

## ***Interactive comment on “Identifying meteorological drivers of extreme impacts: an application to simulated crop yields” by Johannes Vogel et al.***

### **Anonymous Referee #1**

Received and published: 21 August 2020

Vogel et al. showcase how a regression models for regional crop failures can be obtained from a multitude of potential climatological drivers while minimizing the number of relevant variables via LASSO, a statistical method. It's a nicely and thoroughly done analysis that deserves publication after some minor revisions.

####Minor Comments: 1.I. 3 understanding and forecasting of..?

2.I. 8. 'predict' -> 'determine'?

3.I. 31 'depend'

4.I. 85 '..the climatology was defined to be the mean plus the first three annual har-

C1

monics.' Can the authors further explain what is meant by that?

5.Figure 5 what is the growing season (GS) here could this possibly highlighted in the section to the left, which shows the months?

6.I. 167 and later on: Nice to see R-packages explicitly cited.

7.I. 190 is s segregation threshold and the local cut-off value? Maybe it would be better to use one term only?

8.Figure 7 sub-panels are not enumerated.

####Major Comments: 9.Could the authors provide an estimate on how many data points would be necessary at minimum to apply the LASSO method? Is there a relationship between total number of suggested variables and necessary datapoints?

10.Do I understand correctly that LASSO, avoids autocorrelation by checking for the variability of variables only? Could the authors provide evidence for the reliability of such approach?

11.As prediction is mentioned as one application in the abstract, how would the CSI change if growing season variables were not included? In what region is it possible to build a useful prediction model using Months outside the growing season only, in what regions are yields not predictable without including conditions from within the Growing season itself

12.Overall but specifically in Fig.8 It might be more insightful to show results relative to the grid-point dependent growing season? Could this explain the differences in the shape of the histogram between the North America / Europe and Asia?

13.On data sampling: The 5th percentile is a rather low threshold, to increase the number of events (and come closer to a real-world applicability) is the method sensitive to the exact choice? What production loss does a 5th percentile correspond to regionally/on average?

C2

