Response to Editor and reviewer comments of Henny van Lanen and Niko Wanders

(comments in bold, reply in italic)

Editor comments

In general: Carefully check the use of 'basin' and 'catchment' throughout the text (i.e. that the terms are used consistently). For instance, on P5, L14 you state 'modeled basins', whereas on P6, L1 you say 'Austrian catchments'.

We replaced every use of the word "basin" with "catchment".

Abstract: I recommend to add the number of climate classifications used in the paper (four) rather than stating just 'different' or 'several' climate classification systems (more informative for the reader). Then it follows more natural to refer to e.g. the aridity index or e.g. the Köppen-Geiger (please add e.g. to the latter in line 25 of the abstract).

This also applies to other parts of the text, for instance:

P12, L5: 'of all classes of the four climate …'

P16, L3: 'Four different climate classification systems and five individual controls were …'

P22, L2: 'This study evaluated four different …..'

Applied this suggestion in the abstract (page 2), the suggested lines and in several other cases. Also mentioned it for the individual controls (i.e., five individual controls).

P3, L23: I find it a bit confusing to list the aridity index here (rather it should list its parameters).

Replaced Aridity index with precipitation and (potential) evaporation

P8, L13: I suggest adding some more information regarding the calculation of the BFI as its use in seasonal snow regime (using the entire record) has been criticized. At least it should say: 'The BFI was calculated from daily streamflow records using the entire streamflow record based on …'. You may also consider to add this aspect (i.e. the BFI was developed for rainfall regimes) to the discussion.

Added to P8, line 13

"The BFI was calculated from __the entire__ daily streamflow records…"

Also added to the discussion (page 21, line 5-11):

For the final selection of catchments, the BFI was calculated following the approach of Gustard and Demuth (2009). It should be kept in mind that this approach (which uses turning points in minimum flow of a 5 day moving filter to define base flow) was originally designed for rainfall dominated regimes and might represent base flow differently for some of the snow or glacier melt dominated catchments with long-lasting seasonal melt peaks and recessions that are thus more related to climate than to catchment controls (Gustard and Demuth, 2009). Although out of the scope for this research, a more catchment control-specific representation of base flow could be obtained with other calculation procedures.

P10, L10: replace '5' with 'five'.

Replaced

P11, L5: remove '.' before (1) and add 'and' before (2): '…statistical comparison; (1) The Kolmogorov…(..distribution), and (2) the Mann…….

Changed as suggested

Table 1: Here I suggest referring to the abbreviations in the figure legend or at least make the distinction between the two groups clear in the table itself (e.g. add a row above the abbreviations saying 'climate classification systems' and 'individual controls', thus splitting the table in two columns making it clear that one contains four, the other one five subtypes).

Added to the table caption and above the table (page 31)

Figure 1, Legend: Suggest replacing 'corresponding' with 'climate'.

Replaced

Reviewer comments Henny van Lanen

pg. 2, line 3: I suggest to phrase this more general and delete "modeling"; -

Deleted

pg. 3, line 2: Add at the end of the sentence: "... deficit of water or abnormal temperatures."; -

Added: "or temperature anomalies"

pg. 4: line 21: It is good to also address earlier papers. Quote from PhD thesis Anne van Loon (2013): "The first research addressing changes in the drought signal due to propagation through the hydrological cycle was done in Illinois, USA, by Changnon Jr [1987] and Eltahir and Yeh [1999]. The latter were the first to use the word 'propagation' in the context of the translation from 2 meteorological to hydrological drought. The work of Changnon Jr [1987] and Eltahir and Yeh [1999] was continued by Peters [2003] who published a study on the propagation of drought in groundwater. In recent years, drought propagation has been studied by Tallaksen and Van Lanen [2004], Peters et al. [2006], ..."; -

As this is not directly related to the work in this paper, but rather a very broad account of what drought characteristics are used for generally, we added a reference on pg. 4, line 21 to: Anne van Loon (2015): Hydrological droughts explained, where an overview of drought propagation research, including these benchmark papers, is given, and removed the other two references since they are included in this overview as well. Otherwise too much weight would be given to this issue of propagation, which is not really dealt with further in the paper.

pg. 9, line 21: okay to leave out DDCs of the entire dataset (Fig. 2a), because you revised the text accordingly by saying "As an example, ...". However, I suggest to add this also in the caption (see comment pg. 31); -

pg. 31, line 2: Add ".....and the USA (right), as an example. b (left) ....."

