

Dear Editor,

We thank all reviewers and editor for their valuable comments. The major comments were implemented in the new manuscript as suggested.

Extra analysis:

- We applied the same analysis that was done for Europe and the USA to the entire dataset and included this in the new manuscript (Figures 3, 4, 5, 6, 7).
- We included an extra analysis on the differences in DDC for the same classes of individual controls between Europe and the USA (Figures 4 and 6).

Clarifications/justifications in the manuscript:

- We better explained that our focus was on climate classification in the introduction.
- We clarified why we focused on long duration droughts in the methodology section.
- We judged any use of the word significant and changed it when the meaning could be understood in the wrong way.
- We now discuss the importance of deficit volume in the discussion.
- We clarified in the discussion that some catchment characteristics (such as the BFI) are an important control on drought duration as well. However, for the considered basins, this control was not dominant and we emphasize that we show that climate classification systems are still suitable to differentiate basins according to long duration droughts for a wide range of catchment properties. We further mentioned that catchment controls might be stronger in other (e.g. larger) basins and mention the potential bias of near-natural basins towards head water catchments.
- We justified why we included catchment controls Area and Elevation in the introduction.
- We removed the statement that climate classification systems are often used in large scale drought monitoring and early warning systems from the abstract.

A more detailed point-by-point response to all major and minor reviewer comments, including the previously missed reply to the comment of reviewer 1 about the intermittency time of long duration drought events, is presented below. We believe that the manuscript has improved considerably after the implication of all reviewer comments.

Kind regards,

Erik Tisdeman, on behalf of the co-authors

Response to reviewer 1

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

Major comments

The authors aim to provide a global assessment of the control on the drought duration using near-natural catchments. However, they only use data from Europe and the US. I understand that the authors are limited by the data availability and that a real global analysis might be difficult to perform. This in itself is not a real problem. I think the authors could do a better job on generalizing the conclusions. At the moment they separate the two continents in all analysis, while I think the manuscript might be helped by a better comparison between the two continents. Why not use one analysis and pool all catchments for a specific climate or BFI and perform the analysis on the combined data. If the goal is to find the impact of climate on the drought duration, I think it would be better to group catchments with that climate, independent of their geographical location. This would strengthen the analysis and make it more general and hence also more applicable to other catchments on other continents.

The analyses of the total dataset are now presented in Figures 3, 4, 5 and 6. Please see the reply to the reviewer comments of Henny van Lanen.

Linked to the previous comment, I miss one reference to analyse the differences between the drought durations. For each continent a separate reference is used. This does not allow the reader to compare similar climate, BFI or other controls that are located in different continents. I think, that for example Figure 3, would benefit from 1 reference so that I can compare similar KG, AI or other indicators directly.

Now in Figure 3 and 4 as suggested.

Although information on the uncertainty is mentioned in the discussion, I do miss that information in the Figures or in a table. I think the manuscript would be strengthened if this information is provided so that the reader can see how significant the difference between KG or AI are instead of just providing the ensemble mean for the class. The authors mention that the obtained results might help in understanding the catchment behavior in drought conditions. However, I'm not convinced that only information on the drought duration of the long droughts would provide sufficient information. As stated by the authors they leave out the information on the frequency of drought with this analysis. I'm aware that this would require some work, but I was wondering if the authors could not add information on the intermittency of the events (the time between drought events). If this information is provided the reader would also know if the long drought tend to follow one another or that a long drought is always an isolated and rare event. Maybe this is beyond the scope of the paper, but I was wondering if the authors have any ideas regarding this question.

We applied statistical tests to provide a measure of similarity and hence consider the uncertainty quantitatively through the statistics.

Fig. 4 and the corresponding minor comment in the response to Henny van Lanen illustrate that the long duration droughts (red line) match well with previously described major drought events. The short duration droughts (blue) do not really reflect these, but provide more or less a constant base-

signal over the entire period of record. Based on this, we conclude that these long duration droughts provide sufficient information on the historical drought events that we are interested in.

The median intermittency time of the selected long-duration droughts is 26 weeks (about half a year) for both regions; the IQR is 7-72 weeks. This means that only a small proportion of long duration may not be independent, a larger proportion is separated from each other by at least a few months.

Another reason to focus on long duration droughts was based on the larger variation in DDC after the 81st percentile (Fig. 2a, current manuscript). We clarified our focus on long duration droughts in this study in the revised manuscript in Section 2.2

Finally, how could the obtained results be used in an early warning system, like mention in the abstract? Maybe this could be discussed in the Discussion. I think if the authors can show how to use the obtained results can be used in these systems; it would increase the social relevance of the paper.

We present statements about the potential use of the results in Drought Monitoring and Early Warning Systems and comparative drought studies in the Conclusion.

Minor comments

Page 12878 Line 5-6; currently lacking is a large-scale evaluation of the relation between climate and hydrologic drought characteristics, I do not agree. Multiple studies have tried to tackle this topic and the first author is part of some of these studies.

Thanks for pointing this out. We were aware of the use of modeled data in the cited studies and intended to refer to “observed hydrologic drought characteristics” but mistakenly did not do it. We changed this in the abstract

page 12879 Line 7-9 Add reference

We revisited this statement and added a reference to the introduction.

Page 12884 Line 9 Why is the Koeppen classification from Kottek (derived from global forcing data) and not compute the KG class based on the local catchment forcing? This would remove potential problems with the global data compared to the local conditions.

For this analysis, we used the method (not the map) described in Kottek et al., 2006 (pg 12884, line 9) to calculate the KG for each basin based on meteorological data for that basin (corresponding to individual controls P and T). We clarified this in Section 2.1.

Page 12889 Line 3-5 Why could the difference in the DDC for both E climates not be related to the topography. In the US the topography in the E climate is rather flat while in Europe this is not necessarily the case.

We agree. Furthermore, the difference in correlation between precipitation and elevation could contribute to this difference (negative for the USA, positive for Europe, presented in Figure 1 of the response to the review comments by Henny van Lanen). Although we acknowledge the point of the reviewer, we choose not to mention this in the discussion since we did not include slope as an individual control in our analysis.

I miss a proper caption. You need the main manuscript to understand the Figure. I think the reader would benefit if a longer caption would be provided to inform the reader on all the complex figures and information that is provided in Figure 3.

We clarified the captions of the new Figures 3, 4, 5, 6.

Response to reviewer 2

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

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. We had reasons to separate the two regions for this analysis (described in the response to major comment 1 by Henny van Lanen) and did not specifically aim to provide a global analysis. However, in terms of generalizing the results, we agree that it is beneficial to include the analysis for the entire dataset. Therefore, we added the analysis of the entire dataset to the new Figures (3, 4, 5, 6, 7) and changed the Results + Discussion Section accordingly.

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 and therefore added this analysis to the new Figures 4 and 6.

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 and therefore reconsidered any phrasing of significant within the revised manuscript and rephrased where it could be misinterpreted.

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 add this to the revised manuscript in Section 2.1.

Page 12883: The aridity index should be better explained.

We described the Formula to calculate the AI in Section 2.1.

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 revised it in the new version of the manuscript (Section 2.3).

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

We added this to the specific sentence (Section 2.3)

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

Revised (Section 2.4)

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

In the methodology section we explain how classes are defined and mention that resulting class intervals shown in the results.

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 added a more specific note on which climate and catchment controls these studies found to modify drought duration to the discussion.

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 rephrased this sentence in the discussion. It now gives a better introduction to the Figure. (Section 4.1)

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

Done (Section 4.1)

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 and changed this in the new manuscript (Section 4.1).

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 rephrased this in the new manuscript (Section 4.1).

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 applied this in the revised manuscript (See Figures 5 and 6)

Figure 5: If you define when two DDC can be considered as significantly different, it would 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 clarified this in the caption of Figures 5 and 6.

Response to Henny van Lanen

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

Comment 1

An important justification for the paper is that large-scale studies (many gauging stations) are needed based on observed streamflow data. Generic results are needed on drought duration and controls. It is strange that from the beginning (except for distinction of the 5 classes for the individual controls, Section 2.3) the USA and Europe are separated. First results of all 808 gauging stations together should be studied and presented (e.g. extension of Figs. 3, 5 and 6), which then can be followed by a separate treatment of the USA and Europe, as done in the current manuscript. I realize that the authors eventually will show that there are some differences between USA and Europe (e.g. in the higher $PET > P$ or AI classes).

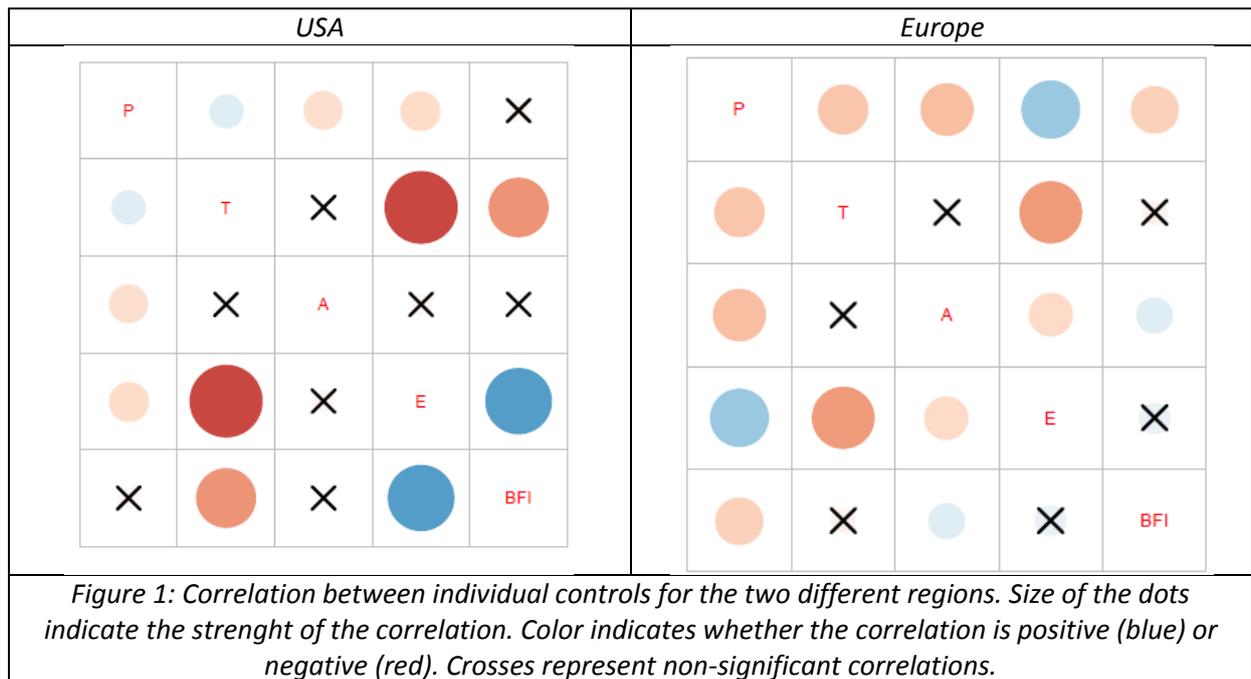
The authors are thankful for the suggestion to additionally conduct a combined analysis based on stations from both regions combined. It was not included in this first version of the manuscript because our research questions were phrased in a comparative way, i.e., putting the comparison to use between regions that have separate operative drought monitoring systems (European Drought Observatory, US Drought Monitor):

- *Which climate classification systems work best within each of the two regions?*
- *Are the same classes of a given climate classification system comparable between the two regions?*
- *What are (possible) reasons for the differences between continents?*
 - o *For this third objective, we look at differences between individual controls that are not pre-classified.*

However, it is certainly possible to rephrase these slightly and also present the same analysis for the entire dataset. However, it must be considered that due to the larger number of USA records, combined results will be biased accordingly.

The following differences exist between the two datasets:

- *Köppen-Geiger climates (See Fig. 1, manuscript)*
- *Correlation between individual control variables (Fig. 1)*
 - o *Correlation between Elevation and Precipitation: In Europe high elevations are wet, whereas in the USA high elevations are dry*
 - o *Correlation between Temperature and Precipitation: Positive Precipitation-Temperature correlation for USA, negative Precipitation-Temperature correlation for Europe*
 - o *Correlation between Elevation and BFI: Strong positive correlation between BFI and Elevation for the USA, non-significant BFI-Elevation correlation for Europe*
- *Latitude range between the two regions*



We acknowledge that including a combined analysis adds additional insights. Therefore, we also applied the methodology on the whole dataset. In general, we found the following:

- Classifications/Controls that showed similar differences between classes for the two regions also show similar differences for the whole dataset.
- Classifications/Controls that were different for each continent fail to describe the differences in the total dataset (e.g. elevation).

Overall, we think it is worth to present both the combined analysis and the analysis for each region. To accomplish this, we replaced Figure 3 and 5 of the previous manuscript with new Figures (3 and 5 for the different climate classification systems (KG, AI, T<0, PET>P) and Figures 4 and 6 for the different individual controls (P, T, A, E, BFI)) that include the overall analysis.

Comment 2

Authors have decided to select the drought duration as a drought characteristic. In the Discussion (Section 4.2) drought frequency is mentioned as another characteristic, although there is a strong link between average duration and frequency when using the threshold approach. In the discussion the (standardized) deficit volume or intensity should also be addressed. These two characteristics are as important as long duration droughts in their effect on natural and socio-economic systems (lacking water for water resources).

We agree with this comment and the importance of deficit volume is now mentioned in the discussion (4.2). Furthermore, the potential to use this method for other drought characteristics, such as deficit volume, is now mentioned.

Comment 3a

Previous more limited studies (e.g. few catchments, only simulated flow) have shown that both climate and catchment properties control drought duration. This is confirmed by the current paper in the Abstract (pg. 12878, line 20), Discussion (pg. 12891, lines 5-11), Conclusions (pg. 12896,

line 18). However, in the Discussion (pg. 12894, lines 6-8), it is suggested that climate classification systems only can be used to discriminate drought durations. This cannot be concluded based on the selected catchment characteristics.

