

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

5 Michael Peichl^{1*}, Stephan Thober¹, Luis Samaniego¹, Bernd Hansjuergens², and Andreas Marx^{1**}

¹UFZ-Helmholtz Centre for Environmental Research, Department Computational Hydrosystems, Permoserstrasse 15, D-04318 Leipzig, Germany

²UFZ-Helmholtz Centre for Environmental Research, Department Economics, Permoserstrasse 15, D-04318 Leipzig, Germany

10 * michael.peichl@ufz.de

** andreas.marx@ufz.de

We thank the reviewers for their detailed comments, which helped us to improve our manuscript substantially. In particular, the main comment from reviewer 1 was very helpful because the corresponding analysis of why the clusters are identical to the state boundaries helped us identify an issue within the data provided to the clustering algorithms that was not valid.

15 Unfortunately, this analysis is the basis of the statistical analysis and therefore has profound implications for all of the results presented in this manuscript.

Reviewer 1

Summary

20 *The authors present an assessment on the relationship of wheat yield anomalies to hydrometeorological features using cluster analysis and random forest together with accumulated local effects. They achieve good model performance and additionally present an in-depth investigation of the most important driver variables, which adds additional value to the article. They highlight that their results serve well for the regional identification of harmful seasonal hydro-meteorological conditions. Furthermore, they state the potential usage of their research e.g. for the identification of harmful features and their related thresholds and also carefully underline the limitations regarding out-of sample predictions, underestimations of extremes and missing inclusion of interaction effects in the model. I suggest that the manuscript should be accepted with moderate revision.*

Dear Reviewer, thank you for your detailed summary and suggestions for accepting the manuscript with moderate revisions. We have responded to the major comments and the minor comments throughout the manuscript. Please find our responses below.

Major comments

- 30 *The fact that the clusters are identical with the borders of the federated states of Germany (or groups of them) without a single exception (as far as I can see) is not discussed and does not seem obvious to me. The authors should provide an explanation for this, e.g. stating which of the included variables in the cluster analysis are so strikingly different between the federated states that it leads to a perfect match of clusters and federated states (while any environmental / climatic differences do not matter in the clustering). It seems plausible that differences between eastern and western Germany due to the different political systems*
- 35 *in place in the past can have a significant influence, but I wonder why this should hold true at the administrative level of federated states. The variables included in the cluster analysis are average yield and monthly averages and daily observations of the meteorological data for the entire year and SMI for both the upper layer and the entire soil column, right? I cannot see how these variables should primarily be superposed by the shapes of the federated states of Germany, even though I understand your argument that average yield is connected to farm size (which differs strongly between eastern and western Germany). We*
- 40 **sincerely thank you for this comment, as the corresponding thorough review of the cluster approach helped us identify a problem within the data provided for the cluster algorithms. Unfortunately, the data contained a variable that included state information, which was an artifact from the previous processing of the data. This data is not scaled and thus dominates the clustering results. For this reason, the resulting areas correspond to the state boundaries. We sincerely apologize for this error. It unfortunately affects all subsequent results as well.**

45 Minor comments

[l. 11] “R-squared” or preferably “R²”. Make this consistent throughout the paper.

Thank you for bringing attention to this unsightly habit. We use now “R²” throughout the text.

[l. 16] “I assume you mean “crop yield variations”.

Yes, we do.

- 50 *[l. 79] Is the trend statistically significant?.*

Yes, the trends in all three configurations (loess is defined as natural splines with 4 degrees of freedom) are significant. An ANOVA test shows, that the loess model with non-linear trend is significantly better than the linear model.

[Figure 1, 2, 5, A1] There are various grey counties, I assume mostly cities and/or counties without nonirrigated arable land (Peichl et al. 2018). If all these counties are the ones with 0 years of data, this should be clearly visible and stated in Fig. A1.

- 55 **Thank you, we adapted the Fig. A1 and the description accordingly.**

[Figure 1 caption] Consider specifically naming the exceptional years 2003, 2014 and 2018 here.

Thanks, the years are now named.

[Table 1] It should say “max. T>3°C” for alternating frost according to Gömann et al. 2015

Indeed, that is how we defined it. Thank you for pointing out this error to us.

- 60 *[Table 1] The presented variables in Table 1 seem plausible, but I wonder how exactly you came up with them? You sometimes depart from the months suggested by Gömann et al. 2015, e.g. they state the usage of Jan-Apr for Alternating Frost, while you*

also included May.

Thank you for drawing attention to this discrepancy. This research project also incorporated expert knowledge from interviews with farmers, which suggested the need to extend these time periods. Because the machine learning algorithm used here inherently performs feature selection, we therefore extended the time periods suggested by G6mann et al. (2015) rather than restricting them.

