equations of SMM, after calibration, could reproduce the response of most GGCMs so that differences between GGCMs can be attributed to the parameters related to processes captured by the emulator. Despite huge differences between GGCMs, we show that if we fit both a parameter describing the thermal requirement for leaf emergence by adjusting its value to each grid-point in space, as done by GGCM modellers following the GGCMI protocol, and a GGCM-dependent globally uni-

form RUE, then the simple set of equations of the SMM emulator is sufficient to reproduce the spatial distribution of the original aboveground biomass simulated by most GGCMs. The grain filling is simulated in SMM by considering a fixed-in-time fraction of net primary productivity allocated to the grains (frac) once a threshold in leaves number (n_{thresh}) is reached. Once calibrated, these two parameters allow for the capture of the relationship between potential yield and final aboveground biomass of each GGCM. It is particularly important as the divergence among GGCMs is larger for yield than for aboveground biomass. Thus, we showed that the divergence between GGCMs can be summarized by the differences in a few parameters. Our simple but mechanistic model could also be an interesting tool to test new developments in order to improve the simulation of potential yield at the global scale.

Published by Copernicus Publications on behalf of the European Geosciences Union.

1   Introduction

Potential yield corresponds to the yield achieved when an adapted crop cultivar grows in non-limiting environmental conditions (i.e., without water and nutrient stresses and in the absence of damages from weeds, pests, and diseases) under a given crop management (e.g., plant density). Fundamentally, it is determined by a reduced number of environmental variables: prevailing radiation, temperature, and atmospheric CO_2 concentration. Biotic variables such as cultivar characteristics (e.g., maturity group, leaf area index, root depth, harvest index), plant density and sowing date modulate how the environmental conditions are converted into yield. At the local scale (field, farm, or small region), potential yields can be estimated from field experiments, yield census, or crop growth models (Lobell et al., 2009). Crop simulation models provide a robust approach because they account for the interactive effects of genotype, weather, and management (van Ittersum et al., 2013). These models are mathematical representations of our current understanding of biophysical crop processes (phenology, carbon assimilation, assimilate allocation) and of crop responses to environmental factors. Such models have been designed to separate genotype–environment–management interactions, for example via factorial simulations where one driver is varied at a time. Models require site-specific inputs, such as daily weather data, crop management practices (sowing date, cultivar maturity group, plant density, fertilization, and irrigation amounts and dates), and soil characteristics, with some of them being not useful for the purpose of simulating potential yield. At the local scale, crop models can be calibrated to account for local specificities, in particular for specificities related to the cultivar used at these sites (Grassini et al., 2011)

Potential yield is also a variable of interest at large (country, global) spatial scales, in particular as it is required for yield gap analyses (van Ittersum et al., 2013; Lobell et al., 2009). Such analyses are necessary to get a large-scale picture of yield limitations and to investigate questions related to production improvement strategies, food security, and management of resources with a global perspective. However, while crop models used at local scale can be calibrated to account for local specificities, it is much more complicated to model the spatial variations of yields at the global scale. Local crop models have been applied at the global scale either directly or through the implementation of some of their equations into global vegetation models (Elliott et al., 2015). Global Gridded Crop Models (GGCMs in the following) provide spatially explicit outputs, typically at half-degree resolution in latitude and longitude. Their simulations are prone to uncertainty. In particular, it is quite difficult to get reliable information about the diversity of cultivars (Folberth et al., 2019) or crop management at the global scale with large effects on the crop behavior (Drewniak et al., 2013).

Increasing our confidence in potential yield simulated by GGCMs is required to improve our ability to reply to societal questions mentioned above. To do so, we first need to understand how and why GGCMs potentially diverge in potential yield simulation. The GGCM Intercomparison (GGCMI phase I) provides a framework relevant to the investigation of the differences between GGCMs, as all GGCM modellers followed the same protocol (Elliott et al., 2015). Model outputs are available on the GGCMI data archive (Müller et al., 2019). In the GGCMI framework, a simulation performed with a harmonized growing period, absence of nutrient stress, and irrigated conditions (see below) is particularly adapted to simulate potential yield. Figure 1 displays the average and coefficient of variation (CV) of such simulated aboveground biomass (biom) and yield (grain) for maize among eight GGCMs participating in GGCMI that have been used in our following analysis. While the GGCM divergence under potential conditions is lower than the GGCM divergence when limiting factors are represented (Fig. S2 in the Supplement), it remains relatively high. Figure 1 shows that the CV in potential conditions is higher for grain than for biom and that the CV for grain can locally reach more than 50 %. To understand what could explain these differences, we built a mechanistic emulator of GGCMs based on generic processes controlling the accumulation of biomass (phenology from the sum of growing degree days, light absorption, radiation use efficiency) and the transformation of biomass into grain yield (trigger of yield formation, allocation of net primary production, NPP, to yield). We then calibrated the parameters of the emulator independently for each GGCM against GGCM-simulated biom and the GGCM-simulated relationship between biom and grain. Our assumption is that a simple set of equations, with calibrated parameters, can reproduce the outputs of most GGCMs and could be used to explore the sources of differences between them. We choose to use a process-based (if very simple) model, as we expect that this model could propose interesting perspectives as explained in Sect. 4. In particular, if able to reproduce the results of an ensemble of GGCMs, it could be an alternative to the model ensemble mean or median usually used in intercomparison exercises (Martre et al., 2015). Running much faster than GGCMs, it would also be an interesting tool to test new developments, such as the implementation of cultivar diversity, to improve the simulation of potential yield at the global scale.

2   Methods

2.1   GGCM emulator

For any given day d of the growing season (defined here as the period between the planting day, t_p, and the crop maturity, t_m), we used the Eqs. (1)–(7), which rely on concepts commonly used in modeling of ecosystem productivity to compute the potential aboveground biomass (biom, in $t\,DM\,ha^{-1}$, with $1\,ha = 0.01\,km^2$). Variables and parameters are summa-
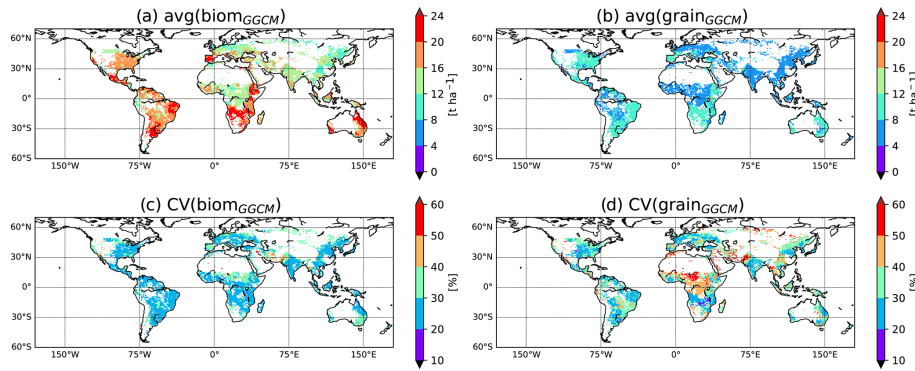
Figure 1. GGCM divergence in simulation of potential aboveground biomass and yield. Average and coefficient of variation for both aboveground biomass (biom) and yield (grain) computed among eight GGCMs used in the current analysis for simulations approaching potential yield in GGCMI (i.e., *harmnon × irrigated* for LPJ-GUESS, LPJmL, CLM-crop, pDSSAT, pAPSIM, GEPIC, and EPIC-IIASA and *default × irrigated* for CGMS-WOFOST; see Sect. 2.2.1). Only grid cells common to the eight GGCMs are considered for this figure. For GEPIC and EPIC-IIASA, the variable biom has been corrected (see Sect. 2.2.1). For the comparison including GGCMs that provide biom and grain for *harmnon × irrigated* simulation in GGCMI but which are not part of the current paper (i.e., EPIC-BOKU, PEPIC, and PEGASUS), see Fig. S1 in the Supplement.

rized in Table 1, and a simplified flow chart is given in Fig. S3 in the Supplement.

For any d in $[t_p, t_m]$, the thermal time (TT) is computed from the daily mean temperature (tas, in °C) by using a reference temperature (T_0):

$$\text{TT}(d) = \text{tas}(d) - T_0. \tag{1}$$

GDD is the sum of growing degree days, defined as follows:

$$\text{GDD}(d) = \sum_{i \leq d} \max(0, TT(i)). \tag{2}$$

The number of leaves per plant (n_{leaf}) is computed from GDD and one parameter representing the thermal requirement for the emergence of any leaf ($\text{GDD}_{1\text{leaf}}$):

$$n_{\text{leaf}}(d) = \min\left(\max_{\text{nleaf}}, \text{GDD}(d)/\text{GDD}_{1\text{leaf}}\right), \tag{3}$$

where $\max_{\text{nleaf}}$ is the maximum number of leaves. In our model, as soon as one leaf emerges, we assumed that it reaches its fully expanded leaf area, which is the same for all leaves (individual leaf area, called S_{leaf} hereafter). The incoming photosynthetic active radiation (PAR_{inc}) is derived from the short-wave downwelling radiation (rsds in $\text{MJ}\,\text{m}^{-2}\,\text{d}^{-1}$) and its active fraction, f:

$$\text{PAR}_{\text{inc}}(d) = f \times \text{rsds}(d). \tag{4}$$

The absorbed PAR by the canopy (APAR) is determined by assuming an exponential function according to the Beer–Lambert law:

$$\text{APAR}(d) = \text{PAR}_{\text{inc}}(d) \times (1 - \exp(-C \times n_{\text{leaf}}(d))), \tag{5}$$

where C is a constant (see below). The net primary productivity dedicated to the aboveground biomass (NPP_{biom}) is computed from APAR with a constant radiation use efficiency (RUE in $\text{g}\,\text{DM}\,\text{MJ}^{-1}$):

$$\text{NPP}_{\text{biom}}(d) = \text{RUE} \times \text{APAR}(d). \tag{6}$$

The aboveground biomass corresponds to the sum over time of NPP_{biom}:

$$\text{biom}(d) = \sum_{i \leq d} (\text{NPP}_{\text{biom}}(i)). \tag{7}$$

The parameter C of Eq. (5) can be decomposed in different parameters:

$$C = k \times S_{\text{leaf}} \times d_{\text{plant}}, \tag{8}$$

where k is the light extinction coefficient, S_{leaf} is the individual leaf area, and d_{plant} is the plant density. The product of S_{leaf}, d_{plant} and the number of leaves of a given day d (i.e., $n_{\text{leaf}}(d)$), corresponds to the leaf area index (LAI) of the same day:

$$\text{LAI}(d) = S_{\text{leaf}} \times d_{\text{plant}} \times n_{\text{leaf}}(d), \tag{9}$$

in such a way that Eq. (5) can be re-written as follows:

$$\text{APAR}(d) = \text{PAR}_{\text{inc}}(d) \times (1 - \exp(-k \times \text{LAI}(d))) \quad \text{(5bis)},$$

We preferred Eq. (5) over Eq. (5bis) as we cannot separately calibrate the different parameters composing C and because we do not have any information about the LAI from GGCMs (see Sect. 4).

