Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2017-130-AC3, 2017

Biogeosciences
Discussions
Open Access

Interactive comment on "Detecting impacts of extreme events with ecological in-situ monitoring networks" *by* Miguel D. Mahecha et al.

Miguel D. Mahecha et al.

mmahecha@bgc-jena.mpg.de

Received and published: 3 August 2017

Bold text are the comments from the reviewers. Standard text is our response. *Italics* are text elements from the revised paper.

This manuscript addresses how the size and distribution of an observational network impacts the ability to detect the occurrence of extreme events. This is a highly relevant topic, which has not been given sufficient analysis in light of the expected changes in the underlying distributions of these events due to climate change. Thus the manuscript is timely and important.

Dear Dr. Brunsell

Many thanks for the overall appreciation.

Printer-friendly version

Discussion paper

I do have a few concerns on the methodology, both from questions about the actual analysis and a few suggestions that might help increase the utility of the analysis for a broader readership. I suspect that many of my concerns about the actual analysis can be addressed by clarifications within the text.

We tried to include the suggested clarifications wherever possible – see our specific responses below.

For example, I wonder how the impact of the event is actually determined. How does the event intensity, geographic extent and duration relate to the rank? I can imagine more intense events of small geographic extent having a more significant impact than larger extent, longer duration, less intense events. Also, given the carbon and water emphasis of the introduction, the physiological impacts could be significantly different. This should be commented on at some point.

Yes, we agree that this point is not understandable from the current version of the paper alone and requires reference to papers published earlier by some of the co-authors of this paper. When we introduce Fig. 5 (which in the revision is now Fig 4), we expand the explanation to clarify this point as follows:

...In contrast, investigating the event durations (Fig. 4d) did not reveal such a clear pattern, which could be explained by the fact that we are dealing with a relatively short time series, in which only a few discrete duration classes can be recognized. The fact that global impacts of extreme events in the terrestrial biosphere behave similarly to those at smaller spatial extents is expected because these properties are known to be strongly correlated as shown e.g. in Reichstein et al. (2013). This study also reported that the duration of extreme events is less strongly correlated with their impact, as we would also suspect from Fig. 4.

Specifically in Figure 5, does the top line in each subplot represent the same case?

Yes, the lines represent cases of a given network size, e.g. the darkest blue is the largest random network placed in Europe.

So does the middle point in duration, extent and impact lead to the middle of the rank?

We don't exactly understand what the reviewer means with "middle point", but we suspect that the concern is that some rank x derived from an FAPAR impact would not correspond to rank x derived from the spatial extent of the extreme, or duration. This observation is correct, but – as explained above – the surprising fact is that at least extent and impact are highly correlated, such that the ranks would only differ is a few cases. The case for duration is very different tough.

How would a long, large geographic extent with less 'impact' appear on the rank?

The impact is defined here as the sum of voxel areas times the FAPAR anomalies. It could happen indeed, that a stronger FAPAR anomaly over a smaller area leads to the same total impact compared to a less intense anomaly that is spread over a larger area. Again, if this would be serious concern we would not see the strong correlations between impact and extents. We work here with the extents as these are very good approximations for the impacts . . . (see also next response)

Assuming that the rank and impact aren't a simple linear relationship, this makes me wonder about the scaling analysis with respect to rank. While the analysis itself is interesting, I expect that the detection probability with respect to impact might be more useful to the broader community. If the rank and impact are related, you should state that relationship (same for geographic extent and duration).

Rank and impact are directly related because rank is defined by sorting the events by impact.

It would also be useful to place some of the discussion within the general dis-

cussion on how extreme events are classified (i.e. exceedence probabilities, etc.). There is some discussion of the 2003 Europe and the 2012 US drought/heat wave events, it would be interesting to see where these actual events appear in the rank/impact plots of the analysis for perspective given the media/scientific coverage of these events.

We have indeed restructured the discussion to consider this aspect in more detail. We now start with a consideration of our event detection method compared to other approaches and referred to the discussion on the known historical events.

More related to the actual analysis, I wonder about the 'clustering' in the PCA space. Why choose an arbitrary mesh and not apply an actual clustering algorithm? What's the benefit of using the PCA approach rather than a normalized/standardized approach on the time series? The benefit stated is with respect to the smaller magnitudes in some bioregions (e.g. semi-arid), but this could be addressed through a proper standardization and the general probability thresholds in the pixel. You could then use a spatial/temporal clustering algorithm that could address the local/regional issue you discuss. To be clear, I'm not suggesting you do that analysis, but some discussion on why this is necessarily better would be useful for context given the simplicity of that more traditional approach.