This message was not intended. We agree that text is misleading and added an extra sentence to the discussion to emphasize that long duration droughts are modified by both climate and catchment controls.

Comment 3b

The BFI shows a substantial control. We know that storage processes are important in the propagation of a drought in a catchment, but the two other selected catchment characteristics (i.e. catchment area, elevation) do not address storage properties.

We agree that elevation and area do not directly or not necessarily describe storage related processes. Nevertheless, they are often used in regionalization approaches as proxies for controls on streamflow dynamics and indeed our study shows that they represent proxies for different controls depending on geographical settings. In some areas, there is a relation between aridity and elevation and in others there is a relation between snow processes and elevation (Salinas, 2009). For a set of Austrian catchments, elevation was found to reflect seasonal storage in snow and glaciers (Van Loon, 2015). Catchment area possibly provides an indication for the amount of storage within the catchment (e.g. Salinas, 2009). We added a few sentences on why elevation and area are included to the introduction.

Comment 3c

If soils or lakes would have been included then likely stronger catchment controls would have been found.

We agree. This was a matter of limitation in data availability. At the beginning of this research, we explored which variables could have been used as an indication for the amount of lake storage. The Gages-II metadata provided the fraction of open water in the basins, however, only little variation in the fraction of open water was found for the considered basins with near-natural streamflow conditions. For the non-considered Gages-II basins (that were not indicated as near-natural), more variation in percentage of open water was found. However, these basins were not necessary free of anthropogenic influences (like reservoir operations) that could have a dominant effect on the streamflow drought duration.

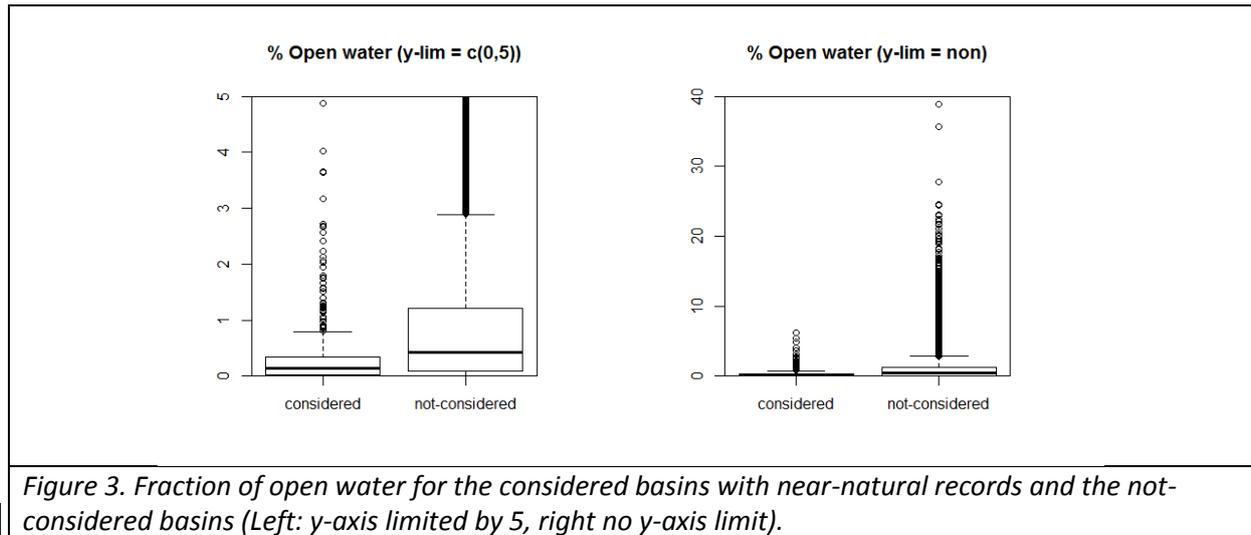


Figure 3. Fraction of open water for the considered basins with near-natural records and the not-considered basins (Left: y-axis limited by 5, right no y-axis limit).

Comment 3d

In the Discussion (Section 4.2) representativity of the selected catchment is discussed. For good reasons only near-natural catchments have been selected (almost no human disturbances), but probable these are biased to headwaters, which have lower storage (steeper topography, thinner soils, less aquifers). For instance, the BFI of 80% of the selected catchments is < 0.7 . I wonder, if the percentage of catchments with a $BFI < 0.7$ would not have been lower, if not only near-natural catchments were selected (headwaters).

Good point. We added a few sentences to the discussion in section 4.2, referring to:

- The potential bias of near-natural catchment towards headwater catchments.
- Possible stronger catchment controls if lakes and soil properties were included.

Comment 3e

In summary, I believe that a catchment classification system that adequately discriminates drought duration should include both climate and catchment controls.

We share this opinion of the referee. In the revised version of the manuscript, we:

- Revisited the objectives and emphasized more that the focus of this study is on climate classification systems (in the introduction)
- Mention the complementary role of individual controls in the introduction
- Discuss the representativeness of near-natural catchments in the Discussion (4.2)
- Discuss a possible stronger catchment control in larger basins with lakes or other soil types in the Discussion.

Comment 4

It is strange that the manuscript makes a difference between climate classification systems (incl. Köppen–Geiger, Aridity Index, number of months with $T < 0$ and number of months with $PET > P$) and individual controls (long-term P, long-term T, Area, Elevation and BFI). I believe it is confusing that climate-related controls (number of months with $T < 0$, number of months with $PET > P$, long-term P and long-term T) are in two different groups. I recommend to make two different groups along other lines, i.e. climate-related controls (incl. Köppen–Geiger, Aridity Index, number of months with $T < 0$, number of months with $PET > P$, long-term P and long-term T) and catchment-related controls (incl. Area, Elevation and BFI).

Our objective was to evaluate existing climate classification systems with predefined classes. For individual controls, we needed another grouping approach due the lack of absolute class boundaries that are globally accepted (described in Section 2.3 of the manuscript). We therefore treat them separately and focus on the evaluation of climate classification systems.

Comment 5

I wonder if climate classification systems, such as Köppen–Geiger, are often used in drought monitoring and early warning systems to stratify regions with similar hydro-climatic drought properties, as mentioned in the Abstract (pg. 12878, lines 1-3) and Conclusions (pg. 12897, line 4). I do not believe that the manuscript needs such mandate. The results on the relationships between drought duration and climate and catchment controls derived from observed flow already justify the paper.

We removed this statement from the abstract, however, we still mention the potential use of these climate classification systems in drought monitoring and early warning systems in the conclusion

Minor comments:

pg. 12878, lines 4-6: I do not believe that what is currently lacking is a large-scale evaluation of the relation between climate and hydrologic drought characteristics. There are a number of papers to which you also refer which deal with this topic. What is missing, is the use of observed flow from many basins rather than simulated data.

We replaced hydrologic drought characteristics with "observed streamflow drought characteristics" in the abstract.

pg. 12880, line 1: "their" can be removed.

Removed

pg. 12882, lines 20-21: Add a reference for "This study focuses on long duration droughts since they most severely affect natural and socio-economical systems."

We revisited this statement. We removed the original statement from the introduction and better justified why we focus on long duration droughts at the end Section 2.2.

pg. 12883, lines 2-3: there is no justification / hypothesis for using the Area (see also pg. 12888, line 29) and the Elevation as catchment characteristics that control drought duration. Add reference(s)

We added a few sentences and references to the introduction on why we hypothesize area and elevation to be a control. (see response to comment 3b.)

pg. 12883, line 15: add how many of the 808 gauging stations are in the USA and how many in Europe.

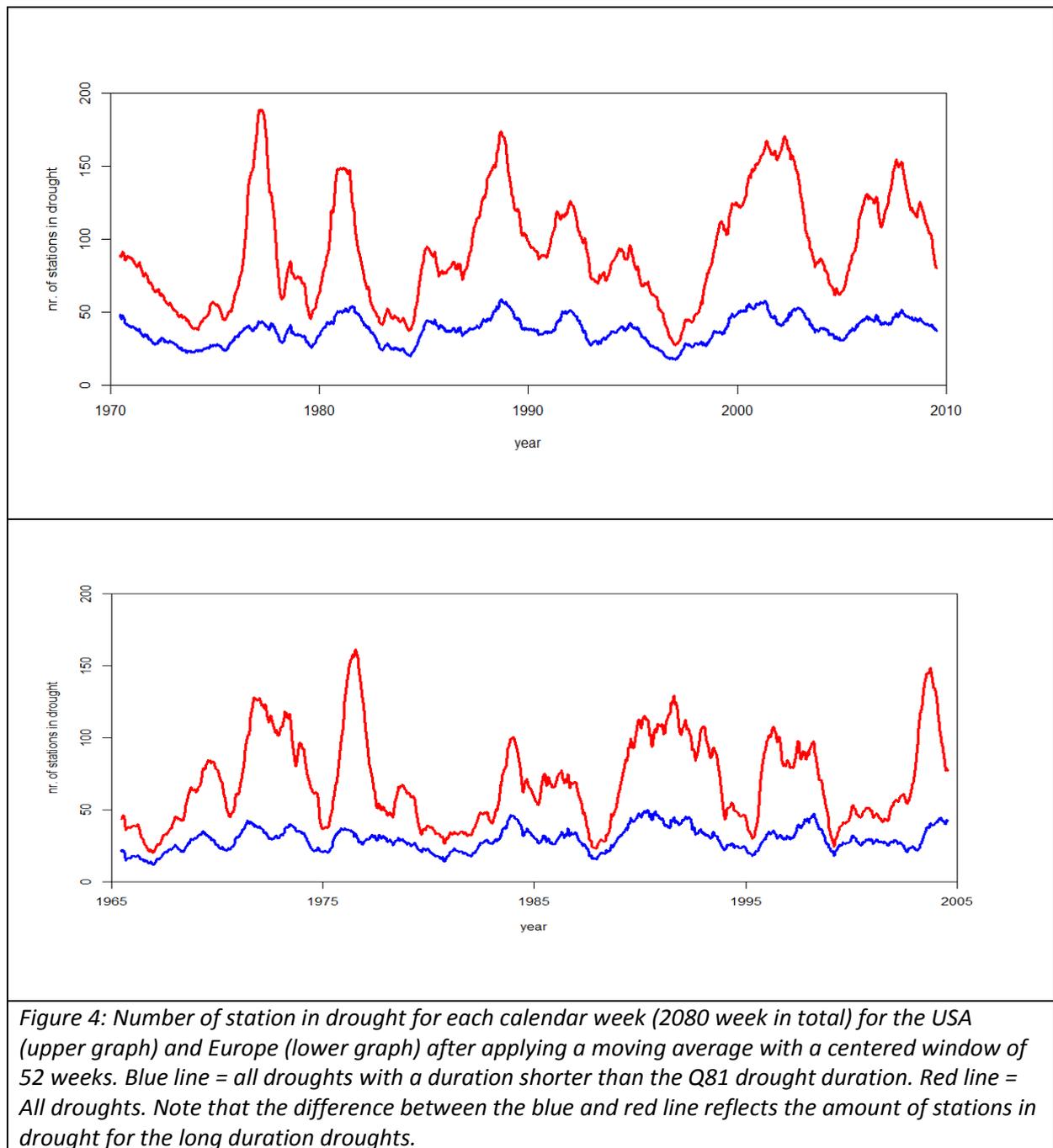
461 for the USA, 347 for Europe. We added this in the revised manuscript (Section 2.1)

pg. 12883, line 19: do you use 40 year of data or for some gauging stations more than 40 year of data?

Always 40 years of data. We clarified this in the revised manuscript (Section 2.1).

pg. 12883, line 20: Elaborate in the Discussion whether there is influence of using different periods (1965–2004 for Europe and 1970–2009 for the USA). Are there more long duration droughts than normal in the periods 1965-1999 and 2005-2009?

Since we compared two separated regions, we did not deem it necessary to reflect similar time periods. We see more value in including a larger number of stations while reflecting recent times. We tested if occurrence of long duration droughts was higher in the non-overlapping time periods (Fig. 4). This figure shows that 2006-2008 had a relatively large proportion of long duration droughts in the USA whereas 1965-1969 did not have notably more long duration droughts in Europe. We did not include this Figure or a literature review about differences in occurrence of major drought events between the two regions because we think it distracts from the main message (also taking into account that the revised manuscript got quite a bit lengthier with all other revisions).



pg. 12883, line 22: “time step” not defined. It becomes clear in following sections that the time step is a week.

We added “weekly time steps” to this Section (2.1). We kept the explanation of how daily data was transformed to weekly in the methods section.

pg. 12884, line 25: transformation from daily to weekly flow is a kind of smoothing. Does this not contradict with the remark in the Discussion (pg. 12895, line 16) that no smoothing has been applied.

Correct, we revised the sentence on pg. 12895, line 15-17 to: "Therefore, procedures that influence this fraction like smoothing of the threshold, pooling of drought events or exclusion of minor drought events were not applied in this study."

pg. 12885, line 15: what type of interpolation (linear, spline, ..?)

Linear interpolation. We added this information to the revised version of the manuscript (Section 2.2).

pg. 12885, line 18: Add a sentence which describes that "long duration droughts" are not defined in an absolute way (minimum number of weeks) but in a relative way (81-100 percentile).

In Section 2.2., we now mention that long duration droughts are defined in a relative way. We did not repeat that they are not defined in an absolute drought.

I recommend to calculate your own Köppen–Geiger class for each gauging station, like it has been done by Wanders (Figure 2, 2015), which makes the KG class consistent with the climate data.

For our analysis we calculated the KG classes according to the method described in Kottek et al. (2006) (pg. 12884, line 9) for each basin based on local meteorological data (individual controls P and T). We clarified this in Section 2.1.

pg. 12885, lines 28 – pg. 12886 (line 6): the procedure is not fully clear. Pg. 12886 (line 3): "equal size", do you mean that each class consist of 20% of all (808) basins?

