Review of Bachmair et al. "Exploring the link between drought indicators and impacts"

This paper is one of the first attempts to link drought indicators to text-based drought impact information. It used the case study region of Germany because of data availability. I think the manuscript is very well written and addresses an important topic. I appreciate the effort that goes into gathering the data necessary for this study and the careful data analysis considerations displayed by the authors. I see this paper as a first step in judging the usefulness of indicators for drought monitoring and early-warning to text-based drought impact information. Clearly, future research is needed on this topic, especially on the spatial variability and event specificness of the relationships. When more impact reports will become available this approach should be repeated per impact category and linking drought severity to impact severity.

I have only one major issue with this research and some minor suggestions for improvement of the paper. My major doubt is related to the selection of the years to include in the analysis. On page 7591 it is mentioned that only years were included with one or more impacts. The authors argue that that is done to prevent inclusion of drought years in which no impacts were reported, but that assumes that impacts always occur during a drought event. Impacts might not have been reported in drought years, they might also just not have occurred. I would like the authors to do a hit rate / false alarm rate analysis, analysing years in which there was a drought in the indicators but no impacts and years in which there were impacts but no drought. Of course this analysis is largely influenced by the biased temporal coverage of the drought impact report database, as the authors suggest, but it is very valuable if one wants to predict impact occurrence from drought indicator values.

Abstract:

Although I see the difficulties, I would encourage the authors to include more tangible results in the abstract. I think the authors are now too cautious and hide very interesting results in the text of the manuscript. I suggest to include some of the following: the higher performance of SPEI compared to SPI, the highest correlation with intermediate accumulation periods, the comparable results for streamflow and SPI, the only minor difference between all impacts and hydrological drought impacts (even when compared to streamflow percentiles), the spatial variability in the relationships (e.g. the difference between north and south in terms of indicator thresholds for drought onset), event-specific relationships (difference between long vs. short events and spatially concentrated vs. widespread events), and the spatial and inter-event differences in "best" indicator and "best" threshold limiting their use in large-scale drought monitors.

The word "threshold" is used differently in this research then in other papers. Please include a short definition in the abstract explaining that it is the indicator value concurrent with impact onset.

Methods:

In some states only a few data points are available, especially in small states. How was the 10th percentile of SPI / SPEI calculated from only a few grid cells and the 10th percentile of Q / G from 3 or 4 stations? I have some doubts in comparing states with a high and small number of data points for the calculation of indicators.

In Figure 2 we see quite a clear spatial distribution of impact report data. This might have a considerable influence on the results, e.g. higher correlations in states with more reported impacts or more bias toward certain impact types in states with less reported impacts. You should explore this in the discussion.

Please mention how you did the differentiation of impacts to get Ih. An example would be helpful.

Results:

The boxplots in Fig. 4 and 7 need more explanation. What are the blue dots? How should we read the figure?

Why were SPEI10-3 en Q selected in Sect. 3.2?

I don't see the conclusion on page 7596 that drought events with geographical concentration have a higher number of indicators with moderate or strong correlation. Same for page 7599. Why do the years 1976, 1992, and 2011 have a higher spatial variability of indicator and/or impact data? I do not see that in Fig. 5. At the least the authors should explain how the figures should be read in order to come to this conclusion.

Small and textual comments:

general: use chronological order for references in text

Fig. 3 & 6: could a colour scale be found with more difference between moderate, strong and perfect correlation? Sometimes the differences pointed out in the text are hard to spot in the figures, for example the slightly better performance of the 10th percentile compared to the mean in Fig. 6.

Fig. 7: Ih onset used for Q and G?

p. 7587, l.2-3: "common tool" > include references

p.7588: which period was used as reference period for SPI and SPEI calculations? Also 1970-2011 like for streamflow and groundwater?

p.7594, l.10: inverse direction, but not significant

p.7595, l.5: later > latter

p.7595, l.26: negative > negative

p.7596: "better the mean" > insert than