The mesh is so fine (grid size 4% of the length of PC1) that it actually does not affect the resulting regional anomalies. We have tried this in many ways. But we see that there are some concerns regarding the usage of a linear clustering/binning of the PCA space. So yes, we could indeed cluster the MSCs and considered this at the very beginning but we found two complications that we were able to avoid now: 1) Clustering is leading to a non-uniform partitioning of the space spanned by the MSCs and computationally more expensive. 2) it does not allow us to easily assess which are the neighbouring clusters in order to consider these values for the assignment of the thresholds. Our approach has the advantage that we can compute it on a subset

Printer-friendly version

Discussion paper

of values (rendering it very efficient), and leading to a uniform grid that allows us to efficiently include the neigoughing “clusters” for the estimation of the threshold of the central cell and is controllable in terms of the variance represented and computationally extremely efficient even on this very large data set. MODIS PFTs or Koeppen-Geiger classes would be possible to use, but are very very coarse and don’t reflect the details we see in Fig 1. Hence, these classifications would lead to a very coarse thresholding that is by no means comparable to the continuous threshold as shown in Fig 3 (Fig 2 in the revised version of the paper). Note that this answer was almost identically given to a similar question by reviewer 2.

What is the role of the underlying resolution (spatial and temporal) of the data on the detection/clustering algorithm. Given the emphasis of the manuscript on the development of the method, these would seem to be important considerations.

The coarser the resolution, the easier it becomes. Our personal challenge in this paper (compared to earlier studies like the ones by Zscheischler et al. 2013, 2014 or Reichstein et al. 2014) was indeed to deal with a high resolution EO data set. This made it necessary to deal with e.g. a relatively complex search for identifying contiguous extremes. But all these efforts didn’t really change the overall picture which has certainly to do with the scaling relationships in the event size distributions that remain unaffected. We now added the following sentence to section . . . to clarify the role of the resolution.

In summary, I find the approach interesting and potentially informative, but I think the manuscript needs additional details on the methodology and some more real world applications to illustrate these benefits and help make the authors case on the significance.

We have substantially revised the methodological descriptions and also reorganized the discussion to clarify the relevance of the paper.

There are a number of minor comments/textual issues that should be addressed

as well (page/line number):

We have addressed all minor comments. Where questions appeared we provide responses below.

What if different extreme events are occurring at the same time in different regions, are they classified as the same event with a larger extent?

No, if an event happens to occur as far as z spatial elements away it is considered a different event.

9/24: is event size in terms of impact? How is this determined? "Size" is the area affected.

We clarified this here again as follows:

To better understand expected extreme event detection rates, we initially explore random networks and their hypothetical capability to detect extreme FAPAR reductions. We focus on Europe and vary the network sizes from $n = 5, \ldots, 10000$ *sites on a logarithmic scale, asking how many of the detected extremes can be identified for each size class. More precisely, we investigate the probability that an extreme event of a given size* m *(measured in terms of affected area) will be detected by* n *hypothetical towers* $P(m, n)$*. All following analyses are based on repeating the tower placement 100 times per size class. We mimic real site placement by assuming that a tower is not mobile, i.e. it remains active at a given location over the entire period covered by the FAPAR observations.*

10/4: 'largest' in what sense? I could imagine some networks sizes being better at getting largest impact, but perhaps different for largest extent or duration.

"Largest" refers here to the integrated FAPAR anomaly and we refer to the ranks of these events. This is explained now in the paper.

Figure 5: how does the FAPAR impact relate to the magnitude of the FAPAR data?

Printer-friendly version

Discussion paper

Compared to some transformed (zero mean/unit variance) version of the data, it seems unlikely that 4 orders of magnitude of FAPAR impact (Figure 5b) are all 'extreme'

Here, we would like to respectfully disagree with the reviewer comment. The voxels are all flagged as extreme or not according to the same regional threshold criterion. The spatiotemporal extends of the contiguous extremes then emerges from the analysis of contiguous extremes and the impacts are resulting from integrating across the contiguous spatiotemporal extremes. As shown in earlier paper (e.g. by Reichstein et al. (2013), Nature; Zscheischler et al. (2013)) these resulting extremes and impacts of extremes are typically power-law distributed. This implies by definition that the size distribution spans across various orders of magnitude.

10/14-16: This would be helpful above prior to the discussion of the figure.

We don't exactly understand this point, as the figure is introduced exactly here.

Figures 5 and 6 should have consistent units (e.g. km2). Figure 6 caption should include description of solid and dashed lines.

The reviewer is absolutely right! The figures have been created by different co-authors. Now, we have recomputed them with the same units. In this context we also found a unit-conversion issue in Fig. 5 (Fig. 4 in the revised paper) that has been corrected.

We have carefully worked on all other editing comments. We acknowledge the reviewer for his efforts!

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-130, 2017.