Reply to Referee 1

We would like to thank reviewer 1 for his/her constructive comments and feedback on this manuscript. We think that the suggested revisions based on the reviewers' comments will improve the article. Please find our responses (in blue) to each point raised by the reviewer (shown in black) below.

Specific comments

1. Why not shorten the title to 'A multivariate drought catalogue for southwestern Germany dating back to 1801'? Would be a bit easier to read.
   Response: Thank you for the suggestion, we will change the title to the version you suggested.

2. I think that the manuscript needs a clear research question (see e.g. abstract, where it could be added in l.16). Currently, the aim is to present a long-term drought collection. This is a methodological goal, which is fair enough. However, the paper could go further than that by asking: 'is the clustering of extreme events during the past decade unprecedented in a historical context?' I personally would frame the introduction in a way that highlights the need of a long-term dataset to answer this question. This would provide motivation for the study and highlight the practical relevance and value of the long-term dataset. The results presented allow for answering this question and lead to the conclusion that the past decade experienced frequent extreme events which is, however, not historically unprecedented if looking back into the 19th century.
   Response: We think your suggestion of highlighting this thematic research question about the current drought cluster compared to the cluster in the past fits perfectly to our study. We will therefore emphasize this point in the next version of the manuscript.
   Most of the Authors of this paper work in the same interdisciplinary research project about droughts in Southwestern Germany (DRIeR). So, our first motivation was to collect all the different datasets we work with in our interdisciplinary team and create a common database where we can compare different drought events.

3. The introduction would profit from a clearer structure. I would first talk about the hazard component and the different drought types. In this first part, I would also shortly mention different drought indices (indices such as SPI but also duration, deficit, intensity, see e.g. [Van Loon and Laaha, 2015; Brunner et al., 2019]). In a second part, I would transition to the vulnerability and impact component. Then, one could highlight the necessity of long term records to determine the rarity of certain events or periods of events. This would nicely transition to the aim of the study of providing a long-term dataset. And I would definitely talk about the value of long-term datasets in the context of trend analyses.
   Response: In the introduction, we aimed to highlight the need of multidisciplinary and long-term research on drought. We will apply the suggested change to first write about the hazard component, then the vulnerability and impact component, then the necessity of long-term records, this will help to further clarify our aim of this study.

4. Could you please provide a short overview of the homogenization procedure for precipitation and temperature data (l. 79)?
   Response: The homogenisation procedure was conducted by HISTALP. We will add this information in the revised manuscript with the link to the study of Auer et al. (2007), where this homogenisation procedure is described in detail.

5. What is the temporal resolution of the tree-ring series (l. 102)?
   Response: Tree-ring series are annually resolved. Please see line 93 in the manuscript.

6. The description of the impact dataset is a bit confusing and needs clarification (l. 105-118). Do you mean to say: 'Dataset 4 is based on reported textual information on the impacts of drought events contained in two databases'? What do you mean by 'additional reports recently collected (l. 116)? Would it be possible to provide a reference here? Could you provide a bit more information on the reasoning behind the choice of the three impact categories agriculture, ecology, and hydrological systems? Where do e.g. hydropower production and industrial water use belong to?
   Response: Yes, Database 4 is based on reported textual information on the impacts of drought events contained in the European Drought Impact Report Inventory and tambora.org. We used some additional sources, which are not yet included in the database tambora.org. We will add the reference of the sources to the description. The textual information was coded after the categorisation system of the EDII.
   The categorisation scheme of the EDII is elaborate and requires detailed information on the impact, which was not always available for impacts dating further back in time. Furthermore, with respect to historic drought events, certain impacts (such as tourism, ecology) did not play a major role in daily life, and hence have not been reported as such. Accordingly, we adapted the categorisation schemes to the "major" areas of drought hazards and concern: soil moisture, hydrology and ecology (including forestry) to pool comparable impact reports. We will include a table in the Appendix, where the specific impact categories and impact types used in this study, such as hydropower production and industrial water use, are listed.

7. Could you please pay attention to a consistent use of the terms 'variable', 'characteristic', 'index',... while revising the manuscript? In l. 125 e.g. do you really mean to talk about 'variables' or rather 'indices'? Or line 127: weren't indices computed from time series of anomalies?
   Response: We will make sure that we follow the terminology in Figure 1. Indices were computed based on the variables.