Added to the Figure caption

<u>Reviewer comment of Niko Wanders</u>

I do have one remaining concern with regard to revised Figure 3 and 4. The final column, with the difference between USA and Europe, shows some strange patterns. For example, the black line in Figure 3, row 1, column 4. When I look at the 100th percentile and read the y-axis I get that there is a 4 week difference between the two continents. While if I look at the USA and Europe subset they end at something like -4 and -3 respectively. That lead me to a difference of at most 1 week. This is the case for other scenarios as well (e.g the Dfc climate). Does the fourth column have a different scale at the y-axis, has the reference been changed or is something else going on. It feels counterintuitive and if the figure is correct it should be adjusted in the caption.

Graphs in column 2 and 3 might look comparable since column 2 is displayed as departure from the average of US catchments and column 3 as departure from the average of EU catchments. However, there is also a difference between average DDC of the USA and Europe (Figure below). Therefore, column 4 is not column 2 – Column 3. We agree that this is counterintuitive, and therefore changed the font of part of the y-axis and heading of the fourth column, two highlight that this difference.
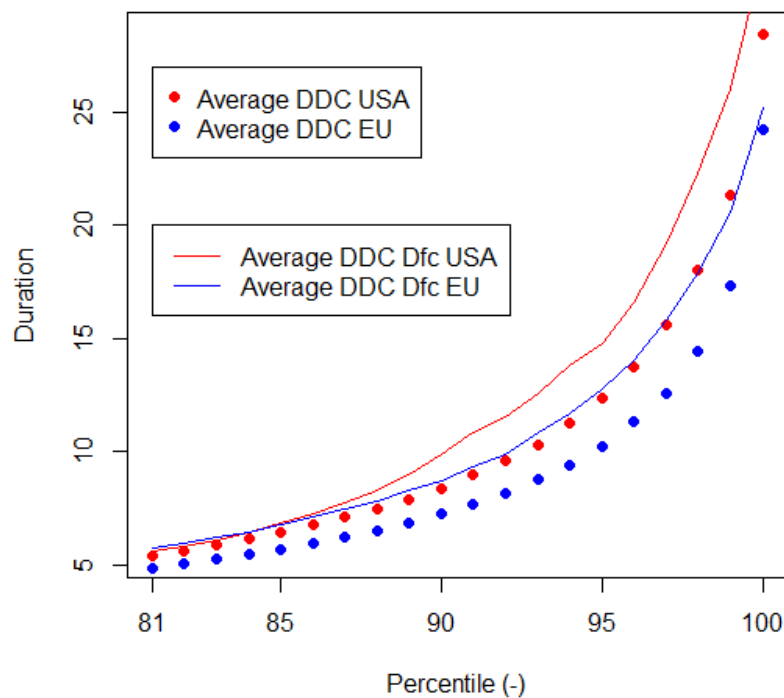
Controls on hydrologic drought duration in near-natural streamflow in Europe and the USA

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

Faculty of Environment and Natural Resources, University of Freiburg, Fahnenbergplatz, 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 used in large-scale drought ~~modeling~~ studies to stratify regions with similar hydro-climatic drought properties. What is currently lacking is a large-scale evaluation of the relation between climate and observed streamflow drought characteristics. In this study we explored how suitable common climate classifications are for differentiating ~~river basins~~ catchments 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 four climate classification systems and five 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~~ catchments according to drought duration. 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~~ catchments with a low aridity index. However, stratification of ~~basins~~ catchments 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 or abnormal temperatures. They are multifaceted phenomena and are often grouped into four main types; meteorological, agricultural, hydrologic and socio-economic. Hydrologic drought
5 relates to "effects of dry spells on surface and subsurface water" (Wilhite and Glantz, 1985). In the absence of human influences, hydrologic droughts are often triggered by anomalies in climatic conditions. Their duration regularly depends on the persistence 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
10 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
15 a hydrologic drought event in a ~~basin~~ catchment with little storage whereas a ~~basin~~ catchment 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~~ catchments with similar hydrologic
20 behavior, partly because it fails to distinguish between ~~basins~~ catchments 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 amount, seasonality and timing of precipitation, ~~the aridity index~~(potential) evaporation, timing of maximum runoff
25 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 clusters of ~~basins~~ catchments and soil, ecosystem and vegetation classes.

3

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~~ catchments with similar hydrologic behavior. For example, variability in precipitation and temperature is dampened when it propagates to streamflow
[5] (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
[10] 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
[15] 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)
[20] 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~~ overview in Van Loon, 2015) and to investigate the impact of climatic and catchment controls on droughts (e.g., Van Lanen et al., 2013; Van Loon et al., 2014).