To compute the grain biomass at maturity, we first define the day l as follows:

$$n_{\text{leaf}}(l) \geq n_{\text{thresh}}, \tag{10}$$

Table 1. List of variables and parameters in SMM.

	Definition	Unit	Status
tas	Daily mean temperature	°C	Input variable
TT	Thermal time	°C	Internal variable
GDD	Sum of growing degree days	°C	Internal variable
n_{leaf}	Number of leaves per plant	–	Internal variable
rsds	Short-wave downward radiation	$\text{MJ m}^{-2}\,\text{d}^{-1}$	Input variable
PAR_{inc}	Incoming photosynthetic active radiation	$\text{MJ m}^{-2}\,\text{d}^{-1}$	Internal variable
APAR	Canopy-absorbed incoming PAR	$\text{MJ m}^{-2}\,\text{d}^{-1}$	Internal variable
NPP_{biom}	Net primary productivity dedicated to aboveground biomass	$\text{g DM m}^{-2}\,\text{d}^{-1}$	Internal variable
biom	Aboveground biomass	g DM m^{-2}	Internal variable. The study focuses on biom at the end of the growing season, called biom_{SMM}, and converted to t DM ha^{-1}.
$\text{NPP}_{\text{grain}}$	Net primary productivity dedicated to grains	$\text{g DM m}^{-2}\,\text{d}^{-1}$	Internal variable
grain	Grain biomass (yield)	g DM m^{-2}	Internal variable. The study focuses on grain at the end of the growing season, called $\text{grain}_{\text{SMM}}$, and converted to t DM ha^{-1}
T_0	Zero of vegetation	°C	Parameter
$\text{GDD}_{\text{1leaf}}$	Sum of growing degree day required for each leaf (phyllochron)	°C	Parameter
$\text{max}_{\text{nleaf}}$	Maximum number of leaves per plant	–	Parameter
f	Active fraction of short-wave downward radiation	–	Fixed parameter ($f = 0.48$)
C	$C = k \times S_{\text{leaf}} \times d_{\text{plant}}$ with k: coefficient of extinction of radiation in canopy; S_{leaf}: the specific leaf area of any leaf and d_{plant}: the plant density	–	Parameter
RUE	Radiation use efficiency	g DM MJ^{-1} (Here, MJ refers to absorbed PAR)	Parameter
n_{thresh}	Number of leaves from which the formation of reproductive structures starts, the grains form, or the grain filling starts	–	Parameter
frac	Fraction of NPP_{biom} going towards the variable grain when $n_{\text{leaf}} > n_{\text{thresh}}$	–	Parameter

From day l, a fixed fraction (frac) of NPP_{biom} constitutes the net primary productivity dedicated to the variable grain (called $\text{NPP}_{\text{grain}}$):

$$\begin{cases} \text{if } d \geq l, & \text{NPP}_{\text{grain}}(d) = \text{frac} \times \text{NPP}_{\text{biom}}(d), \quad (11) \\ \text{if } d < l, & \text{NPP}_{\text{grain}}(d) = 0. \quad (12) \end{cases}$$

and finally,

$$\text{grain}(d) = \sum_{i \leq d} \left(\text{NPP}_{\text{grain}}(i) \right). \quad (13)$$

The variable grain (in t DM ha^{-1}) could be considered either as reproductive structures + grains or grains only. The parameter n_{thresh} is a threshold in the number of leaves from which either the formation of reproductive structures starts,

the grains form, or grain filling starts. The above equations aim to be generic and to reproduce the diversity of approaches in GGCMs. That is why we do not distinguish between the production of reproductive structures and the accumulation of assimilates in grains after anthesis.

Equations (1)–(7) and (10)–(13) are called SMM (for simple mechanistic model) in the following. We focused on biom and grain at maturity, i.e., computed on the last day of the growing season t_{m}. They are called biom_{SMM} and $\text{grain}_{\text{SMM}}$

in the following:

$$\text{biom}_{\text{SMM}} = \text{biom}(t_{\text{m}}),\tag{14}$$

$$\text{grain}_{\text{SMM}} = \text{grain}(t_{\text{m}}).\tag{15}$$

The variable $\text{grain}_{\text{SMM}}$ is used to approach the potential yield. Our analysis focuses on maize because of the importance of cereals in human food and because of the widespread distribution of this crop across latitudes.

2.2 Set-up

We focused first on the computation of biom_{SMM} and then on the relationship $\text{grain}_{\text{SMM}}$ vs. biom_{SMM}. The sensitivity of SMM to each parameter involved in the computation of biom_{SMM} was first studied. Then, we calibrated SMM against each GGCM to make the spatial distribution of biom_{SMM} mimic the spatial distribution of aboveground biomass at maturity simulated by each GGCM (hereafter called $\text{biom}_{\text{GGCM}}$). This calibration happened in two steps. The first step concerned C and RUE, which have one value at the global scale. The second step concerned $\text{GDD}_{1\text{leaf}}$, which we made varying in space to mimic procedure used by GGCM modellers in GGCMI (see below). The choice of focusing on C, RUE, and $\text{GDD}_{1\text{leaf}}$ is justified below. In the last step (step 3), we calibrated n_{thresh} and frac to make SMM mimic the relationship grain vs. biom of each GGCM.

2.2.1 GGCMs and GGCMI simulations considered

The eight GGCMs considered in our approach were LPJ-GUESS (Lindeskog et al., 2013; Smith et al., 2001), LPJmL (Bondeau et al., 2007; Waha et al., 2012), CLM-crop (Drewniak et al., 2013), pDSSAT (Elliott et al., 2014; Jones et al., 2003), pAPSIM (Elliott et al., 2014; Keating et al., 2003), CGMS-WOFOST (Boogaard et al., 2014), GEPIC (Williams et al., 1995; Folberth et al., 2012; Izaurralde et al., 2006; Liu et al., 2007), EPIC-IIASA (Williams et al., 1995; Izaurralde et al., 2006). GGCMs simulations are provided in the framework of the GGCM Intercomparison (GGCMI) and described in Müller et al. (2019). GGCMI is an activity of the of the Agricultural Model Intercomparison and Improvement Project (AgMIP; Rosenzweig et al., 2013) and is an element of a broader AgMIP effort to explore cropping system responses to climate conditions and climate changes to facilitate applications including toward integrated assessment (Ruane et al., 2017). Six other GGCMs also participated in GGCMI but were not considered here as necessary output variables (timing and duration of the growing season for EPIC-BOKU, PEPIC, EPIC-TAMU; aboveground biomass for ORCHIDEE-crop) or simulations (for PRYSBI2) were not provided on the data archive of GGCMI.

In GCCMI, all GGCMs followed a common protocol and were forced by the same weather datasets. We focused here on the simulations forced by one of them, the AgMERRA dataset (Ruane et al., 2015). We used simulations forced by the AgMERRA dataset as all GGCMs performed these simulations. Three levels of harmonization have been used in GGCM simulations: *default*, *fullharm*, *harmnon*. In *fullharm* simulations, all GGCMs have been forced by the same prescribed beginning and end of the growing season, which were derived from a combination of two global datasets (MIRCA, Portmann et al., 2010; and SAGE, Sacks et al., 2010; Elliott et al., 2015). In *harmnon* simulations, in addition to forced timing and duration of the growing season, all GGCMs experienced no nutrient limitation through prescribed fertilizer inputs. Besides this harmonization level, two water regimes have been considered: *irrigated* and *non-irrigated*. For our analysis focusing on the simulation of potential yield, we decided to select the configuration (*harmnon* and *irrigated*). This is true for all GGCMs considered except CGMS-WOFOST. In fact, the *harmnon* simulation was not provided for CGMS-WOFOST but because (i) this model does not consider nutrient limitation and (ii) the growing season was prescribed in the *default* simulation, we assumed that the potential yield could be approached by the (*default* and *irrigated*) simulation.

For EPIC family models (here, GEPIC and EPIC-IIASA), we used a corrected $\text{biom}_{\text{GGCM}}$ computed as below, as it has been noticed that some issues related to the variable biom appeared in the outputs available on the GGCMI data archive that are likely related to the output time step of specific variables (Folberth, personal communication, 2019):

$$\text{biom}_{\text{GGCM}} = \text{grain}_{\text{GGCM}}/\text{HI}_{\text{max}},\tag{16}$$

where HI_{max} is the maximum harvest index (no unit), varying in space as a function of cultivars. In EPIC, the actual HI at harvest only differs to HI_{max} if a drought stress occurs in the reproductive phase. Because this stress was virtually eliminated by sufficient irrigation in the *harmnon* × *irrigated* simulations, Eq. (16) provides the most accurate estimate of aboveground biomass at harvest. Maps of cultivar distribution, used as input to the EPIC models (Fig. 1 and Table D in Folberth et al., 2019), have been considered here to compute the corrected $\text{biom}_{\text{GGCM}}$.

2.2.2 Input variables for SMM

We focused our analysis on the growing season starting in calendar year 2000 (and potentially finishing in calendar year 2001). SMM was forced by the short-wave downwelling radiation (rsds in $\text{MJ m}^{-2}\text{d}^{-1}$) and the daily mean temperature (tas, in °C) from the AgMERRA weather dataset (Ruane et al., 2015). SMM also needs the beginning and end of the maize growing season, and we used the planting day (t_{p}) and the timing of maturity (t_{m}), respectively, both being provided in the output of each GGCM. Despite the fact that all GGCMs are forced by the same growing season in *harmnon*, some GGCMs allow flexibility in regards to t_{p} and t_{m} prescribed as input (Müller et al., 2017), as suggested by the GGCMI protocol: "crop variety parameters (e.g., required

growing degree days to reach maturity, vernalization require-
ments, photoperiodic sensitivity) should be adjusted as much
as possible to roughly match reported maturity dates". Thus,
we cannot use t_p and t_m from GGCM input files (Text S1 in
the Supplement).

