Weather-induced crop failure events under climate change: a storyline approach

Henrique M.D. Goulart, Karin van der Wiel, Christian Folberth, Juraj Balkovic, and Bart van den Hurk

= Response to the reviewers =

Reviewer #1:

Goulart et al. provide an interesting analysis of soybean failures in the US as modeled using Random Forest in the present climate and in future climates. The research is well conducted and uses appropriate methods. My main concern is not with the methods or results, but with the interpretation of those results. The manuscript is well written and the figures describe the data well. I thank the authors for presenting the research in such a complete and coherent manner.

We would like to genuinely thank the reviewer for their constructive review of our manuscript. In this document we respond to these comments and highlight the modification in the revised text/figures.

Major comments

I rather like your use of Random Forest and you apply the model rigorously. Your use of shuffling data is also clever and provides a nice analysis structure, but your interpretation of the results could be improved. See my two comments below

- The interpretation of the compound factor becoming less important needs contextualization. I don't think it's quite correct to say that it becomes less important, but only that it becomes more dependent on precipitation. I think this relates to your test of importance being to shuffle the data in an attempt to break the correlation structure as compared to the ordered data. But in a warming climate, when all years are hot years, the dependence of joint hot-dry extremes becomes dependent mostly on rainfall, making the joint failure and shuffled failure years similar to one another at high return periods. I think it is misleading then to say that the compound nature of these events becomes less important, only that their occurrence depends primarily on precipitation anomalies.

The reviewer makes a good point. First we decided to change the "compound factor" for "relative compound contribution" to avoid confusions. The reduction of the relative compound contribution is because more years are warm, and the definition of failure, as defined in the paper, becomes more dependent on rainfall. Therefore, in the future climate simulation the failure risk in the ordered data increases less than in the randomized data. We interpret it as follows:
- Statistically the reduction of the relative compound contribution shows the events are tending towards univariate behaviour, however
- From a biophysical perspective, the events are still characterised by a combination of hot and dry conditions.

    - L 429-438 - There is no real evidence (from figure 7) that precipitation contributes to the increase in soybean failures overall. There is only evidence that it contributes in a secondary way to truly extreme events, such as 2012 (Figure 10), although its influence is much smaller than that of temperature

We agree with the reviewer. The distributions of precipitation for the warmer scenarios do not show large changes, perhaps a slightly drier tail. We updated the text to reflect this:

Line 469: "The main findings of the paper suggest soybean failures in the Midwestern United States are likely to increase with global warming mainly due to warmer atmospheric conditions during summer. With more frequent warmer years, the joint hot-dry conditions leading to crop failures become more common. Conversely, the increase in frequency of warmer years renders the occurrence of joint hot-dry extremes more dependent on summer precipitation anomalies, reducing the importance of the relative compound contribution. With global warming, the crop failures approximate statistically to a more univariate behaviour, despite still being physically the results of compound events."

Line 408: "In the EC-Earth runs, we see a drying trend for extreme years in the Midwest US (Figure 10d), which explains the different results."

- There is an unintended consequence of using a threshold to define failures that leads to the conclusion that joint dependence will become less important in the future. When you set a threshold to identify "failure" years arising from hot years, you lose the sensitivity of yields to warming above that threshold. For example, in the future even if all years meet the criteria for hot years, drought years may be exceptionally hot. This joint temperature-precipitation correlation structure would show up if you were modeling yields as a continuous variable but is lost once you have converted your dependent variable to a binary (e.g. no discrimination between "failure" and "exceptional failure").
    - In particular, your analysis cannot say that "the correlation between contributing meteorological variables becomes less relevant" (L392-393). You can only say that it becomes less relevant for your chosen yield threshold. If you were to up your threshold definition of "failure", it would once again become important.

The reviewer is correct in their point about events just above the threshold. In our analysis we have chosen to define crop failure events as those years with a yield below the historical standard deviation levels. We have updated the text to be more precise and explicit. A similar use of binary data, without qualification of the strength of failure, is quite common. It is, for example, used in Ben-Ari et al. (2018), Vogel et al. (2021) and Zhu et al. (2021).

Line 429: "... the correlation structure of the variables contributes to the occurrence of compound events, as previously shown by (van den Hurk et al., 2015) and (Santos et al., 2021). Yet, we observe a decrease in its importance under global warming conditions for the failure yield threshold adopted here (Figure 6)."

- It is not true that "non-critical values of precipitation […] can lead to failures" (L394). We can see in Figure 7 that in fact all of the 3C failures still occur at very low levels of precipitation, but it is true that now low levels of precipitation lead to crop failures at higher rates**.** My qualm here is really with the interpretation of the precipitation not DTR as you already discuss the limitations of the model with respect to DTR well in the discussion section.

This is true, precipitation is not changing much between the scenarios. We have updated the text.

Line 416: "... In a warmer world, the same levels of precipitation lead to higher rates of crop failures than under the current climate conditions. … A higher frequency of years with critical temperature during summer makes crop failures mostly dependent on precipitation values. Therefore, while still physically a compound event, the soybean failures under global warming become statistically similar to a univariate event based on precipitation"

Minor comments

Does EPIC-IIASA include changing management in the historical run or is it static?

EPIC treats management as static. We have updated the text to make this more explicit.

Line 106: "... Another advantage of using simulated crop models is that management and technology trends are static, whereas they are intrinsically embedded in the observed yield datasets."

Why do there only appear to be nine values in your partial dependence plots (Figure 4)? Or am I misinterpreting the ticks at the bottom of the plots on the x-axis?