Yes, as suggested by reviewer 2, we replaced "equal class size" with "equal number of basins" (Section 2.3).

Pg. 12886 (line 5): "class size" do you mean number of basins (there should be 10 or more basins in a class)?

Yes, we replaced class size with number of basins (Section 2.3).

Is the smaller number than 10 caused by the separate investigation of the USA and Europe?

Yes, class boundaries are based on the entire dataset. We clarified this in the revised manuscript (Section 2.3).

pg. 12886 (lines 13-14): meaning of "average". I suggest the following phrasing: ".....of the average DDC per class, we plot them as departures from the overall average to make differences easier...".

We rephrased this in the revised manuscript (Section 2.4)

pg. 12886 (lines 13-14): motivate why the average has been used instead of the median.

We acknowledge the subjective decision to use the average DDC over the median DDC. However, differences in ranks between groups are covered by the statistical analysis. We did not deem it necessary to include this information in the revised version of the manuscript.

pg. 12889 (lines 8-15): It is bit strange to start with “It reveals for the KG that basins in the Cfb climate in the USA have lower average DDC compared to Europe...”. The general impression by looking at Fig. 4 (upper) is that the DDCs for the USA are larger than for Europe. I would start with this finding.

The suggested finding is mentioned in the beginning of Section 3.1.

pg. 12890 (lines 21-29): Figure 4 needs to be split up in two separate graphs. The upper graph is about the visual comparison approach (Section 3.1), whereas the lower graph is about the statistical approach (Section 3.2). In between you describe Figure 5 (pg. 12889, line 16 - pg. 12890, line 20).

As described in comment 1, we will re-arrange the figures. DDC of Fig. 4 will then be combined with the DDC of new Figure 3 and 4 (see example Figure 2 of this response). We also

- *removed Figure 5 in the previous version of the manuscript*
- *created two new figures to present the statistics (new figures 5 and 6)*
 - *one for the different climate classification systems (KG, AI, T<0, PET>P)*
 - *one for the individual controls (P, T, A, E, BFI)*

We realize that the two figures are now more difficult to interpret and therefore provided an exemplary figure in the legend of the newly proposed Figure 5 of the manuscript.

pg. 12893 (line 16): the phrasing “...annual actual evaporation calculated with the Thornthwaite formula...” is incorrect. Thornthwaite provides an estimate of the PET. The actual evapotranspiration that is mentioned by Van der Schrier et al. (2011) is from a simple water balance model that uses Thornthwaite PET.

Thanks for pointing out this mistake. We corrected this in the discussion (Section 4.1).

pg. 12893 (line 27): replace “a suitable” with “suitable”.

Done

pg. 12902 (Table 1, caption): replace “class size” with “number of basins”, or “class size (number of basins)”.

Done

pg. 12902 (Table 1, AI column): replace “90” with “90+”.

Done

pg. 12904 (Figure 2): parts are hard to read, too small.

We enlarged some labels within Figure 2, esp. of panel c2

pg. 12904 (Figure 2, B, left): the x-axis label “USA Europe Region” is confusing. It can be left out.

We left these labels out in Figure 2.

pg. 12904 (Figure 2, C2): Duplication of the x-axis label (-----81-----/ /-----91-----/ /-----100---/) would improve readability. Add x-axis label below the box plots.

We duplicated these labels and place them below the boxplots (Fig. 2).

pg. 12905 (Figure 2, caption): (a), (b) etc. Capital in the graph. Make it consistent.

We changed this in Figure 2

pg. 12905 (Figure 2, caption): replace “....values for basins in both Europe (red) and the USA (blue)...” with “....values for basins in both the USA (blue) and Europe (red) ...”. Use same sequence as in graph.

We changed this in the caption of Figure 2

pg. 12905 (Figure 2, caption): replace “....exemplary ensembles of DDC groups for classes 1, 2 and 3 for the USA...” with “....exemplary ensembles of DDC groups for precipitation classes 1, 2 and 3 for the USA...”

We changed this in the caption of Figure 2

pg. 12907 (Figure 4): needs to be split into two figures, Figure 4 (only upper graph) and new Figure 6 (lower rows). Revise caption, hard to understand.

This information is split in the revised version of the manuscript. Figures are changed and captions are revised / clarified.

pg. 12908 (Figure 5): add set of figures that show the similarities for climate classification systems and individual controls for the USA and Europe (all basins together); see previous major comment.

We added this to the new Figure 5 and 6.

pg. 12909 (Figure 6): add box and whiskers for the USA and Europe (all basins together); see previous major comment.

We added these boxplots in white to the new Figure 7.

pg. 12909 (Figure 6, caption): replace “End of lines: percentiles 5 and 95” with “End of whiskers: percentiles 5 and 95”.

We changed this in the caption of Figure 7.

pg. 12910 (Figure 7, caption): replace “..(left column)..” and “..(right colum)..” with “..(left)..” and “..(right)..”.

We changed this in the caption of Figure 8.

References:

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, *Hydrol. Earth Syst. Sci.*, 17, 2637-2652, doi:10.5194/hess-17-2637-2013, 2013.

Van Loon, A. F., and G. Laaha. "Hydrological drought severity explained by climate and catchment characteristics, *J. Hydrol.*, 526 ,3-14, doi:10.1016/j.jhydrol.2014.10.059, 2015.

Manuscript prepared for Hydrol. Earth Syst. Sci. Discuss.
with version 2015/04/24 7.83 Copernicus papers of the L^AT_EX class copernicus.cls.
Date: 22 May 2016

Controls on hydrologic drought duration in near-natural streamflow in Europe and the USA

E. Tijdeman, S. Bachmair, and K. Stahl

Faculty of Environment and Natural Resources, University of Freiburg, Fahrenbergplatz, 79098 Freiburg, Germany

Correspondence to: E. Tijdeman (erik.tijdeman@hydrology.uni-freiburg.de)

Abstract

Climate classification systems, such as Köppen–Geiger and the aridity index, are often used in large-scale drought modeling studies and in drought monitoring and early warning systems to stratify regions with similar hydro-climatic drought properties. What is currently lacking is a large-scale evaluation of the relation between climate and hydrologic-observed streamflow drought characteristics. In this study we explored how suitable common climate classifications are for differentiating river basins according to their characteristic hydrologic drought duration and whether drought durations within the same climate classes are comparable between different regions. This study uses a dataset of 808 near-natural streamflow records from Europe and the USA to answer these questions. First, we grouped drought duration distributions of each record over different classes of climate classification systems and individual climate and catchment controls. Then, we compared these drought duration distributions of all classes within each climate classification system or classification based on individual controls. Results showed that climate classification systems that include absolute precipitation in their classification scheme (e.g., the aridity index) are most suitable to differentiate basins according to drought duration within both the USA and Europe. However, differences in duration distributions were found for the same climate classes in Europe and the USA. These differences are likely caused by differences in precipitation, in catchment controls as expressed by the base flow index and in differences in climate beyond the total water balance (e.g., seasonality in precipitation), which have shown to exert a control on drought duration as well. Climate classification systems that include an absolute precipitation control can be tailored into drought monitoring and early warning systems for Europe and the USA to define regions with different sensitivities to hydrologic droughts, which, for example, have been found to be higher in basins with a low aridity index. However, stratification of basins according to these climate classification systems is likely to be complemented with information of other climate classification systems (Köppen–Geiger) and individual climate and catchment controls (precipitation and the base flow index), especially in a comparative study between Europe and the USA.

1 Introduction

Droughts are natural disasters that originate from a temporary deficit of water. They are multifaceted phenomena and are often grouped into four main types; meteorological, agricultural, hydrologic and socio-economic. Hydrologic drought relates to “effects of dry spells on surface and subsurface water” (Wilhite and Glantz, 1985). These-In the absence of human influences, hydrologic droughts are often triggered by anomalies in climatic conditions and their-. Their duration regularly depends on the persistence and magnitude of these anomalies and on seasonal transitions, such as a shift from the rain to snow season or a shift from the wet to dry season (Van Loon and Van Lanen, 2012). However, climatic conditions alone do not determine the onset, persistence and recovery of a hydrologic drought. Storage related processes (like snow accumulation or groundwater storage) play an important role as well (e.g., Haslinger et al., 2014; Staudinger et al., 2014; Van Loon and Laaha, 2015).

Knowledge of a region’s hydro-climate is important for drought related research (Tallaksen and Van Lanen, 2004), e.g., short term precipitation deficits can lead to a hydrologic drought event in a basin with little storage whereas a basin with a lot of storage is likely to be little affected by such a dry spell. The Köppen–Geiger climate classification system (Geiger, 1961) is a popular way to describe a region’s (hydro-)climate in a broad range of disciplines (Rubel and Kottek, 2011). However, it may not be the most optimal way of grouping basins with similar hydrologic behavior, partly because it fails to distinguish between basins with different “filtering behaviors” (Coopersmith et al., 2012). More recent hydro-climatic classification schemes build on the ideas of the Köppen–Geiger climate classification system. For the USA, such classification schemes are based on controls like seasonality and timing of precipitation, the aridity index, timing of maximum runoff and fraction of precipitation falling as snow (e.g., Berghuijs et al., 2014; Coopersmith et al., 2012). The latter two studies suggest that in the USA, climate is the dominant control on hydrologic behavior, however, Berghuijs et al. (2014) also found similarity between their clusters of basins and soil, ecosystem and vegetation classes.

Apart from climatic controls, catchment controls also play a role in the propagation from climatic input to streamflow (e.g., Barker et al., 2015; Haslinger et al., 2014) and could thus be useful to group basins with similar hydrologic behavior. For example, variability in precipitation and temperature is dampened when it propagates to streamflow (Gudmundsson et al., 2011b). The latter study suggests that this is related to physical catchment characteristics. Gudmundsson et al. (2011a) found support for stronger control of physical catchment characteristics during situations of low flow, which was shown by reduced cross-correlation of low vs. high flows.

In order to improve our understanding of these climatic and catchment controls on hydrologic droughts, the drought characteristics of interest need to be quantified. Commonly, hydrologic droughts are characterized by duration, deficit volume, frequency and areal extent (Andreadis et al., 2005). Quantifying these properties helps to compare historical drought events and can be used to place current and predicted drought events in a historical context. One method to compare these characteristics is by Severity Area Deficit (SAD) curves, which have been used to compare major soil moisture and runoff drought events in the USA (Andreadis et al., 2005) and major soil moisture drought events on a global scale (Sheffield et al., 2009). Knowledge about past drought characteristics can further be used to create probabilistic return periods of hydrologic drought events with certain characteristics, using so-called Severity Area Frequency (SAF) curves (e.g., Hisdal and Tallaksen, 2003). Furthermore, these drought characteristics have been utilized to study the propagation of drought through the hydrologic cycle (e.g., Tallaksen et al., 2009; Van Loon et al., 2014) and to investigate the impact of climatic and catchment controls on droughts (e.g., Van Lanen et al., 2013; Van Loon et al., 2014).

Climate related differences in modeled drought characteristics were found between the major classes of the Köppen–Geiger climate classification system, where droughts in snow, polar and arid climates have longer durations compared to the equatorial and temperate climates (Van Lanen et al., 2013). The different major classes of the Köppen–Geiger classification can be further divided into different sub-classes that take into account seasonality in precipitation and the occurrence of cold or hot seasons (Kotteck et al., 2006).

Van Loon et al. (2014) found that for these sub-climates, droughts with long durations occurred more often within classes with seasonal properties. Droughts starting before annual recurring periods of low precipitation or high or low temperature are less likely to recover due to either a low influx of precipitation, temporary storage of precipitation as snow or a high level of evaporation (Van Loon and Van Lanen, 2012). Climate classification systems, like the Köppen–Geiger climate classification, are based on long term average climatic conditions. However, drought durations are modified when meteorological droughts propagate through the hydrologic cycle. For example, drought duration increases with an increasing groundwater response time (Van Lanen et al., 2013; Van Loon et al., 2014). Both these studies showed that this drought prolonging effect was visible for different climates, suggesting a combined influence of both climatic and catchment controls on drought duration where neither climate nor physical catchment structure seemed to be dominant.

Studies based on modeled basins may lead to a better theoretical understanding of controls on hydrologic droughts since they enable isolated research on the effect of one control at a time. However, modeling incorporates uncertainties, e.g., in climatic forcing and due to modeling assumptions (Sheffield et al., 2009). It is therefore questionable how representative models are of the real world. This highlights the importance of using observed streamflow data in research about controls on hydrologic droughts. However, outside the modeling environment, a comparative study on the isolated effect of one individual control is nearly impossible due to the unique combination of catchment and climate properties of each real-world basin. For example, in Austria, propagation of drought (from precipitation to streamflow) was found to be more dependent on climatic forcing under humid conditions and on storage properties under more arid conditions (Haslinger et al., 2014). Therefore, research about controls on observed hydrologic drought durations is limited to finding the dominant ones. Tallaksen and Hisdal (1997) showed for a set of 52 Nordic basins that the distribution of drought durations is variable over different basins, which they hypothesized to be controlled by climate. In contrast, Van Loon and Laaha (2015) showed that storage related processes mainly control the duration of drought for

a set of Austrian catchments. They showed that the base flow index (BFI, representing several different storage related processes), has the highest correlation with average streamflow drought duration, ~~however, annual precipitation showed a strong negative correlation with average drought duration as well.~~ Elevation is another catchment control that is hypothesized to exert a control on streamflow droughts since it can be related to seasonal snow storage (Van Loon and Laaha, 2015). However, the influence of elevation might not be uniform around the world due to differences in geographical settings. For example, in some areas, there is a relation between aridity and elevation and in others there is a relation between snow processes and elevation (Salinas et al., 2013). Catchment area is negatively correlated with the variance in catchment runoff (Skøien et al., 2003). It is therefore hypothesized that low flow conditions are generally more persistent in larger catchments, although the latter study also found proof that the temporal smoothing of catchment runoff when it propagates from precipitation is mainly attributed to runoff generating processes. Catchment area also showed a positive correlation with mean drought duration, although it was not the most dominant catchment control (Van Loon and Laaha, 2015).