[25] 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

4

seasonality in precipitation and the occurrence of cold or hot seasons (Kottek 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
5 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
10 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.

15 Studies based on modeled ~~basins~~ catchments 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
20 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~~catchment. For example, in Austria, propagation of drought (from precipitation to streamflow) was found to be more dependent on climatic
25 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~~ catchments that the distribution of drought durations is variable over different ~~basins~~catchments, which they hypothesized to be controlled by

5

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. 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 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~~ four climate classification systems to differentiate ~~basins~~ catchments 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~~ these four classification systems are transferable between these regions. A similar analysis was done for five 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~~ the four climate classification systems to differentiate ~~basins~~

catchments 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~~ the four climate classification systems and classes of individual controls. Duration is preferred over other drought
5 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:

10 – 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 four climate classification systems are therefore hypothesized to be
15 suitable for differentiating ~~basins~~ catchments 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
20 ($E_{POT} > P$). However, none of these climate classification systems considers catchment controls so their suitability to differentiate ~~basins~~ catchments according to drought duration in observed streamflow was investigated in this study under a wide variability of catchment characteristics.

7

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
5 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) 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
10 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
15 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 the entire daily streamflow records based on the calculation procedure described in Gustard and Demuth (2009). Different Four climate classification systems were
20 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.

25 - **–** $E_{POT} > P$: sum of months with average E_{POT} (calculated following the method of Thornthwaite, 1948) above the average P.

8

The KG classification system classifies ~~basins with 2 or 3~~ catchments with two or three 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 $> 3\,^\circ$C and D for minimum temperature $\leq 3\,^\circ$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~~ catchments 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. As an example, the drought duration curves of all ~~basins~~ catchments (or drought duration curve ensembles) for the USA and Europe are presented in Figure 2a. In this study we

only take into account long duration droughts 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
5 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 differ between classes of ~~different~~ the four climate
10 classification systems and five individual controls we grouped them accordingly. For the four climate classification 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
15 precipitation) and divided these values into ~~5~~ five classes with an equal number of ~~basins~~ catchments (Fig. 2b, left). In a second step, these classes were used to group the drought duration curves into ~~5~~ five different ensembles for the entire dataset and ~~5~~ five different ensembles for the two regional subsets (Fig. 2b, right; only three classes of the USA are shown in this example). The minimum number of ~~basins~~ catchments in a class was set to 10
20 for both classes of climate classification systems and individual controls. Classes of the two regional subsets with less ~~basins~~ catchments were excluded from the analysis. An overview of all remaining classes of drought duration curves (abbreviated to DDC when referring to subsets) with corresponding number of ~~basins~~ catchments in each class is presented in Table 1. Class ranges can be found in the results section of this study.

10

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) and (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 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).

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

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

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 $P_{\mathsf{KS},i}$ and $P_{\mathsf{MWU},i}$ 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 and individual control of both the entire dataset and the two regional subsets, DDC of the same climate classes were

5 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

10 used as a measure of statistical similarity between DDC.

3 Results

3.1 Visual comparison of DDC

Figure 3 presents average DDC (for long duration droughts) of all classes of ~~different~~ the four climate classification systems. In general, the patterns displayed for the entire dataset

15 and for the two regional subsets (USA and Europe) are comparable. However, average DDC of ~~basins~~ catchments 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 Europe is mostly positive (Fig. 3, right column)).

The KG reveals lowest average DDC for ~~basins~~ catchments in the non-seasonal

20 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~~ catchments in the hot summer, cold and seasonal climates (Cfa, Dfa, Csb, Dfc, Dsb, Dsc). ~~Basins~~ Catchments in the Dfc and Dfb climate of the USA have higher average DDC compared to Europe, whereas average DDC of ~~basins~~ catchments in the Cfb climate in

25 Europe are higher. The AI shows highest average DDC for ~~basins~~ catchments in the lowest (most arid) AI classes. Generally, average DDC decrease with increasing AI classes, apart

