Machine learning methods to assess the effects of a non-linear damage spectrum taking into account soil moisture on winter wheat yields in Germany

General comments

The authors apply a random forest procedure to explain observed yield anomaly in Germany thanks to meteorological predictors. The central result of this paper is the quantification of individual non-linear contributions of meteorological variables/indices and especially the important role of soil moisture. The ALE plots are valuable material in this study (I am not sure about the interpretation of confidence intervals though, see specific comment n.3). To my opinion, this paper deserves publication after minor revisions.

Specific comments

1. Fig 1b. You mention the identification of significant trends, did you perform some trend test or shift test?
2. l141. Is the clustering performed on raw yield directly, or anomalies? How should one interpret the clusters obtained: are the counties gathered in terms of yield magnitude or variability or occurrence of extremes?
3. Figure 4: the confidence intervals you obtain are related to the smoothing function, and not the robustness of the RF model itself. I am wondering to what extent it is possible to interpret it as an uncertainty of the variable effect (ex. l.238). To my understanding, this confidence interval tells us about the uncertainty of the smoothed curve, but not about the uncertainty of the local effect.
   (small remark: it is nice to specify the package for the ALE).
4. l. 405 Do I understand correctly that "the feature is shuffled" means that the variable, for which we want to compute the importance, is shuffled?. I.e., to get the importance of e.g. SMI3, the RF is re-run on the exact same data, except for SMI3, which is shuffled in time, and then the MAE is computed?

Technical corrections

1. l233. missing ")".
   l240 error in the reference to the figure
   l257 missing ")"

It seems like there are some confusions in the figure captions:
2. Caption figure 3 not "observed and predicted data for the *two* clusters" but "*four* clusters"?
3. Figure 5: you chose to conduct the analysis with PAM 4 clusters, so in Fig5 it is four and not eight subregions, right?
4. Figure A5 and A7: the captions don't seem to fit with the figures (clustering method and number of clusters).
5. l.378: aren't the results in the appendix about PAM 4 and 6, and hierarchical 2 and 6?