The ticks represent the decile marks (9 ticks to divide the space in 10). All reviewers found it confusing, so we have decided to remove them.

Figure C7 - not required, but I'd suggest using different colors. It's nearly impossible to distinguish the differences in the PDFs at the top of the graph using orange/red and yellow

Thank you for the suggestion, it has been updated.

L 355 - I'd qualify that the change in temperature is much greater than the change in precipitation

Would it be possible that there was a typo in the line number? We cannot link the statement to the text at line 355.

L288-289 - It is not true that you see an increase in extreme dry years. The two blue PDFs are nearly identical, and the difference seems marginal. You certainly see an increase in the failure rate at the low-end of the rainfall distribution though (e.g. "Failure 3C" PDF larger than "Failure PD" PDF). And because heat is increasing you will still see increasingly frequent joint warm and dry conditions.

The reviewer makes a good point, the sentence was badly written. It has been updated now:

Line 313: "... we see an increase in the failure rate at the low-end of the rainfall distribution for 2C and 3C ..."

Figure 7: change the axis labels to be interpretable to the reader instead of being the variable names used in the code

Thank you for the suggestion. We have updated the axis labels.

L 282 -283 (and relevant to 390-392) - The interpretation that compound weather extremes become less important in a warming world because extreme temperature alone drives failures is not necessarily true. It may be (and probably is) rather that compound weather extremes become more likely in your shuffled data when you uniformly warm all years. For example, if every year is an exceptionally hot year then your crop failures depend almost exclusively on rainfall, making the joint failure and shuffled failure years similar to one another at high return periods. In fact, we can see that all of the 3C failures still occur at very low levels of precipitation in Figure 7, but it is true that now low levels of precipitation lead to crop failures at higher rates.

Thank you for the comment. This is true, there is a difference between the statistical 'compound factor' and the more physical 'compound event' that has led to confusion. We updated the term to 'relative compound contribution'. We also updated the text, explaining that the relative compound contribution is reducing, which means the correlation structure between the meteorological variables becomes less important, but that the compound conditions leading to crop failures are increasing due to higher temperatures during summer.

Line 305: "Thus, while the number of failure seasons increases with global warming for the original data, the increase in the number of failure seasons for the shuffled data is larger. A reduction in the relative compound contribution for the 2C and 3C scenarios suggests the correlation structure between meteorological variables becomes less relevant for our definition of crop failure."

L 255 - I don't think the data support a robust increase of failure probability at higher precipitation levels unless you are willing to put equal weight on the corresponding dips in higher maximum temperatures leading to lower rates of failure and higher diurnal temperature ranges leading to lower rates of failure. I'd at least caution the reader that this may be noise rather than signal in the data.

Thank you. This is correct, we have updated the text.

Line 279: "Precipitation shows a general inverse proportion to crop failure probability, suggesting low values of precipitation to increase failure probability, as indicated by Figure 4."

L243 - define DTR at first use in the text. While it's defined in the table it is not defined before first use in the text

Correct, it has been updated.

L159-161 - what does assigning weights mean in this case? Was the dependent variable weighted? Are you running RF on binary data instead of crop yields generally? If so, justify this decision.

The dependent variable has not been weighted. The weights are related to the penalty function associated with the random forest model. We run the RF on binary data, being divided into failure years (any year with yields below the historical standard deviation) and non-failure years (all the other years). This is a common approach, with previous works, such as Ben-Ari et al. (2018), Vogel et al. (2021) and Zhu et al. (2021), also following this approach.

Line 148: "Here we explore the probability of soybean failure, which is defined by means of a threshold, similarly to Ben-Ari et al. (2018); Vogel et al. (2021); Zhu et al. (2021). Every season with yield one standard deviation below the mean was considered a failure."

- Also, why does it matter if the failure observations are less frequent?

RF have a natural bias towards the majority class and suffer from performance issues when the dataset is imbalanced.

Line 173: " Because the crop yield dataset has less failure seasons than non-failure seasons, the dataset is imbalanced, affecting the model's capacity in identifying the minority class. To address this issue and improve the model performance, we assigned weights to the predictions of each class with values inversely proportional to their frequency. This increases the penalties for underrepresentation of the minority class, balancing the model."

L158 - need to better define for the reader what a "data split" and a "shuffle" is in your cross validation procedure

- Described on L190

Thank you for the comment. We added the following information to better explain the training and validation procedure:

Line 169: "We tuned the random forest's parameters following a resampling technique called cross-validation. It consisted of dividing the data in 10 different splits, where 9 splits are used to train the model and the remaining one is used for validation. The process is run 10 times so that every split is used once for testing. In addition this process was repeated 5 times with different random divisions, leading to 50 runs in total."

L100 - note that soybean yield data is available for this entire period going back to 1900 from USDA, but it is difficult to remove the management and technology changes trend. So there is still reason to use crop models and the work provides a useful complement to observation-based analyses.

Thank you for the comment. As this is related to the first minor comment, on the EPIC use of management and technology trends, we addressed in the same section:

Line 106: "Another advantage of using simulated crop models is that management and technology trends are static, whereas they are intrinsically embedded in the observed yields datasets"

L 37: "the majority of climatic shocks are compound events" - is this true?

Thank you for the comment. We tried to be more specific according to the referred literature as:

Line 41: "... the majority of climate-driven societal or natural shocks are the result of compound events (Zscheischler et al., 2017; Zampieri et al., 2017)"