We performed SMM simulations (and thus, computed
$biom_{SMM}$ and $grain_{SMM}$) for each GGCM, i.e., for each
GGCM growing season. For a given GGCM, SMM simu-
lations were performed only for grid cells considered in the
given GGCM. In addition, grid cells for which information
about the growing season from MIRCA and SAGE was not
available are masked to prevent to consider grid cells where
internal GGCM computation was performed.

The maps of cultivar distribution used by EPIC models
(Folberth et al., 2019) were also used as inputs to SMM in the
simulation aiming to mimic the biom vs. grain relationship of
EPIC models (see Sect. 2.2.4).

2.2.3 Sensitivity of global $biom_{SMM}$ to SMM parameters

Except the active fraction of short-wave downward radiation
(f in Eq. 4) the value of which is physically well-known,
other parameters involved in the computation of $biom_{SMM}$
(T_0, max_{nleaf}, GDD_{1leaf}, C, RUE) are relatively uncertain.
The sensitivity of the global averaged $biom_{SMM}$ to these pa-
rameters was assessed by performing 3125 (i.e., 5^5) SMM
simulations, allowing for the combination of five different
values for each parameter. In each of these SMM simula-
tions, a given parameter was constant in space. The initial
estimate of each parameter was provided in Table S1. While
the initial estimate of each parameter was based on the liter-
ature, we chose quite arbitrarily the same range of variation
of [50 %–150 %] (in % of the initial estimate) for all param-
eters, with the five values tested uniformly distributed within
the range of variation (i.e., 50 %, 75 %, 100 %, 125 %, 150 %
of initial guess).

Following our current knowledge based on observations,
it would be partly possible to choose a different uncertainty
range for each parameter: for instance, the literature tends
to show that RUE is relatively well constrained for maize
(Sinclair and Muchow, 1999), while the C parameter, which
depends on plant density, is expected to vary a lot as a func-
tion of the farming practices (Sangoi et al., 2002; Testa et al.,
2016) (Table S1). However, SMM aimed to mimic GGCMs
and not observations, and it is quite difficult to know if pa-
rameter values used in GGCMs reflect our current knowl-
edge well. For instance, there is some confusion in values
of RUE reported in the literature following the diversity of
experimental approaches and units of expression that have
been used (Sinclair and Muchow, 1999). Some confusion in
RUE values exists in the literature between RUE expressed in
$g DM MJ^{-1}$ of intercepted solar radiation (called here RUE′)
or in $g DM MJ^{-1}$ of intercepted PAR (RUE″, with RUE″ =
RUE′/0.5) or in $g DM MJ^{-1}$ of absorbed PAR (RUE‴, with

RUE‴ = RUE′/0.425) (Sinclair and Muchow, 1999), and
this could lead to erroneous values in GGCMs. In the follow-
ing we used MJ of absorbed PAR to be consistent with our
Eq. (6). Observations also showed that RUE decreases during
grain filling following the mobilization of leaf nitrogen to the
grain (Sinclair and Muchow, 1999). Thus, RUE is larger dur-
ing vegetative growth (3.8–4.0 g DM $(MJ of absorbed PAR)^{-1}$;
Kiniry et al., 1989) than averaged over the whole season
(3.1–4.0 g DM MJ^{-1}; Sinclair and Muchow, 1999). It is likely
that some GGCMs used RUE values which are not represen-
tative to the whole growing season. Thus, we used the same
range of uncertainty for all parameters in our calibration pro-
cedure. Exploring a wider range of values also allows for a
more complete assessment of GGCM performance.

Potential confusion in units mentioned above also led
us to choose an initial estimate of RUE (2 g DM MJ^{-1})
lower than values commonly reported in the literature (3.1–
4.0 g DM MJ^{-1}) (Sinclair and Muchow, 1999); however, note
that the highest values of RUE tested during our calibration
(3.0 g DM MJ^{-1}) reach the literature-based range.

The global mismatch between each GGCM and SMM
was quantified thanks to the root-mean-square error (RMSE),
computed as follows:

$$RMSE(u) \qquad\qquad (17)$$
$$= \sqrt{\frac{1}{N}\sum_{g=1}^{N}(biom_{SMM}(u,g) - biom_{GGCM}(g))^2},$$

where u is a combination of parameters and g is a grid cell
among the N grid cells considered for each GGCM. All grid
cells are assumed independent and have the same weight in
the RMSE computation. RMSE has the same unit as biom
($t DM ha^{-1}$).

2.2.4 SMM calibration against each GGCM

SMM was calibrated following three steps. The first two
steps aimed to mimic $biom_{GGCM}$ distribution, while the last
step aimed to make SMM reproduce the relationship grain vs.
biom of each GGCM. The procedure of calibration was sum-
marized in Table 2. "Emulated GGCM" is used from now to
define SMM output after SMM calibration, aiming to mimic
a given GGCM.

Parameters involved in the computation of $biom_{SMM}$

Regarding the simulation of biom, we restricted the cal-
ibration to RUE, C, and GDD_{1leaf} as follows: f is well
constrained, max_{nleaf} has a small effect on global simu-
lated $biom_{SMM}$ (see the results of the analysis prescribed
in Sect. 2.2.3), and T_0 co-varies with GDD_{1leaf}, and we de-
cided to focus on GDD_{1leaf} (see below). These parameters
(f, max_{nleaf}, T_0) were prescribed equal to their initial esti-
mate and were the same for all SMM simulations.

We choose to make C and RUE globally constant and
GGCM-dependent. The decision to not make C and RUE

Table 2. Strategy of SMM calibration for each parameter.

Step of calibration	Parameters	Values used in SMM simulations	GGCM variable used for the calibration
	f, T_0, $\max_{\mathrm{nleaf}}$	One value at the global scale and same for all GGCMs	None
1	C, RUE	One value at the global scale and GGCM-dependent	Global averaged biom
3	n_{thresh}, frac*		Relationship grain vs. biom
2	$\mathrm{GDD}_{1\mathrm{leaf}}$	Variable in space and GGCM-dependent variability	Spatial variability of biom

* n_{thresh} and frac are variable in space as a function of the cultivar when SMM aims to mimic EPIC family models as these latter consider some cultivar diversity in harvest index.

vary in space is consistent with the rule of parsimony, which we used to build SMM. It also follows the procedure commonly used in GGCMs that involved a similar approach. For instance, GEPIC is based on a biomass-energy conversion coefficient that does not vary in space (Folberth et al., 2016). Plant density (hidden in C) is constant in space in LPJmL (Schaphoff et al., 2018a). We calibrated C and RUE at the same time to assess potential compensation between these parameters in SMM. The three pairs (C, RUE) that minimized the global RMSE computed following Eq. (17) the most among the pairs tested were chosen. A fourth pair corresponding to (C, RUE), where C is equal to its initial estimate, has been used. The use of four different pairs aimed to assess the sensitivity of our conclusions to the parameter values. For each (C, RUE) pair, we finally calibrated $\mathrm{GDD}_{1\mathrm{leaf}}$. We made $\mathrm{GDD}_{1\mathrm{leaf}}$ vary in space as it is allowed in the GGCMI exercise. In the GGCMI protocol, accumulated thermal requirements were adjusted to catch the growing season (duration and timing) prescribed as input in the *harmnon* GGCM simulation. In SMM, the procedure slightly differs as we calibrated thermal requirements to match $\mathrm{biom}_{\mathrm{GGCM}}$: for each grid cell, $\mathrm{GDD}_{1\mathrm{leaf}}$ is chosen among its five possible values to minimize the absolute difference between $\mathrm{biom}_{\mathrm{GGCM}}$ and $\mathrm{biom}_{\mathrm{SMM}}$. Grid cells were considered independently.

The ability of SMM to match the spatial distribution of $\mathrm{biom}_{\mathrm{GGCM}}$ for each GGCM after SMM calibration was measured through the bias, RMSE, and Nash–Sutcliffe model efficiency coefficient (NS), defined as follows:

$$\mathrm{NS} = 1 - \frac{\sum_{g=1}^{N} (\mathrm{biom}_{\mathrm{SMM}}(g) - \mathrm{biom}_{\mathrm{GGCM}}(g))^2}{\sum_{g=1}^{N} \left(\mathrm{biom}_{\mathrm{GGCM}}(g) - \overline{\mathrm{biom}_{\mathrm{GGCM}}}\right)^2}, \quad (18)$$

where g refers to any grid cell and $\overline{\mathrm{biom}_{\mathrm{GGCM}}}$ is the average of $\mathrm{biom}_{\mathrm{GGCM}}$ over grid cells. NS = 1 means that SMM perfectly matches the spatial distribution of $\mathrm{biom}_{\mathrm{GGCM}}$.

To assess the mismatch between $\mathrm{biom}_{\mathrm{GGCM}}$ and $\mathrm{biom}_{\mathrm{SMM}}$ after SMM calibration for a given GGCM, we aimed to assess how a variable related to climate or soil type can contribute to this mismatch. To do so, we separated all grid cells within two sub-groups according to the value of this variable (e.g., one sub-group corresponding to grid cells with high temperatures and one sub-group with low temperatures) and assess if the RMSE is different for the two sub-groups. If yes, it would suggest that a process related to this variable (e.g., heat stress) could be missing in SMM.

Parameters involved in the computation of grain$_{\mathrm{SMM}}$

C, RUE, and $\mathrm{GDD}_{1\mathrm{leaf}}$ determine biom simulated by SMM at each time step. Two SMM parameters are involved in the computation of grain for any day from biom, namely n_{thresh} and frac. The calibration of these parameters aims to make SMM able to mimic the relationship between grain and biom at the end of the growing season from each GGCM. One global and GGCM-dependent pair (n_{thresh}, frac) was chosen by using the following criteria:

$$\begin{cases} \text{if } A_{\mathrm{GGCM}} = 0, \text{ find } u \text{ that minimizes} & (19) \\ \quad R_{\mathrm{slope}}(u) = |a_{\mathrm{GGCM}} - a_{\mathrm{SMM}}(u)|, \\ \text{if } A_{\mathrm{GGCM}} \neq 0, \text{ find } u \text{ that maximizes} & (20) \\ \quad R_{\mathrm{areas}}(u) = \left(\frac{A_{\mathrm{GGCM}} \cap A_{\mathrm{SMM}}(u)}{\max(A_{\mathrm{GGCM}}, A_{\mathrm{SMM}}(u))}\right), \end{cases}$$

where u corresponds to a given pair (n_{thresh}, frac), A_X is the area defined by the grid cell clouds in the grain vs. biom space for X, and a_X is the slope of the linear regression $\mathrm{grain}_X \sim \mathrm{biom}_X$, with X in {GGCM, SMM}.

