Anonymous Referee #2

Received and published: 9 September 2019

This is a well-written study comparing the European heatwaves of 2003 and 2018. Comparison of climatological data and vegetation indices lead to the conclusion that the 2018 heatwave was more severe than the 2003 heatwave. However, substantial regional differences occur. The idea behind this study is interesting, but the study misses out on several aspects needed to support the conclusions. Especially the lack of the temporal patterns in weather data makes it hard to evaluate the results. No time series for temperature, precipitation or drought indices are shown to illustrate the heatwave patterns of both years. Moreover, end of July was chosen as the study period for the impact on vegetation, hence ignoring any change that took place in August (e.g. the massive forest fires in Portugal early August 2018).

Thank you for your constructive review. We agree with most of your comments and will modify our manuscript accordingly. Please find our detailed reply in the attached pdf-file.

Specific comments:

L 96ff: I suggest to simply argue that you focused on March-Oct because that is the period of interest for vegetation dynamics and leaving out winter helps avoiding artefacts (e.g. snow cover, but also defoliation in deciduous systems).

Thank you for this suggestion. We will modify the corresponding text accordingly.

L 104-105: interpolation may create artefacts when searching for anomalies – especially when there are gaps during the drought episode under study. I suppose this is of minor importance for this study, because gaps are less likely during periods of drought (i.e., no clouds), but I wonder if the interpolation can be avoided. If not, the possibility for such artefacts should at least be discussed.

As you pointed out correctly, the likelihood of gaps is very low in 2003 and 2018 since large parts of Europe were free of clouds during these events. However, in 2018 the Mediterranean experienced above average precipitation, which may have resulted in higher cloud coverage and thus missing values. To quantify the amount of missing values, we will compute maps for 2003 and 2018 which show the number of gaps that were filled in each of the two seasons and moreover add a paragraph to the discussion related to the influence of associated artefacts.

L 119ff: NDVI and EVI are mainly greenness indicators. They may reflect photosynthesis, but not if photosynthesis changes without changing greenness. This is particularly relevant for drought. In this sense, EVI is better than NDVI (see Vicca et al 2016, Scientific Reports). I therefore advise to use the EVI results rather than the NDVI in this manuscript. It should also be clearly indicated what these VIs can reflect (and what not!). This is completely missing from the discussion of the current manuscript, but needs to be discussed (i.e., are we looking at green biomass/browning/defoliation. . . and what are the implications for e.g. legacy effects).

In the revised version we will elaborate the corresponding methods section as proposed. We agree, that EVI has certain advantages over NDVI regarding its ability to detect specific changes in GPP as in Vicca et al., 2016. However, the study by Vicca et al., 2016 focused on GPP-reduction where no defoliation or leaf-coloration was observed. In contrast, both in 2003 and 2018 early leaf shedding and coloration was observed (see also below regarding additional phenological analyses undertaken). The question which VI to use for drought indices is widely *debated and depends on the purpose. For instance, Li et al. (2010 in Procedia of Environmental Sciences) found stronger correlations between NDVI and field observations in comparison to EVI. Moreover, NDVI is usually more chlorophyll sensitive and mainly reflects the photosynthetic activity (which we were mainly interested in) while EVI is more responsive to canopy structural variations. Also, in light of the findings by Vicca et al., 2016, it is hard to interpret the fact that the area with lowest quantiles was much larger for NDVI compared to EVI (Fig. 4 compared to Fig. S 4). That is, if a non-visible drought-response would have been missed by NDVI, the areas indicating a severe drought response should be larger for EVI, which is not the case. Finally, all of the cited studies which monitored drought-impacts by means of remote sensing made use of NDVI (Anyamba and Tucker, 2012; Bastos et al., 2017; Xu et al., 2011), wherefore concentrating on NDVI makes our work more comparable to previous studies. Since we from the beginning were aware of that the choice of VI matters, we included EVI in the supplementary to provide the full picture. To account for potential issues with regard to the chosen metric for quantifying drought we will add a paragraph to the discussion about NDVI-EVI comparison and mention the possibility to also assess other remotely sensed indices such as the photochemical reflectance index (Vicca et al., 2016) and solar induced fluorescence in future studies.*

L 141-142: awkward phrasing: Subsequently, we for 2003 and 2018 determined . . . should be: Subsequently, we determined the difference between 2003 and 2018 for the respective metric. . .

Thank you for this suggestion. We will modify the corresponding text accordingly.

L149ff: the timing of the heatwaves should be demonstrated with data to justify the choice to focus on end of July. Time series of temperature and CWB for e.g. France, Germany (which suffered from the heat in both 2003 and 2018), or even for the different regions (N, W, S, Central Europe). How sensitive is your analysis to the time choice? Are results similar if the analyses were repeated for end of August for example?