To extend the knowledge about controls on ~~hydrologic~~ streamflow droughts and to evaluate the suitability of climate classification systems for describing regions with different hydrologic drought characteristics, large scale studies are needed based on observed streamflow data. Therefore, we evaluated the suitability of several climate classification systems ~~and individual controls~~ to differentiate basins according to hydrologic drought duration in near-natural streamflow records from Europe and the USA. Furthermore, we tested if drought duration distributions of the same climate classes were comparable between the USA and Europe, which answers the question whether or not climate classification systems are transferable between these regions. ~~For this analysis, we~~ A similar analysis was done for different individual climate and catchment controls. However, these controls do not have commonly accepted grouping approaches, i.e., we needed another (more arbitrary) grouping approach for these individual controls. Therefore, individual controls are complementary in the interpretation of the suitability of different climate classification systems to differentiate basins according to drought duration. For both

analyses, we used a hypothesis testing approach to systematically compare cumulative drought duration distributions (hereafter called drought duration curves) between classes of different climate classification systems and classes of individual controls. ~~This study focuses on long duration droughts since they most severely affect natural and socio-economical systems.~~ Duration is preferred over other drought characteristics like severity or magnitude since this characteristic is less influenced by systematic measurement errors and relies on ranks of data rather than on accurate gauged quantities.

Based on the above mentioned studies, we hypothesize that the following climate or catchment characteristics exert a control on drought duration:

- Occurrence and length of a precipitation deficit season
- Occurrence and length of a cold season
- Climatic controls (precipitation (P) and temperature (T))
- Catchment controls (base flow index (BFI), area (A) and elevation (E)).

The following climate classification systems are ~~also~~ therefore hypothesized to be suitable for differentiating basins with different hydrologic drought duration characteristics since they include one or more of these controls: The Köppen–Geiger climate classification system (KG), the aridity index (AI), the number of months with an average temperature below zero ($T < 0$) and the number of months with a climatic water deficit, i.e., when the average potential evaporation is larger than the average precipitation ($PET E_{pot} > P$). However, none of these climate classification systems considers catchment controls so their suitability to differentiate basins according to drought duration ~~needs to be investigated for a~~ in observed streamflow was investigated in this study under a wide variability of catchment characteristics.

2 Data and methods

2.1 Streamflow data and potential controls

The analysis was based on 808 near-natural streamflow records from Europe ([n=347](#)) and the contiguous USA ([n=461](#)). The streamflow records for the USA were selected from the Hydro-Climatic Data Network (HCDN-2009, Lins, 2012) and for Europe from the European Water Archive (EWA, Stahl et al., 2010). Only records meeting the following criteria were selected for further analysis: (1) ~~at least~~ 40 years of continuous daily data for the time period 1965–2004 for Europe and 1970–2009 for the USA. Different time periods were chosen to optimize the number of stations while incorporating recent times. (2) Percentage of zero streamflow occurrence at each [weekly](#) time step is ≤ 20 , since the chosen drought identification method was not designed to deal with more frequently occurring zero streamflow.

Individual controls were assembled from various sources for both regions. Climatic (annual and monthly P and T) and topographic (mean E and A) controls were obtained for the USA from the GAGES-II dataset (Falcone, 2011). For Europe, climatic controls were obtained from the E-OBS dataset (Haylock et al., 2008) and topographic controls originate from the pan-European River and Catchment Database CCM2 (Vogt et al., 2007). The BFI was calculated from daily streamflow records based on the calculation procedure described in Gustard and Demuth (2009). Different climate classification systems were ~~determined~~ [calculated from the individual climatic controls](#) as follows:

- KG: according to the method of Kottek et al. (2006).
- AI: following the method of de Martonne (1926) (P divided by $(T + 10)$) with a grouping interval of 10 (similar to the map presented at the FAO website; Grieser et al., 2006).
- $T < 0$: sum of months with average T below zero.
- $E_{\text{POT}} > P$: sum of months with average E_{POT} (calculated following the method of Thornthwaite, 1948) above the average P .

The KG classification system classifies basins with 2 or 3 letter codes. For the considered regions, distinctions are made based on the minimum of the average monthly temperature (first letter C for a minimum temperature $> 3^{\circ}\text{C}$ and D for minimum temperature $\leq 3^{\circ}\text{C}$), seasonality in precipitation (second letter f for precipitation all year round and s for a relatively low amount of precipitation in summer) and summer temperatures (third letter a stands for hot summers, b for warm summers and c cool summers). Figure 1 shows the locations of the selected basins and their classification according to the KG and AI climate classification systems.

2.2 Drought duration curves

The goal of this step is to extract drought durations distributions from the streamflow records. Daily streamflow records were transformed to weekly data (sum of total streamflow volume per week). Defining droughts at this temporal resolution is in line with other studies (e.g., Tallaksen and Stahl, 2014) and with the US drought monitor classification scheme (Svoboda et al., 2002). Hydrologic drought events were identified from these weekly records using the threshold level approach following the principals of Zelenhasić and Salvai (1987); a drought event starts when the streamflow record is at or below a certain threshold level and ends when this record passes the threshold again. The threshold level used in this study was the 20th percentile of streamflow, which was calculated for each week. This is a common threshold used in various other large scale drought studies (e.g., Andreadis et al., 2005; Tallaksen and Stahl, 2014; Van Lanen et al., 2013; Van Loon et al., 2014). Drought durations, defined as the sum of weeks the streamflow record is continuously at or below the threshold, were extracted for each record. Similar to flow duration curves, these weekly values of drought durations were sorted from shortest to longest. For each drought duration, the fraction of non-exceedance was calculated. The resulting drought duration curves were calculated by [linear](#) interpolation of these cumulative drought duration distributions in such a way that each percentile (ranging from 1 to 100) has a value. [The As an example, the](#) drought duration curves of all basins (or drought duration curve [ensembleensembles](#)) for the USA and Europe ~~is presented in Fig.~~ [are presented in](#)

Figure 2a. In this study we only take into account long duration droughts ~~;~~ ~~we~~ which are defined in a relative way. Reasons to only focus on long duration droughts are related to the hypothesis that these droughts affect natural and socio-economical systems more severely. Furthermore, drought duration curves are more different from each other after the 81st percentile (Fig. 2a). We hence only consider the drought duration curves between the 81st and 100th percentile for further analysis. For simplicity, we hereafter use the term drought duration curves when referring to drought duration curves between the 81st and 100th percentile.

2.3 Grouping drought duration curves

To test whether drought duration curves ~~significantly~~ differ between classes of different climate classification systems and individual controls we grouped them accordingly. For the climate classification systems this means that drought duration curves were grouped according to the predefined classes. Since no such straightforward classification systems exist for the selected individual controls, we had to use another approach. In a first step, we combined all values of an individual control of both the USA and Europe (e.g., annual precipitation) and divided these values into 5 classes ~~of equal size~~ with an equal number of basins (Fig. 2b, left). In a second step, these classes were used to group the drought duration curves into 5 different ensembles for ~~each region~~ the entire dataset and 5 different ensembles for the two regional subsets (Fig. 2b, right; only three classes of the USA are shown in this example). The minimum ~~class size~~ number of basins in a class was set to 10 for both classes of climate classification systems and individual controls. ~~Smaller classes~~ Classes of the two regional subsets with less basins were excluded from the analysis. An overview of all remaining classes of drought duration curves (abbreviated to DDC when referring to subsets ~~of drought duration curves~~) with corresponding ~~class sizes~~ number of basins in each class is presented in Table 1. Class ranges can be found in the results section of this study.

2.4 Comparing DDC

DDC of the different classes were compared with each other both visually and statistically. For visual comparison, the DDC ensemble average per class (e.g., per KG class) was calculated. Instead of showing the absolute values of the average DDC per class, we plot them as departures from the overall average to make differences easier to discern (Fig. 2c1).

For the statistical analysis, we systematically compared, for each climate classification system or individual control, the DDC values of each class at each percentile between 81 and 100 with all other classes (boxplots Fig. 2c2). This percentile based comparison was preferred over a statistical comparison of average DDC ensembles because the latter does not take into account the variability in DDC ensembles at the different percentiles (Fig. 2a). Two different non-parametric tests were used for this statistical comparison. (1) The Kolmogorov–Smirnov test (KS, Wilks, 2011), which is sensitive to differences in shape, spread and median of distributions (H_0 : DDC values of two classes at percentile i follow a similar distribution) (2) the Mann–Whitney U test (MWU, Wilks, 2011), which is sensitive to differences in mean ranks (H_0 : mean ranks of DDC values of two classes at percentile i are similar). Non-parametric tests were used since different groups of DDC values were not always normally distributed. As final measure of statistical similarity in DDC of the different classes we use-used the number of percentiles with non-significant differences ($P \geq 0.05$) according to either the KS or MWU test (Eqs. 1 and 2).

$$S_{KS} = \sum_{i=81}^{100} \begin{cases} 0 & \text{if } P_{KS,i} < 0.05 \\ 1 & \text{if } P_{KS,i} \geq 0.05 \end{cases} \quad (1)$$

$$S_{MWU} = \sum_{i=81}^{100} \begin{cases} 0 & \text{if } P_{MWU,i} < 0.05 \\ 1 & \text{if } P_{MWU,i} \geq 0.05 \end{cases} \quad (2)$$

where S_{KS} and S_{MWU} are the number of similar percentiles ranging between 0 and 20 (0 = 0 percentiles similar and 20 = all percentiles similar) and $P_{KS,i}$ and $P_{MWU,i}$ are the

P values of the two tests at percentile i (Fig. 2c2). A high value of S_{KS} and S_{MWU} thus indicates more similarity between the DDC of two classes. In addition to the comparison of DDC between all classes of each climate classification system ~~of each region and individual control of both the entire dataset and the two regional subsets~~, DDC of the same climate ~~classification~~ classes were compared between Europe and the USA (e.g., DDC of KG class Cfb in the USA vs. DDC of the same class in Europe). For the visual comparison, the difference in average DDC of the same classes between the USA and Europe was used (average DDC USA minus average DDC Europe). For statistical comparison, number of percentiles with similar DDC values between classes with the same classification (according to both S_{KS} and S_{MWU}) was again used as a measure of statistical similarity between DDC.

3 Results

3.1 Visual comparison of DDC

Figure 3 (~~left two columns~~) presents average DDC (for long duration droughts) of all classes of different climate classification systems. ~~The Kppen–Geiger climate classification system (KG) in the USA show lowest average DDC for basins in the non-seasonal temperate climate with warm summers (Cfb), followed by the~~ In general, the patterns displayed for the entire dataset and for the two regional subsets (USA and Europe) are comparable. However, average DDC of basins from the same climate classes in the USA are mostly higher, i.e., biased towards longer drought durations (average DDC of the USA minus average DDC of basins in Europe is mostly positive (Fig. 3, right column)).

The KG reveals lowest average DDC for basins in the non-seasonal snow climate with warm summers (Dfb) temperate and snow climates (Cfc, Cfb and Dfb) for both the entire dataset and the two regional subsets of the USA and Europe. Higher average DDC are displayed for basins in the hot summer, cold and seasonal climates (Cfa, Dfa, Csb, Dfc, Dsb, Dsc). ~~For Europe, highest average DDC are visible for basins in the cold winter climate with cool summers (Dfc) and lowest average DDC for~~ Basins in the Dfc and Dfb climate of

the USA have higher average DDC compared to Europe, whereas average DDC of basins in the temperate climate with cool summers (Cfc). The aridity index (AI) Cfb climate in Europe are higher. The AI shows highest average DDC for basins in the lowest (most arid) AI classes for both regions. Generally, the average DDC decreases average DDC decrease with increasing AI classes, apart from an occasional exchange between some of the neighboring classes. For the number of months with an average temperature below zero ($T < 0$), Average DDC are higher for basins in the same AI classes in the USA (USA minus Europe is positive), especially for basins in the lower AI classes. For $T < 0$, average DDC are generally highest for basins with most months $T < 0$, intermediate for basins that have least months $T < 0$ and lowest for basins that have 3 or 4 months $T < 0$. This ordering of DDC was found for both regions the entire dataset and the two regional subsets, however, differences in average DDC between classes are small compared to the differences in average DDC between classes of other climate classification systems. Number of months with average potential evaporation larger than the precipitation ($PET E_{POT} > P$) displays an ordering of average DDC with a general pattern of higher average DDC for the basins with a high number of months $PET E_{POT} > P$ and lower average DDC for basins with a low number of months $PET E_{POT} > P$. Similar to the ordering of average DDC of the AI, the systematic ordering of average DDC (from high for low classes too low for basins in low classes to low for basins in high classes of $PET E_{POT} > P$) is occasionally interrupted due to an exchange between average DDC of basins in neighboring classes. Basins in lower classes of $E_{POT} > P$ are comparable between the two regions whereas basins in classes with more months $E_{POT} > P$ show distinct higher average DDC for the USA.

Figure 3 (right two columns) 4 presents the average DDC of basins grouped by individual controls. For individual control precipitation (P) of both regions Average DDC of basins in the same classes are again most of the times higher for the USA compared to Europe. However, in contrast to climate classification systems, not all individual controls exert a similar control on drought duration in both regions.

