We thank A.C.Ruane and and two anonymous Referees for their comments. Answers are provided below in green. The Main Text and Supp Inf initially submitted are defined as «original» while the ones after taking into account the Associate Editor and Referee's comments are defined as « new ». All line numbers given below refer to the lines of the new Main Text with tracked changes.

Comment of A.C. Ruane (alexander.c.ruane@nasa.gov)
Received and published: 1 July 2020

A short comment on an interesting manuscript:
It would be helpful if the authors could please note that GGCMI is an activity of the Agricultural Model Intercomparison and Improvement Project (AgMIP; Rosenzweig etal., 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).
This connection is particularly important given the voluntary nature of these efforts and as an indication of community willingness to systematically compare, open inputs and outputs for broader scientific inquiry, and facilitate cross-scale and cross-disciplinary applications of crop models for basic research and societal benefit.

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

Ruane, A.C., C. Rosenzweig, S. Asseng, K.J. Boote, J. Elliott, F. Ewert, J.W. Jones,P. Martre, S. McDermid, C. Müller, A. Snyder, and P.J. Thorburn, 2017: An AgMIP framework for improved agricultural representation in IAMs. Environ. Res. Lett., 12,no. 12, 125003, doi:10.1088/1748-9326/aa8da6.

We thank A.C. Ruane for this comment and to point us the need to underline the connection with AgMIP. We totally agree in particular as it promotes the open-science aspects of AgMIP. We added the two above references and the first sentence given by A.C. Ruane in the new version of the manuscript (L197).

Anonymous Referee #1
Received and published: 19 August 2020

Great Job!

We thank the Referee #1 for his/her positive comment !

Anonymous Referee #2
Received and published: 5 December 2020

1.Overview

Review of “Potential yield simulated by Global Gridded Crop Models: a process-based emulator to explain their differences” by Bruno Ringeval et al. Bruno Ringeval et al. build an emulator SMM with generic equations describing the processes to reproduce the simulation of aboveground biomass and potential yield of maize. They showed that with few carefully calibrated parameters, the SMM can capture the spatial distribution of aboveground biomass and potential yields, and the variations between different GGCMs. This SMM can thus provide a useful tool to compare different GGCMs. For this manuscript, the texts are well written, the methods are clearly described, and the logic are easy to follow, thus I only have some minor comments.

We thank the Referee for his/her positive comments.

2. Comments:

1) A GMD paper must “include the title (concise but informative, including model name and version number if a model description paper)” in the title (https://www.geoscientific-model-development.net/submission.html#manuscriptcomposition), so I suggest the authors to follow this instruction.

We understand this comment. However, our manuscript’s type is « Methods for assessment of models » (and not « Model description paper »). For this type, the rule for the title are less clear than for model description papers. We are a little bit reluctent to give a name + version number in the title as our model will be used in a near future to catch the spatial distribution of potential yield derived from empirical approach, and not to emulate GGCMs. Nevertheless, if the Editorial team asks us to follow this comment, we will add name and version (SMM version 1.0) in the title.

2) Ln 222:..., varying in space as “a” function…
3) Ln 278:...It is likely “that” some GGCMs…

We correct the two mistakes underlined by the Referee in the new version of the draft (now L227 and L283).

4) Ln 285: the authors admit that their RUE values are smaller than the commonly reported values, but why not increase the range of RUE tested? In the results, they show that for SMM of some models (e.g. pDSSAT), the RUE can reach the higher end of the tested values. If RUE is allowed to be larger than 3.0g DM MJ-1, will the best RUE and the results change?

In the previous version of the draft, 5 values of RUE were tested : 50, 75, 100, 125, 150 % of its initial estimate (equal to 2 gDM MJ-1), i.e. : 1.0, 1.5, 2.0, 2.5, 3.0 gDM MJ-1.

As we can see in the Fig.3 of the original draft, the optimized RUE values for each GGCM (e.g. in the simulation with C equal to its initial estimate : i.e. the magenta symbol in Fig.3) are :
1.0 gDM MJ-1 for : LPJ-GUESS
1.5 gDM MJ-1 for : GEPIC, EPIC-IIASA, pAPSIM, CLM-crop,
2.0 gDM MJ-1 for : CGMS-WOFOST and LPJmL
2.5 gDM MJ-1 for : pDSSAT.
The highest value of RUE allowed in the optimization procedure (i.e. 3.0) was not chosen during the optimization for any GGCM. Thus, it is likely that increasing the highest RUE allowed during

the optimization does not change the optimized RUE and results. Only for LPJ-GUESS, the calibrated RUE reach one boundary (the lowest) of the range of values allowed.

