

Response to reviewer 2

- The reviewer's comments are **bold**, our response is in *italic*.

**The presented study mainly aims at evaluating common climate classification systems for their ability to stratify regions with similar hydrological drought characteristics. It is indeed an interesting and relevant study. However, I think the study can be improved in some aspects.**

*We are grateful for the reviewer's relevant comments and suggestions on how to improve the manuscript. In this reply, we respond to each comment in order of appearance. The final implementation of the comments will be presented in the revised version of the manuscript. To avoid duplication in responses to the reviewers, we sometimes refer in this response to the response on reviewer comments of Henny van Lanen.*

**It is interesting to compare the varying drought characteristics between catchments of the same climate class in Europe and the USA. And showing that drought duration characteristics of the same climate class differ between Europe and the USA clearly adds to the evaluation of the climate classification systems. However, for a general large-scale evaluation of the suitability of the climate classification systems to stratify regions with similar hydrological drought characteristics, it would be useful to also evaluate the classification systems for the complete data set as a whole and not only for regionally predefined sub-sets (i.e. Europe and the USA separately).**

*We are thankful for pointing this out. Please see the response to comment 1 of the review of Henny van Lanen for the suggestion on how to include this in the new manuscript.*

**I understand that the main objective of your study was to evaluate the climate classification systems. However, since you compare the suitability of the climate classification systems with classified individual controls, it would be interesting to compare whether also the DDCs of the same class of an individual catchment characteristic differ between Europe and the USA. This would serve both of your objectives, the evaluation of the climate classification systems as well as extending the knowledge about the controls of hydrological drought.**

*We agree. Please see the response to the review off Henny van Lanen (comment 1) for the suggestion on how to include this.*

**In the section "3.2 Statistical comparison" and also in the discussion, you frequently write that two DDS differ / do not differ "significantly". In a section on "statistical comparison" this could be easily understood as "statistical significance". However, currently you only assess the statistical significance of differences between individual percentiles and not between the DDCs as a whole (i.e. the part above the 81st percentile). Please make this clear in the text or specify when the whole DDCs are statistically significantly different / similar.**

*We agree. In the revised manuscript we will clearly distinguish tested statistical significance from e.g. "notable differences" or other phrasing).*

**Minor comments:**

**Page 12883, lines 15-16: How many catchments are in Europe and how many in the USA?**

*There are 461 catchments in the USA and 347 in Europe. We will add this to the revised manuscript.*

**Page 12883: The aridity index should be better explained.**

*We will better explain this in the revised manuscript*

**Page 12886, line 3: It might be better to say “equal number of catchments” instead of “equal size”. “Equal size” can also be understood as classes with equal interval widths.**

*Thanks for pointing this out. We will revise it in the new version of the manuscript.*

**Page 12886, line 4: When referring to figure 2b, mention that only three classes are shown as an example instead of five.**

*We will mention this in the revised manuscript*

Page 12887, line 1: Write “we used” instead of “we use” to be consistent with the tense used.

*We will revise this in the new version of the manuscript.*

**Page 12886-12887: For the individual controls the class intervals could be mentioned to give the reader a bit better understanding of the catchments.**

*Class intervals are presented in Fig. 3 of the current manuscript. In the revised version of the manuscript we will specifically make a reference in the text that class intervals can be found in the corresponding figure.*

**Page 12891, lines 10-11: A bit more detail to the studies of Van Lanen et al. and Van Loon et al. would be useful.**

*We will provide more detail on these studies in the revised version of the manuscript.*

**Page 12891, lines 13-16: This is a long and complicated sentence, which even introduces a new comparison. Please rephrase the sentence and introduce figure 6 a bit better.**

*We will rephrase this part of the text in the revised the manuscript and give a better introduction to Fig. 6 in the text.*

**Page 12892, line 11: Use a comma instead of a semicolon after “(2012)”.**

*Ok*

**Page 12893, line 7: Should it be “higher PET>P classes” instead of “higher AI classes”?**

*Thanks for pointing out this mistake. We indeed meant PET>P instead of AI.*

**Page 12894, lines 3-5: You write that the lower AI classes in the USA “mainly consist of the hot summer climates (Cfa, Dfa)”. However, these two climates together represent clearly less than 50% of the catchments in these classes. I would rather say that the catchments in the lower AI classes are represented in all of the climate zones in the USA.**

*We agree and will rephrase this in the next version of the manuscript.*

**Figure 5: As some classes in the KG-system are not represented in either Europe or the USA, it would be useful to mark in the figure for which combinations of classes the similarity was not assessed (e.g. by shading those cells in grey instead of white, which also stands for “no similarity”).**

*Thanks for this suggestion. We will apply this in the revised version of the manuscript. In Fig. 5 of the response to the review of Henny van Lanen, we have provided an example how this grey-shading will look.*

**Figure 5: If you define when two DDC can be considered as significantly different, it would be helpful to adjust the color coding accordingly, i.e. that it can be clearly seen from the figure which DDC are significantly different or where differences are not significant.**

*We use the number of similar percentiles (percentiles with no statistically significant difference) as a measure of statistical similarity. We will clarify this in the text (see also response to your major comment on the use of significant in Section 3.2).*