Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2017-130-AC1, 2017

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

The manuscript describes a very well conducted study of the potential to detect extreme events by in situ observation networks. I agree with the analytical approach and the analysis results. There are however many language and style errors and the manuscript requires a serious editing effort. I provide meticulous editing comments below. These refer to the test until about page 13. There are problems past it, I just didn't have more time.

Dear Dr. Bohrer:

Thank you very much for your detailed comments on our manuscript. Indeed, none of the coauthors is a native speaker and we apologize for the inconveniences. We have now asked a native speaker to revise the manuscript for language issues and addressed all other comments. Please find below a point-by-point response to your remarks:

P3L31 All MODIS datasets have a code (typically starting with three letters, often MOD or MYD with lots of numbers that encapsulate the product type, resolution, return period Pinty references MCD43B3.005, but I think that is what he used to make the next level product). Please list the code of the exact dataset you used. If it is not from one of the MODIS DAACs and doesn't have a MODIS code, please provide the reference to where you downloaded the data from.

We have now revised the paper citing the original MODIS code. The text now reads as follows: *Here we use FAPAR data derived by the JRC-TIP approach (TIP-FAPAR, Pinty et al. 2011), available at1 km spatial resolution. These estimates are based on the MODIS broadband visible and near-infrared surface albedo products from NASA Collection 5 at 1 km spatial resolution (MCD43B.005, Schaaf et al. 2002, available on demand from co-author T. Kaminski). These satellite data cover the entire surface every 16 days and the data range from 2000 to 2014; in this study we use data covering Europe and the continental US (excluding Alaska).*

Eq. 1 is trivial and can be removed. The explanation of space-time voxel you provide in P4L3-4 is enough (and can easily be revised so that it doesn't need the equation).

We agree, this is not worth an equation. We now simply introduce the notation in the text without labeled equation.

Section 2.2 – please emphasize that you only used the location of each existing flux tower and did not use any of the data collected by the tower. Explaining what types of fluxes and other measurements are done is each site (P4L11-14) is

rather confusing and to some degree, misleading

Yes, we agree and modified the text accordingly.

P4L17 Please add the link to Ameriflux, (https://ameriflux.lbl.gov/) and ICOS (https://www.icos-cp.eu/) in the same way you added the links to all other networks you used (fluxnet, euroflux, neon...)

Yes, sorry for the inconsistency - this is now done.

P4L24 Are you sure that the NEON sites "that can be moved for dedicated studies" include EC towers? Did you use any? I think not (for either questions), and if I am correct, you should remove their mention from here as these were not used and create a false impression that they were.

Indeed we did not use the translocatable sites and removed any reference to them in the text.

P6L7 Appendix A (or any other appendix) does not include any details of the event detection method. It only contains supplementary figures. Please provide the detailed method description, as it is potentially some of the most exciting and applicable parts of this manuscript.

Indeed, reference to "Appendix A" at this point referred to an earlier version of the paper. In response to the reviewers request we not only revised the description of the event detection methods, but also included a more in depth methodological description as Appendix A, which reads as follows:

In the following we develop a strategy for defining thresholds of regional relevance that are computationally suitable for dealing with high-resolution remote sensing data like the 1 km FAPAR data considered here. Our aim is to find regions of comparable phenology. Our assumption is that the expected seasonal cycle in FAPAR is a good representation of overall phenology, and hence ecosystem type.

The first step considers the data set of mean seasonal FAPAR patterns $\mathbf{F} = \{f_{n,s} : \forall n \in 1, \ldots, N; s \in 1, \ldots S\}$, where each point n is pointing to a geographical location u, v and contains the local mean of seasonal observations s.

In the second step, we use principal component analysis (PCA) to reduce this S-dimensional data set. In other words, we seek orthogonal components that represent the main gradient along the covariances of the seasonal cycles. More formally, the covariances of these centered mean seasonal cycles are given as

$$\mathbf{C} = \mathbf{F}^t \mathbf{F} \,. \tag{1}$$

Common patterns of seasonality are identified by first estimating the k leading eigenvectors,

$$\mathbf{C} E_k = \lambda_k E_k \tag{2}$$

where E_k the kth eigenvector of length S, and λ_k the corresponding eigenvalue. The scores of the kth principal component are given by

$$A_k = \mathbf{F} E_k \,. \tag{3}$$

and k leading A_k can be interpreted as a proxy for the characteristic patterns underlying the mean seasonal cycles across space. Figure REFERENCE TO FIGURE ! visualizes the three leading principal components as an RGB-color composite, revealing a distinct map of European phenological regions.