To assess further, we performed a sensitivity test in which we increased the range of values allowed in the optimization for RUE : **0.5**, **0.75**, 1.0, 1.5, 2.0, 2.5, 3.0, **3.5**, **4.0** gDM MJ-1 (with new values tested in bold). In such case, the Figure 3 would be Fig.R1. As expected, only LPJ-GUESS has a calibrated RUE different to the one of the original draft (0.75 in the sensitivity test vs 1 in the original version of the draft). This would have a minor effect on the later results as the fit data vs model is alredy very good for LPJ-GUESS once a spatial variability in GDD_{1leaf} is introduced (Fig.5). Also, we prefer not modifying the range of values allowed as it is difficult to justify such low values of RUE. The range of values allowed during the optimization and related issue of compensation are already discussed in the original version of the manuscript (L515-520).

To respond precisely to the question of the Referee, allowing larger values for RUE in the optimization procedure (3.5, 4.0 gDM MJ-1) has no effect on the calibrated RUE and on the results.

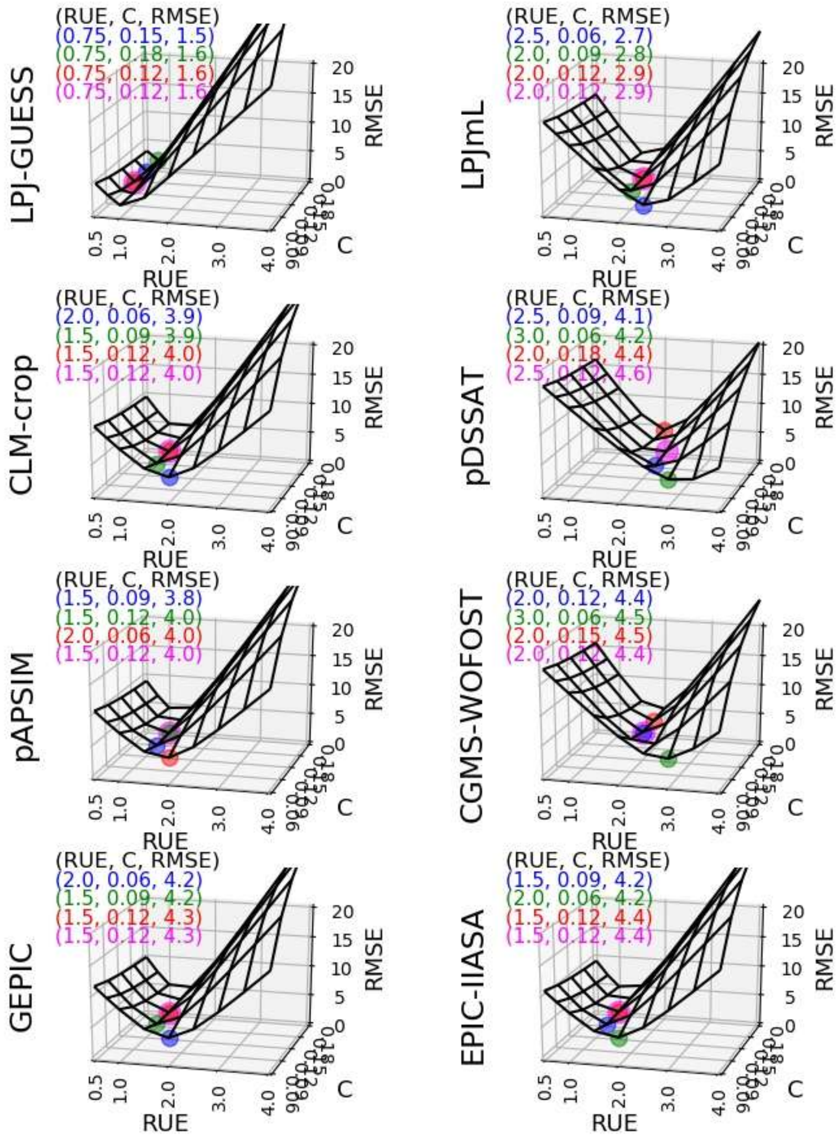

Fig.R1 : update of the Figure 3 when the range of RUE values allowed during the optimization procedure is increased.

5) Ln335: please clarify what are the two sub-groups, and what is a third variable?

To assess the mismatch between $biom_{GGCM}$ and $biom_{SMM}$ 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 temperature (e.g. heat stress) could be missing in SMM. We clarified this point in the new version of the draft (L340).

6) Fig. 3: the listed values should be (RUE, C, RMSE) rather than (C, RUE, RMSE)?