In other words, if $\mathrm{grain}_{\mathrm{GGCM}}$ vs. $\mathrm{biom}_{\mathrm{GGCM}}$ is a line, (n_{thresh}, frac) is chosen to make the relationship between $\mathrm{grain}_{\mathrm{SMM}}$ vs. $\mathrm{biom}_{\mathrm{SMM}}$ linear with the same slope as the one of the GGCM. If $\mathrm{grain}_{\mathrm{GGCM}}$ vs. $\mathrm{biom}_{\mathrm{GGCM}}$ is not a line, the grid cells in the space $\mathrm{grain}_{\mathrm{GGCM}}$ vs. $\mathrm{biom}_{\mathrm{GGCM}}$ define a non-null area, called A_{GGCM}, and (n_{thresh}, frac) is chosen to make A_{SMM} as similar as possible to A_{GGCM}. EPIC family GGCMs introduced a cultivar diversity in parameters related to grain filling and, in that case, the calibration of (n_{thresh}, frac), instead of being done at the global scale, was made for each cluster of grid cells corresponding to a given cultivar. The distribution of cultivars from EPIC was used as input of SMM in that case.

2.3 Contribution of different processes to yield in SMM

We computed ratios between some SMM internal variables to assess the global contribution of different processes represented in SMM to the achievement of grain$_{SMM}$. The following ratios have been computed: n_{leaf}/tas, APAR/rsds, APAR/n_{leaf}, biom/APAR, and grain/biom. The ratio n_{leaf}/tas represents the phenology sensitivity to temperature, APAR/rsds reflects how radiation is absorbed by the canopy, APAR/n_{leaf} represents the absorption sensitivity to phenology, biom/APAR reflects the conversion from absorbed radiation to biomass, and grain/biom represents the harvest index.

We also investigated how the contribution of the different processes to the achievement of grain$_{SMM}$ varies between emulated GGCMs. Variations in these ratios reflects the difference in global averaged key parameters between emulated GGCMs. For instance, variations in grain/biom between emulated GGCMs reflects differences in calibrated n_{thresh} and frac (Fig. S3).

To compute the different ratios, averages over the growing season were used for tas, rsds, APAR, and n_{leaf}, while the end of the growing season was used for biom and grain (so-called biom$_{SMM}$ and grain$_{SMM}$, Eqs. 14 and 15). A given ratio was computed for each grid cell, and its grid-cell distribution was plotted in the following as a bar plot (Sect. 3.4 and Fig. 8).

3 Results

3.1 Global averaged biom$_{SMM}$: sensitivity to each parameter and calibration of (C, RUE)

As expected when looking at Eqs. (6) and (7), the global averaged biom$_{SMM}$ sensitivity to RUE was large (Fig. 2). Varying RUE was the only way possible to capture the global averaged biom$_{GGCM}$ for LPJ-GUESS (Fig. 2). The global averaged biom$_{SMM}$ was only slightly sensitive to other parameters as compared to the sensitivity to RUE. When all parameters are equal to their initial estimate, RUE minimizing the RMSE computed following Eq. (17) was 1.0 (LPJ-GUESS), 2.0 (LPJmL), 1.5 (CLM-crop), 2.5 (pDSSAT), 1.5 (pAPSIM), 2.0 (CGMS-WOFOST), 1.5 (GEPIC), and 1.5 g DM MJ^{-1} (EPIC-IIASA).

In Fig. 3, we plotted how RMSE changes according to both C and RUE varying at the same time, with all other parameters being equal to their initial estimate. Figure 3 shows that C and RUE can compensate in SMM. Calibrating (C, RUE) (with one global value for each parameter) allows the value to reach RMSE around 4 t DM ha^{-1} for all GGCMs except LPJ-GUESS and LPJmL (around 2 and 3 t DM ha^{-1} respectively) (Fig. 3). We chose 3 pairs (C, RUE) among the 25 tested couples that minimized the RMSE to assess the sensitivity of our conclusions to the pair chosen. Using a fourth pair with the same C for all GGCMs equal to its initial es-

timate decreased the ability of SMM to match the GGCMs only slightly (magenta dots in Fig. 3) and did not drastically change the RUE compared to the pairs where both C and RUE were calibrated.

3.2 Calibration of GDD$_{1leaf}$

Once (C, RUE) was globally chosen, a spatially varying GDD$_{1leaf}$ was calibrated. After calibration, SMM was able to catch the spatial variability of biom$_{GGCM}$ for most GGCMs (Figs. 4 and 5a). Difference in percent can be large, especially for regions with small biom, but the global distribution was relatively well captured (Fig. 4).

Global RMSE reaches between ~ 1 (LPJ-GUESS) and 3.3 t DM ha^{-1} (EPIC-IIASA) (Fig. 5a). The Nash–Sutcliffe coefficient (NS) is large (≥ 0.6) for all GGCMs but CLM-crop (0.46) and pAPSIM (0.41). RMSE is greater if computed for grid cells that experience some days with temperature above 30 °C than if computed for grid cells without such days, i.e., for LPJ-GUESS (1.5 t DM ha^{-1} vs. 0.8), GEPIC (4.4 vs. 2.3), EPIC-IIASA (4.2 vs. 2.8), and pAPSIM (3.5 vs. 2.3) (not shown). Nevertheless, the implementation of a heat stress within SMM (Text S2 in the Supplement) only slightly increases the fit of SMM vs. GGCM for these GGCMs: for example, NS increases from 0.41 (without heat stress) to 0.52 (with heat stress) for pAPSIM (Fig. S5 in the Supplement). The limited increase can be explained by the fact that optimized GDD$_{1leaf}$ in the SMM simulation without heat stress encompasses a part of the heat stress for these grid cells.

EPIC family GGCMs simulate some other stresses, such as stresses related to salinity and aeration, that could have an effect on the potential yield even in the (*harmnon* and *irrigated*) simulations (Müller et al., 2019). The intensity of some of these stresses (aeration) depends on soil orders and should be particularly important in Vertisols (Christian Folberth, personal communication, 2019). However, RMSE is only slightly different for grid cells characterized by Vertisols vs. others soil orders for GEPIC (3.4 for Vertisols vs. 3.2 for other soil orders) and EPIC-IIASA (3.8 vs. 3.3).

CLM-crop (NS = 0.46) and pAPSIM (NS = 0.52 for SMM configuration with heat stress) are the two GGCMs for which the GGCM vs. SMM agreement remains relatively poor.

When using other (C, RUE) pairs, the fit SMM vs. GGCM overall slightly decreases for most GGCMs as the (C, RUE) chosen tends to lower global fit when GDD$_{1leaf}$ is constant (see RMSE for the different pairs given in Fig. 3) but the same conclusions as above are reached: the fits are relatively correct, except for CLM-crop and pAPSIM (Fig. S6). Calibrating GDD$_{1leaf}$ when the fourth (C, RUE) pair is used leads to reasonable fit between SMM and GGCM (Fig. 5b): calibrating C is of the second-order importance compared to the calibration of RUE.

The distribution of calibrated GDD$_{1leaf}$ is provided in Fig. 6. This distribution varies between GGCMs. Most of

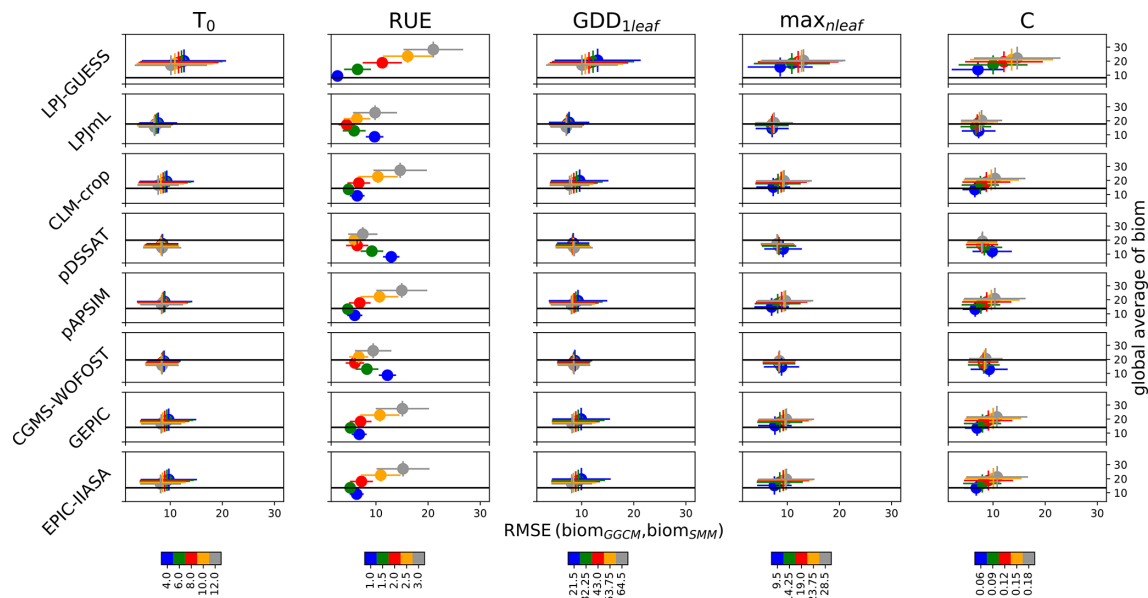**Figure 2.** Global RMSE ($biom_{SMM}$, $biom_{GGCM}$) and global averaged $biom_{SMM}$ sensitivity to SMM parameters. Five different values are used for each SMM parameter, and thus 5^5 SMM simulations have been performed. The different lines correspond to the different GGCMs. For a given line and for a given column (i.e., for a given SMM parameter), each dot represents all SMM simulations sharing the same value for this parameter, which defines its color. Error bars correspond to the standard deviation among these SMM simulations. Each dot is defined through the RMSE($biom_{GGCM}$, $biom_{SMM}$) (x axis, in $t\,ha^{-1}$) and the global averaged $biom_{SMM}$ (y axis, in $t\,ha^{-1}$). The black horizontal line reports the global averaged $biom_{GGCM}$ on the y axis. The difference of $biom_{SMM}$ between rows for a given column results from differences in growth periods between GGCMs (and to a lesser extent from the difference in the spatial coverage of maize between GGCMs). Global values (average, RMSE) are computed by giving the same weight to each grid cell.

the grid cells are characterized by extreme GDD_{1leaf} values. The distribution of GDD_{1leaf} is also sensitive to the chosen (C, RUE) pair, particularly for LPJ-GUESS and LPJmL. For these GGCMs, the difference ($biom_{GGCM} - biom_{SMM}$) is small and has the same sign almost everywhere (Fig. 4, last column). The sign is sensitive to the chosen (C, RUE): for instance, the difference is negative for the first (C, RUE) pair and positive for the second one for LPJ-GUESS. The calibration of GDD_{1leaf}, as it happens after the calibration of (C, RUE), tends to compensate this systematic bias and varies between pairs.