For the individual control P of both the entire dataset and two regional subsets (USA and Europe), the class of basins with the highest average DDC is the class with the lowest P and

Discussion Paper | Discussion Paper | Discussion Paper

vice versa. Average DDC decrease from lowest to highest P class. Classes of ~~temperature (T) in both regions~~ T show highest average DDC for ~~basins in~~ both the lowest and highest ~~temperature~~ T class. Longer drought events are thus found for basins with temperatures from the tails of the temperature distribution. However, differences in average DDC ~~between~~ ~~of basins in~~ different classes of T are not as distinct as for precipitation classes. Even smaller differences in average DDC are found for ~~area (A)~~ ~~basins in the different classes~~ ~~of A~~ . In Europe, small basins display lowest average DDC, and large basins ~~the~~ highest average DDC. This is different in the USA, where both small and large basins exhibit the highest average DDC. Similar to A , ~~elevation (E)~~ E shows differences in ordering of average DDC between the two regions. For the USA, highest average DDC are displayed for ~~basins in~~ the highest E class whereas the highest average DDC of Europe are displayed for ~~basins in~~ the lowest E class. ~~For the base flow index (BFI) in both regions~~ ~~These distinct differences are averaged out for the entire dataset.~~ ~~For the BFI~~; high BFI coincides with higher average DDC and low BFI with lower average DDC.

~~Figure 4 (upper row) displays the average DDC of the same climate classification system classes of the~~

3.2 Statistical comparison

~~Figure 5 shows the measures of statistical similarity (S_{KS} and S_{MWU}) between ensembles of DDC for basins in different climate classes. Patterns are again most of the time comparable between the entire dataset and the two regional subsets (USA and Europe. It reveals for the KG that). Differences occur for some specific combinations (e.g. DDC of basins in the Gfb climate in the USA have lower average DDC compared to Europe whereas the average DDC Dfc climate are comparable with DDC of basins in the Dfb and Dfc climate are higher for the USA. For the AI, Dsb climate within the USA according to S_{KS} , however, DDC of basins in these two climates are not comparable according to the same measure of similarity for the entire dataset where the DDC in the USA show higher averages for all classes compared to Europe, especially for the lower AI classes. Average DDC for $T < 0$ are generally higher for the same classes in the USA as well. For $PET > P$, low classes~~

display similar average DDC whereas the USA has higher average DDC for classes with more months $PET > P$.

3.3 Statistical comparison

Figure 5 (left two columns) shows the results of of basins in the Dfc climate of the USA are combined with the statistical comparison of DDC for the different climate classification systems (lower DDC of basins in the European Dfc climate).

For the KG, DDC of basins in the Cfc climate have significantly lower DDC values at most percentiles compared to all other climates. DDC of the USA, DDC of basins in the Cfb climate are only similar with DDC of basins in the Dfb climate according to both S_{KS} and S_{MWU} . DDC of rivers-basins in this Dfb climate show little similarity with DDC of the other climates (basins in the other, seasonally influenced, climates again) indicating the distinction between shorter droughts for the non-seasonal warm summer basins in climates affected by no or small seasonal influences (Cfc, Cfb and Dfb climates) and longer droughts for basins in the other climates. However, DDC of basins in these other climates (Cfa, Dfa, Csb, Dfc, Dsb) mostly do not differ significantly show notable differences among each other according to both tests. From measures of statistical similarity. Out of these climates, basins in the Dsb climate, which has reveal the highest average DDC, is the most different also have the most distinctive DDC and only shows similarity in DDC with the Dfc and basins in the Dsc climate (and at some percentiles with basins in the Csb and Dfa climates) : For Europe, DDC of the Cfb, Dfb and to a lesser extent Dfc climate are similar to each other. The Cfc climate has significantly lower DDC compared to these other climates for the entire dataset and with the Dfc climate for the regional subset of the USA. Regarding the differences between the USA and Europe, basins in the Dfb and Cfb climate have similar DDC between the two regions according to both S_{KS} and S_{MWU} : (presented in the diagonal of the matrices in the right two columns of Figure 5). Basins of the Dfc climate of the USA show significantly higher DDC values for most percentiles. The differences in DDC between of basins in different AI classes is most distinct for are most distinct between the lowest AI classes. The higher the AI class, the more neighboring classes of basins show similarity

in DDC, whereas for ~~these basins in the~~ lower AI classes, only ~~direct neighbors~~ DDC of basins in direct neighboring classes occasionally show similarity. For the comparison between Europe and the USA, the lower AI classes (< 50) shows basins with higher DDC in the USA according to both measures of similarity, whereas basins of higher AI classes did not show many notable differences between the two regions. The small differences in average DDC ~~between of basins in~~ different classes of $T < 0$ ~~is-are also~~ reflected by the corresponding ~~statistics~~ measures of statistical similarity, especially for Europe. For this region, DDC of basins in almost all classes are similar to each other. Basins in the same classes of $T < 0$ are mostly comparable between the USA and Europe. Differences in DDC ~~between for basins in~~ different classes of $PET_{E_{POT}} > P$ are ~~mostly significant~~ notable. S_{KS} and S_{MWU} indicate similarity only in DDC ~~between of basins in~~ neighboring classes. Differences between the USA and Europe are only found for the DDC of basins in the two highest classes of $E_{POT} > P$. For the other classes the DDC are similar.

Figure ~~5 (right two columns)~~ 6 displays the statistical comparison of DDC grouped by individual controls. ~~Average DDC of DDC of basins in~~ different classes of P are mostly ~~significantly~~ different from each other. ~~T displays a high number of similar DDC classes in Europe. For the USA, according to both S_{KS} and S_{MWU} . Classes 3 and 5 (higher P) are comparable between the two regional subsets whereas classes 1 and 2 (lower P) have higher DDC for basins in the USA according to both measures of similarity.~~ DDC of basins of intermediate T classes are similar to each other as well as DDC of basins of the lowest and highest temperature classes for the entire dataset and for the regional subset of the USA, confirming that long duration droughts ~~in this region are significantly are~~ longer in both colder and warmer basins. ~~Area (A) hardly shows significant differences~~ These differences are less distinct for Europe; both S_{KS} and S_{MWU} indicate a high number of similar DDC classes. Differences in DDC between classes for both Europe and the USA. Basins of the are found for classes of basins with a lower T . Basins grouped by A hardly show differences in DDC. Only for the entire dataset, the largest basins have different DDC. According to both S_{KS} and S_{MWU} , basins in the highest E class for of the USA have significantly higher DDC compared to DDC of basins in the other E classes of this region, whereas for Europe, basins

~~of~~ in the lowest E class have ~~significantly~~ higher DDC. The patterns of statistical similarity specific for the two regional subsets are not found for the entire dataset. For the BFI, DDC of basins in different classes are often ~~significantly~~ different from each other according to both measures of statistical similarity besides some similarity between neighboring classes.

~~Figure 4 (lower rows) presents the statistical comparison of DDC between Europe and the USA. Basins in the Dfb and Cfb climate have similar DDC between the regions according to both S_{KS} and S_{MWU} . Basins of the Dfc climate of the USA show significantly higher DDC for most percentiles. The lower AI classes (< 50) consist of basins with significantly higher DDC in the USA for most percentiles, whereas basins of higher AI classes did not show many significant differences between the two regions. Differences in DDC between classes for $T < 0$ are in most cases not significant. For $PET > P$, DDC of basins of the highest classes are significantly different from each other between the two regions.~~

4 Discussion

4.1 Evaluation of climate classification systems

Different climate classification systems and individual controls were evaluated for their suitability to differentiate basins according to long duration droughts in observed streamflow in Europe and the USA. From the individual controls, precipitation (P) and the base flow index (BFI) were most suitable to differentiate basins according to their characteristic drought duration distribution, which is in line with the results found in Barker et al. (2015) and Van Loon and Laaha (2015). These individual controls could therefore be seen as dominant control on the drought duration, which confirms the findings of Van Lanen et al. (2013) and Van Loon et al. (2014) that drought duration is modified by both catchment ~~and climate~~ (groundwater response time) and climate (seasonality in precipitation and the occurrence of hot or cold seasons) controls. Our ~~result results~~ also fit with findings by Zaidman et al. (2002), who found that the 1976 drought in Europe was more persistent in regions with a high BFI or low P . ~~These dominant~~ The distributions of dominant individual controls,

however, are not ~~the same always comparable~~ between the classes of different climate classification systems (~~Fig. 6~~), ~~which in the end affects~~, as can be seen in the boxplots of Figure 7. In the end, these differences in dominant individual controls over different classes of climate classification systems affect their overall suitability to differentiate basins according to drought duration ~~in observed streamflow~~. Furthermore, it partly explains why DDC of basins in the same climate classes are not always comparable between the two regional subsets (USA and Europe).

For the KG climate classification system ~~in the USA, the only climate that was~~, basins that were located in the two climates that were not influenced by seasonality in precipitation nor the occurrence of a cold or hot season, Cfb and Cfc, show the lowest average DDC (shortest droughts) ~~and was~~. According to the two measures of similarity used in this study, basins in the Cfc climate (generally wetter than most other climates (Fig. 7)) were distinctly different from DDC of basins in the other climates and the basins in Cfb climate were only comparable with DDC of basins in the Dfb climate. ~~This Dfb climate was~~ Basins in this Dfb climate were expected to have longer drought durations due to the occurrence of a cold season causing low streamflow due to temporary snow storage (Van Loon et al., 2014). Our tests show that although this influence is visible in the average DDC, it is not often statistically significant when comparing ~~the percentiles of the DDC~~. DDC values at the different considered percentiles. Further notable was the difference in average DDC for basins in the Cfb climate between Europe and the USA. This was the only combination of climate classes where average DDC of basins in Europe were distinctively higher, possibly explained by wetter condition in the Cfb climate for the basins in the USA (Fig. 7). Basins in the Dfb climate, on the other hand, have higher average DDC for the USA compared to Europe, which is likely related to differences in dominant climate and catchment controls between the two regional subsets (lower P and higher BFI for basins in the USA (Fig. 7)).

~~The hot~~ Hot summer climates without seasonality in precipitation (Cfa, Dfa) ~~have consist~~ of basins with higher average DDC ~~than their warm summer variations~~ compared to the DDC of basins with warm summer climates (Cfb, Dfb), which is in contrast with Tijdeman et al. (2012). This difference could possibly be attributed to the fact that the study by

Tijdeman et al. (2012) is based on global data whereas this study only deals with basins in the Dfa and Cfa in the USA. The differences in P between the hot and warm summer climates (Fig. 67) in the USA (Cfa and Dfa have lower P values) may not reflect those on a global scale. Other reasons might be related to modeling assumptions needed in large scale gridded models. Nevertheless, results of this study indicate that the occurrence of a hot summer is an important control on long duration droughts as well. ~~Within the USA, basins of Measures of statistical similarity show little differences between DDC of basins in the hot summer climates show above average DDC. However, their DDC show similarity with the DDC and DDC of basins in the cold and seasonal climates, which makes KG a less suitable climate classification system to differentiate basins with different drought duration characteristics other seasonal climates (Csb, Dfc, Dsb, Dsc). Results thus indicate that the KG is mainly suitable to make the distinction between basins in climates with and without seasonal influences.~~

~~KG climates~~ Basins in the KG climate classes that showed highest average DDC were basins in the snow climates with cool winters or seasonality in precipitation (Dfc, Dsb and Dsc), which matches finding by ~~Tijdeman et al. (2012); Van Lanen et al. (2013)~~ Tijdeman et al. (2012), Van Lanen et al. (2013) and Van Loon et al. (2014). Therefore, a climate classification system that specifically aims to reflect the length of the cold season (months with an average temperature below zero ($T < 0$)) was expected to be suitable to differentiate basins according to drought duration. However, this was not the case and differences between average DDC were small and ~~often not statistically significant~~ the measures of statistical similarity did not indicate strong differences between classes of basins, especially for Europe. These European basins with most months of $T < 0$ are partly located in Scandinavia and the Alps, which have been related to short drought durations before (Hannaford et al., 2011). Altogether, a climate classification system that only includes cold season dynamics while ignoring other drought prolonging processes (e.g., total amount and seasonality in precipitation or the occurrence of hot summers) is not the most suitable to differentiate basins with different drought duration characteristics.

More suitable for such a differentiation of basins are the climate classification systems that take into account the dominant annual precipitation control (months with average potential evaporation larger than the precipitation ($PET_{E_{POT}} > P$) and the aridity index (AI); note that the KG does not have such an annual precipitation term). $PET_{E_{POT}} > P$ does not only take into account the total precipitation, it is also influenced by seasonality in precipitation and the occurrence of hot summer temperatures. This climate classification system shows a sorting of average DDC over the different classes of $PET_{E_{POT}} > P$ that followed the hypothesized pattern of higher DDC for the higher $PET_{E_{POT}} > P$ classes and lower DDC for the lower $PET_{E_{POT}} > P$ classes, which makes it a suitable climate classification system to differentiate basins according to drought duration for both regions. The same classes for Europe and the USA show similarity in DDC for basins located in the lower $PET_{E_{POT}} > P$ classes, however, basins located in the higher $PET_{E_{POT}} > P$ classes show significantly higher DDC values at most percentiles for the USA. One possible explanation could be the difference in distribution of KG climates between these regions for the higher AI these higher $E_{POT} > P$ classes (Fig. 78). Basins located in high $PET_{E_{POT}} > P$ classes of Europe mainly are from the Cfb climate whereas basins of in these higher classes of the USA mostly consist of hot summer (Dfa and Cfa) and seasonal (Csb, Dsb) climates, which have shown to have longer drought durations.

Another possible factor that might explain these differences in classes is the difference in latitude between Europe and the USA, where for the same $PET_{E_{POT}} > P$ classes, the lower latitude USA has shorter summer days with higher temperatures compared to longer summer days with lower temperatures in Europe. In addition, Van der Schrier et al. (2011) showed that annual actual evaporation calculated with a simple water balance model that uses the Thornthwaite formula to compute E_{POT} leads to an underestimation of evaporation in parts of the USA and an overestimation in North-Western Europe. Defining evaporation with another method may therefore lead to more comparable classes between the USA and Europe.