8. Drought definition section (l. 124-168): It remains unclear to me how exactly the drought events were determined based on the time series of indices (meteorological droughts) and percentiles (hydrological droughts). Currently, I see two aspects discussed: computation of index time series, and classification of years. Is it correct that the classification step corresponds to a threshold approach, in your case with three different thresholds? If so, could this be clarified?
   Response: We will emphasize that we determined drought years and not drought events with a fixed start and end date. Drought years were derived from the anomaly time series of the different indices (SPI_n, $SPEI_n$, Q_P, $TRI_{species}$). A year was defined to be in drought whenever the variable of interest was abnormally low, in this study below the 20th percentile. Drought years were further classified according to three different severity classifications: D1 (moderate; <20th percentile), D2 (severe; <10th percentile) and D3 (extreme <5th percentile). We will clarify how we defined drought years in section 2.2.

9. Computation of SPI and SPEI: why did you not use hydrological years for the computation of the index time series (l. 138)? This would be more consistent with a hydrological perspective than the use of calendar years.

Response: Thank you for the question. As you mentioned in comment 20, we have already included quite some meteorological drought indices, and did not want to give more weight to meteorological drought. Deriving meteorological drought indices for the hydrological year makes sense from a hydrological perspective, but probably not for other perspectives (e.g., tree-rings are time-stamped based on the calendar year). In fact, in hydrology some countries use a 'low flow year', starting in April for routine analyses (CH) - this is close to some of our accumulation periods. Further, our aim was to show the differences between the different perspectives and not to find the best link between the different drought types.

10. Choice of distribution functions for derivation of SPI and SPEI: please provide a reason for the specific choices made or a suitable reference (l. 140 and 145).
    Response: For the SPI we selected the gamma distribution, because it best fits precipitation sums of different accumulation periods for Europe (Stagge et al., 2015). We will clarify this and add a reference in the revised manuscript. For the SPEI we used the generalized logistic distribution as suggested by Beguería et al. (2014). We will add the reference to the sentence.

11. The vegetation drought section needs some additional explanation (l. 151-161) for non-dendrochronologists: provide a reference to the 'standard methods' (l. 151), explain what a 50% frequency cutoff is (l. 154), explain what a bi-weight robust mean is (l.156), explain what an expressed population signal is (l.159).
    Response: We will provide additional information on the dendrochronological methods and statistics used.

12. Drought severity classification scheme (l. 169-178): In my understanding, this corresponds to the actual drought identification step. Could you please clarify this?
    Response: We will clarify how we defined drought years in section 2.2: Please see our response to comment 8.

13. I think that the term 'frequencies of indices' (l. 176, l.240, l.251) is confusing (applies throughout the manuscript. If I understand this correctly, this is not a frequency but rather number of indices that co-detect a certain event. This whole part on the moving window is a bit unclear (l. 175-178). Why is this moving window approach even necessary?
    Response: With our frequency analysis we want to analyse how many D1, D2 and D3 Droughts occured in a specific unit of time. For this analysis, we were not interested in the different drought types (we assume that every perspective on drought is equally important) but instead we were interested in whether drought occurrence clusters in time, i.e., whether there were decadal hotspots of increased drought occurrences. By using a moving window instead of fixed decadal time blocks, we assure that we don't miss decadal drought hot-spots that happen at the end of one decade and the beginning of the following. With your comment in mind, we see that we have to clarify that step in the revised version of the manuscript.

14. Choice of Pearson correlation for correlation analysis (l. 186). Why use a linear correlation measure and not just a monotonic one, e.g. Kendall's or Spearman's rank correlation coefficient. Maybe there is a relationship which is just not linear.
    Response: We will now repeat the analysis using Spearman's Rank-Order Correlation and compare the results obtained from the two analyses.

