Increased vulnerability of European ecosystems to two compound dry and hot summers in 2018 and 2019

Bastos, A., Orth, R., Reichstein, M., Ciais, P., Viovy, N., Zaehle, S., Anthoni, P., Arneth, A., Gentine, P., Joetzjer, E., Lienert, S., Loughran, T., McGuire, P. C., O, S., Pongratz, J., and Sitch, S. *Earth Syst. Dynam. Discuss*

Response to Reviewer #1

R1C1: The manuscript is clearly written and easy to understand its message, and the results will be of great interest to a wide community of climate and ecosystem researchers. However, I suggest a few points be addressed before it is considered for publication, as follows.
We thank the reviewer for the overall positive evaluation of our study and for the constructive comments. Below, we provide a point-by-point reply to the reviewer comments.

R1C2: My biggest concern is that the "increased vulnerability" can be exaggerated. First, the authors only focused on the area with negative EVIanom in DH2018 (area_clusters), while the rest area (white color pixels in Fig. 3) was not considered, including core regions such as scattered parts in Germany and Poland, and large assembled parts in southern Sweden and the southeast of the study region. Based on the climate-vegetation relationship estimated from this area_clusters, the 2018 and/or 2019 EVIanom of some ecosystem clusters appears to be deviated from the long-term relationships. However, including the rest area can both affect the departures of 2018/2019 EVIanom as well as the long-term relationship. First, the slopes (EVIanom-SManom slope and EVIanom-Tanom slope) can be larger as modulated by vegetation susceptibility in the rest area during 2001-2017, for instance, the southeastern part (southern Sweden) during 2003 (2015) heat wave. Second, including 2018 EVIanom in the rest area can surely offset some of the 2018 departures since it is a positive EVI anomaly. Therefore, in my opinion the approach to just look at DH2018 EVIanom-negative region in this manuscript does not allow to conclude the increased vulnerability of European ecosystems to the compound events (also given the fact that it only covers 20% of the area_clusters).
We agree with the reviewer that the expression "increased vulnerability" does not fully reflect the results of our study, which shows differentiated responses to DH2018 and DH2019. We propose dropping the expression "increased" from the title. We have added one sentence on regional compensation effects in the discussion as follows:

> *It should be noted, though, that we focused on pixels which were negatively impacted by DH18, but some pixels in the regional domain selected showed greening, even in DH18 (Fig. 1) These regional asymmetries result in partial regional compensation of the DH18 impacts, as shown in Bastos et al. (2020b).*

Nevertheless, we would like to justify our choice to focus on negative anomalies from DH2018.
Different perspectives can be used to analyse the impacts of a given extreme event, from a hydrometeorological perspective (Flach et al., 2018; Bastos et al., 2020a, b) to an impact-centered perspective (Smith, 2011; Bastos et al., 2014; Zscheischler et al., 2014; Reichstein et al., 2013). We agree with the reviewer that the choice of perspective matters, and we argue that this choice should be based on the specific question(s) one intends to answer.
First, we are interested in evaluating the impacts of the two consecutive extreme summers on vegetation activity, particularly how the impacts of the first event might precondition the response to the second event (Ruehr et al., 2019; Anderegg et al., 2020). Therefore, we follow an impact-centered compound event approach (Zscheischler et al., 2020), where we separate the impact of the concurrent hazard (high temperature and drought) from preconditioning or modulating factors. The preconditioning/modulating factors are expressed by departures of observed impacts from those predicted based on the hazard intensity only. This is an analogue approach to that used in (Bastos et al., 2020a), but their study was only model-

