

We would like to thank John Bloomfield for his feedback and constructive suggestions on this manuscript. We think that revisions based on the reviewers' comments will further improve our work. Below we give point-by-point responses to the comments (blue and italic).

The paper presents an analysis of baseflow 'to characterize groundwater drought on a catchment scale'. Trends in observed baseflow minima and derived drought descriptors are identified and investigated in the context of 'climatic and catchment controls'. A 'scenario-neutral' approach is adopted to characterise the sensitivity of the drought descriptors to future changes in. The study uses flow data from 338 gauges in headwater catchments across Germany.

General comments

Introduction:

The aims of the study could be set out more clearly in the Introduction. The two following statements 'we use a baseflow approach to characterize groundwater drought on a catchment scale' (p2, l19) and 'Employing a data-based approach, in this study we assess future changes in drought hazard on catchment scale across Germany' (p3, l29) describe what has been done, but there is no unambiguous aim or research question stated in the Introduction.

Thanks for pointing this out. We agree and will set the main aim and the more specific objectives more clearly in the revised version. The main aim may perhaps read:

This study aims to develop a data-based approach to estimate the potential direction of future changes in hydrological drought hazard under climate projection uncertainty.

Study area and data:

Data from 338 gauges on headwater catchments were used in this study. In this context, what constitutes a headwater catchment?

The term "headwater catchment" was chosen as all catchment areas are below 200 km². We will state this more clearly in the revised version.

It's not easy to tell from Figure 2, but it is possible that some of the gauges are nested catchments. Is this the case? If so, what biases if any might this introduce into the data? What are the implications of those potential biases for the trend analysis (section 4.2) and the results of the potential future drought hazard assessment (section 4.3)?

We agree that nested catchments would introduce biases and hence did not use any nested catchments in this study. We will add this information in the revised version.

Streamflow data for the period 1970-2009 was analysed. Was the data complete? If there was any missing data how was it accounted for in the pre-processing of the data?

As we report in section 2, any records with data gaps were not considered for this study.

The data 'were visually screened for signs of anthropogenic influence' and 'four of the gauges showed spurious changes ... and were subsequently removed' (p4, l7-8). Please could you justify their exclusion in more detail. What anomalies were present in the data that caused you to exclude the sites?

For all records, summed precipitation was plotted over summed streamflow for the entire period. Under constant conditions, the plot is expected to show seasonal varying slopes but no sudden knee. However, four of the gauges had a knee, indicating some kind of anthropogenic influence. In a Master thesis, L. Gerke was able to relate the knee in the plot of one catchment to a change of the gauges' location. We will include these details in the revised version.

Why was a 2/3rds fraction used to define mixed catchment (p4, l21)? Is there a citation for this?

The fraction is arbitrary and was chosen to get the relatively homogenous groups “fractured” and “porous”. Due to the high heterogeneity of the catchments we think it is hard to define more detailed groups with sufficient sample size. Using another reasonable threshold is not expected to change the results as the groups all cover a broad range of values. However, significance of the results could slightly change. We suggest to test this on a few examples in the course of the revision process and then decide what needs to be included based on the results.

Is there any information on the distribution of low permeability superficial deposits across the region? It's not uncommon for such deposits to play an important role in stream flow generation and so the proportion of such deposits in catchments might be an interesting parameter to investigate in the context of the study, and should at least be commented on, either here or in the Discussion.

We don't know of data on superficial deposits across the region that are detailed enough to use them as predictors. However, we will explore this further and in lack of data add this point to the discussion in the revised manuscript.

Methods:

The justification for the use of Mann-Kendall (MK) test could be more robust. There is a significant literature on the application of this test to hydroclimatic time series. However, it's application to such time series is also contentious when there are underlying auto and cross-correlations present. Consider adding to the justification of use of the zstatistic [also see comment below about identification of significant trends based on MK test].

Results:

Figure 4c is a map of the MK z-statistic indicating the direction and magnitude of the trend in Qb7. However, it is stated that 'according to the MK-Test, 40 out of the 338 catchments show a significant trend in Qb7' (p8, 17). What was the level of significance that was used? Please justify this and link this justification back to the Methods, section3.2?

The Mann-Kendall test is a common test for statistical analysis of time series in hydrology. The detailed description of the method applied is given in Appendix B of the manuscript. Since MK overestimates significance for autocorrelated time series, we correct in this study for serial correlation using pre-whitening (see equation A6 in the manuscript). As we think that the detailed procedure is only interesting for part of the readers, we prefer to leave most of this information in the Appendix, however we agree that the significance level ($\alpha=0.05$) may be mentioned in the main text as well.

Discussion:

It is noted in Section 5.1 of the Discussion that 'a baseflow approach does not allow for conclusions on groundwater storage in snow-dominated catchments' (p9, 130). In this context, how do you define snow-dominated catchments, and which if any of the 338 headwater catchments that you have analysed fall into this category? If any of the catchments are 'snow-dominated' should they be excluded from your analysis?

Groundwater recharge and thus groundwater baseflow as well depend not only on precipitation but also on temperature if precipitation falls as snow. As our analyses were performed on monthly scale, this is only relevant for catchments where temperatures are regularly below zero degrees for several weeks. In our study these are mainly the catchments in the Alpine foothills where snow accumulates over winter and smelts in spring/summer so that annual minimum flows occur in winter/spring. For the low mountain ranges of central Germany there are only few catchments in our study where annual low flows occur occasionally in winter/spring. As we discuss, the results for these “snow-dominated” catchments have to be interpreted carefully. In accordance with the comment of Mr. Sawaske we will add the changes for summer low-flows in these catchments to our analysis.

Specific comments

P1, 110: replace reflexion with reflection

P2, l28: resp. is not a normal abbreviation to use in articles like this. If it is and abbreviation for respectively please re-write [see also p5, l28, p10, l9 and p13, l25]

P9, l9, replace Marchant and Bloomfield (2013) with Bloomfield and Marchant (2013)

P10, l5, delete comma after revealed

Will be corrected.