Response to

V. Arora (Editor)

---The editor's review comments are quoted in black, and our responses in *blue italics*.---

Authors evaluate the potential of a simple modelling framework to reproduce the geographical distribution of changes in crop yield based on solely the change the global mean temperature (ΔGMT) and change in atmospheric CO2 concentration. This simple modelling framework is calibrated based on results from five spatially explicit process-based crop models that are driven with data from five Earth system model (ESM) for four future RCP scenarios. The potential uses of such a tool are obvious for impact studies but in its current form the manuscript is somewhat difficult to follow. As a reader, I felt several clarifications are needed. In particular, I wasn't able to completely follow the three emulator models used to reproduce the results from the full crop models and still unclear how the different temperature bins come into picture. In addition,it would be helpful for a reader if the plots and graphics were bit more readable as suggested in comments below.
I also have comments marked on an annotated version of the manuscript which I attach as a supplement. Most comments from my hand written notes are summarized below except the minor comments for which I request you to see the attached annotated version of your manuscript. Hope my hand writing is easily readable.

Reply: Regarding the temperature bins: We expect that yield changes may not necessarily scale linearly with temperature over a large temperature range (i.e. 0.5 to 5°C of global warming) but we expect yields to change similarly for similar warming. That's why the temperature-yield relation is not derived from one linear regression of the full scenario period (2005-2100) but from smaller subsets. These could be time slices of a fixed length which, however, would create a scenario dependence. Temperature bins combine yield changes from years with similar global warming, in our case years with global warming in a 0.5°C-wide range. For example, the 2°C bin contains all years with a global warming between 1.75 and 2.25°C. This allows to combine yield data independent of time (scenario), under the assumption that global mean temperature is the primary driver of long-term yield changes.

Specific comments

Page 3, lines 100-105. I find it difficult to believe that spatially explicit crops models (driven with spatially explicit climate information even though it may require some kind of scaling) will be worse off than the approach presented in this manuscript.

Reply: The main advantage of our approach compared to that in Blanc (2017) is that our approach does not require spatially explicit climate scenarios nor access to a crop model (and thus, considerably less computational resources). We will remove the statement in line 101 – 105 that using a two-step emulator approach (crop yield emulator as proposed in Blanc (2017) together with emulated climate) would lead to higher deviations than our one-step approach since we do not have concrete evidence for this.

Page 4, lines 133-135. "The simulated impacts of climate and CO2 changes on global and regional crop yields are shown to be related to global mean temperature change, and to be largely independent of the emissions scenario". There is a bit of circular argument here and then later on in Figure 4 where most of the variance in crop yield is explained by crop models. I think, by choosing a specified ΔGMT you have made ΔGMT independent of

emission scenarios and ESMs. What remains scenario and ESM dependent is the year when this specified ΔGMT is reached. Results from climate models show that the typical spatial pattern of temperature change is also similar across ESMs – that is higher warming over land than over ocean and higher warming at high latitudes than in tropics. Combined with specified ΔGMT this means that it is somewhat expected that most of the variance in crop yield will be explained by crop models.

Reply: The finding cited above may not be surprising, but is a necessary condition for the derivation of the emulators, which is the central part of our study. The goal of the emulators presented in this paper is to provide a method to quickly estimate yield changes for any level of global warming regardless of the specific emissions scenario (i.e. the timing of warming, but within the limits of maximum warming used to train the emulator). For this to work, yield changes need to be largely independent of the emissions scenario. Simulation results show that yield changes for a specific level of global warming depend mostly on the crop model and (to a lesser degree) the ESM which is why we derive crop model-specific and ESM-specific emulators.

Section 2. page 4.) Please 1) specify the time period of projections, 2) introduce the four RCPs and what they imply (i.e. low, medium and high emissions scenarios), 3) introduce CMIP5 a bit more and 4) provide 1 or 2 sentence about how bias correction is done.

Reply: Thank you, we will add more information on all the four mentioned points.

Table 1. page 5. The description of “Fertilizer use” needs to be made more consistent across the five crop models. For example, for LPJ-GUESS I am not sure what does “no consideration of spatial and temporal changes in nutrient limitation” means. Does this mean nutrient constraints are not considered or a specified fertilizer application rate is assumed for all times and all parts of the world.

Reply: The LPJ-GUESS model does not consider nutrient constraints on plant growth. Upon reflection, we have decided that much of the information in Table 1 is not directly relevant for our study: We develop and compare several emulator approaches to approximate yield changes simulated by the crop models. The differences between the crop models participating in ISIMIP have been studied elsewhere (e.g. Rosenzweig et al., 2014, Frieler et. al. 2017) and are not the focus of our study. Therefore, we will remove some of the information in Table 1 and refer the interested reader to the relevant literature. The remaining information in Table 1 will be clarified.

Page 7, lines 200-221. What is MIRCA 2000? Is this an observation-based product?

Reply: The MIRCA2000 dataset provides gridded growing area information for 26 individual crops or crop groups for the year 2000, distinguishing between rainfed and irrigated crops. It is based on a combination of remote-sensing-based and census-based information. The dataset is documented in Portmann et al., 2010 (cited on line 220).

Figure 3 and for other similar figures. Please consider putting the figure titles (Maize, Rice, Soybeans and Wheat) in horizontal format and a bigger font size as suggested in the annotated manuscript. Also, as a reader I was wondering what are the limits on the horizontal colour bar. I found the colour bar a bit difficult to interpret. If the white colour represents yield change of less than ± 5% then how can light blue and light red colours represent changes of more than 50%. For example, +6% will likely be indicated by light blue colour but this is not more than 50%.