The AI also showed to be suitable to differentiate basins according to drought duration, with a sorting of average DDC over the different AI classes that clearly followed the expected

pattern of higher average DDC for basins ~~of in~~ lower AI classes and lower average DDC for basins ~~of the in~~ higher AI classes. The AI was applied in previous studies focusing more on the arid spectrum (low values) of this index (e.g., Spinoni et al., 2015), where all non-arid regions (higher AI) are generalized to one humid class. Nevertheless, results of this study indicate that the wetter range of this index is also ~~a~~ suitable to differentiate basins according to drought duration. When comparing DDC of ~~basins in~~ Europe with the USA, ~~basins in~~ the lower three AI classes (< 50) of the USA have ~~significantly~~ higher average DDC. This difference was not explained by differences in dominant controls P (lower in Europe) and BFI (higher in Europe) for basins in these ~~climate~~ classes (Fig. 6). ~~Differences~~ 7). ~~The difference~~ in KG climates falling into the lowest three AI classes (Fig. 78) is more likely to explain this difference. ~~The in DDC. Basins in the~~ lower AI classes of Europe mainly encompass the Cfb climate whereas ~~basins~~ in the USA ~~these classes mainly consist of the hot summer climates (Cfa, Dfa) that have showed longer drought durations are represented by a mixture of different climates, including the climate classes that have shown a drought prolonging control.~~

Overall, ~~results of this study show that long duration droughts are modified by both climate and catchment controls. Still,~~ different climate classifications systems have shown to be suitable to differentiate basins according to ~~drought duration for long duration droughts in~~ observed streamflow under a wide range of catchment properties. This suggests that, for the selected basins, catchment controls were not dominant over climatic controls, which is in line with ~~the~~ previous catchment classification studies of Berghuijs et al. (2014) and Coopersmith et al. (2012). Climate classification systems are thus useful to identify regions with different sensitivities to long duration droughts in observed streamflow, but they do not necessarily distinguish regions with unique hydrologic drought duration characteristics. ~~This is confirmed by differences in DDC of basins in the same climate classes in Europe and the USA (e.g. the KG climates Cfb and Dfc), likely to be caused by differences in dominant individual controls P and BFI.~~ Most suitable in differentiating basins according to drought duration within both Europe and the USA are ~~the~~ climate classification systems that include an absolute water balance term (AI or ~~PET~~ $E_{POT} > P$). However, both these classification

systems show differences in DDC ~~between of basins in~~ the same classes ~~for of~~ Europe and the USA for low AI and high $PET_{E_{POT}} > P$ classes. Combining information of the different climate classification systems and individual climate and catchment controls suggests to be the most suitable way for large scale drought studies to stratify regions, especially when comparing the USA with Europe.

4.2 Evaluation of the method

This study compared DDC of basins of classes of a variety of climate classifications systems and individual controls using a ~~data-set~~ dataset of near-natural streamflow records. Based solely on observations means that basins in this ~~data-set~~ dataset are not uniformly distributed for the two regions. For example, for Spain, only a small number of streamflow records was available that met the selection criteria of being near-natural without falling dry too often. Despite this unequal coverage, the data set used includes basins with a large variety of climatic and catchment properties, which allowed for a detailed comparison within and between ~~groups of basins in the different regions~~ classes of basins. Furthermore, this study only considered near-natural basins, which are potentially biased towards smaller headwater catchments. For larger basins, catchment controls such as lakes and wetlands, might have a stronger effect. However, the anthropogenic controls on streamflow drought characteristics in these basins might dominate the natural ones and therefore, these basins were excluded in this study.

Droughts were identified from ~~the~~ near-natural streamflow records using a ~~-drought identification method that~~ threshold based approach. This study focused solely on drought duration. However, there are other characteristics that quantify properties of hydrological drought, such as (standardized) deficit volume (which is of interest for, e.g., the water supply sector). Although other drought characteristics were out of scope for this research, the proposed method lends itself to investigate the effect of climate and catchment controls on other drought properties such as deficit volume.

The drought identification method was specifically chosen to avoid artificial drought events caused by ~~the~~ methodological choices rather than by water deficits (Beyene

et al., 2014). ~~These drought durations~~ Drought durations computed with this method were transformed to cumulative distributions and displayed as a function of their fraction of non-exceedance (comparable to Tallaksen et al., 2009). Another approach would be to show these cumulative drought duration distributions as a function of the total number of drought events as ~~is done~~ in Fleig et al. (2011). This approach conserves the frequency of drought events, but for this research, the used approach was preferred to allow for a systematic comparison between all ~~groups~~ classes of DDC. However, since the used approach loses information about the frequency, it is essential to have a drought identification method that does not introduce artificial drought events and thus conserves an equal fraction of time in drought for all streamflow records. Therefore, procedures that influence this fraction ~~like smoothing and pooling procedures (described in~~, e.g., ~~?)~~ or smoothing of the threshold, pooling of drought events or the exclusion of minor drought events were not applied in this study.

For the statistical comparison of DDC, both the KS and MWU test were applied. Using two tests increases the robustness of the analysis as they focus on different aspects of the distribution. However, one assumption of the MWU test (equal shape in distribution of DDC values of two classes) did not hold true for all combinations of classes and percentiles. Therefore, results of this test were interpreted as difference in mean ranks and not as a difference in median (Bergmann et al., 2000). The strength of the statistical design of this study is that it indicates whether differences occur between neighboring classes (possibly related to our grouping criteria) or non-neighboring classes. This systematic statistical comparison also provides more insight about which classes are similar to each other for predefined climate classification systems, e.g., which KG climates have similar DDC. This information would be lost if, for example, a Kruskal–Wallis test was applied, which only detects if one group is different from the total.

5 Conclusions

This study evaluated climate classification systems and classified individual controls for their suitability to differentiate basins according to drought duration characteristics ~~within the USA and Europe~~. Results show that from the individual controls, precipitation and the base flow index were most suitable differentiators ~~for both the USA and Europe~~. Climate classification systems that included an absolute precipitation term, the aridity index and months with average potential evaporation larger than the precipitation, were most suitable to differentiate basins according to drought duration ~~within the two regions~~. The Köppen–Geiger climate classification system was able to differentiate basins according to drought duration between seasonally influenced climates (dry, cold or hot season) and climates with no or little seasonal influences. However, the high number of seasonal climate classes with similar DDC does not make this climate classification the most suitable differentiator.

DDC of basins of the same climate classes were not always comparable between Europe and the USA. For the Köppen–Geiger climate classification system, this is likely related to differences in dominant controls (precipitation and base flow index) over the same Köppen–Geiger classes. ~~For the aridity index and months with average potential evaporation larger than the precipitation, the high number of climates~~ The higher number of basins located in climates that are influenced by seasonality in precipitation and temperature in the USA for low aridity index classes and classes with a high number of months with average potential evaporation larger than the precipitation is likely the cause of differences in DDC between these classes of basins in the two regions.

Although climate classification systems that include an absolute precipitation control are most suitable to differentiate basins according to drought duration ~~within Europe and the USA~~, their power to differentiate is likely to be improved when complemented with information of other climate classification systems and individual climate and catchment controls. Furthermore, such a combination of information of different climate classification and individual controls likely results in a better comparability of the same classes between Europe and the USA. Knowledge about differences in sensitivities to hydrologic

drought events can be applied in drought monitoring and early warning systems, e.g., through tailoring such systems to regions with a similar sensitivity to hydrologic drought. Furthermore, being able to better differentiate basins according to drought duration allows for more accurate stratification in comparative drought studies. However, further research is needed to combine these insights into one classification system that is specifically designed to classify the sensitivity to observed hydrologic drought duration.

Acknowledgements. This study is an outcome of the Belmont Forum project DrIVER (Drought Impacts: Vulnerability thresholds in monitoring and Early warning Research). Funding to the project DrIVER by the German Research Foundation DFG under the international Belmont Forum/G8HORC's Freshwater Security programme (project no. STA-632/2-1) is gratefully acknowledged. The article processing charge was funded by the German Research Foundation (DFG) and the Albert Ludwigs University Freiburg in the funding programme Open Access Publishing. We further acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>) and the data providers in the ECA&D project (<http://www.ecad.eu>).

References

- Andreadis, K. M., Clark, E. A., Wood, A. W., Hamlet, A. F., and Lettenmaier, D. P.: Twentieth-century drought in the conterminous United States, *J. Hydrometeorol.*, 6, 985–1001, doi:10.1175/JHM450.1, 2005.
- Barker, L. J., Hannaford, J., Chiverton, A., and Svensson, C.: From meteorological to hydrological drought using standardised indicators, *Hydrol. Earth Syst. Sci. Discuss.*, in press, 2015.
- Berghuijs, W. R., Sivapalan, M., Woods, R. A., and Savenije, H. H. G.: Patterns of similarity of seasonal water balances: a window into streamflow variability over a range of time scales, *Water Resour. Res.*, 50, 5638–5661, doi:10.1002/2014WR015692, 2014.
- Bergmann, R., Ludbrook, J., and Spooren, W. P. J. M.: Different outcomes of the Wilcoxon Mann Whitney test from different statistics packages, *Am. Stat.*, 54, 72–77, doi:10.1080/00031305.2000.10474513, 2000.

- Beyene, B. S., Van Loon, A. F., Van Lanen, H. A. J., and Torfs, P. J. J. F.: Investigation of variable threshold level approaches for hydrological drought identification, *Hydrol. Earth Syst. Sci. Discuss.*, 11, 12765–12797, doi:10.5194/hessd-11-12765-2014, 2014.
- Coopersmith, E., Yaeger, M. A., Ye, S., Cheng, L., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves – Part 3: A catchment classification system based on regime curve indicators, *Hydrol. Earth Syst. Sci.*, 16, 4467–4482, doi:10.5194/hess-16-4467-2012, 2012.
- de Martonne, E.: L'indice d'aridité, *Bulletin de l'Association de géographes français*, 3, 3–5, 1926.
- Falcone, J.: GAGES-II: geospatial attributes of gages for evaluating streamflow, Digital spatial data set 2011, available at: http://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml (last access: 11 October 2015), US Geological Survey, 2011.
- ~~Fleig, A. K., Tallaksen, L. M., Hisdal, H., and Demuth, S.: A global evaluation of streamflow drought characteristics, *Hydrol. Earth Syst. Sci.*, 10, 535–552, doi:, 2006.~~
- Fleig, A. K., Tallaksen, L. M., Hisdal, H., and Hannah, D. M.: Regional hydrological drought in north-western Europe: linking a new Regional Drought Area Index with weather types, *Hydrol. Process.*, 25, 1163–1179, doi:10.1002/hyp.7644, 2011.
- Geiger, R.: Überarbeitete Neuauflage von Geiger, R / Köppen–Geiger/Klima der Erde, Wandkarte 1 : 16 mill., Klett-Perthes, Gotha, 1961.
- Grieser, J., Gommel, R., Cofield, S., and Bernardi, M.: Data sources for FAO worldmaps of Koeppen climatologies and climatic net primary production, available at: http://www.fao.org/nr/climpag/globgrids/kc_commondata_en.asp (last access: 11 October 2015), 2006.
- Gudmundsson, L., Tallaksen, L. M., and Stahl, K.: Spatial cross-correlation patterns of European low, mean and high flows, *Hydrol. Process.*, 25, 1034–1045, doi:10.1002/hyp.7807, 2011a.
- Gudmundsson, L., Tallaksen, L. M., Stahl, K., and Fleig, A. K.: Low-frequency variability of European runoff, *Hydrol. Earth Syst. Sci.*, 15, 2853–2869, doi:10.5194/hess-15-2853-2011, 2011b.
- Gustard, A. and Demuth, S.: Manual on low-flow estimation and prediction, Operational hydrology report no. 50, wmo-no. 1029, WMO, Geneva, 2009.
- Hannaford, J., Lloyd-Hughes, B., Keef, C., Parry, S., and Prudhomme, C.: Examining the large-scale spatial coherence of European drought using regional indicators of precipitation and streamflow deficit, *Hydrol. Process.*, 25, 1146–1162, doi:10.1002/hyp.7725, 2011.
- Haslinger, K., Koffler, D., Schöner, W., and Laaha, G.: Exploring the link between meteorological drought and streamflow: effects of climate-catchment interaction, *Water Resour. Res.*, 50, 2468–2487, doi:10.1002/2013WR015051, 2014.

- Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.*, 113, D20119, doi:10.1029/2008JD010201, 2008.
- Hisdal, H. and Tallaksen, L. M.: Estimation of regional meteorological and hydrological drought characteristics: a case study for Denmark, *J. Hydrol.*, 281, 230–247, doi:10.1016/S0022-1694(03)00233-6, 2003.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World map of the Köppen–Geiger climate classification updated, *Meteorol. Z.*, 15, 259–263, doi:10.1127/0941-2948/2006/0130, 2006.
- Lins, H. F.: USGS Hydro-Climatic Data Network 2009 (HCDN-2009), US Geological Survey Fact Sheet 2012–3047, available at: <http://pubs.usgs.gov/fs/2012/3047/> (last access: 11 October 2015), US Geological Survey, 2012.
- Rubel, F. and Kottek, M.: Comments on: “The thermal zones of the Earth” by Wladimir Köppen (1884), *Meteorol. Z.*, 20, 361–365, 2011.
- [Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins—Part 2: Flood and low flow studies, *Hydrol. Earth Syst. Sci.*, 17, 2637–2652, doi:10.5194/hess-17-2637-2013, 2013.](#)
- Sheffield, J., Andreadis, K. M., Wood, E. F., and Lettenmaier, D. P.: Global and continental drought in the second half of the twentieth century: severity-area-duration analysis and temporal variability of large-scale events, *J. Climate*, 22, 1962–1981, doi:10.1175/2008JCLI2722.1, 2009.
- [Skøien, J. O., and Blöschl, G., and Western, A. W.: Characteristic space scales and timescales in hydrology, *Water Resour. Res.*, 39, 1304, doi:10.1029/2002WR001736, 2003.](#)
- Spinoni, J., Vogt, J., Naumann, G., Carrao, H., and Barbosa, P.: Towards identifying areas at climatological risk of desertification using the Köppen–Geiger classification and FAO aridity index, *Int. J. Climatol.*, 35, 2210–2222, doi:10.1002/joc.4124, 2015.
- Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., van Lanen, H. A. J., Sauquet, E., Demuth, S., Fendekova, M., and Jódar, J.: Streamflow trends in Europe: evidence from a dataset of near-natural catchments, *Hydrol. Earth Syst. Sci.*, 14, 2367–2382, doi:10.5194/hess-14-2367-2010, 2010.
- Staudinger, M., Stahl, K., and Seibert, J.: A drought index accounting for snow, *Water Resour. Res.*, 50, 7861–7872, doi:10.1002/2013WR015143, 2014.
- Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., Miskus, D., and Stephens, S.: The drought monitor, *B. Am. Meteorol. Soc.*, 83, 1181–1190, 2002.