A SMM simulation where the range of variation allowed for GDD_{1leaf} during the step two of the calibration is increased (from [50 %–150 %] of the initial estimate in the default calibration to [25 %–200 %] of the initial estimate) allows us to significantly improve the match between GGCM and SMM; the NS increases for CLM-CROP (from 0.46 to 0.66) and pAPSIM (from 0.41 to 0.60) (second column of Fig. S7 in the Supplement). Increasing the sensitivity to temperature by letting both GDD_{1leaf} and T_0 vary at the same time during the calibration gives similar results (third column of Fig. S7). These results are obvious, as allowing more variation in SMM allows a best fit to GGCM in more grid cells. This underlines the difficulty of making an SMM that functions as a mechanistic model (see Sect. 4).

3.3 Calibration of parameters involved in grain computation (n_{thresh}, frac)

Varying (n_{thresh}, frac) allows the dots (corresponding to the different grid cells) to define different shapes in the space $yield_{SMM}$ vs. $biom_{SMM}$ (e.g., Fig. S8 in the Supplement for pDSSAT). $n_{thresh} = 0$ leads to a linear relationship between $yield_{SMM}$ and $biom_{SMM}$ with a slope equal to frac (left panels of Fig. S8). Non-null n_{thresh} make some grid cells deviate to this linear relationship and the number of such grid cells increases with n_{thresh} (Fig. S8). For all GGCMs, we found a (n_{thresh}, frac) combination that allows the relationship $yield_{SMM}$ vs. $biom_{SMM}$ to fit the relationship $yield_{GGCM}$ vs. $biom_{GGCM}$ (Fig. 7). For CLM-crop, we are not able to reproduce the cloud of dots corresponding to grid cells where the potential yield is below the line grain $= 80 \% \times biom$. For EPIC family GGCMs, a calibration per cluster of grid cells sharing the same cultivar is required.

3.4 Contribution of different processes to the achievement of $grain_{SMM}$

The ratio n_{leaf}/tas is relatively constant among the emulated GGCMs, and this is true whatever the (C, RUE) pair is chosen (Fig. 8a). The ratio n_{leaf}/tas reflects GDD_{1leaf}. The calibrated GDD_{1leaf}, even if its spatial distribution varies from

LPJ-GUESS

(RUE, C, RMSE)
(1.0, 0.09, 1.7)
(1.0, 0.06, 1.7)
(1.0, 0.12, 2.5)
(1.0, 0.12, 2.5)

LPJmL

(RUE, C, RMSE)
(2.5, 0.06, 2.7)
(2.0, 0.09, 2.8)
(2.0, 0.12, 2.9)
(2.0, 0.12, 2.9)

CLM-crop

(RUE, C, RMSE)
(2.0, 0.06, 3.9)
(1.5, 0.09, 3.9)
(1.5, 0.12, 4.0)
(1.5, 0.12, 4.0)

pDSSAT

(RUE, C, RMSE)
(2.5, 0.09, 4.1)
(3.0, 0.06, 4.2)
(2.0, 0.18, 4.4)
(2.5, 0.12, 4.6)

pAPSIM

(RUE, C, RMSE)
(1.5, 0.09, 3.8)
(1.5, 0.12, 4.0)
(2.0, 0.06, 4.0)
(1.5, 0.12, 4.0)

CGMS-WOFOST

(RUE, C, RMSE)
(2.0, 0.12, 4.4)
(3.0, 0.06, 4.5)
(2.0, 0.15, 4.5)
(2.0, 0.12, 4.4)

GEPIC

(RUE, C, RMSE)
(2.0, 0.06, 4.2)
(1.5, 0.09, 4.2)
(1.5, 0.12, 4.3)
(1.5, 0.12, 4.3)

EPIC-IIASA

(RUE, C, RMSE)
(1.5, 0.09, 4.2)
(2.0, 0.06, 4.2)
(1.5, 0.12, 4.4)
(1.5, 0.12, 4.4)

Figure 3. Sensitivity of global RMSE (biom$_{SMM}$, biom$_{GGCM}$) (t DM ha^{-1}) to C ($-$) and RUE (g DM MJ^{-1}) for each GGCM. For each GGCM, the shape of RMSE is derived from 25 SMM simulations (five values of C combined with five values of RUE). Other SMM parameters are equal to their respective initial estimate. For each GGCM, three (C, RUE) pairs that minimize the most RMSE among the 25 pairs tested are plotted in blue, green, and red. A fourth (C, RUE) pair minimizing the RMSE but with C equal to its initial estimate is shown in magenta. The values of the four (C, RUE) pairs and the corresponding RMSE are given in top left of each panel.

one GGCM to the other (see Sect. 3.2), remains relatively constant at the global scale between GGCMs.

The ratios APAR/rsds and APAR/n_{leaf} (Fig. 8b and c) vary a lot between GGCMs, but this variation is of the same order of magnitude as the one between (C, RUE) pairs. These ratios reflect C, which is highly variable between GGCMs and between pairs.

The ratio biom/APAR (Fig. 8d) reflects global RUE. Calibrated RUE varies a lot between GGCMs and only slightly between pairs for a given GGCM.

The ratio grain/biom (Fig. 8e) varies a lot between GGCMs and is only slightly sensitive to the (C, RUE) pair. This ratio reflects a combination of n_{thresh} and frac. GGCMs with n_{thresh} equal to 0 (LPJmL, EPIC-IIASA) have no grid cell variability in grain/biom (Fig. 8e). Overall, regardless of the GGCM variability at grid cell scale, we can distinguish (i) emulated GGCMs that convert a large fraction of biom to grain, such as CLM-crop; (ii) emulated GGCMs that convert around 50 % of biom to grain, such as LPJmL, pDSSAT, GEPIC, pAPSIM, and EPIC-IIASA; and (iii) emulated GGCMs that convert around 30 %–40 % of biom to grain, such as LPJ-GUESS and CGMS-WOFOST. Large variation of grain/biom between GGCMs is consistent with the fact that difference in grain among GGCMs is larger than the one in biom (Fig. 1).

4 Discussion and conclusions

We have shown that a simple set of equations with one GGCM-dependent global RUE and spatially variable thermal requirement (GDD$_{1leaf}$) is able to mimic spatial distribution of aboveground biomass of most GGCMs. Calibrating one additional global parameter at the same time as RUE (namely C) only slightly improves the fit between SMM and GGCM and only modified the calibrated value of RUE to a small extent. RUE represents canopy photosynthesis and GDD$_{1leaf}$ determines crop duration, i.e., the two main drivers of crop productivity (Sinclair and Muchow, 1999). The relationship between potential yield and aboveground biomass of GGCM is captured by the calibration of two additional global parameters: one that triggers the start of grain filling and one corresponding to a time-invariant fraction of NPP allocated to the grain. These two parameters allow us to capture the relationship between biom and grain from all GGCMs. This feature of SMM is particularly important, as we have shown that the divergence between GGCMs is larger for grain than for biom (Fig. 1). Cultivar diversity regarding these latter parameters is nevertheless required to catch the behavior of some GGCMs. Despite the apparent complexity of GGCMs, we have shown that differences between them regarding potential yield can be explained by differences in a few key parameters.

Our approach has a few caveats. First, SMM could be able to fit individual GGCMs for the wrong reasons, i.e., following a compensation between SMM internal processes that is not representative of the considered GGCM. We think that this issue is nevertheless minimized in our approach. First, we investigated how parameters can compensate, e.g., by calibrating RUE and C at the same time. We showed that calibration of C is of second-order importance and that getting calibrated RUE to vary less among emulated GGCMs would require extreme values for C, well outside the range
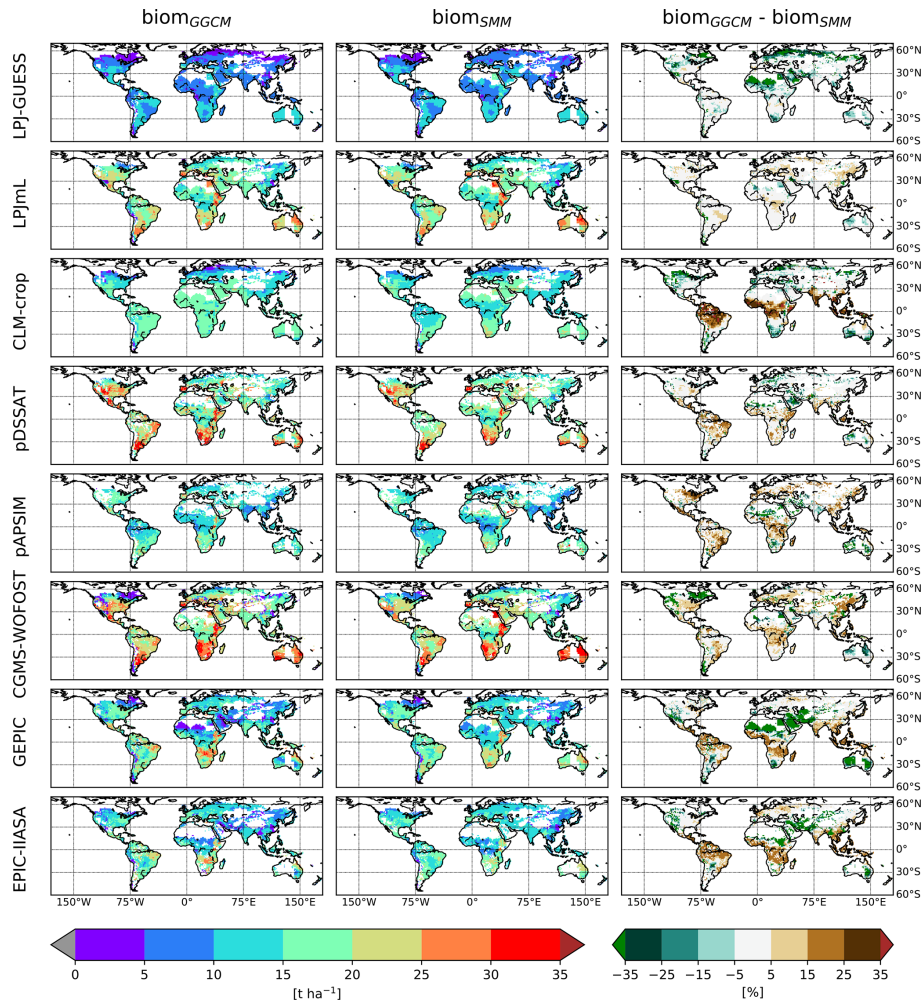
Figure 4. Comparison of the spatial distribution of simulated aboveground biomass (biom) between GGCM and SMM after its calibration. For each GGCM (row), the following variables are displayed: $\text{biom}_{\text{GGCM}}$ (left column), biom_{SMM} after SMM calibration (i.e., calibration of global GGCM-dependent C and RUE and spatially varying $\text{GDD}_{1\text{leaf}}$), and the difference between the two ($\text{biom}_{\text{GGCM}} - \text{biom}_{\text{SMM}}$) (expressed in percent of $\text{biom}_{\text{GGCM}}$). Only biom_{SMM} after SMM calibration with the first (C, RUE) pair is shown.

of values allowed in our calibration. The parameter C encompasses different parameters (see Eq. 8), and a better alternative would be to separate them as well as to explicitly simulate the leaf area index (LAI) variable. SMM-simulated LAI would be compared to GGCM output and this comparison would further reduce the risk of compensation between processes in SMM. However, LAI was neither available from GGCMI data archive nor upon request to GGCM modellers. We stress the need to incorporate this output variable in the next intercomparison exercise. It is also important to note that the average over the growing season of LAI or LAI at set fractions of the growing season (including anthesis) would be more interesting than LAI at harvest as, under potential conditions, LAI at harvest is very likely close to maximum LAI allowed by the different GGCMs.