[Table 1] It would be beneficial to have all predictor variables for the random forest in one overview table, so consider including SMI here, too.

For a better overview, all variables are now included in Table 1.

[l. 100] "each grid point depends"

Thanks, we use cells instead of points to improve clarity here.

[l. l. 101-102] You compare your soil moisture index to simulations, but the soil moisture index used here is also simulated, so I do not understand this sentence fully. Could you elaborate please?

Soil moisture is presented here as an index because an index configuration supports the reduction of systematic errors of data that are simulated as well as spatially processed, such as in the present study (Auffhammer et al., 2013; Lobell, 2013)."

[l. 113, 350, 368] adjust citation brackets

Thank you for pointing this out.

[l. 113:] I understand the line of argumentation of the authors, that soil moisture is a slow-responding variable with a long memory, but relating this to autocorrelation per se might be a bit misleading here, because temperature is of course also auto-correlated. So consider writing something like "The autocorrelation / long-term temporal persistence of soil moisture is comparatively high in comparison to temperature".

Thank you, we rewrote the entire passage to clarify the points you made. Cumulative measures are widely used for meteorological variables as for instance growing degree days and killing degree days (Schlenker and Roberts, 2009) .

[l. 116] State here also specifically that on the other hand for SMI all months of the growing season were used (only SMIA is reduced to four months).

This is elaborated here and shown in Table 1 now.

[l. 119] My understanding is that masked grid cells are excluded, but you mean the opposite that you kept only non-irrigated agricultural land, right?

Grid cells that are not non-irrigated agricultural land are excluded and the remaining cells are used to calculate the respective county average.

[l. 145] Did you run the random forest model for each cluster size from 2 to 16 for each algorithm as you did for your standard internal validation?

Yes, we ran it for all possible combinations of cluster size (2-16) and clustering algorithms.

- 95 *[l. 147-49] I wonder whether the statements on the data included in the cluster analysis should be rather mentioned adjacent to the data included in the random forest model in section 2 to have the complete overview at once.*
- We placed the data information here to avoid confusion with the data used in the Random Forests.**
- [Table 2] While SMIa stands for the entire soil column, usage of the term SMI is ambiguous. You use it to refer to the uppermost 25 cm (e.g. l.111), but also as a general term for both depths (e.g. l. 148, Table 2). Consider using separate terms*
- 100 **We now use the abbreviation SMI only for the top 25 cm and when it is a general term we do not use the abbreviation SMI but write soil moisture index.**
- [Figure 2] You use dark grey for missing data and light grey for cluster 2. Consider taking another color for cluster 2 to make it better distinguishable.*
- The figure is now revised and the gray areas should now be easier to see.**
- 105 *[Figure 2 and A1:] coordinate degree sign as superscript*
- Thank you, we have not noticed this yet as it happened when converting the respective files.**
- [l. 189] You mention structural differences between western and eastern Germany. Consider mentioning briefly here the large differences in farm sizes due to different political systems in place, as an international audience might not be aware of this.*
- 110 **Thank you, we added this information backed by data from the Agrarstrukturerhebung in 2016.**
- [Figure 3] Indicate that the regression line is shown in bold black in the figure description. I assume the ellipses are point densities? Please specify this.*
- We added those information in Figure 3.**
- [l. 227] I assume this reduction is to be compared with an average month, right?: “for each month with an SMI value lower*
- 115 *than 0.125 compared to an average month.”*
- By definition, an SMI of 0.5 indicates the 50 percent quantile of the empirical distribution of water content for that county and thus serves as a benchmark for the average month (assuming that the empirical distribution is symmetric). The value of 0.125 is taken here as representative of a threshold at which yield falls rapidly relative to average yield expectations. We have modified this section to make the interpretation clearer.**
- 120 *[Figure 4] Is there a reason why there are many more red dots for the SMI compared to Heat and PS in a given subpanel?*
- Meteorological features are discrete values (counting days above or below a predefined threshold), while SMI is a continuous feature between 0 and 1.**
- [Figure 4] You state for Fig. A4 and in the text section describing Fig. A5 that the average of 50 repetitions is taken. Is this also the case here?*
- 125 **Thank you, the 50 in Fig. A5 refers to the number of repetitions of the permutation to define the variable importance. Instead of using the phrase interval size, we now use grid size in the context of ALE-plots to avoid further confusion. The plot is based on the feature importance assessment with 50 repetitions and then the ALE plots are shown for the most important features. These plots are based on a grid size of 100.**

130 [Figure 4 and A4] You chose an interval size of 100, whereas for Figure A4 it is 50. Why did you choose different interval sizes?

The grid size in Figure A4 of 50 was an artifact of previous procedures. The ALE plots shown here are also based on a grid size of 100.