True. Thanks for pointing to this mistake. We re-do Fig.3.

7) Ln 555-570: The RUE in SMM are much lower than GGCMs. Three SMMs havea best RUE to be 1.5. Comparing with the RUE (for the total biomass) at the lower end of the commonly reported values of 3.1 g DM MJ-1, the RUE of SMMs is two-fold smaller. The authors imply that the RUE corresponds to the aboveground biomass, such that it should be smaller than the RUE for total biomass. However, they show that the root:total ratios varying from 0.4 to 0.2 in GEPIC, which seems cannot explain the two-fold differences in RUE from SMM and GGCMs. The second argument is about the LAImax. But from Fig. 2, it seems that the best maxnleaf of the SMM for GEPIC is 9.5, which should translate to a LAImax in this SMM to be 1.8, again two-fold smallert han the original GEPIC (3.5 as the authors wrote), which contradicts to the authors'claim that LAImax is 5 and such that the LAImax can compensate lower RUE.

The two arguments mentioned by the Referee (aboveground vs total biomass ; LAI_{max}) are used in our manuscript to explain the mismatch between the RUE calibrated for GEPIC in our SMM optimization (1.5 g DM MJ-1) and the true RUE prescribed to GEPIC (4.0 g DM MJ-1). We come back on these arguments below.

RUE prescried to GEPIC is for total biomass while our RUE is for aboveground only. In GEPIC, root:shoot ratios vary between 0.4 at germination to 0.2 at maturity. Thus, RUE prescribed to GEPIC corrected to represent only aboveground biomass should vary between 2.4 at germination ((1-0.4)*4.0) to 3.2 at maturity ((1-0.2)*4.0). This cannot explain totally the mismatch but can contribute to it, in particular in stages close to germination. We modified L577 to make this point clearer :
« 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 (in particular in first growth stages), to our calibrated RUE for GEPIC."

About LAI_{max} : Fig.2 quoted by the Referee is for assessing the sensitivity of the global-averaged SMM biomass to each parameter. But during the calibration procedure, only C, RUE, GDD_{1leaf} has been calibrated (C and RUE are constant in space ; RUE varies in space) while other parameters are equal to their initial estimates (see L309). Thus, during the calibration of (C, RUE, GDD_{1leaf}), max_{nleaf} (maximum number of leaves) is set to 19, corresponding to a LAI_{max} of 5.
To go further, we performed a new sensitivity test, where we modified LAI_{max} within SMM towards lower values and check how it modifies the calibrated RUE. The aim is to assess if it increases the consistency with the RUE prescribed to GEPIC.

The LAI is not used in SMM. Instead, it is a pronostic variable (Eq.9 of the original draft). LAI_{max} can be derived from Eq.9 by prescribing n_{leaf} (the number of leaves) equal to max_{nleaf}, i.e. :

$$LAI_{max} = S_{leaf} * d_{plant} * max_{nleaf}$$

with S_{leaf} is the individual leaf area and d_{plant} is the plant density. Thus, in SMM, LAI_{max} can be modified either by modifying max_{nleaf} or by modifying one of the other parameter S_{leaf} or d_{plant}, with different effect on the aboveground biomass simulated : max_{nleaf} modifies the seasonal cycle in n_{leaf} (Eq.3) while modifying S_{leaf} or d_{plant} have a similar effect on APAR through their implication in the computation of C (Eq.5 and 8).

We modified max_{nleaf} by keeping C (thus S_{leaf} and d_{plant}) at its initial estimate (C=0.12). We performed some SMM simulations in which different RUE values are tested for 3 different values of max_{nleaf} :
max_{nleaf}=19 (initial estimate as in the original draft)
max_{nleaf}=13, which is consistent with a LAI_{max} of 3.5 as in GEPIC
max_{nleaf}=10, corresponding to LAI_{max} of 2.7, that could be considered as a more extreme test. Figure R2 shows how the RMSE varies for each set of SMM simulations with RUE (line) and which RUE would be chosen as minimizing the RMSE (dot). As expected, calibrated RUE increases when max_{nleaf} decreases (i.e. from black to yellow then to red). The same effect is seen when modifying C and keeping max_{nleaf} constant (and equal to its initial estimates) (Fig.R3). Nevertheless, in both cases (Fig.R2-R3), the effect of reducing LAI_{max} in SMM to be more consistent with GEPIC (from black to yellow) on calibrated RUE is small. We added one sentence about this point in the discussion of the draft (L585):
"Some additional sensitivity tests with varying *LAI_{max}* (not shown) suggest nevertheless that the mismatch in *LAI_{max}* between GEPIC and SMM contributes only slightly to the mismatch in *RUE*."

