

Referee comment to “Controls on hydrologic drought duration in near-natural streamflow in Europe and the USA” by Tijdeman et al.

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.

Main comments

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).

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.

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.

Minor comments

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

Page 12883: The aridity index should be better explained.

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.

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

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

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

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.

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.

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

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

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.

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”).

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.