Third, the question is how to identify regions of similar phenology in this continuous space spanned by the principal components. One could use, for instance, some clustering algorithm. However, given the high density of spatial points and the continuous sampling, an equivalent approach is to choose an equidistant grid in the space of the *principal components. We choose a very dense grid, such that each cell is as wide as 4% of the range of the first PC. We then define an FAPAR anomaly threshold as a predefined quantile based on the distribution of FAPAR values separately for each grid cell and its 26 neighbours in the space of the leading 3 PCs. This threshold is assigned to all points in the respective grid cell represented herein. This threshold is assigned to the all points represented therein. Figure REF TO FIGURE illustrates this approach in detail.*

PLACE FIGURE HERE: THIS WAS FIGURE 2 IN THE ORIGINAL SUBMISSION.

We have now proposed a FAPAR threshold for each point and can map this threshold back to the geographical space by remapping each point to the known geographical coordinates u, v. This is shown in Fig. REF TO FIGURE 3 WHICH WILL BE FIG 2 IN THE RESUBMISSION.

P10L1 "shown before for the US in Fig. 8." ??? Fig. 8 did not appear yet and nothing was shown for the US

This reference to Fig. 8 was inherited from a very old structure of the paper where Fig 8 came first. We have removed this reference here now as it is indeed not helpful.

Section 4.1 – please make a conclusive statement (I ended the section confused and wondering)– after all the tests you conducted and results you show in appendix A, is the discrepancy presented in figure 6 explained by the spatial/temporal autocorrelation of the extreme event or the discrepancy must have another explanation. Does it indicate a weakness of random in-situ networks?

Thanks for highlighting these previous inconsistencies in the manuscript. In a revised manuscript, a more consistent and stringent explanation is provided: We restructured Section 4.1: The previous subsection 4.1.2 on Spatiotemporal Correlations is now in the Appendix (where also the artificial simulations are shows); and the explanation why for the real-data case does underestimate detection probabilities for large extremes, but overestimate detection rates for small extremes is given in Subsection 4.1.1. We have now also tested our explanation in much more detail (see revised manuscript and Appendix B). In fact, the reason for the underestimation of the detection probabilities for large extremes lies in the fact that spatio-temporal correlations lead to a clustering effect: large extremes are more likely found with increasing distance from the boundaries of the domain (or, in other words, the coasts). This effect leads to a higher occurrence probability of a large extreme in the middle of the domain; and due to the fact that the random networks are indeed placed randomly, these networks tend to have a lower chance to detect large extremes because the sampling probability is not adjusted. If these edge effects are small or negligible, the theoretical predictions work very well (as shown in Appendix B). The overestimation of detection rates of small extremes is due to the search radius and explained in Subsection 4.1.

Table 1 – lines are strangely discontinuous. Fix it to look like a proper table. In any case, I am not sure that I truly understand the details of what you are trying to convey with this table. Can there be a better way to explain it? Is it necessary and are you using all the categories listed in the table?

The issue of discontinuous lines comes from the LaTeX typesetting. But you are right, finally we don't need all these categories in this very paper. Hence, we have decided to remove it and extend the text where it was introduced.

P12L3 what is betta? Is it coming from some equation that you did not provide?

We have now explained this in the text (moved to the appendix B) in more detail as follows:

The idea is that the Fourier coefficients of artificial data (spatial white noise) are forced to decay as a power law function across frequencies i.e. proportionally to $f^{-\beta}$. An inverse transformation to space yields a correlated data field. If we choose $\beta = 0$, it corresponds to uncorrelated, $\beta = -\frac{3}{5}$ to moderately correlated, and $\beta = -\frac{8}{5}$ to highly correlated data. Hence, β is the decay exponent of the Fourier coefficients. We also

Printer-friendly version

Discussion paper

provide reference with respect to these data generation schemes.

Editing comments: . . .

The review provided many very helpful suggestions on the text and we have carefully worked through all of them. We acknowledge the reviewer for his efforts!

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