

Review for manuscript “Exploring the added value of a long-term multidisciplinary dataset in drought research – a drought catalogue for southwestern Germany dating back to 1801”

Authors: Mathilde Erfurt, Georgios Skiadaresis, Erik Tjeldeman, Veit Blauhut, Jürgen Bauhus, Rüdiger Glaser, Julia Schwarz, Willy Tegel, and Kerstin Stahl

Journal: Natural Hazards and Earth System Sciences

Summary

Erfurt et al. present a long-term drought catalogue for southwestern Germany for the period 1801 – 2018 collected using four types of datasets: precipitation and discharge time series, tree-ring datasets, and drought impact information. They identify meteorological, hydrological, vegetation, and impact drought events using standardized time series of the four variables. They show that not all variables indicate the same events and that there are overall three periods of events with clustered drought occurrence including 1857-1870, 1947-1964, and 2003-2018. While the most severe events are visible through most variables, certain important events would not be detected if looking at just one specific time series. The study nicely highlights that the cluster of extreme events observed in the last few decades are historically not unprecedented.

General comments

The study compiles a long-term drought dataset considering different types of droughts. It therefore addresses two important problems in hydrology: (1) the limited record length usually available for trend analyses in extremes and (2) the multifaceted nature of drought events, which propagate through the hydrological cycle. The study is generally well written and organized even though it may profit from rephrasing and slight restructuring now and then. The datasets chosen are suitable for the analysis and the methods chosen mostly appropriate. The results are presented in clear and nicely designed figures. I below point out the need for strengthening the research question, discuss a few methodological points that should in my opinion be improved, point out a few passages that need clarification, and provide some suggestions of how the conclusions could in my view be strengthened.

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.
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.

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.
4. Could you please provide a short overview of the homogenization procedure for precipitation and temperature data (l. 79)?
5. What is the temporal resolution of the tree-ring series (l. 102)?
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?
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?
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?
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.
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).
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).
12. Drought severity classification scheme (l. 169-178): In my understanding, this corresponds to the actual drought identification step. Could you please clarify this?
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?
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.

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 $n_{\text{extreme}}/n_{\text{all}}$? If so, did you compute this ratio for both indices and then compare the ratios to determine similarity? Please clarify.
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.
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?
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).
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.
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?
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?
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?
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.
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.
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.

26. Could you provide some information on how the community can access the dataset?

Minor points

I. 26: with 'potentially' widespread negative consequences ...

I. 44: which 'can' affect all components ...

I.70: drought types such as ...

I.71: drought indices such as ...

Figure 1: I would slightly extend the caption and provide a bit more information on the content of the figure. The equation within the blue box on streamflow percentiles is strange. I would remove it from the figure. Indexvalue should be 'index value'. Variable names should be in italic (e.g. T should be *T*).

I.75: 'the study employs and assembly...'. Rephrasing needed.

Figure 2: I would use a gray scale for the relief as you are just displaying one variables which does not require the use of a rainbow color scheme. By 'Stand', do you refer to 'Standort'? If yes, I would use an English abbreviation instead such as 'Loc'.

I.138: can you also provide a reason for June as you did with the other periods, analogous to the tree growth example for September?

I. 140: The distribution is fitted to the data and not the data to the distribution. Sentence needs rephrasing.

I. 142-143: reference period for what? I do not understand the meaning of this sentence.

I. 149: provide a reference for the Weibull plotting position: [*Weibull*, 1938]

I.181: do you mean from 1801-1900?

I. 184: two instead of 'three' metrics? I only see the Pearson correlation and the 'similarity index'.

I. 196: by extreme, do you refer to set D3?

I. 197: which datasets were grouped?

I. 208: I see a few more years in Figure 3b: 1964, 1949, 1991.

I. 222: Link this ET statement to literature on changes in temperature.

I. 271-271: with most, do you mean more than 10 (see caption Figure 5)? And what does the sentence 'in all cases more than 25%...' mean?

I. 288: which 'two' drought types?

Figure 5: Would it be possible to make the red color a bit more purplish to better fit into the overall color scheme?

I.304: When talking about the two different accumulation periods, do you refer to meteorological droughts?

I. 377: and due to more frequent reporting?

I. 405: how was this increase in reporting taken into account?

I.424f: use of word 'distinct', do you mean 'index-specific'. To me, the term distinct looks odd in this context.

I. 433: specify which two periods.

I.460: 'all versions of the paper'. The readers just see one.

References used in this review

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.

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.