15. ‘Similarity index’ (l.186-187): It remains unclear to me what exactly this index does, and why it is called similarity index. Is the ratio you are talking about n_extreme/n_all? If so, did you compute this ratio for both indices and then compare the ratios to determine similarity? Please clarify.
    Response: We will clarify in the revised manuscript that the similarity index is calculated as the total number of extreme drought years identified by two considered indices divided by the number of extreme drought years identified by index separately

    $$Similarity = \frac{total\ number\ of\ extreme\ droughts/2}{number\ of\ common\ extreme\ droughts}$$

    It is used to describe the similarity between two indices to identify extreme drought events.

16. It would be nice to compute the similarity measures r and s not only for two periods but using a moving window approach allowing for an actual trend analysis (l.189-192). The problem with the two-period as opposed to a moving window approach is that one may compare a period located at the high end of an oscillation with one at the low end of an oscillation and therefore mistakenly interpret a trend even though these two periods are just located in two different parts of a cycle.
    Response: We agree that the selection of two periods might seem arbitrary and a moving window approach could be more appropriate here. The problem with a moving window approach, however, would be to demonstrate these results for all indices (153 combinations) and the long period examined here. We will try this approach with a reduced number of indices.

17. I do not understand why this second grouping is necessary (l. 196-200). Do you mean that you assign one or several reasons to the choice of an event?
    Response: Figure 5 and 7 grouping was necessary to identify similarities and differences among the datasets. Additionally we want to distinguish between short- and long-term meteorological drought indices (SPI and SPEI). There was almost no difference between the long-term SPEIs and the long-term SPIs, therefore we excluded the latter one for Fig. 5 and 7. We will add more information on the grouping in order to clarify that point.

18. No actual trend analysis is performed in this study. I would therefore not talk about ‘become more frequent’ (l.209) but rather say that extreme droughts happened in clusters (e.g. 1860s and recent decade). Similarly I would say ‘the last decade shows a high (not higher) severity of events’ (219).
    Response: Indeed we did not perform any trend analysis, so we changed the sentence to “several clusters of increased drought occurrence were identified”. With the sentence “the last decade shows a higher severity of drought events” we mean, that more drought events are classified as extreme (class D3). We will change the sentence accordingly.

19. Event clustering (l.240-249): I think that this temporal clustering aspect as opposed to a trend is interesting and deserves some more attention.
    Response: We will emphasize more on the temporal clustering in the revised manuscript.

20. Figure 3: Following the methods description, would it not be more logical to present the impact panel after panel b)? Why does panel a) not have a grey background for ‘no events’? In the calculation of the percentages presented in panel c, aren’t the meteorological indices getting much more weight than the other indices because there is so many of them?
    Response: We decided to present the catalogue in this order, because the impacts were not included in panel c). Based on your comment, we see that this seems to be confusing, so we will change the order of the plot. In the revised manuscript we will present Figure 3 in the following

order: (a) individual indices, (b) composite information, (c) impacts.
We decided against a grey background for "no events" in the impact dataset, because unlike the other datasets used in this study, we cannot guarantee that we have evaluated all available written sources. That means, we cannot say for sure if there was "no event". We will expand the point on challenges with textual information in the discussion. In panel c) the meteorological indices make up a large proportion of the number of events per year, but this proportion does not change through time, so the 'weight' stays the same. The percentage of indices indicating droughts was calculated based on the number of available indices per year, i.e. adjusted for the period following 1900 when Q became available.

21. Drought frequency (section 3.2): I do not see the added value of this moving window approach. What does it allow to demonstrate which is not already shown in Figure 3c? Wouldn't some temporal clustering approach be more beneficial here? E.g. group all events separated by less than 2 years without a drought?
Response: Please see also our response to comment 13. In Fig. 3c we only show the number of drought events per severity class per year. In the next step, we were interested in the temporal clustering of droughts, i.e., decadal drought hot spots. Therefore we decided to calculate how many droughts per severity class occurred in 5 and 15-year periods. Instead of the moving windows, we could have used fixed timespans (e.g. for 5 years starting in 1801: 1801 to 1806, 1807 to 1811 and so on) but this is then biased by the starting point, so we found it more objective to use the moving window approach.

