Earth System Dynamics Discussions

Earth Syst. Dynam. Discuss., referee comment RC2
https://doi.org/10.5194/esd-2021-24-RC2, 2021

Comment on esd-2021-24

Anonymous Referee #2

Referee comment on "Impacts of compound hot–dry extremes on US soybean yields" by Raed Hamed et al., Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2021-24-RC2, 2021

Thank you for the opportunity to review this paper.

The study by Hamed et al. investigates the effect of growing season hydro-climatic conditions, including hot-dry compound extremes, on US soybean yield variability. In a first step, the authors identify a set of most important climate and hydrological predictors that affect soybean yield variability across the US. In a second step, they fit statistical models to county-level yield time series to examine the strength and direction of the relationship between hydro-climatic predictors and yield outcomes. In particular, the study finds that the co-occurrence of hot and dry events leads to more negative yield outcomes than the effect of hot or dry conditions alone would predict. The authors finally investigate the effect of historical hydro-climatic trends on soy yields. The authors show that historically, there have been wetting and cooling trends across important production regions in the US. However, in the same regions, compound hot-dry extremes increased in frequency. These results highlight that the effect of compound events may be masked when looking at statistical relationships of individual variables alone, without considering interactions between hydro-climatic extremes.

The paper is clearly and well written. From my perspective the manuscript is largely suitable for publication as it stands. I only have a few suggestions for the authors to consider which will hopefully help improve this paper for publication.

General comments:

Overall, the statistical approach is robust, and limitations are clearly presented in the text.

However, I would ask the authors to consider the following suggestions:

1) Predictor selection

The authors apply a strict predictor selection process, which eliminates the occurrence of highly-correlated predictors – both at the same time as well as in subsequent months.

However, I wonder whether this approach eliminates predictors that do have an important effect on soy yields. The example presented in the text is: "we excluded soil moisture in September as August soil moisture was already selected". I understand the reasoning to avoid collinearity, but it appears a little arbitrary – likely soil moisture would be relevant in both August and September (and potentially across the whole season).

Would it be more suitable to consider three aggregations for each variable (monthly, seasonal and the whole growing season) and select only one temporal aggregation per predictor in the final model? In this configuration, a predictor of "growing season soil moisture" could have been selected by the algorithm, if it was found to have the highest correlation with yields. This would lead to more interpretable results in the context of understanding climate influences on soy yields.

Similarly, it was not entirely clear to me why the authors selected two predictors per season (spring, summer and autumn) instead of selecting – for example – three heat and three moisture related predictors based on their individual predictive skill, irrespective of in which season they occur.

I have the impression, with the current way of how the predictors are selected, important predictors may be missed and less important predictors are selected. It would be great if the authors could test this or add a few clarifications in the text.

2) Cross-validation

I think it is great that the authors present the overall R2 and cross-validated (out-of-sample) R2 for their statistical models (given many studies only present an overall R2). However, the cross validation does not include the predictor selection step using the individual BICs and subsequent stepwise regression. Hence, the out-of-sample R2 will likely be over-estimated for new observations. Ideally, the cross-validation would include a predictor selection step for each iteration to obtain a true “out-of-sample” R2.

I understand the need to obtain one shared set of predictors to keep the results interpretable and to assess the influence of these predictors across all counties. I am not too concerned of overfitting, because the authors applied a very strict predictor selection process – only five predictors were selected based on all data points for the US (i.e. not selecting predictors for each county which would likely lead to overfitting) and the selected predictors are plausible. However, it should be mentioned somewhere in the paper that the cross-validation does not include the predictor selection step and may therefore lead to a potential overestimation of the out-of-sample R2.

3) Climate and hydrological data

The authors used climate data from a global bias-corrected reanalysis dataset (WFDE5). Generally, reanalysis data can contain uncertainties and biases stemming from the earth system model used to generate the reanalysis dataset. At the same time, observation-based datasets also contain uncertainties due to the interpolation applied to the data. I would ask the authors to add a comment on why reanalysis data were used here and how the WFDE5 global reanalysis compares to observational datasets available for the US (or globally). Do the authors see any biases in the reanalysis data that could influence the results of their study?

The hydrological indicators (actual evapotranspiration and root-zone soil moisture) are described as satellite-based, obtained from the GLEAM dataset. However, the description of the GLEAM dataset indicates that GLEAM uses a hydrological model to simulate soil moisture and actual evapotranspiration (instead of, for example, directly using satellite-based observations of soil moisture). I think a clarification in the text that soil moisture and actual evapotranspiration are not observed directly, but simulated, would help understand the data – as simulations and remote-sensed data have different uncertainties and potential sources of errors.

Specific comments:

- Line 20-21: "Moreover, in the longer term, climate models project substantially warmer summers for the continental US which likely creates risks for soybean production."
  Given the effect of future trends using climate model outputs was not within the scope of this study, I would suggest removing this sentence, as it is a bit vague. It might be best if the abstract includes a sentence on potential future research, e.g. future studies are needed to understand the frequency of hot-dry compound extremes under climate change (similar to what was mentioned in the discussion).
- Line 74-75: "(ii) have 75 common planting dates (i.e. April-May)"
  Why was there a need for common planting dates? Would it not be better to include as many yield observations as possible, and instead subset the growing season into first, second and last third? Can you please explain this in the text?
- Line 81-82: You applied a linear trend to the yield time series. Previous studies have applied cubic trends or more complex trend fitting algorithms to account for non-linear trends. Can you please confirm (possibly with a plot in the appendix) that visual examination showed that county-level yields follow a relatively linear trend?
- Line 145-146: "To overcome this limitation, we used precipitation and temperature minimum and maximum variables from the CRU V4 global dataset (Harris et al., 2020) covering the period 1901-2019 at a spatial resolution of 0.5°."
  Why was this dataset not used to fit the statistical models, instead of the reanalysis dataset? This way you could be certain that the same data used for fitting the model is used to assess trends. Could you show the correlation between monthly Tmin, Tmax and precipitation in this observational dataset compared to the WFDE5 reanalysis (in addition to the correlations you show in Figure A.1)?
- Line 146-147: "Minimum temperature in the early season was used as a proxy for early season actual evapotranspiration…"
  Why did you choose minimum temperature instead of daily mean temperature (or the average of Tmin and Tmax)? Would this not capture the relationship with evapotranspiration more accurately, as it includes information on maximum daily temperatures as well?

Technical Comments:

- Line 168: I think it should by "county-level" instead of "country-level".