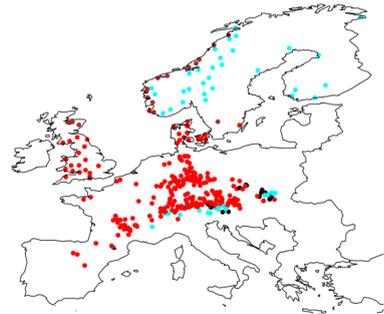
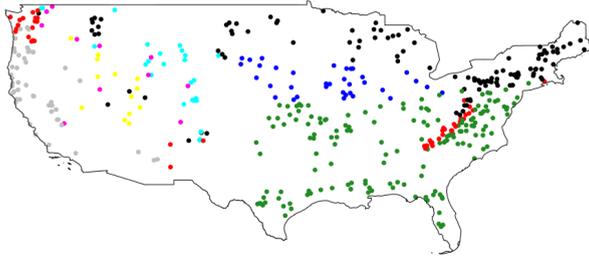
- Tallaksen, L. M. and Hisdal, H.: Regional analysis of extreme streamflow drought duration and deficit volume, FRIEND '97 – regional hydrology: concepts and models for sustainable water resource management, in: Proc. International Conference, Postojna, Slovenia, 30 September 1997–4 October 1997, 246, 141–150, 1997.
- Tallaksen, L. M. and Stahl, K.: Spatial and temporal patterns of large-scale droughts in Europe: model dispersion and performance, *Geophys. Res. Lett.*, 41, 429–434, doi:10.1002/2013GL058573, 2014.
- Tallaksen, L. M. and Van Lanen, H. A. J.: Hydrological drought: processes and estimation methods for streamflow and groundwater, *Developments in Water Science* 48, Elsevier Science B.V., Amsterdam, 2004.
- Tallaksen, L. M., Hisdal, H., and Van Lanen, H. A. J.: Space–time modelling of catchment scale drought characteristics, *J. Hydrol.*, 375, 363–372, doi:10.1016/j.jhydrol.2009.06.032, 2009.
- Thornthwaite, C. W.: An approach toward a rational classification of climate, *Geogr. Rev.*, 38, 55–94, 1948.
- Tijdeman, E., Van Loon, A. F., Wanders, N., and Van Lanen, H. A. J.: The effect of climate on droughts and their propagation in different parts of the hydrological cycle, DROUGHT-R&SPI Project, Tech. Rep. 2, available at: <http://www.eu-drought.org/> (last access: 11 October 2015), Wageningen University, Wageningen, 2012.
- Van der Schrier, G., Jones, P. D., and Briffa, K. R.: The sensitivity of the PDSI to the Thornthwaite and Penman-Monteith parameterizations for potential evapotranspiration, *J. Geophys. Res.*, 116, D03106, doi:10.1029/2010JD015001, 2011.
- Van Lanen, H. A. J., Wanders, N., Tallaksen, L. M., and Van Loon, A. F.: Hydrological drought across the world: impact of climate and physical catchment structure, *Hydrol. Earth Syst. Sci.*, 17, 1715–1732, doi:10.5194/hess-17-1715-2013, 2013.
- Van Loon, A. F. and Laaha, G.: Hydrological drought severity explained by climate and catchment characteristics, *J. Hydrol.*, 526, 3–14, doi:10.1016/j.jhydrol.2014.10.059, 2015.
- Van Loon, A. F. and Van Lanen, H. A. J.: A process-based typology of hydrological drought, *Hydrol. Earth Syst. Sci.*, 16, 1915–1946, doi:10.5194/hess-16-1915-2012, 2012.
- Van Loon, A. F., Tijdeman, E., Wanders, N., Van Lanen, H. A. J., Teuling, A. J., and Uijlenhoet, R.: How climate seasonality modifies drought duration and deficit, *J. Geophys. Res.-Atmos.*, 119, 4640–4656, doi:10.1002/2013JD020383, 2014.
- Vogt, J. V., Soille, P., de Jager, A. L., Rimaviciute, E., Mehl, W., Foisneau, S., Bodis, K., Dusart, J., Paracchini, M. L., Haastруп, P., and Bamps, C.: A pan-European River and Catchment Database,

- Ec-jrc report eur 22920 en, available at: http://ccm.jrc.ec.europa.eu/documents/CCM2-Report_EUR-22920-EN_2007_STD.pdf (last access: 29 October 2015), EC-JRC, Luxembourg, 2007.
- Wilhite, D. A. and Glantz, M. H.: Understanding: the drought phenomenon: the role of definitions, *Water Int.*, 10, 111–120, doi:10.1080/02508068508686328, 1985.
- Wilks, D. S.: *Statistical methods in the atmospheric sciences*, Academic press, Oxford, 2011.
- Zaidman, M. D., Rees, H. G., and Young, A. R.: Spatio-temporal development of streamflow droughts in north-west Europe, *Hydrol. Earth Syst. Sci.*, 6, 733–751, doi:10.5194/hess-6-733-2002, 2002.
- Zelenhasić, E. and Salvai, A.: A method of streamflow drought analysis, *Water Resour. Res.*, 23, 156–168, doi:10.1029/WR023i001p00156, 1987.

Table 1. Considered classes of climate classification systems and individual controls and corresponding number of basins in each class (USA/Europe).

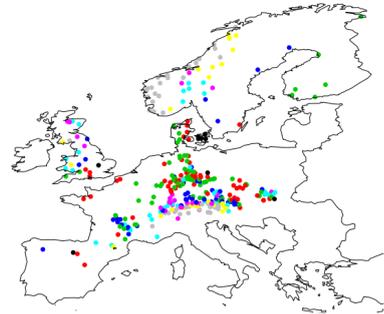
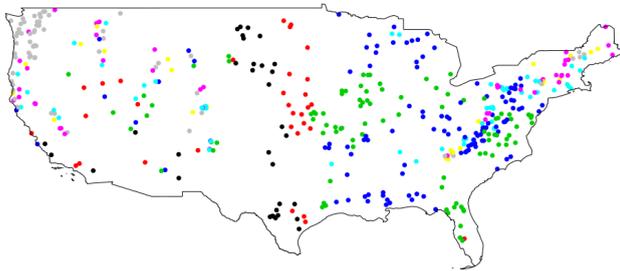
KG	AI	$T < 0$	$E_{\text{POT}} > P$	P	T	A	E	BFI
Dfb(114/15)	20–30(33/11)	0(184/118)	0(20/83)	1(68/94)	1(84/78)	1(87/75)	1(100/62)	1(134/29)
Cfb(48/247)	30–40(32/59)	1(31/30)	1(27/22)	2(75/86)	2(73/88)	2(77/84)	2(101/60)	2(110/50)
Cfa(156/–)	40–50(92/78)	2(14/33)	2(83/33)	3(98/64)	3(47/115)	3(77/85)	3(84/78)	3(67/95)
Dfa(35/–)	50–60(114/45)	3(100/98)	3(140/37)	4(115/46)	4(96/65)	4(105/56)	4(71/89)	4(70/90)
Dfc(29/49)	60–70(56/45)	4(46/18)	4(128/61)	5(105/57)	5(161/–)	5(115/47)	5(105/58)	5(80/83)
Dsc(11/–)	70–80(47/29)	5(64/25)	5(37/94)	–	–	–	–	–
Dsb(13/–)	80–90(24/28)	$\geq 6(22/25)$	$\geq 6(26/17)$	–	–	–	–	–
Csb(48/–)	90+(63/52)	–	–	–	–	–	–	–
Cfc(–/25)	–	–	–	–	–	–	–	–

Köppen-Geiger



- Dfb
- Cfb
- Cfa
- Dfa
- Dfc
- Dsc
- Dsb
- Csb
- Cfc

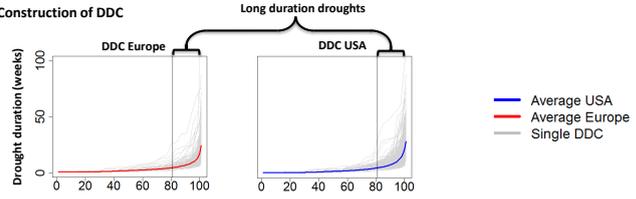
Aridity index



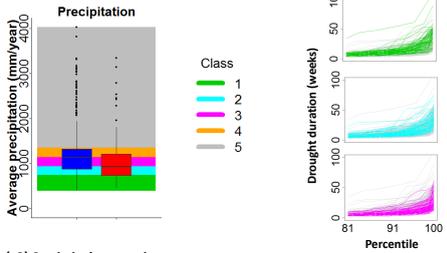
- 20-30
- 30-40
- 40-50
- 50-60
- 60-70
- 70-80
- 80-90
- 90+

Figure 1. Basin locations and two corresponding classifications (Köppen-Geiger and the aridity index). A description of these two climate classification systems is presented in Section 2.1.

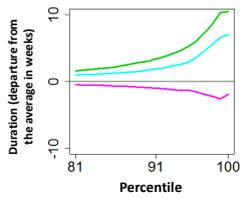
(a) Construction of DDC



(b) Grouping of DDC



(c1) Visual comparison



(c2) Statistical comparison

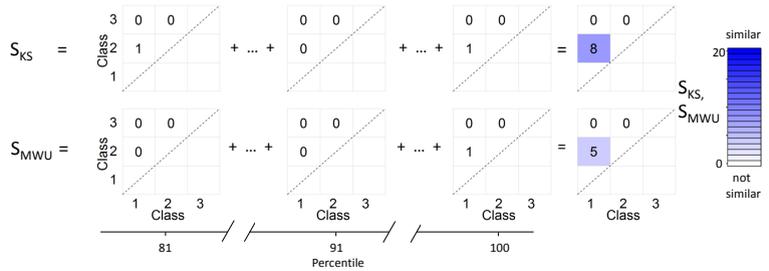
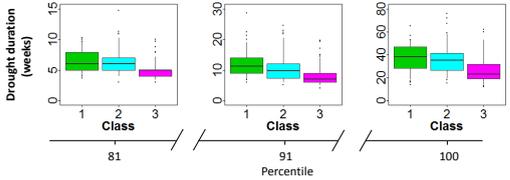


Figure 2. Conceptual approach. a: total ensemble of drought duration curves for both Europe (left) and the USA (right). b (left): example of the grouping of drought duration curves based on precipitation classes with boxplots of precipitation values for basins in both the USA (blue) and Europe (red) and background colors indicating the class ranges. b (right): corresponding exemplary ensembles of DDC groups for precipitation classes 1, 2 and 3 for the USA. c1: visualization of average DDC of basins in the three exemplary classes displayed as departures from the overall average of DDC of the USA. c2: Statistical comparison of distributions of DDC at each percentile between 81 and 100 (in the boxplots displayed for percentile 81, 91 and 100). Significance of differences in DDC values per percentile are indicated in the matrices below (1=significant, 0=not significant). The final measure of similarity (sum of significance scores over the 81st-100th percentile) is shown on the right.

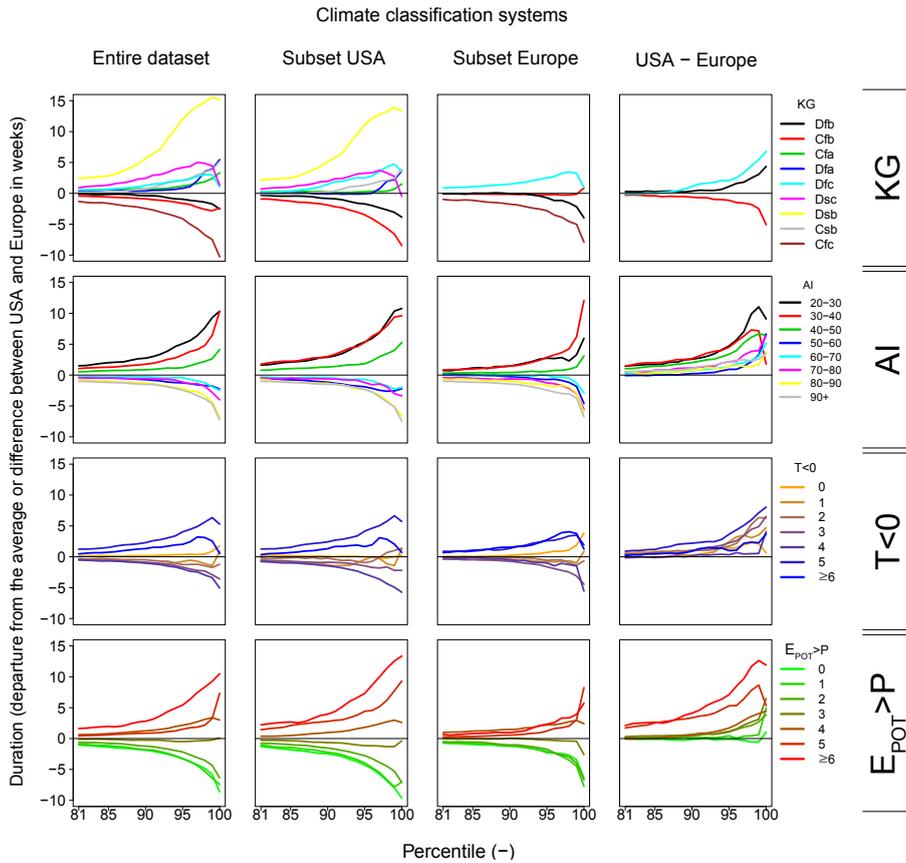


Figure 3. Averages of the ensembles of subsets of drought duration curves between the 81st and 100th percentile (average DDC) for basins in different classes of climate classification systems (rows) for: the entire dataset (first column), the USA (second column) and Europe (third column). Average DDC are displayed as departures from the overall average of DDC for the specific selection of basins, i.e., average of: all basins (first column), all basins in the USA (second column) and all basins in Europe (third column). The fourth (right) column shows the difference in average DDC of basins in the same climate classes for the USA and Europe (average DDC USA minus average DDC Europe).

