Response to

Anonymous Referee #1

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

General comments:

The manuscript "Changes in crop yields and their variability at different levels of global warming" tests an emulator approach to estimate crop yield impacts for arbitrary CO2 emission scenarios based on ISIMIP multi-impact model projections. From a topical point of view, I think this study falls within the general scope of Earth System Dynamics journal. Clearly, the novelty of this paper lies in designing statistical methods to emulate crop yield impacts described in terms of dGMT, going beyond the existing emission scenarios included in the ISIMIP data-cube. Such emulator could provide a shortcut for fast assessment of yield impacts for a range of mitigation targets, thus extending the capacity of ISIMIP ensemble. The authors demonstrate that both mean yields and yield variability are, to a certain degree, directly scalable across emission scenarios, without additional explicit RCP-GCM-GGCM simulations. I believe that the ISIMIP data cube alone, as described by the authors, already provides a powerful tool for linking crop yield changes with arbitrary dGMTs, but has limitations with respect to arbitrary CO2 levels. This is definitely a valuable work, with clear and sound conclusions, and after a moderate revision it would be an original contribution to community of crop impact modellers and for Integrated Assessment Models in general.

Overall, the manuscript is of high technical standard. However, I must say that I had problems with following the methods at some places since various methodical aspects are scattered across the whole document, which makes data reproducibility more difficult. The main "weakness" comes with using the linear-regression (or weighted-average) emulators at a grid basis without further testing its significance. Since the emulators for individual dGMT are based on intercept (a0) and regression slope (a1) of the fitted linear models (determining climate- and emission-induced yield changes, respectively), it would be useful to add a step of testing whether or not these values and the fit overall are statistically significant, and exclude those grids for which it is not so. One could expect that yield response along a narrow range of CO2 increase (such as for a low dGMT) may be quite noisy, and not necessarily captured by linear models in a significant way. This might be the case for some GGCMs and regions where crop yield changes in a response to shifts in T and CO2 are limited by other constrains, such as nutrient deficiency for example.

Reply: We have added a test whether the derived fits are statistically significant. We have also changed the regression for method b) (equation 2) slightly, which increases the number of grid-cells with significant fits. We will add language to the paper to discuss the implications of limited (and noisy) training data for the emulator.

Specific comments:

P3, paragraph 115: soya is not a cereal crop

Reply: Thanks for pointing this out. We will rephrase it.

P4, p.150: "...generated for four RCPs..." which RCP scenarios were used here? It would be nice to list the emission scenarios here as this is the first time the scenarios are mentioned in the Methods section.

Reply: We will add the RCPs used (which are all RCPs considered in CMIP5).

P4, p.165: "...were forced by climate change projections from HadGEM2-ES, RCP8.5..." I wonder if this statement is correct. The authors refer to "YnoCO2" also for other RCPs in Figure 5, if I am not mistaken.

Reply: Only fixed CO2 simulations (YnoCO2) driven by HadGEM2-ES RCP8.5 were used for emulator approach b) (linear regression according to equation 2). However, some crop models such as LPJmL also provided fixed CO2 simulations for other RCPs. We will clarify this.

P5, Table 1: I agree that the model-related assumptions have to be short, but some statements in Table 1 are jargon difficult for others to understand. For example, for LPJ-GUESS, is the thermal time needed for maturity adjusted over time? It is not clear from the explanation in Table 1 … if so, there is probably no shortening of the growing season effect in this model, which may explain more positive effects of warming on wheat in Figure 2, in contrast to other models. Next, what does "decadal adjustment of winter and spring wheat sowing areas based on temperature" mean? Then, do LPJ-GUESS and LPJmL use no information on fertilization, meaning there is no limitation by nutrients at all? If the same assumption applies for both LPJ-GUESS and LPJmL, why not using the same description for both? Please, use harmonized and clearer descriptions that would allow readers to understand fundamental differences in the models, and which implications it might bring. Overall, I understand that the differences in yield projections due to unique specifications of individual GGCMs are not topic in this manuscript, but it would be very useful to discuss some results with a deeper insight into individual models where appropriate (such as that some models simulate nutrient-unlimited yields while the others not).