based. Here, using a data-driven approach, we find consistent results with their modelling study, particularly the higher vulnerability of cropland dominated regions to DH2018 (as well as DH2019).
Of the four clusters, only cluster 4 does not show departures from the long-term relationship, which can be associated with higher resistance and resilience of forests in response to DH2018 and DH2019. The other three clusters all show signs of increased vulnerability (L259-L262, Fig. 3), C1 to DH2019 only, C2 to both events, and C3 to DH2018 only. These three clusters encompass the majority of the studied pixels (56% of the area with negative EVI anomalies in DH2018), not 20%.
Second, indeed some regional compensation is expected, which would necessarily affect the relationships found *if* we would include the pixels with positive EVIanom in DH2018 in our regression analysis, which we do not. Moreover, regional asymmetries and compensation effects during DH2018 have been thoroughly analysed in (Bastos et al., 2020a, b).
Finally, we do not limit our analysis to the regional aggregated values, but also perform the same analysis at pixel level, so that regional compensation effects do not affect our results based on Figures 4 and 5. We noted however, that in Fig. 4 and 5, the pixels with positive anomalies in DH2018 were not masked out. This has now been corrected for consistency.

R1C3: Soil moisture anomaly is a useful indicator for the climate impacts of DH events. I feel the correlation is not enough to assess the model skills in simulating soil moisture variability (Table 1). First, the record can be too short for a robust correlation (Apr-Sep, sample size=6?), and this can be an issue not only for soil moisture validation, but also for GPP (e.g. L310). Second, even if the correlation is high, the magnitude of the changes in the absolute soil moisture values can be underestimated in the model due to too shallow soil. Therefore, calculating RMSE or normalized-RMSE using the absolute soil moisture values can be useful.
Here we were particularly interested in how the models simulate the dynamics during the DH events, not the model skill to simulate soil-moisture variability in general. We agree with the reviewer though, that this information is also important, to include the RMSE. We now compare the modeled variables with observed SM and EVI using both the correlation and RMSE for the two DH events (i.e. Table 1, updated).

Other comments
L16 Should it be dominated by forests and grasslands? It is inconsistent with results shown in L240.
Yes, thanks for pointing out, it's corrected now.

L179&L289 Please add a few lines to justify the use of the RF regression. What's the advantage compared to a linear multivariate regression?
In the original manuscript (L178-180) we explained that:

> *"Because impacts on EVI could result from non-linear interactions between soil-moisture and temperature or from legacy effects from spring (Bastos et al., 2020a; Lian et al., 2020), we extend this analysis by random-forest (RF) regression […]".*

We have now reformulated the paragraph to improve clarity:

> *"However, departures from a linear model could also result from non-linear interactions between soil-moisture and temperature or from legacy effects from spring (Bastos et al., 2020; Lian et al., 2020). Therefore, we extend this analysis to the pixel scale and further include non-linear effects and interactions by fitting a random-forest (RF) model with SManom and Tanom in spring (MAM) and summer SManom and Tanom as predictors.*

L181 Please explain more on the 3x3 and 17x9 used here.
We have restructured the description of the method. The sentence now reads:

> *To reduce the risk of over-fitting due to the small sample size (17 years) and large number of predictors (4), we fit the RF model on 3 pixels × 3 pixels moving windows centered around each pixel (i.e. 17×9 samples).*

Sec 3.2.2 & Sec 3.2.3 There are some repetitions in these two sections, such as the definition of residuals.

We have removed redundancies in the revised version.

L245 Is it a correct statement? It seems like a few exceptions can be there (SManom and Tanom of cluster C4).

Thanks for pointing out the inconsistency. We were referring to the centroids of the clusters, but indeed some pixels show different behavior. We re-wrote the paragraph:

> *All clusters align along proportional SManom and Tanom in DH2018 vs DH2019, with predominantly negative SManom and positive Tanom in both DH events but alleviation of soil-moisture deficits and heat stress in DH2019 compared to DH2018 (Fig. 2). […] in spite of drought and heat stress alleviation (Fig. 2). Furthermore, the distributions of climate anomalies for each cluster overlap each other and, in some cases, the 1:1 line, indicating that the intensity of the hazard (temperature, drought) cannot account for the resulting impacts alone.*

L260-261 This may be inaccurate. See my major comment.