Therefore, the reasonable match between SMM-simulated aboveground biomass and GGCM-simulated biom is made possible only because of the large range of variation allowed for RUE and $\text{GDD}_{1\text{leaf}}$ during the calibration, i.e., [50 %–150 %] of their initial estimate. This range should have a meaning in term of values commonly used in GGCMs. Otherwise, the calibrated parameters could implicitly encompass different mechanisms considered in GGCMs but not in SMM, and this issue should become more likely as we choose a larger range of variations. For instance, calibrated $\text{GDD}_{1\text{leaf}}$ in SMM could artificially encompass the sensitivity to temperature of processes not considered in SMM, as we discussed for heat stress in Sect. 3.2. It is likely that the variation of $\text{GDD}_{1\text{leaf}}$ also encompasses a spatial variation of emergence in GGCMs, as in SMM we did not compute emergence and plants start to grow from the planting day. In such cases SMM should not be considered as a pure mechanistic model but more as a meta-model, and purely statistical meta-models should be more appropriate than our simplified
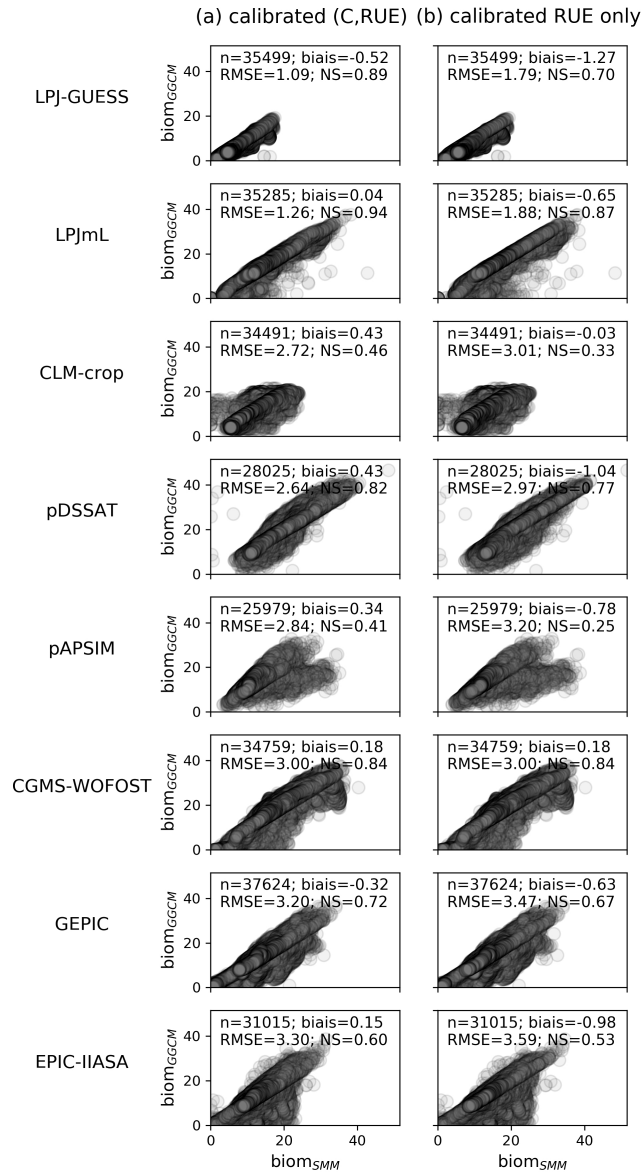
Figure 5. Scatterplots of $biom_{GGCM}$ (y axis) vs. $biom_{SMM}$ (x axis) after SMM calibration, i.e., calibration of global GGCM-dependent (C, RUE) and spatially varying GDD_{1leaf}. In **(a)** the (C, RUE) pair that minimizes the most the RMSE is used, while in **(b)** the (C, RUE) minimizing the RMSE with C equal to its initial estimate (the so-called fourth pair in Sect. 2.2.4) is used. Each dot corresponds to one grid cell. In the top left of each panel, the number of grid cells considered (n) and the different quantifications of the SMM vs. GGCM agreement (bias, RMSE, and NS) are given. See Fig. S6 in the Supplement for scatterplots of the four (C, RUE) pairs.

Figure 6. Calibrated spatial distribution of GDD_{1leaf}. Five values of GDD_{1leaf} are allowed (color scale and Table S1). Global GGCM-dependent (C, RUE) have been calibrated prior to GDD_{1leaf} and for a given GGCM; grid cells for which calibrated GDD_{1leaf} is different for the four (C, RUE) pairs are in grey.

process-based model. However, the range of variation that we used for GDD_{1leaf} ([50 %–150 %] around the initial estimate of 43 °C, corresponding to $\sim$ [22–65 °C]) is consistent with ranges reported by observations focusing on the sensitivity of phyllochron (thermal requirement for the emergence of any leaf) to temperature (fig. 2 of Birch et al., 1998) and

Figure 7. Relationship between grain and biom: comparison between GGCM and SMM after its calibration. The grain vs. biom relationship for GGCMs (left column) and for SMM after independent calibration against each GGCM (right column) are both displayed. For SMM, the relationship is given for the (n_{thresh}, frac) combination that maximizes the match between GGCM and SMM (see Sect. 2.2.4). The calibrated combination is given at the top of each SMM panel. When multiple cultivars are considered, one calibrated (n_{thresh}, frac) combination is given per cultivar. In the right-hand panels, pink shading defines A_{GGCM} (see Sect. 2.2.4).

cultivar (Padilla and Otegui, 2005). Our range of GDD_{1leaf} cannot be straightforward compared to the range of heat units commonly used in GGCMs to catch the prescribed growing season (e.g., $\sim [10\text{--}225\,°C]$ in GEPIC if we divide the values of heat units given in Minoli et al., 2019, by a maximum number of leaves of 19, as in our study) or computed in van

Bussel et al. (2019b) ($\sim [25\text{--}160\,°C]$ if we divide the values given in Fig. 2 of that latter reference by 19). Indeed, in our approach, $GDD_{1leaf} \times \max_{nleaf}$ corresponds to the thermal requirement up to the emergence of all leaves, while the sum of heat unit used in GEPIC or van Bussel et al. (2015b) is required to reach plant maturity and thus encompasses both the emergence of all leaves and the period from flowering (concomitant to the end of leaf emergence) up to plant maturity.

Some discrepancies remain between SMM and some GGCMs, especially CLM-crop and pAPSIM. This could be explained by differences between GGCM and SMM in the choice of processes represented (e.g., net productivity in SMM instead of a balance between gross productivity and plant respiration in some GGCMs) or in the choice of equations used to represent a given process (e.g., Farquhar (CLM-crop) vs. RUE (SMM) for assimilation). The representation of stomatal conductance and CO_2 assimilation rate within Farquhar equations introduces a sensitivity to variables not considered in the RUE-based approach (e.g., water vapor pressure deficit), which is in line with observations that show that RUE is sensitive to many variables (Sinclair and Muchow, 1999). This would lead to differences in the spatial variability of simulated aboveground biomass as compared to one simulated with spatially constant RUE. The succession of phenological stages with different parameterizations in some GGCMs (e.g., in pAPSIM; Wang et al., 2002) can also partly contribute to differences with SMM because plant development is continuously simulated in SMM, as it is in other GGCMs (e.g., LPJmL, Schaphoff et al., 2018b). Another potential mismatch is related to some limiting factors (nutrients, water, etc.) that could exist in the *harmnon* and *irrigated* simulation for GGCMs (despite the protocol of these simulations) or through different biases. First, GGCMs that do not explicitly simulate nutrient limitations may have integrated these stresses implicitly in their parameterizations (discussed below). Second, irrigation in some GGCMs ensures that plants are not limited by water supply. However, plants can still experience water stress if the atmospheric demand is higher than the plant hydraulic structure can service. This could likely explain the lower yield for the group of grid cells below the line corresponding to $grain_{GGCM} = 80\,\% \times biom_{GGCM}$ for CLM-Crop in Fig. 7. Finally, some other stresses (salinity, aeration, etc.) are present in a few GGCMs (e.g., EPIC family models) and are not alleviated in *harmnon* simulation. However, it seems that these stresses, restricted to a few grid cells, cannot significantly contribute to the GGCM vs. SMM mismatch.