from an occasional exchange between some of the neighboring classes. Average DDC are higher for ~~basins~~ catchments in the same AI classes in the USA (USA minus Europe is positive), especially for ~~basins~~ catchments in the lower AI classes. For $T < 0$, average DDC are generally highest for ~~basins~~ catchments with most months $T < 0$, intermediate for ~~basins~~ catchments that have least months $T < 0$ and lowest for ~~basins~~ catchments that have 3 or 4 months $T < 0$. This ordering of DDC was found for both 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. $E_{POT} > P$ displays an ordering of average DDC with a general pattern of higher average DDC for the ~~basins~~ catchments with a high number of months $E_{POT} > P$ and lower average DDC for ~~basins~~ catchments with a low number of months $E_{POT} > P$. Similar to the ordering of average DDC of the AI, the systematic ordering of average DDC (from high for ~~basins~~ catchments in low classes to low for ~~basins~~ catchments in high classes of $E_{POT} > P$) is occasionally interrupted due to an exchange between average DDC of ~~basins~~ catchments in neighboring classes. ~~Basins~~ Catchments in lower classes of $E_{POT} > P$ are comparable between the two regions whereas ~~basins~~ catchments in classes with more months $E_{POT} > P$ show distinct higher average DDC for the USA.

Figure 4 presents the average DDC of ~~basins~~ catchments grouped by individual controls. Average DDC of ~~basins~~ catchments in the same classes are again most of the times higher for the USA compared to Europe. However, in contrast to the four 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~~ catchments with the highest average DDC is the class with the lowest P and vice versa. Average DDC decrease from lowest to highest P class. Classes of T show highest average DDC for ~~basins~~ catchments in both the lowest and highest T class. Longer drought events are thus found for ~~basins~~ catchments with temperatures from the tails of the temperature distribution. However, differences in average DDC of ~~basins~~ catchments in different classes of T are not as distinct as for precipitation classes. Even

13

smaller differences in average DDC are found for ~~basins~~ catchments in the different classes of A. In Europe, small ~~basins~~ catchments display lowest average DDC, and large ~~basins~~ catchments highest average DDC. This is different in the USA, where both small and large ~~basins~~ catchments exhibit the highest average DDC. Similar to A, E shows differences in

5 ordering of average DDC between the two regions. For the USA, highest average DDC are displayed for ~~basins~~ catchments in the highest E class whereas the highest average DDC of Europe are displayed for ~~basins~~ catchments in the lowest E class. 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.

10 **3.2 Statistical comparison**

Figure 5 shows the measures of statistical similarity (S_{KS} and S_{MWU}) between ensembles of DDC for ~~basins~~ catchments in different climate classes. Patterns are again most of the time comparable between the entire dataset and the two regional subsets (USA and Europe). Differences occur for some specific combinations (e.g. DDC of ~~basins~~ catchments in the

15 Dfc climate are comparable with DDC of ~~basins~~ catchments in the Dsb climate within the USA according to S_{KS}, however, DDC of ~~basins~~ catchments in these two climates are not comparable according to the same measure of similarity for the entire dataset where the DDC of ~~basins~~ catchments in the Dfc climate of the USA are combined with the lower DDC of ~~basins~~ catchments in the European Dfc climate).

20 For the KG, DDC of ~~basins~~ catchments in the Cfc climate have significantly lower DDC values at most percentiles compared to all other climates. DDC of ~~basins~~ catchments in the Cfb climate are only similar with DDC of ~~basins~~ catchments in the Dfb climate according to both S_{KS} and S_{MWU}. DDC of ~~basins~~ catchments in this Dfb climate show little similarity with DDC of ~~basins~~ catchments in the other, seasonally influenced, climates again indicating the

25 distinction between shorter droughts for ~~basins~~ catchments in climates affected by no or small seasonal influences (Cfc, Cfb and Dfb) and longer droughts for ~~basins~~ catchments in the other climates. However, DDC of ~~basins~~ catchments in these other climates (Cfa, Dfa, Csb, Dfc, Dsb) mostly do not show notable differences among each other according

14

to both measures of statistical similarity. Out of these climates, ~~basins~~ catchments in the Dsb climate, which reveal the highest average DDC, also have the most distinctive DDC and only shows similarity in DDC with ~~basins~~ catchments in the Dsc climate (and at some percentiles with ~~basins~~ catchments in the Csb and Dfa 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~~ catchments 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~~ Catchments of the Dfc climate of the USA show significantly higher DDC values for most percentiles. The differences in DDC of ~~basins~~ catchments in different AI classes are most distinct between the lowest AI classes. The higher the AI class, the more neighboring classes of ~~basins~~ catchments show similarity in DDC, whereas for ~~basins~~ catchments in the lower AI classes, only DDC of ~~basins~~ catchments in direct neighboring classes occasionally show similarity. For the comparison between Europe and the USA, the lower AI classes (< 50) shows ~~basins~~ catchments with higher DDC in the USA according to both measures of similarity, whereas ~~basins~~ catchments of higher AI classes did not show many notable differences between the two regions. The small differences in average DDC of ~~basins~~ catchments in different classes of $T < 0$ are also reflected by the corresponding measures of statistical similarity, especially for Europe. For this region, DDC of ~~basins~~ catchments in almost all classes are similar to each other. ~~Basins~~ Catchments in the same classes of $T < 0$ are mostly comparable between the USA and Europe. Differences in DDC for ~~basins~~ catchments in different classes of $E_{POT} > P$ are notable. S_{KS} and S_{MWU} indicate similarity only in DDC of ~~basins~~ catchments in neighboring classes. Differences between the USA and Europe are only found for the DDC of ~~basins~~ catchments in the two highest classes of $E_{POT} > P$. For the other classes the DDC are similar.