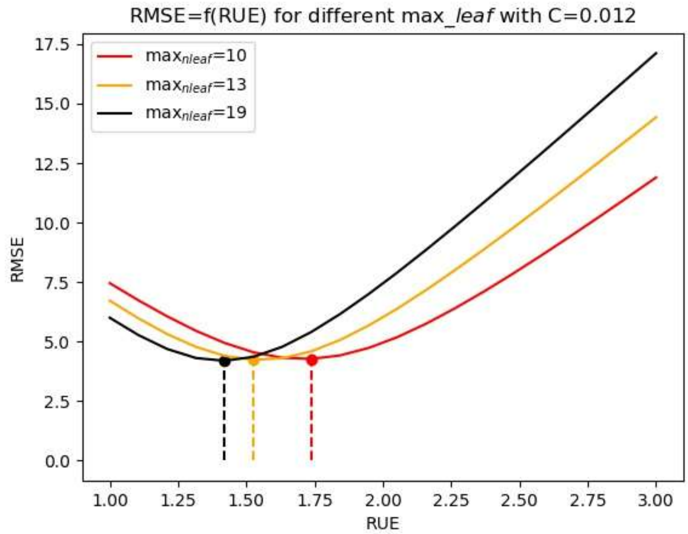

Fig.R2 : Variation of RMSE between SMM and GEPIC as function of RUE for different max_{nleaf} values. With C equal to its initial estimate, the different values of max_{nleaf} tested here correspond to LAI_{max} equal to 5 (as in the original draft, black), 3.5 (as GEPIC, yellow) and 2.6 (extreme case, red). Note that 20 SMM simulations have been done for each curve, i.e. the increment in RUE used here (~0.1 g DM MJ-1) is smaller than the one tested in the original draft (0.25 in Fig.3).

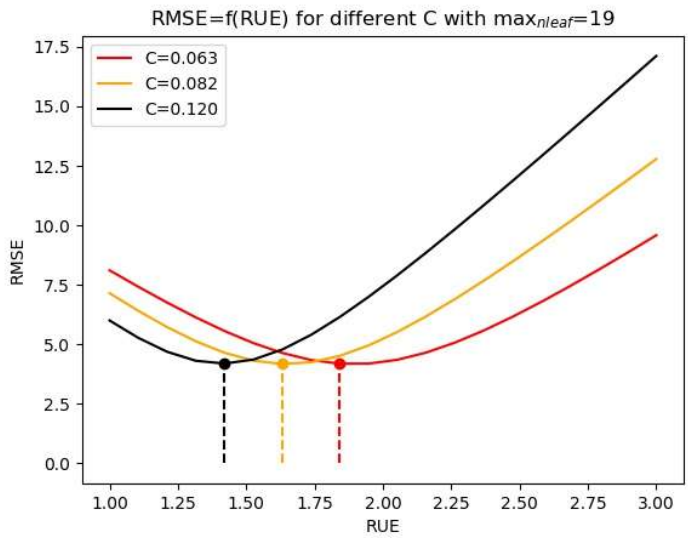

Fig.R3 : Variation of RMSE between SMM and GEPIC as function of RUE for different C values. With max_{nleaf} equal to its initial estimate, the different values of C tested here correspond to LAI_{max} equal to 5 (as in the original draft), 3.5 (as GEPIC, yellow) and 2.6 (extreme case, red).

Overall, this first version of SMM is a simple emulator that only targets at the potential biomass and yields and does not account for the water and nutrient stresses. Accounting for these stresses requires to describe quite a few new processes with more parameters, which will exponentially increase the demand for computing resources for the parameterization. I am keen to see how the next development of such a SMM will emerge in the near future.

We thank the Referee for his/her interest in next developments of SMM. It is true that the increase in number of parameters has a huge effect on the demand for computing resources. In the current stage, we are aiming to optimize SMM against empirical dataset of potential yield (Mueller et al. 2012). This step has its own difficulties as the constraint from this dataset is small (no information about the potential biomass, only about the potential yield) but we think that it is important to confront SMM not only against GGCM (as in the current paper) but also against datasets more related to observations.
Once parameters allowing SMM to fit empirical potential yield have been calibrated, we will use SMM to assess nutrient limitations. We hope that decoupling in time the calibration of parameters related to the achievement of potential biomass/yield and the investigations about nutrient limitation (with additional parameterizations) will prevent us to issues related to computing resources.

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(7419), 254–257, doi:10.1038/nature11420, 2012.