[236] SMI year-round. "entire" could be confused with the entire soil column.

Here, "entire" referred to the entire range of the SMI in March. This has now been corrected.

135 [l. 239] "led to large losses"

Thank you, changed accordingly.

[l. 241] "but need to be"

Thank you, we changed it accordingly.

[l. 243] I assume "crop yield potentials" might be more precise.

140 **Thank you, to include "yield" is indeed more accurate.**

[l. 245] "water governs"

The structure of the entire sentence was odd so we changed it: "The observation that the absence of water governs crop production in eastern Germany is in alignment with recent studies."

[l. 246] delete "this"

145 **See reply above.**

[l. 266] "eleventh day"

We changed it accordingly.

[l. 268] Why especially East Germany? It is valid for non-cluster and cluster 1, too, is it not?

Thank you for bringing this ambiguity to our attention. We have changed the entire paragraph there: "Our approach,

150 **which explicitly controls for plant water supply through soil moisture, generally shows (for the no cluster approach as well as cluster 1 and cluster 2) more negative effects in terms of water deficit compared to heat. Especially for eastern Germany (cluster 2), the water deficit in the upper 25 cm of the soil plays a prominent role in late summer and spring."**

[l. 274] delete "years" in front of "in-sample"

We have deleted years.

155 [l. 274] It did not directly become clear to me that 2019 is an out-of-sample year at the first read. I suggest explicitly naming it an out-of-sample year in this context (as you also do below).

We added this information to clarify in the beginning of this paragraph that 2019 was not used in the training set.

[l.286] "in some of the easternmost"

Thank you, we changed it accordingly.

160 [l. 288] "the highest negative"

Thank you, we adopted this.

[l. 314] You use the terms East / eastern Germany and northeastern Germany. If all these mean the same region, consider using only one term or are you specifically referring only to cluster 7 with northeastern Germany?

165 **Thank you, we changed the term here to eastern Germany. For the rest of the text we use eastern Germany when the spatial scope is the focus, and East Germany when the political system and the history of the region is of particular interest.**

[l. 324] GCM: Abbreviations should be written out at first occurrence.

Thank you, we introduced global climate models.

[l. 355] "a small number" "a rather large number"

170 **Changed accordingly, thank you.**

[l. 367] Do you mean "mean absolute error"?

Yes, indeed, the mean absolute error is meant.

[l. 376/378] Do you refer to SMI11 and SMI12 as "lagged"? Simply because they belong to another calendar year does not make them "lagged" variables in my opinion, so this term might be inappropriate here.

175 **Thank you, to avoid any confusion when it comes to those terms we instead refer to the previous years.**

[l. 377] Is SMI11 or Heat8 meant here?

SMI11 of course.

[Figure A4 caption] "50 repetitions"

Thank you, we adopted this.

180 *[Figure A5 caption] The caption should specify that the range of each variable stems from 50 (or 100?) repetitions.*

Thank you, we included this information now.

References

- Auffhammer, M., Hsiang, S., Schlenker, W., and Sobel, A.: Using Weather Data and Climate Model Output in Economic Analyses of Climate Change, Tech. Rep. 2, National Bureau of Economic Research, Cambridge, MA, <https://doi.org/10.3386/w19087>, <http://www.nber.org/papers/w19087.pdf>, 2013.
- 185
- Gömann, H., Bender, A., Bolte, A., Dirksmeyer, W., Englert, H., Feil, J., Frühauf, C., Hauschild, M., Krengel, S., Lilienthal, H., Löpmeier, F., Müller, J., Mußhoff, O., Natkhin, M., Offermann, F., Seidel, P., Schmidt, M., Seintsch, B., Steidl, J., Strohm, K., and Zimmer, Y.: Agrarrelevante Extremwetterlagen und Möglichkeiten von Risikomanagementsystemen: Studie im Auftrag des Bundesministeriums für Ernährung und Landwirtschaft (BMEL); Abschlussbericht: Stand 03.06.2015., Tech. rep., Johann Heinrich von Thünen-Institut, <https://doi.org/10.3220/REP1434012425000>, 2015.
- 190
- Lobell, D. B.: Errors in climate datasets and their effects on statistical crop models, *Agricultural and Forest Meteorology*, 170, 58–66, <https://doi.org/10.1016/j.agrformet.2012.05.013>, <https://linkinghub.elsevier.com/retrieve/pii/S0168192312001906>, 2013.
- Schlenker, W. and Roberts, M. J.: Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change, *Proceedings of the National Academy of Sciences*, 106, 15 594–15 598, <https://doi.org/10.1073/pnas.0906865106>, <http://www.pnas.org/cgi/doi/10.1073/pnas.0906865106>, 2009.
- 195