

## ***Interactive comment on “The GGCM phase II emulators: global gridded crop model responses to changes in CO<sub>2</sub>, temperature, water, and nitrogen (version 1.0)” by James Franke et al.***

### **Anonymous Referee #1**

Received and published: 31 March 2020

The study is a step forward in crop model emulation and is in principle a useful contribution to the literature. The paper contains a lot of excellent technical work. I focus here on areas for improvement, which I describe as major simply because I think a re-framing is needed in order to ensure that the paper is used well, and is not mis-used in the future.

The uses stated in the abstract for the emulators are: “providing a tool that can facilitate model comparison, diagnosis of interacting factors affecting yields, and integrated assessment of climate impacts.” It would be good to understand more from the paper about how these different usages are envisaged. In particular, the suggestion that the

C1

emulator might be used for integrated assessment lacks evidence. It is far from clear that this would be a sensible step to take, because study is subject to a number of important limitations. Whilst the authors are cognisant of these limitations, not enough attention is paid to them in the way that the work is framed and interpreted.

One limitation is the use of mean yields. “We emulate the climatological mean response, because that is the response of interest in assessments of climate change impacts. ... Emulation then becomes relatively straightforward, since changes in time-averaged yields are also considerably smoother than those in year-to-year yield response.” – L108. Why is mean yield the response of interest? Perhaps it is because it is relatively straightforward, rather than because it is useful per se. Climate variation explains a third of global crop yield variability – Ray et al. (2015) Nature communications. Why do the authors think that mean yields are interesting? There would need to be a clear rationale in the paper.

Assuming a rationale exists for assessing mean yields, over what lead times might the emulators be usefully used? As the authors point out, the emulators cannot be used out of sample, thus implying relatively short lead times, before climate changes significantly. However, over the next couple of decades, changes in mean yields are unlikely to be important relative to extremes.

Assuming a focus on mean yields can be justified for an appropriate lead time, there remains the question of why an emulator is a valid method to use. Two issues need to be addressed here:

i. Whether or not the emulator is fit for purpose. Does it reproduce observed yields well? The link to observed yields is tenuous. Error (which should actually be termed “deviation” – since it is not a true error) is defined relative to yields simulated by the underlying crop models. If the emulators are to be used, then one would need to be sure it captures real historical climate impacts. The language on this is imprecise in many places. For example, in the abstract: “... suggesting that effects of changes

C2

in temperature and precipitation distributions are small relative to those of changing means.” This statement is true only of model space; indeed, it is untrue of observations as Ray et al. (2015) and others have pointed out.

ii. Is there a better method? Statistical regressions would by definition capture to some extent observed yield responses to weather and climate. The resulting emulators [are] lightweight, computationally tractable “ – but so are statistical models. Reasons to use an emulator over a statistical model are presented in the introduction. However, neither the lack of observed yields in calculating skill, nor the lack of model calibration (another limitation; see below), are brought into this discussion. Similarly, what does the focus on yield changes, rather than yields per se, mean for the robustness of the methodology?

The other option discussed briefly in the introduction is the use of process based models. The full set of GGCM simulations is available; surely the emulators are not expected to outperform their masters? Presumably the “lightweight” approach is deemed to be an advantage for integrated assessment. If this is so then the advantage should be clearly presented.

The major revisions needed for the paper will follow on naturally from framing it more clearly to demonstrate the uses the emulators can be put to. As is no doubt clear, I think that the rationale for their use in integrated assessment is extremely difficult to demonstrate; but perhaps I am wrong. It would be worth thinking about the conditions (data availability, crop knowledge, model skill, input data availability, ..) under which the emulators might be a preferred option.

Model comparison and diagnosis are easier to justify – but even here some work is needed to explain how the emulators could be used. The emulators could be used to highlight areas of CTWN-A where there is consensus and where there is not, thus providing clear evidence of where model improvement, and associated observational datasets, are needed.

---

C3

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2019-365>, 2020.

C4