We agree, the sentence was incorrectly formulated. We have now rephrased to:

> *The results correspond to a general summer water-limited regime, especially in clusters $C_{Decline}$, C_{HighV} and C_{PRecov}, which show stronger sensitivities to Tanom and SManom (slopes in Fig. 4) and higher variance explained by both models (R2 0.58–0.68 for SManom and 0.49–0.55 for Tanom). For these clusters, EVIanom is below the 95% confidence interval of the long-term linear relationships for DH18 (C_{PRecov} and C_{HighV}) and DH19 ($C_{Decline}$ and C_{HighV}). SManom and Tanom in DH18 and DH19 are generally similar to those of 2003, but DH18 was drier than 2003 in C_{PRecov} and C_{HighV}.*

The justification to focus on these three clusters can be found in the reply to R1C2 above.

Fig. 4 Anomalies during EVIanom positive years and EVIanom negative years are supposed to be comparable (add up to 0 eventually). There is a trick that 2018&2019 were not included in this long-term relationship. If they were, "abnormal" positive values could show up.

We disagree with the reviewer that this is a "trick". The rationale for not including these years in the modelling approach is to evaluate whether the vegetation response to the hazard in each of these two years are consistent with long-term relationships with climate (2001-2017).

I.e., we are testing whether past climate-vegetation relationships can be used to predict the anomalies in each event. Such an approach has been used in other studies to evaluate legacy effects from droughts (Anderegg et al., 2015), as well as extreme vegetation anomalies (Bastos et al., 2017).

L266 How can different out of bag scores affect explained variability so much?

This refers not to different ways of calculating OOB, but to the differences in OOB between pixels as shown in Fig. A4. We have rephrased for clarity:

> *"The model is able to explain 48-90% (median and maximum out of bag score across pixels) of the pixel-level temporal variability of summer EVIanom in 2001—2017 […]"*

L291-293 Please rephrase. The improved predictivity of RF seems to be contradictory to the comparable residuals.

It is not contradictory since it refers to the long-term RF model out-of-bag scores for the period 2001-2017, versus the ability to predict DH2018 and DH2019. See our reply to the comment on Fig. A4 page above.

Fig.6 The x-axis labels are hard to read.

We apologize for the low resolution of the figure. We have improved the readability of all figures.

Sec 5.2 Is the simulation of 2018 productivity anomalies really so well? At least during the spring season precondition it is not so consistent between model and data.
Thank you for pointing out. Indeed, if we include the full year the correlations are lower. However, we are also interested in how models simulate the impacts of DH2018 and DH2019, therefore we keep the analysis for the 6 months as well. We now include a new table where we compare the correlations and RMSE (R1C3) for the whole year.

Fig. A5 I cannot find anywhere in the text that this figure is discussed.
It is now refered to when analysing the RF model fit (Section 4.2).

L3 though -> through
Corrected.

L72 Modify the citation.
MULNV-NRW refers to Ministerium für Umwelt, Landwirtschaft, Natur- und Verbraucherschutz des Landes Nordrhein-Westfalen. Since it is correctly linked in the reference, we decide to keep it as in the original manuscript, except the reviewer has a specific suggestion.

L180 Double-check the variables used here.
We have rephrased for clarity:

> *"[…] random-forest (RF) model with SM_{anom} and T_{anom} in spring (MAM) and in summer (JJA) as predictors (i.e. four predictors, T^{spr}_{anom}, SM^{spr}_{anom}, T^{sm}_{anom},SM^{sm}_{anom})."*

L226 excepting -> except
Done

L285 Add . before "In DH2019"
Done

L301 ,since -> , since
Done

L310 GPP should be GPPanom?
Yes, it's corrected.

L323 EVIanom, subscript anom
Corrected

L388 related -> be related
Corrected