Reply: Regarding font sizes we will check and improve readability of the figures. Regarding the colour legend there seems to be a misunderstanding: Red and blue colours denote the percentage of model combinations that agree on a change of at least 5%. White denotes cells where either less than 50% of model combinations agree on the sign of change or the change is less than 5%. A change of +6% will be indicated as either white, light blue or dark blue depending on the percentage of model combinations that agree on such change. This is explained in the figure caption. We will change the figure legend from a continuous colour bar to separate legend entries in order to avoid misunderstanding.

Page 12, lines 333-334, "To quantify the extent of the CO2 induced scenario dependence and its potential reduction at each grid point . . .". I am unable to understand what do "CO2 induced scenario dependence" and "potential reduction" refer to. There are two possible aspects here and I am not sure which one is right in his context. The first is that of CO2 fertilization – is this what is being referred to in this sentence. Second, note that just like by choosing a specified ΔGMT of 2.5 degree Celsius the dependence on emissions scenario and ESM has been reduced ΔGMT is also related to change in atmospheric CO2 concentration. Is this what is being referred to here?

Reply: We apologize for the confusion. We will rephrase this sentence to: "We use two methods to determine the CO2 effect on crop yields within each global mean temperature bin at each grid-cell".

Page 13. Equations (1) and (2) are the crux of the paper. Yet, I unable to understand how these equations are used. Perhaps if more equations were used to describe each and every term of these equations it would have been easier to follow them. For example, the change in yield is actually a two-dimensional quantity (i.e. it depends on the geographical location) at a given time (represented by i in equations 1 and 2). Perhaps if (t) can be used to represent time and not as subscript and if these equations were written again properly it would be easier to understand the objective of these equations. I am also unable to appreciate if equation (1) and (2) were applied at each individual grid cell or to the time series of globally-summed yield. In this context, I am also unable to understand what do the different temperature bins refer to.

Reply: Linear regressions according to Equation 1 and 2 are carried out for each grid cell and for each temperature bin separately. As mentioned in our first reply above, a temperature bin includes all years where the smoothed global mean temperature is within +-0.25°C of the respective bin temperature. E.g. the 2.5°C bin includes values between 2.25 and 2.75°C. Indices for grid cell and temperature bin are omitted from all variables in eq. (1) and (2) for the sake of visual clarity. We will improve the clarity of the text describing the equations and check if indices can be added to the equations themselves.

In Figures 6 and 7, the absolute changes in yield are small (since they are mostly white in colour) yet the percentage changes are huge because those percentage changes are corresponding to small absolute values (as mentioned in the manuscript). Perhaps if the percentage changes can be masked over regions of low crop yields then these figures will be much more easier to interpret than the current percentage change figures. Alternatively, maybe the colour scale for figures with absolute yield changes can be changed and the percentage change figures can be removed.

Reply: Thank you for the suggestion. We will use different colour scales for methods a and b (absolute changes) so that details will be better visible. We will move figures with relative changes to the supplementary online material.

Page 18, lines 439-442. “While approach (b) requires a pair of crop model simulations – one with time-varying pCO2 and one with fixed pCO2, approach (a) only requires the default simulations with time-varying pCO2”. I wasn’t able to appreciate this. So perhaps this should be mentioned earlier on where the three emulator approaches are presented.

Reply: We will outline the differences between emulator approaches earlier in the manuscript.

Page 20, line 485-486. While shown in Lotze-Campen et al., 2008 please include the map of the 10 world regions in this manuscript as well.

Reply: We will include a map of the 10 world regions.

Page 20, line 486-487. “Compared to potential yields, using production gives less weight . . .”. Please introduce (if necessary) and differentiate yield from production. Not all readers can be expected to appreciate this difference between the two.

Reply: Yield refers to the harvest per unit area, e.g. tonnes per hectare. To derive production, yields are multiplied by the year-2000 harvested area. We will explain this better in the manuscript.

Figure 10. It is very hard to see the thin lines corresponding to each RCP. Please consider including thick lines for say 8 or 10 year moving average values which would likely yield a more fair comparison with values from the emulators which do not account for the effect of climate variability on crop yield. Also, please consider using a different set of colours for the four RCPs. The green and the blue seem very similar and the orange and red are pretty close colours.

Reply: In Figure 9, 10 and Table 1 (and the corresponding text in the second part of section 4), we compare decadal averages of regional crop production (crop yields multiplied by growing areas in the region) as simulated by the crop models with those calculated by the emulator methods. In Figure 10, these decadal values are marked by crosses (simulated) and circles (emulated production). Annual simulated production (thin lines) was added for illustrative purposes only. We will remove it in the revision, not least in order to avoid misunderstandings regarding the scope of the emulator (emulate long-term changes rather than inter-annual variability).

Page 23, lines 541-544. I am unable to understand this. Also, I am unsure how there can be any variance attributed to CO2 effects. If this refers to the CO2 fertilization effect – doesn’t CO2 changes gradually in all RCP scenarios. So yes, while there is a trend in specified CO2, there is not year-to-year variability. Or, am I misinterpreting this.

Reply: We will rephrase this to “The variance is calculated separately for each RCP-GCM-GGCM combination and compared to the matching GCM-GGCM combination over the historical period (1980-2010).”

I think, it should be made clear in the manuscript that the emulators do not capture the year-to-year variability but rather the long term trend in crop yield.

Reply: We will add language to clarify this.