Reply: Regarding your first question: In LPJ-GUESS, the thermal time needed for maturity is adjusted over time to keep the growing season length constant. This is indeed different from most of the other crop models (except PEGASUS) which keep the heat sum constant, leading to a shorter growing season length under higher temperatures. Regarding your second question: The decision to grow wheat as either winter or spring wheat in any particular grid-cell and year is left to the models. In GEPIC, this can change over time, compared to other crop models which use fixed sowing dates. Regarding the third question: Neither LPJ-GUESS nor LPJmL consider nutrient limitation of plant growth.
After further consideration, we have decided that much of the information in Table 1 is not directly relevant for our study. Our focus is on developing and comparing three emulator methods to approximate yield changes simulated by the crop models; rather than on the differences between the crop models themselves. These have been studied elsewhere for the same ISIMIP crop models (e.g. Rosenzweig et al., 2014). Therefore, we will remove parts of this table and refer to the appropriate literature for readers interested in the crop model details; the information remaining in table 1 will be clarified.

p.7: It is obvious from Fig 1 that different dGMT bins would bring different number of years into analysis (1) and (2). Time intervals for some dGMT bins may be quite short (e.g. 10 years), and probably not sufficiently long to smooth possible short-term inter-annual fluctuations and anomalies in simulated yields. In other words, the mean yield change could be very sensitive to anomalous years. A concept of climatic normal is usually used to eliminate anomalous fluctuations. I am not sure about the implications for this approach, but the authors should consider and possibly discuss this aspect.

Reply: Climatic normals are usually based on 30-year time slices which, especially in the case of RCP8.5, can encompass a significant range of warming (e.g. more than 1.5°C of warming 2071–2100). In comparison, most 0.5°C-wide dGMT bins are much shorter. Since dGMT bins are independent of the emissions scenario this offers an opportunity to combine data from different scenarios and extend the number of years of data in a bin. For example, a total of 66 years from four RCPs fall into the 1°C warming bin. Combining data from different RCPs is not possible for high levels of warming only reached by RCP8.5. We note, however, that all bins *used for training in our study contain at least 7 years, which does allow for interpolation, although potentially with large noise. To clarify this, we will test for statistical significance of derived fits and discuss implications of limited training data for the emulator.*

p.12, Figure 5: please add information whether the linear regressions are statistically significant. Free scaling of y-axis may work better here to see the data scatter more clearly.

Reply: The intention in Fig. 5 is to show that the effect of pCO2 within a given dGMT bin does not depend on the scenario. Figure 5 illustrates this by showing the globally averaged yield difference between the default and the fixed-CO_2 simulations. The linear regression line was added as a visual aid, but is not actually used for any further analysis. We realize that it would probably be more helpful to show separate regression lines for the different RCPs to underscore the argument. We will add these in the revision, and will also use free y-axis scaling. A test of statistical significance is not necessary in our opinion since these regressions are solely for illustration; but we will consider showing confidence intervals along with the regression lines.

p. 16, Fig. 7: as discussed in the manuscript, negative impacts of CO2 are contra-intuitive, even though they may make sense from the statistical point of view. Maybe a rule that all negative impacts of increased CO2 are reverted to zero would make more sense.

Reply: For emulator approach a) (equation 1) both the temperature effect and the CO2 effect are derived from the same regression. In this case, it's not possible to set negative CO2 effects to zero because both effects are linked. For emulator approach b) (equation 2) only the CO2 effect is derived from the regression. Since submitting the first version of the paper, we have changed the regression slightly to derive the correction term a0 in equation 2 more robustly. (We have explicitly accounted for between-model differences in the baseline CO2 value, whereas earlier these were only implicitly included in the estimation of a0.) This has eliminated any negative CO2 effects. We will update the paper with the revised method.

P17. P. 405 and elsewhere: "...change on potential yields..." This statement here (and also elsewhere) is confusing for me. Do the authors estimate yield potentials or yields under historic management?

Reply: All GGCMs except LPJ-GUESS simulate yields assuming present-day management as a starting point. In this case, "potential yields" refers to the fact the yield simulations are not only carried out for areas where the crops are grown today. Instead, yields projections cover all areas where the crop can grow. We will try to clarify this more.

P17.420: (ao(dGMT)) should be (dYCLIM(dGMT))

Reply: Thank you for pointing out this inconsistency with equation 1. We will change it in the revision.

Section 5, p.23: In my opinion, de-trending of simulated yields within a dGMT bin by RCP-specific yield averages provides only an artificial variability, since it combines different RCP projections together. I am not an expert in climate modelling, but I assume that each radiative forcing scenario in a GCM generates a unique temporal variability of meteorological variables, and I have my doubts that mixing RPC-specific simulations together makes sense in terms of dGMT-specific variability. I am not sure that I understand the concept in Section 5 correctly though.

Reply: We do not combine results from different RCP projections. Changes in variance are computed for each RCP-GCM-GGCM combination separately. Figure 11 shows the *percentage of RCP-GCM-GGCM combinations that experience an increase or decrease of yield variance of greater than 5% at 2.5°C warming.*

p. 25: "......implies that projected impacts at different dGMT levels are not substantially dependent on the choice of emissions pathways." This might be true unless lead time is considered. Time horizons associated with individual dGMT levels are of high importance for adaptation and economic processes.

Reply: Time horizons associated with adaptation and economic processes are not considered in the ISIMIP crop simulations. Assumptions on management in the GGCMs are either constant over time or are driven solely by climate.