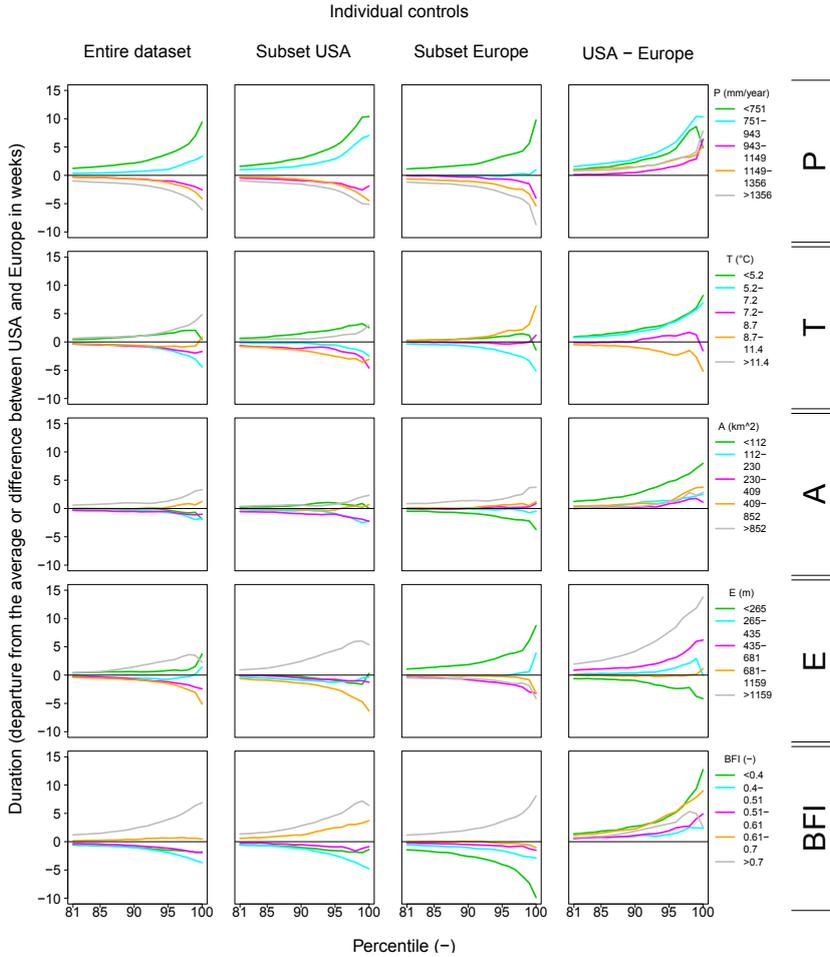


Figure 4. Averages of the ensembles of subsets of drought duration curves between the 81st and 100th percentile (average DDC) for basins in different classes of individual controls (rows) for: the entire dataset (first column), the USA (second column) and Europe (third column). Average DDC are displayed as departures from the overall average of DDC for the specific selection of basins, i.e., overall average of: all basins (first column), all basins in the USA (second column) and all basins in Europe (third column). The fourth (right) column shows the difference in average DDC of basins in the same classes of individual controls for the USA and Europe (average DDC USA minus average DDC Europe).

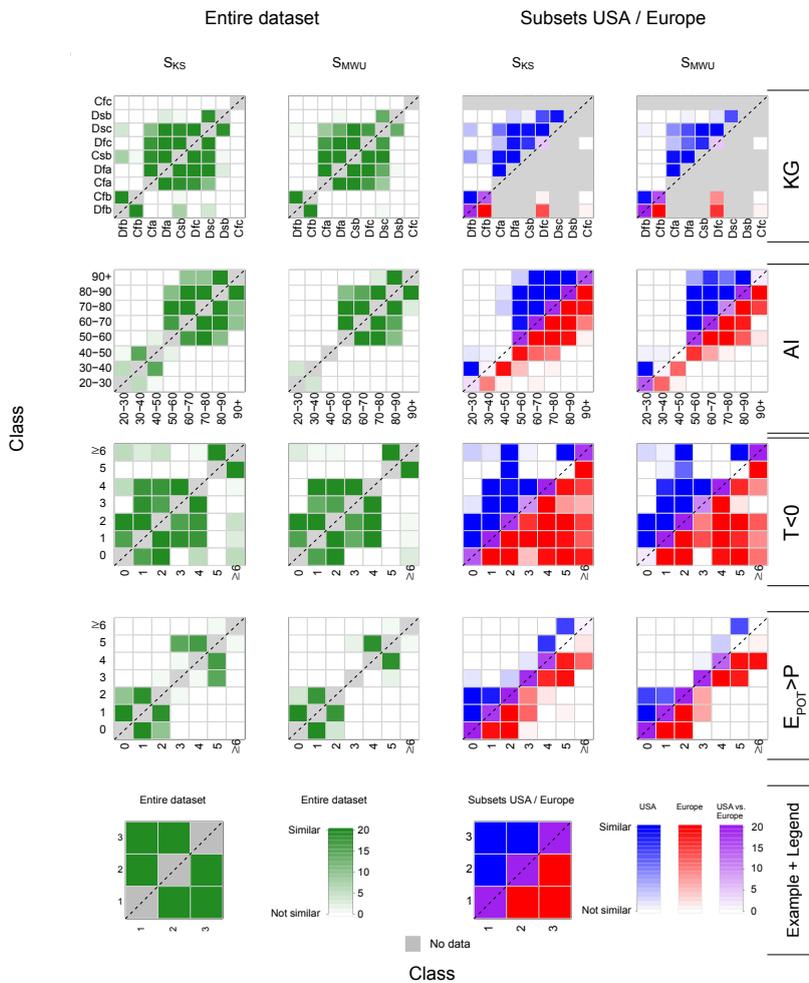


Figure 5. Number of percentiles with similar DDC values of basins in different classes of climate classification systems according to the KS and the MWU test, reflected by two measures of statistical similarity (S_{KS} and S_{MWU}). Left two columns show these measures of similarity for the entire dataset (in green) and right two columns for the two regional subsets: USA (blue, above the diagonal of each matrix) and Europe (red, below the diagonal of each matrix). Measures of similarity between DDC of basins in the same climate classes of Europe and the USA are displayed in the diagonal cells of the matrices (purple). No data (grey) indicates the combinations that were not considered (i.e., when the numbers of basins was smaller than 10 in one of the two regions).

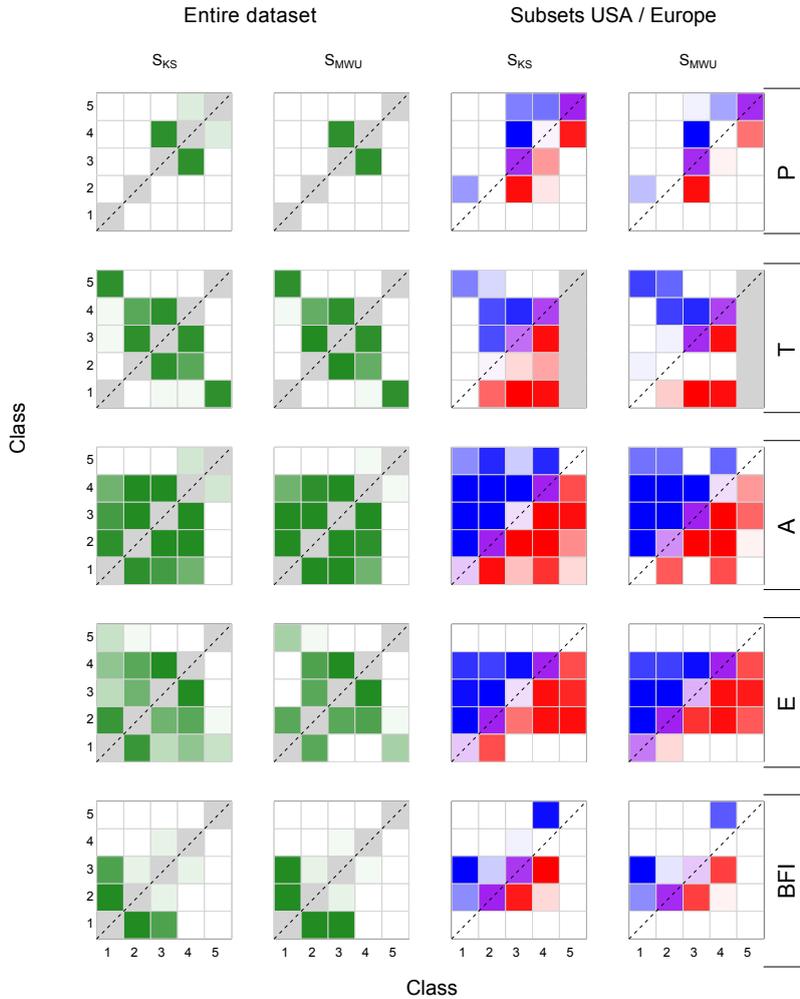


Figure 6. Number of percentiles with similar DDC values of basins in different classes of individual controls according to the KS and the MWU test, reflected by two measures of statistical similarity (S_{KS} and S_{MWU}). The darker the color, the more similar percentiles (legend is presented in Fig. 5). Left two columns show these measures of similarity for the entire dataset (in green) and right two columns for the two regional subsets: USA (blue, cells above the diagonal of each matrix) and Europe (red, cells below the diagonal of each matrix). Measures of similarity between DDC of basins in the same climate classes of Europe and the USA are displayed in the diagonal cells of the matrices (purple). No data (grey) indicates the combinations that were not considered (i.e., when the number of basins was smaller than 10 in one of the two regions).

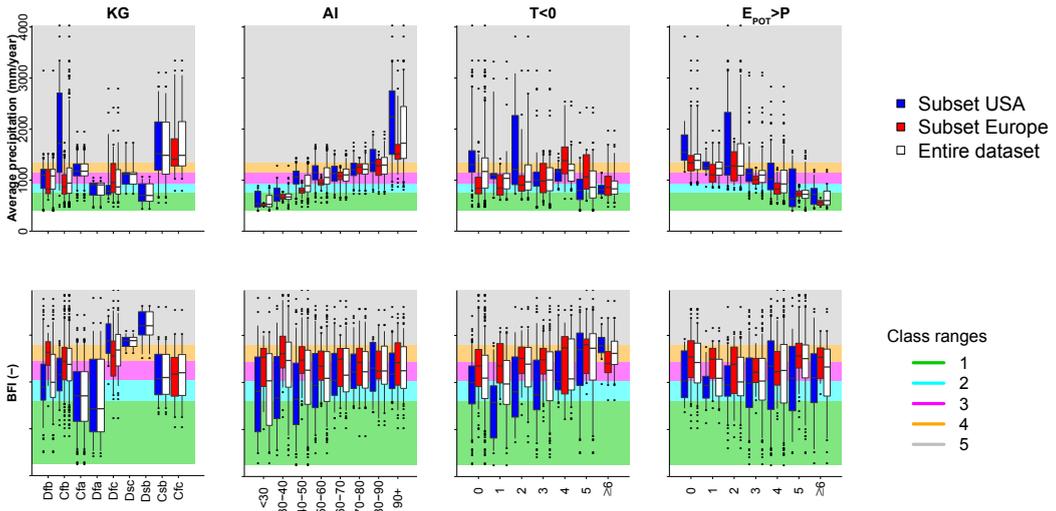


Figure 7. Distribution of individual controls P (upper row) and BFI (lower row) over classes of different climate classification systems for the USA (blue), Europe (red) and the entire dataset (white). Background colors indicate the ranges of classes of the individual controls (see Figure 4 for class ranges). Box: percentile 25, 50 and 75. End of whiskers: percentiles 5 and 95. Points: outliers.

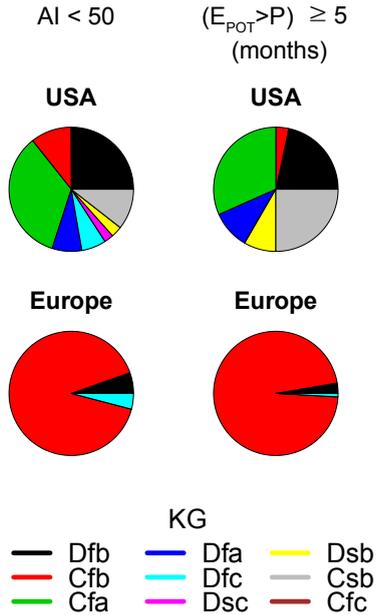


Figure 8. Distribution of different KG climates for all basins with an AI smaller than 50 (left) or $E_{\text{POT}} > P$ of 5 or more months (right) for both the USA and Europe.