Interactive comment on "The GGCMI phase II emulators: global gridded crop model responses to changes in CO_2, temperature, water, and nitrogen (version 1.0)" *by James Franke et al.*

Anonymous Referee #3

Received and published: 13 April 2020

The authors present a highly detailed description and evaluation of newly-developed statistical emulators for global gridded crop model simulations (as being contributed to the GGCMI Phase II), specifically targeting emulation of mean yield changes due to changed climate conditions. The authors construct these emulators by varying over carbon dioxide concentrations, temperature, water, and nitrogen inputs, and also test the effects of adaptation.

In general, this paper is highly useful contribution to the emerging work of global gridded crop modeling primarily due to providing a very well tested, relatively low-error, computationally economical, and low data-input means of reproducing and/or running

GGCM experiments (again, as related to GGCMI Phase II). Given the computational expense and large data requirements of the GGCMs, it is worthwhile to have an option to run climate-crop experiments with comparatively less "overhead" and relatively high confidence that the emulators overall faithfully represent specific model and (thus ensemble?) sensitivities. I also think that the authors generally did well to note some key uncertainties both in the GGCMs and how these influence the emulators, although a couple of aspects could be addressed a bit more (and I note these below).

Ultimately, one of the key strengths of this work is to provide a comparable and easier means of representing geospatial crop responses relative to the GGCMs (which certainly have other uses as full process-models). Thus, with a few minor revisions, this paper makes an interesting and useful contribution to the field, and I anticipate these emulators being put to good use by many researchers exploring mean climate-crop interactions.

I do have just a few questions and remarks that may be useful to the authors as they think about some minor revisions and next steps:

Section 3.2, Lines 215 onward: I found it interesting that several of the carbon-terms dropped out due to their relatively negligible contributions. [CO2] effects on crops (and ecosystems!), and their nonlinear interactions with other changing climate parameters, are still highly uncertain. Crop models also display much variability in their respective [CO2] responses. I noticed that for the simulations emulating HadGEM responses [CO2] was held fixed or not varying with other parameters. Since the authors are emulating, and evaluating against, GGCM outputs, if the GGCMs do not display [CO2], then it follows that neither will the emulators I suppose. However, I wonder if the authors could further comment on this: the fact that [CO2] was negligible for the emulation does not necessarily mean the effects are negligible in reality, correct? I don't see this discussed much elsewhere in the manuscript, so having a bit more commentary on this, with respect to [CO2] and/or more generally, would be useful as readers consider the terms of your model emulator.

Section 4 and elsewhere: This comment is not just relegated to Section 4, but I'm more generally trying to parse out the relative contributions of climate variability and mean climate change, and the arguments provided in the paper that support emulation of the latter. I think there are two types of "variability" (admittedly not the best word, perhaps more "characteristics" other than the climatological mean yield) that the authors address that might be clarified just a bit more in the Discussion to avoid any confusion.

Firstly, in Section 2.2, the authors make the case that year-to-year variability is structurally different than simulation of the climatological mean yields, and that the former doesn't preclude the latter, correct? The authors also highlight (I think) that the emulators are not suitable for a full interpretation of interannual variability and extremes, particularly highly non-linear interactions between climate (and other) parameters, despite the higher order terms of the emulation (which, as the authors note, are geared towards emulating climatological means).

Secondly, in Section 4.3, the authors demonstrate that potential shifts in the distributions of climate parameters do not impact the climatological yield emulations and the results still compare well with the GGCMs (Figure 9). This fact – this shift in distribution and potential changes in variability that result from it (which is where the readers' mind might go, as mine certainly did) – is I believe distinct from the discussion of year-to-year variability discussed above.

I appreciate that the authors have provided detailed explanations of their approach, treatment, and findings wrt to considering these variability and distributional changes. Still, there's a lot of material here to keep track of, and I think it may be useful to reiterate each of the above points clearly in the Discussion (particularly if I've mistakenly represented it, as I think this may be an example of reader confusion!). For example, there is a sentence in the Discussion that minimizes the impact of future variability (particularly at the aggregate level – around Line 445), particularly in the area aggregate, and I think this is in reference to the findings in Section 4.3 However, this doesn't mean that interannual variability, or extremes or nonlinear interactions, won't be impactful to

future (or current) crop impacts.

Discussion: Lastly, I think the major point of this paper is to provide these emulator frameworks as an alternative to climate-crop assessments with the full process-based GGCMs. I therefore understand the authors' approach to evaluate the emulators against the GGCMs – this is quite reasonable.

It might be helpful, though, to take one step beyond this and compare to some observed historical yield changes. I would not expect this to be better than the GGCMs, and such evaluations have already been done for the GGCMs, so I would expect to see a similar response (and this is notwithstanding the applicability and veracity of comparison products). However, I don't think I've seen such an evaluation for GGCMI Phase 2 yet (I expect one is planned), and so pre-empting this with a comparison of the emulators may just be useful to have on hand. If this could be done and stuck into Supplementary, it would be a useful figure for the community moving forward, rather than having to show an intermediary figure of emulator-GGCM comparisons.