22. I would include Figure S5 in the main article and remove Figure 4 instead. What is the difference in the results derived from the correlation and similarity analysis? If both transport the same message, why not remove one of them?
Response: Do you mean Figure 6 instead of Figure 4? In the similarity analysis we used only extreme droughts (severity class D3), while in the correlation analysis (Figure S5) we included all severity classes. We will include the formula we provided in our response to point 15, we hope this will demonstrate the differences between the two metrics.

23. Section 4.1: It is interesting to note that the droughts identified by all indices seem to have a regional extent as illustrated by the references provided. I think it would be interesting to discuss this aspect a bit further.
Response: Indeed, that is an interesting point which we will include in the discussion of the revised manuscript.

24. I do not think that the statement 'the recent period was characterized by higher frequency of extreme droughts' (l. 435) is particularly well supported by the results. The results presented in Figure 3 rather show that there are temporal clusters of extreme events and that the cluster of extreme events observed in the recent decade is not unprecedented (e.g. 1855- 1870). I think that the strength of this study is exactly that it provides this context which is often missing when looking at short records (last 30-40) which bring us to conclusions such as 'extreme events become more frequent'. Your dataset nicely shows that periods of frequent extremes happen now but also happened in the past. I would add a discussion point on this temporal clustering aspect. Ideally, referring to existing literature.
Response: We agree with you. We will change this sentence to: "was characterized by increased occurrence of extreme droughts." Nevertheless, within our dataset especially in the last five years

(see Fig. 4c, last bar: 2014 to 2018) we found more extreme droughts (severity class D3) than in the other timespans. But we agree with you, that the clustering of several years with extreme droughts is not unprecedented. Both points are important, so we will distinguish between them and focus also on the temporal clustering in the discussion.

25. I think that the conclusions could be much stronger than the ones currently presented (l.442- 456). I suggest to add something along the lines of: 'Our long-term dataset shows that (1) extreme droughts cluster in time, (2) the recent decade experienced many extreme droughts similar to a period in the mid 19th century, (3) the last decade is less exceptional in a historical context than when looking at the last 30-40 years as often done in trend analyses.
    Response: Thank you very much for your suggestions. We will include these points in the conclusion along with the value of using a multivariate dataset. We think this will nicely underline the main aspects of this study: I) using a long-term and II) and multivariate dataset for regional drought research.

26. Could you provide some information on how the community can access the dataset?
    Response: Yes, we are currently looking for an option to provide access to the dataset used here.

Minor points

Thank you for the minor comments. We agree with the comments and suggestion of small changes and will address all minor points in the revised manuscript.

References used in this review and the reply to this review

Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K. R., Jones, P., Efthymiadis, D., Brunetti, M., Nanni, T., Maugeri, M., Mercalli, L., Mestre, O., Moisselin, J.-M., Begert, M., Müller-Westermeier, G., Kveton, V., Bochnicek, O., Stastny, P., Lapin, M., Szalai, S., Szentimrey, T., Cegnar, T., Dolinar, M., Gajic-Capka, M., Zaninovic, K., Majstorovic, Z., and Nieplova, E.: HISTALP - historical instrumental climatological surface time series of the Greater Alpine Region, International Journal of Climatology, 27, 17–46, https://doi.org/10.1002/joc.1377, 2007.

Beguería, S., Vicente-Serrano, S. M., Reig, F., and Latorre, B.: Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring, Int. J. Climatol., 34, 3001–3023, https://doi.org/10.1002/joc.3887, 2014.

Brunner, M. I., K. Liechti, and M. Zappa (2019), Extremeness of recent drought events in Switzerland: Dependence on variable and return period choice, *Nat. Hazards Earth Syst. Sci.*, *19*(10), 2311– 2323, doi:10.5194/nhess-19-2311-2019.

Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., van Loon, A. F., and Stahl, K.: Candidate Distributions for Climatological Drought Indices (SPI and SPEI), Int. J. Climatol., 35, 4027–4040, https://doi.org/10.1002/joc.4267, 2015.

Van Loon, A. F., and G. Laaha (2015), Hydrological drought severity explained by climate and catchment characteristics, *J. Hydrol.*, *526*, 3–14, doi:10.1016/j.jhydrol.2014.10.059.

Weibull, W. (1938), *Investigations into strength properties of brittle materials*, Generalstabens litografiska anstalts förlag, Stockholm.