Despite some confusion in values of RUE reported in the literature arising from diversity of experimental approaches and units of expression that have been used (Sinclair and Muchow, 1999; Sect. 2.2.3), RUE is relatively well constrained from observations ($[3.1\text{--}4.0]\,\text{g DM MJ}^{-1}$). Here, we found that calibrated RUE values in emulated GGCMs are lower than values derived from observations and varied a lot among GGCMs: between $1\,\text{g DM MJ}^{-1}$ for LPJ-GUESS and
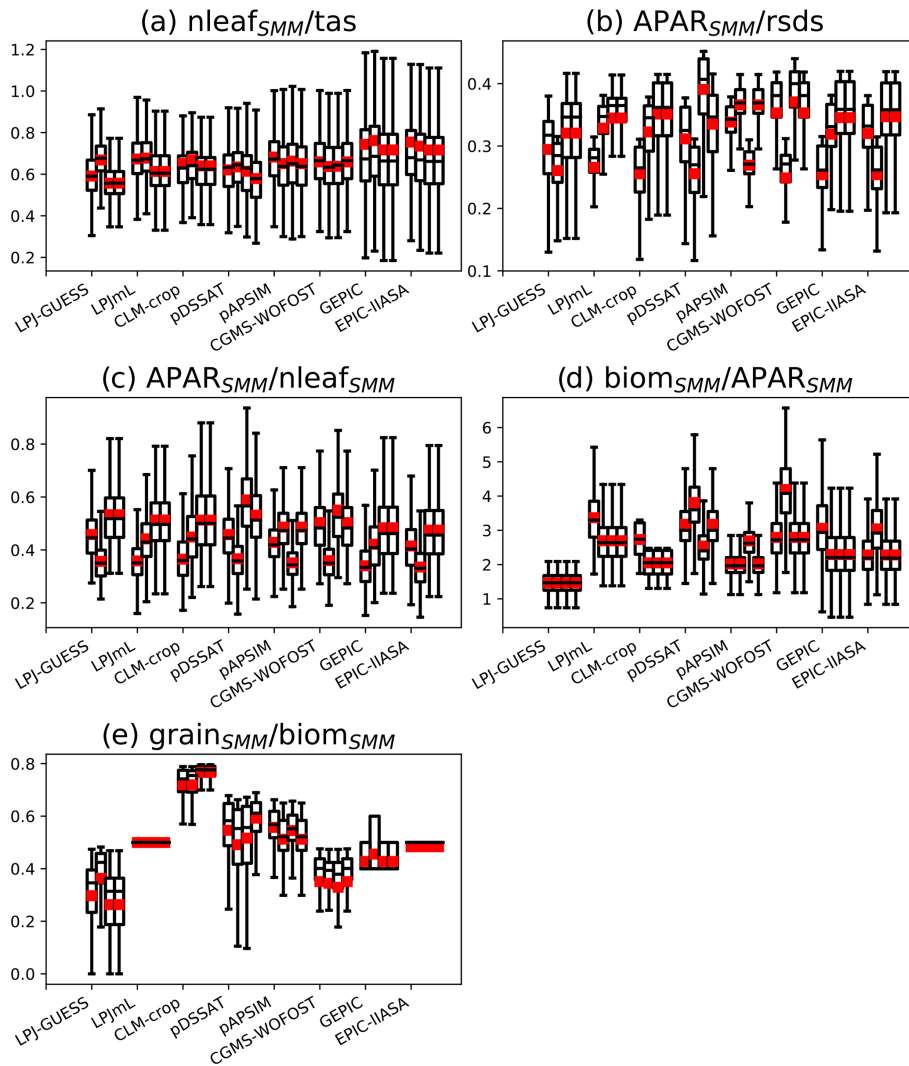
Figure 8. Bar plots of different ratios computed with SMM calibrated against each GGCM. Each panel shows a ratio (**a** n_{leaf}/tas, **b** APAR/rsds, **c** APAR/n_{leaf}, **d** biom/APAR, **e** grain/biom), while for a given panel x axis ticks correspond to emulated GGCMs. Four different (C, RUE) pairs have been tested and correspond to the four bars of each x axis tick.

2.5 g DM MJ^{-1} for pDSSAT. The values of calibrated RUE found in our study can be compared to values actually used in GGCMs based on the same approach of conversion of radiation to aboveground biomass. RUE used in GEPIC is equal to 4.0 g DM MJ^{-1} (Christian Folberth, personal communication, 2020), while our calibrated value for the same GGCM is 2.0 g DM MJ^{-1}. Our calibrated RUE is an apparent RUE and the mismatch with actual RUE prescribed to GEPIC can be explained as follows. First, the daily increment of biomass in GEPIC derived from the conversion of radiation encompasses an increment for both aboveground biomass and roots (with a root : total ratio varying from 0.4 at germination to 0.2 at maturity), while both the RUE values used in SMM and derived from most observations (Sinclair and Muchow, 1999) concern aboveground biomass only. Actual RUE prescribed to GEPIC after correction to make

it represent only aboveground biomass should vary between 2.4 at germination to 3.8 g DM MJ^{-1} at maturity and is closer (particular in the first growth stages) to our calibrated RUE for GEPIC. Second, LAI$_{\text{max}}$ in GEPIC is lower than values used in SMM: in GEPIC, LAI$_{\text{max}}$ varies with plant density and is equal to 3.5 at plant density of 5 plants m^{-2} used in the GGCMI simulations, while LAI$_{\text{max}}$ in SMM is equal to 5 (this value can be derived from Eq. (9) when n_{leaf} reaches its maximum, i.e., max$_{\text{nleaf}}$, whose the value is prescribed during the calibration procedure; max$_{\text{nleaf}}$ = 19). Lower LAI$_{\text{max}}$ in GEPIC than SMM can compensate for higher RUE. Some additional sensitivity tests with varying LAI$_{\text{max}}$ (not shown) nevertheless suggest that the mismatch in LAI$_{\text{max}}$ between GEPIC and SMM contributes only slightly to the mismatch in RUE. Finally, the seasonal dynamic of LAI could be different between GEPIC and SMM, and this can counterbal-

ance the lower increment of biomass computed in SMM than in GEPIC once LAI_{max} is reached. As mentioned above, the values of LAI computed by GGCMs at different set fractions of the growing season would be very helpful.

The diversity in apparent RUE found between GGCMs raises a question about the physical meaning of the parameters used in each GGCM. GGCMs are tools that are primarily dedicated to simulating actual yield and could have been tuned for that purpose against local observations. During this tuning, processes not explicitly represented in a given GGCM could be implicitly incorporated into parameterizations of other processes. For instance, this could be the case for GGCMs that do not explicitly incorporate nutrient limitations. Potential yield is a variable that has been computed in a second step and that could suffer from these implicit incorporations. In the end, the divergence in potential yield between GGCMs raises questions about their ability to reproduce real processes as the basis of the actual yield, as this depends on the combination of potential yield and many limiting factors.

Our study has some implications for GGCM modellers in regards to the simulation of potential yield. We showed that differences between GGCMs can be explained by differences in a few key parameters, namely the RUE and parameters driving the grain filling (n_{thresh} and frac). For RUE, we recommend for GGCM modellers to investigate why each individual GGCM has a such a small (explicit or implicit) apparent RUE. We showed that n_{thresh} and frac vary a lot between GGCMs. For GGCMs with n_{thresh} equal to 0, a parametrization based on a better distinction between the emergence of all leaves and the period from flowering to maturity could be interesting. We also showed that n_{thresh} and frac determines harvest index (HI), and we showed that HI varies a lot between GGCMs. Thus, we advise that more effort needs to be directed toward assembling a global dataset of HI for either parametrization or evaluation of GGCMs. The maximum HI that a plant can reach is a cultivar characteristic, and one possibility to build such dataset could partly rely on information from seed companies. Finally, we suggested that the next intercomparison exercise encompass LAI and the beginning and end of the different growing season periods (vegetative period, flowering, etc.) for GGCMs that distinguish between such stages.

Model ensemble mean or median from an intercomparison exercise is commonly preferred to individual GGCM as it has better skill in reproducing the observations (Asseng et al., 2014; Martre et al., 2015). Our mechanistic model tuned against GGCMs could be a viable alternative to this ensemble mean and median as it allows us to keep track of processes leading to the final variable, namely the potential yield. Once tuned, SMM could be forced by an ensemble of parameters to reproduce the ensemble of GGCMs. Our emulator could also offer a potential for GGCM evaluation and analysis of their structural uncertainty. Its use under different climate and CO_2 conditions would nevertheless require the imple-

mentation of some missing mechanisms (e.g., heat stress and effect of CO_2).

For the purpose of simulating potential yield at the global scale, our emulator forced by daily temperature and radiation, growing season, and a few adjustable parameters could be considered an interesting alternative to GGCMs as they are easier to manipulate and allow much faster simulations. For instance, our model could be used to investigate the implementation of cultivar diversity at the global scale. The introduction of cultivar diversity is a keystone in development of crop models at the global scale (Boote et al., 2013). Cultivar diversity considered at the global scale was mainly related to phenological development through sensitivity to photoperiod and sensitivity to temperature (and vernalization for winter cultivars) (van Bussel et al., 2015a). The effect of cultivar diversity on allometry (e.g., through variability in harvest index) was considered in a less extensive extent at the global scale and restricted to some EPIC models (Folberth et al., 2019) or pDSSAT in a specific study (Gbegbelegbe et al., 2017). Through the protocol of GGCMI in which thermal requirements are tuned to match the growing season, cultivar diversity was implicitly accounted for. The same applies for SMM. The parameters of the emulator, here fitted to reproduce GGCM output, could also be fitted to global datasets based on census data and observations in a manner similar to that of PEGASUS (Deryng et al., 2011) but related to potential yield (against real yield for PEGASUS). For instance, SMM could be calibrated against the global dataset of potential yield based on statistical approach (Mueller et al., 2012), and the spatial variation of calibrated parameters could be compared to existing knowledge about the spatial distribution of cultivars. Finally, because it allows temporal dynamic simulation, our model could be coupled with simulation of limiting factors (water, nutrients) to investigate the limitation of potential yield at the global scale in a simple but mechanistic manner.

References

Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., Kimball, B. A., Ottman, M. J., Wall, G. W., White, J. W., Reynolds, M. P., Alderman, P. D., Prasad, P. V. V., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Challinor, A. J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L. A., Izaurralde, R. C., Jabloun, M., Jones, C. D., Kersebaum, K. C., Koehler, A.-K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J. E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A. C., Semenov, M. A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P. J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., and Zhu, Y.: Rising temperatures reduce global wheat production, Nat. Clim. Change, 5, 143–147, https://doi.org/10.1038/nclimate2470, 2014.

Birch, C. J., Vos, J., Kiniry, J., Bos, H. J., and Elings, A.: Phyllochron responds to acclimation to temperature and irradiance in maize, Field Crops Res., 59, 187–200, https://doi.org/10.1016/S0378-4290(98)00120-8, 1998.

Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Müller, C., Reichstein, M., and Smith, B.: Modelling the role of agriculture for the 20th century global terrestrial carbon balance, Glob. Change Biol., 13, 679–706, https://doi.org/10.1111/j.1365-2486.2006.01305.x, 2007.

Boogaard, H. L., Wit, A. J. W. D., Roller, J. A. T., and Diepen, C. A. V.: WOFOST Control Centre 2.1 and WOFOST 7.1.7. User's guide for the WOFOST Control Centre 2.1 and WOFOST 7.1.7 crop growth simulation model, Alterra, Wageningen University and Research Centre, Wageningen, the Netherlands, 2014.

