The authors present an interesting analysis of UK’s wheat yield variability. They first explore the influence of different climatic conditions on wheat yields to then construct a scoring system for combined climate effects. In their analysis, they separate major plant development stages. Finally, they use climate model projections to estimate potential yields in a warmer climate.

Despite the confined regional focus, I would expect that the findings and the presented approach would be of interest to a wide readership. The manuscript is well written.

**Major concerns:**

The paper is based on statistical analysis and this analysis should be described in more detail including a description of underlying assumptions. Especially the part about the scoring system should be better introduced and potentially justified.

The analysis of climate effects during the plant development phases delivers interesting results. The authors argue that with their scoring system they can assess the combined effect of climatic conditions throughout the plant development. Here the question arises whether the climatic impacts during the production phase are the same irrespective of the climatic conditions throughout the earlier plant development stages. For example, Ben-Ari et al. 2018 describes a compound event where the combination of warm winter and wet
spring lead to a crop failure. As I understand the analysis, it wouldn’t be able to capture such compound events if it is not generally bad for wheat to have warm winters and wet springs. This is just an example, but it might help to understand a limitation that comes from splitting up events. I would find it interesting to read the authors’ view on this concern. These reflections could also be included in the discussion.

The use of only one climate model appears problematic to me. Furthermore, for this type of analysis I don’t see the benefit of high spatial resolution if in the end regional averages are used. I would find it more convincing to see a CMIP6 ensemble instead of one high-resolution model. On the other hand, the climate model projections are not the main part of the analysis. Therefore one could also think of comparing this climate model to the CMIP6 ensemble and discussing the differences and potential biases.

**Minor comments:**

The abstract could be improved. At the moment it reads a bit like a summary of different results and ideas. The aim of the study should be clarified more precisely and not all results have to be included in the abstract.

L9-10: “future impacts of climate projections on wheat”. I think this should be formulated differently.

L30-31: Is this due to climatic conditions only? Or does technology play a role here?

L97-101: Did you consider a different spatial aggregation methods for precipitation? While for temperature it seems reasonable to average over the regions, for precipitation there could be other meaningful choices. As an example, what would you think about area affected by extreme precipitation instead of regionally averaged precipitation?

L111: I think you should mention here, that the scientific community is not considering this scenario as a plausible future. I have seen, that you do so later on. Maybe still worth mentioning earlier.

L102: Although the UKCP Local simulations are surely great, there remains a large uncertainty with respect to forced changes in precipitation. The accurate representation of small features in these simulations does not necessarily reduce the uncertainty concerning
the regional trend in precipitation. Therefore it would be good to compare the precipitation tendency from this model with climate models from other institutes. I have seen that you do so later in the manuscript.

L120: Could you add one or two sentences on the bias correction method? Is it a trend-preserving bias correction?

L173-175: Are these two sentences contradicting each other?

Section 3.2 and Table 2: How would you explain that the effects of climate conditions are different between the regions? I wouldn't have expected different effects for the different regions. If there is a reason for that it would be good to mention it. You explain this in L194-206, right?

L220: What is the advantage of using this “score”. Couldn’t you also work directly with the correlations of table 2?

L246: Is “sample” the correct word here? I would have written “project”. But I’m not a native speaker.

L246: Is this statement true for the UK in particular? And how did you get there? I think it would be good to spend a few more sentences on this aspect to provide a good overview of potential biases over UK.

Figure 7: I think this figure could be improved a bit. What do you think about displaying the ensemble spread by a shaded area and the ensemble median by a line?

L292: “since crop yields” instead of “since inter-annual crop yields”?