Review of “Weather-induced crop failure events under climate change: a storyline approach”

This study explored climate impacts on soybean crop failures in the Midwestern United States under present and future warming scenarios using the random forest model and storyline approach with model simulated crop data and CRU climate data. The findings suggest that the failures are likely to increase with global warming, changes in both univariate and multivariate climate drivers of failure decrease the importance of compound factor, and impact-analogues are significantly increased under global warming compared to event analogues. The manuscript is well in line with the scope of the journal Earth System Dynamics. While the random forest model is not new, it is novel to use that to analyze climate impacts on crop failures. Also, it is very interesting to use storyline approach to study this issue. The findings are relevant. I think, in general, the paper is publishable after revisions. While the statistical analysis is done rigorously, the data may be improved, and method description and the interpretation of the results and findings also need to be improved. Here I provide specific comments as follows.

Major comments:

- It is reasonable to use model simulated crop data to obtain longer time series for model training. However, the study does not show the performance of the simulated data compared to observations. Since the study region has accessible observational yield data over long period, it should demonstrate the simulated crop data can well represent observations. Otherwise, it may require a bias correction to improve the quality of the simulated crop data.
- It can be simple to average all the data over the region, but will discard many useful information. It is arbitrary to state “local scale of impacts are not meaningful for national and global implications”. Using all the data points within the study region can increase the sample size by taking into account of the within region variability and increase the robustness of the model.
- It is not clear how the random forest model was trained and validated. While a description of the evaluation metrics is provided, it is not clear how the data was split as training data and validation data. The validation data needs to be excluded as
unseen data when is used for model training.

- All the results require thorough interpretations such as the mechanisms and biophysical meaning besides just reporting the data.
- Based on the previous literature, the increase in DTR may actually benefit yields, which is contradictory with the finding from this study. Please explain.

Minor comments:

- The approach formulated in this study consists of three parts. The storyline approach is part 3 of the approach (part c) and address part of the research objectives. However, based on the title, storyline approach seems the key message of this study. I think the title needs to be changed since the storyline approach seems not the major focus of this study.
- Little result is presented in the abstract. It can be more informative to shorten the background and methods descriptions and include more results.
- Line 55: it is not clear to say “crop failures are compound events”. Do you mean crop failures can be consequences of compound events? It requires a clear definition on compound events.
- Lines 74-75: “…, a natural next step is to …” It may be not so natural. I suggest the authors better explain their motivations and why such study is needed.
- The last paragraph of the introduction mostly described the method/experimental outline. The majority of this part should be better move to the method section as a general statement of the method prior to detailed descriptions. The introduction section should focus on explaining why the research is needed, is novel, and stating the scope, research questions/objectives.
- Line 119: “… a similar 5-year period …” Please clarify what are the similar 5-year period at 2C and 3C.
- Line 194: Please interpret what is compound factor besides providing the equation.
- Lines 196-197: the definition of return period is confusing. Based on Figures 5 and 6, return period is not a one to one relationship with inverse of the failure probability, how it can be calculated as the inverse of the failure probability? It needs to be clarified.
- It is not clear which month of data was used as candidate meteorological input prior to eliminating and combining monthly data in July and August.
- Variable names and acronyms are inconsistent in the text and tables and do not follow standard practice.
- Figure 3b: Is Pearson’s correlation performed on detrended or categorical yield?
- I would suggest only providing half of the correlation matrix. To limit redundancy
- Figure 4: I am assuming that the ticks on the bottom of the partial dependency plot represent the percentiles of the data. If they are being included, they need to be explained.
- Ln 390 decreasing compound factor. Is this just an artifact of choosing DTR.
- Ln 374 Figure 10d?