Boote, K. J., Jones, J. W., White, J. W., Asseng, S., and Lizaso, J. I.: Putting mechanisms into crop production models: Putting

mechanisms into crop production models, Plant Cell Environ., 36, 1658–1672, https://doi.org/10.1111/pce.12119, 2013.

Deryng, D., Sacks, W. J., Barford, C. C., and Ramankutty, N.: Simulating the effects of climate and agricultural management practices on global crop yield: Simulating Global Crop Yield, Global Biogeochem. Cycles, 25, GB2006, https://doi.org/10.1029/2009GB003765, 2011.

Drewniak, B., Song, J., Prell, J., Kotamarthi, V. R., and Jacob, R.: Modeling agriculture in the Community Land Model, Geosci. Model Dev., 6, 495–515, https://doi.org/10.5194/gmd-6-495-2013, 2013.

Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K., Best, N., Wilde, M., and Foster, I.: The parallel system for integrating impact models and sectors (pSIMS), Environ. Model. Softw., 62, 509–516, https://doi.org/10.1016/j.envsoft.2014.04.008, 2014.

Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C., Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., and Sheffield, J.: The Global Gridded Crop Model Intercomparison: data and modeling protocols for Phase 1 (v1.0), Geosci. Model Dev., 8, 261–277, https://doi.org/10.5194/gmd-8-261-2015, 2015.

Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., and Yang, H.: Regionalization of a large-scale crop growth model for sub-Saharan Africa: Model setup, evaluation, and estimation of maize yields, Agric. Ecosyst. Environ., 151, 21–33, https://doi.org/10.1016/j.agee.2012.01.026, 2012.

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, Nat. Commun., 7, 11872, https://doi.org/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: Martínez-Paz, J. M., PLOS ONE, 14, e0221862, https://doi.org/10.1371/journal.pone.0221862, 2019.

Gbegbelegbe, S., Cammarano, D., Asseng, S., Robertson, R., Chung, U., Adam, M., Abdalla, O., Payne, T., Reynolds, M., Sonder, K., Shiferaw, B., and Nelson, G.: Baseline simulation for global wheat production with CIMMYT mega-environment specific cultivars, Field Crops Res., 202, 122–135, https://doi.org/10.1016/j.fcr.2016.06.010, 2017.

Grassini, P., Thorburn, J., Burr, C., and Cassman, K. G.: High-yield irrigated maize in the Western U. S. Corn Belt: I. On-farm yield, yield potential, and impact of agronomic practices, Field Crops Res., 120, 142–150, https://doi.org/10.1016/j.fcr.2010.09.012, 2011.

Izaurralde, R. C., Williams, J. R., McGill, W. B., Rosenberg, N. J., and Jakas, M. C. Q.: Simulating soil C dynamics with EPIC: Model description and testing against long-term data, Ecol. Model., 192, 362–384, https://doi.org/10.1016/j.ecolmodel.2005.07.010, 2006.

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., and Ritchie, J. T.: The DSSAT cropping system model, Eur. J. Agron., 18, 235–265, https://doi.org/10.1016/S1161-0301(02)00107-7, 2003.

Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., Huth, N. I., Hargreaves, J. N., Meinke, H., and Hochman, Z.: An overview of APSIM, a model designed for farming systems simulation, Eur. J. Agron., 18, 267–288, 2003.

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 Res., 20, 51–64, 1989.

Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., and Smith, B.: Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa, Earth Syst. Dynam., 4, 385–407, https://doi.org/10.5194/esd-4-385-2013, 2013.

Liu, J., Williams, J. R., Zehnder, A. J. B., and Yang, H.: GEPIC – modelling wheat yield and crop water productivity with high resolution on a global scale, Agric. Syst., 94, 478–493, https://doi.org/10.1016/j.agsy.2006.11.019, 2007.

Lobell, D. B., Cassman, K. G., and Field, C. B.: Crop yield gaps: their importance, magnitudes, and causes, Annu. Rev. Environ. Resour., 34, 179, 2009.

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J. W., Rötter, R. P., Boote, K. J., Ruane, A. C., Thorburn, P. J., Cammarano, D., Hatfield, J. L., Rosenzweig, C., Aggarwal, P. K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A. J., Doltra, J., Gayler, S., Goldberg, R., Grant, R. F., Heng, L., Hooker, J., Hunt, L. A., Ingwersen, J., Izaurralde, R. C., Kersebaum, K. C., Müller, C., Kumar, S. N., Nendel, C., O'leary, G., Olesen, J. E., Osborne, T. M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M. A., Shcherbak, I., Steduto, P., Stöckle, C. O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J. W., and Wolf, J.: Multimodel ensembles of wheat growth: many models are better than one, Glob. Change Biol., 21, 911–925, https://doi.org/10.1111/gcb.12768, 2015.

Minoli, S., Müller, C., Elliott, J., Ruane, A. C., Jägermeyr, J., Zabel, F., Dury, M., Folberth, C., François, L., Hank, T., Jacquemin, I., Liu, W., Olin, S., and Pugh, T. A. M.: Global Response Patterns of Major Rainfed Crops to Adaptation by Maintaining Current Growing Periods and Irrigation, Earths Future, 7, 1464–1480, https://doi.org/10.1029/2018EF001130, 2019.

Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., and Foley, J. A.: Closing yield gaps through nutrient and water management, Nature, 490, 254–257, https://doi.org/10.1038/nature11420, 2012.

Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., and Yang, H.: Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications, Geosci. Model Dev., 10, 1403–1422, https://doi.org/10.5194/gmd-10-1403-2017, 2017.

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, Sci. Data, 6, 50, https://doi.org/10.1038/s41597-019-0023-8, 2019.

Padilla, J. M. and Otegui, M. E.: Co-ordination between Leaf Initiation and Leaf Appearance in Field-grown Maize (Zea mays): Genotypic Differences in Response of Rates to Temperature, Ann. Bot., 96, 997–1007, https://doi.org/10.1093/aob/mci251, 2005.

Portmann, F. T., Siebert, S., and Döll, P.: MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling: Monthly Irrigated and Rainfed Crop Areas, Global Biogeochem. Cycles, 24, GB1011, https://doi.org/10.1029/2008GB003435, 2010.

Ringeval, B.: Process-based emulator to explain the differences in simulated potential yield between Global Gridded Crop Models, Portail Data INRAE, V1, https://doi.org/10.15454/9EIJWU, 2020.

Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P., Antle, J. M., Nelson, G. C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., and Winter, J. M.: The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies, Agric. For. Meteorol., 170, 166–182, https://doi.org/10.1016/j.agrformet.2012.09.011, 2013.

Ruane, A. C., Goldberg, R., and Chryssanthacopoulos, J.: Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation, Agric. For. Meteorol., 200, 233–248, https://doi.org/10.1016/j.agrformet.2014.09.016, 2015.

Ruane, A. C., Rosenzweig, C., Asseng, S., Boote, K. J., Elliott, J., Ewert, F., Jones, J. W., Martre, P., McDermid, S. P., Müller, C., Snyder, A., and Thorburn, P. J.: An AgMIP framework for improved agricultural representation in integrated assessment models, Environ. Res. Lett., 12, 125003, https://doi.org/10.1088/1748-9326/aa8da6, 2017.

Sacks, W. J., Deryng, D., Foley, J. A., and Ramankutty, N.: Crop planting dates: an analysis of global patterns, Glob. Ecol. Biogeogr., 19, 607–620, https://doi.org/10.1111/j.1466-8238.2010.00551.x, 2010.

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 Res., 79, 39–51, 2002.

Schaphoff, S., Forkel, M., Müller, C., Knauer, J., von Bloh, W., Gerten, D., Jägermeyr, J., Lucht, W., Rammig, A., Thonicke, K., and Waha, K.: LPJmL4 – a dynamic global vegetation model with managed land – Part 2: Model evaluation, Geosci. Model Dev., 11, 1377–1403, https://doi.org/10.5194/gmd-11-1377-2018, 2018a.

Schaphoff, S., von Bloh, W., Rammig, A., Thonicke, K., Biemans, H., Forkel, M., Gerten, D., Heinke, J., Jägermeyr, J., Knauer, J., Langerwisch, F., Lucht, W., Müller, C., Rolinski, S., and Waha, K.: LPJmL4 – a dynamic global vegetation model with managed

land – Part 1: Model description, Geosci. Model Dev., 11, 1343–1375, https://doi.org/10.5194/gmd-11-1343-2018, 2018b.

Sinclair, T. R. and Muchow, R. C.: Radiation Use Efficiency, Adv. Agron., 65, 215–265, 1999.

Smith, B., Prentice, I. C., and Sykes, M. T.: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space, Glob. Ecol. Biogeogr., 10, 621–637, https://doi.org/10.1046/j.1466-822X.2001.t01-1-00256.x, 2001.

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

van Bussel, L. G. J., Grassini, P., Van Wart, J., Wolf, J., Claessens, L., Yang, H., Boogaard, H., de Groot, H., Saito, K., Cassman, K. G., and van Ittersum, M. K.: From field to atlas: Upscaling of location-specific yield gap estimates, Field Crops Res., 177, 98–108, https://doi.org/10.1016/j.fcr.2015.03.005, 2015a.

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, Glob. Ecol. Biogeogr., 24, 1018–1029, https://doi.org/10.1111/geb.12351, 2015b.

van Ittersum, M. K., Cassman, K. G., Grassini, P., Wolf, J., Tittonell, P., and Hochman, Z.: Yield gap analysis with local to global relevance – A review, Field Crops Res., 143, 4–17, https://doi.org/10.1016/j.fcr.2012.09.009, 2013.

Waha, K., van Bussel, L. G. J., Müller, C., and Bondeau, A.: Climate-driven simulation of global crop sowing dates: Simulation of global sowing dates, Glob. Ecol. Biogeogr., 21, 247–259, https://doi.org/10.1111/j.1466-8238.2011.00678.x, 2012.

Wang, E., Robertson, M. J., Hammer, G. L., Carberry, P. S., Holzworth, D., Meinke, H., Chapman, S. C., Hargreaves, J. N. G., Huth, N. I., and McLean, G.: Development of a generic crop model template in the cropping system model APSIM, Eur. J. Agron., 18, 121–140, https://doi.org/10.1016/S1161-0301(02)00100-4, 2002.

Williams, J. R.: Computer models of watershed hydrology, ed. Singh, V. P., 909–1000, Water Resources Publications, Highlands Ranch, 1995.