We agree, that the timing of drought as well as the location of the drought epi-centers differs between 2003 and 2018. To better visualize the temporal development in 2018 and 2003 we will add supplementary material showing the temporal development of CWB and NDVI in the two years. Fig. 6 was particularly designed for this purpose, but we understand that we have to elaborate this part of the manuscript further. We will also perform the analyses depicted in Fig. 5 and 7 on the basis of a different selection of time slices for 2003 (DOY 241) and 2018 (DOY 209) to account for the differing timing of the droughts and discuss the results in light of the temporal development of the two events.

L 150-151: VIs cannot be lower than 0. (anomalies can)

Here, we disagree. Given the formulation of VIs in the numerator (NIR-RED)/(NIR+RED), VIs range from -1 to +1 and will become smaller than zero if the reflectance in the red spectrum is larger than in the NIR. Also, we want to point out that we were not using anomalies of VIs, since VIs have a bounded distribution (See also below). That is, instead we computed quantiles. All of this is described in detail in lines 150-156 of the initial submission.

L 159: What was done with pixels where land cover changed between 2003 and 2018? Was that even considered? (I don't think it will have a big impact on the analyses, but it’s worth a mention).

Thanks for this suggestion. We did not consider the changes in land-cover between 2003 and 2018 but will add a map to the supplementary to indicate which pixels of the considered CLC-classes changed their class from 2000 to 2018 and exclude all corresponding pixels from the analyses to avoid artefacts.

L191: 0.55 should be 55% I suppose.

Thanks for spotting the error. Will be changed.

Fig.1: why was the timing April-July chosen for these figure? This is not motivated in the text I think a time series with weather data would be very helpful to evaluate this choice (see earlier comment). I noticed that this is briefly mentioned in the discussion (l. 320), but data are not shown. Please do show these data.

We agree with the referee and we will improve our manuscript regarding this aspect. The end of the period was set to July since we were assessing end of July VI-quantiles. The beginning was set to April, since the integration over this period revealed the strongest relationship with end of July VI-quantiles. That is, adding or removing months before or after April lowered the relationship between CWB and VI-quantiles. As mentioned earlier, we will add several time-slices to our analyses as supplementary material and elaborate the analyses using CWB of different time-steps for modelling VI quantiles (e.g. if assessing NDVI at DOY 241 for 2003 consider the period May-Aug instead of Apr-July).

Fig.4: It is unclear where VI-deviations from the mean were significant. Please clarify, also in the text.

This seems to be a misunderstanding of the shown values. Since we used quantiles and not anomalies, it is not possible to derive p-values for the presented values. The reason for using quantiles instead of anomalies (bounded distribution of VIs) is given in lines 150-156.

l.240ff: A map with vegetation types is missing to illustrate where the different vegetation types occur and how the differences in impact for the different vegetation types correspond with the regional differences (e.g. Scandinavia being dominated by conifer forests).

In the map depicting which pixels changed their CLC between 2000 and 2018 (see above) we will also depict the representation of CLC-classes.

Fig.6: consider moving to appendix, adding instead a figure with time series for weather data.

We think that it is important to show the temporal development of NDVI-quantiles and will complement Figure 6, according to your suggestion, with time series for weather data.

L 251: Fig. 7 shows EVI, not NDVI. The text is about NDVI. (I suggest to focus on EVI for in the main document and move NDVI to appendix – see earlier comment).

Thank you for spotting an error in the axis labelling. Actually, Fig. 7 shows NDVI and not EVI (the same results for EVI are shown in the supplementary, Fig S7). Regarding your suggestion to focus on EVI, please see our detailed reply above.

L328ff: Portugal suffered from severe wildfires in 2018. This is not included in the analyses because the fires occurred mostly in August and the analyses are only for April-July. Other important events may be missed out because of the choice for April-July.

*As mentioned above, we will add supplementary information which will allow to see the impact of forest fires in NDVI and EVI. This will allow to spot other possibly important events. However, we believe that, given their comparably low relative spatial extent (please don't get us wrong **-** we are aware that these fires were massive but in relation to all European forests rather contribute a lower proportion of forest area,) it seems likely that the overall impact of these regional wildfires on our analyses is low.*

L 340ff: I suggest to include in this part of the discussion some text on the relationship between ecosystem types and climatic regions and how this may/may not influence the interpretation (see also earlier suggestion for figure addition).

In complementation to Fig. 5, we will add a figure to the supplementary depicting the spatial contributions of the landcover types to the drought-affected areas. Moreover, we will elaborate the discussion in section 4.2.2 dealing with the drought response of ecosystem types across different climate zones.

L358: public news references are not appropriate for this statement.

We agree, but to date, we have not found any scientific publication that reports an earlier leaf shedding for broadleaved trees in 2018. To provide a more robust background to this statement, we will add results from a new analysis of DWD-phenological data to the supplementary which clearly highlights the earlier leaf-shedding of beech in summer 2018 in Germany.