Figure 6 displays the statistical comparison of DDC grouped by individual controls. DDC of ~~basins~~ catchments in different classes of P are mostly different from each other 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~~

catchments in the USA according to both measures of similarity. DDC of ~~basins~~ catchments of intermediate T classes are similar to each other as well as DDC of ~~basins~~ catchments of the lowest and highest temperature classes for the entire dataset and for the regional subset of the USA, confirming that long duration droughts are longer in both colder and
5 warmer ~~basins~~catchments. 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 Europe and the USA are found for classes of ~~basins~~ catchments with a lower T. ~~Basins~~ Catchments grouped by A hardly show differences in DDC. Only for the entire dataset, the largest ~~basins~~ catchments have different DDC. According to both S_{KS} and S_{MWU}, ~~basins~~ catchments in the
10 highest E class of the USA have higher DDC compared to DDC of ~~basins~~ catchments in the other E classes, whereas for Europe, ~~basins~~ catchments in the lowest E class have 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~~ catchments in different classes are often different from each other according to both measures of statistical similarity besides some
15 similarity between neighboring classes.

4 Discussion

4.1 Evaluation of climate classification systems

~~Different~~ Four different climate classification systems and five individual controls were evaluated for their suitability to differentiate ~~basins~~ catchments according to long duration
20 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~~ catchments 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,
25 which confirms the findings of Van Lanen et al. (2013) and Van Loon et al. (2014) that drought duration is modified by both catchment (groundwater response time) and climate

16

(seasonality in precipitation and the occurrence of hot or cold seasons) controls. Our 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. The distributions of dominant individual controls, however, are not always comparable between the classes of the four different climate classification systems, 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~~ catchments according to drought duration in observed streamflow. Furthermore, it partly explains why DDC of ~~basins~~ catchments in the same climate classes are not always comparable between the two regional subsets (USA and Europe).

For the KG climate classification system, ~~basins~~ catchments 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). According to the two measures of similarity used in this study, ~~basins~~ catchments in the Cfc climate (generally wetter than most other climates (Fig. 7)) were distinctly different from DDC of ~~basins~~ catchments in the other climates and the ~~basins~~ catchments in Cfb climate were only comparable with DDC of ~~basins~~ catchments in the Dfb climate. ~~Basins~~ Catchments 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 DDC values at the different considered percentiles. Further notable was the difference in average DDC for ~~basins~~ catchments in the Cfb climate between Europe and the USA. This was the only combination of climate classes where average DDC of ~~basins~~ catchments in Europe were distinctively higher, possibly explained by wetter condition in the Cfb climate for the ~~basins~~ catchments in the USA (Fig. 7). ~~Basins~~ Catchments in the Dfc 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~~ catchments in the USA (Fig. 7)).

17

Hot summer climates without seasonality in precipitation (Cfa, Dfa) consist of ~~basins~~ catchments with higher average DDC compared to the DDC of ~~basins~~ catchments 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

5 based on global data whereas this study only deals with ~~basins~~ catchments in the Dfa and Cfa in the USA. The differences in P between the hot and warm summer climates (Fig. 7) 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

10 is an important control on long duration droughts as well. Measures of statistical similarity show little differences between DDC of ~~basins~~ catchments in the hot summer climates and DDC of ~~basins~~ catchments in the other seasonal climates (Csb, Dfc, Dsb, Dsc). Results thus indicate that the KG is mainly suitable to make the distinction between ~~basins~~ catchments in climates with and without seasonal influences.

15 ~~Basins~~ Catchments in the KG climate classes that showed highest average DDC were ~~basins~~ catchments 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) 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

20 ($T < 0$)) was expected to be suitable to differentiate ~~basins~~ catchments according to drought duration. However, this was not the case and differences between average DDC were small and the measures of statistical similarity did not indicate strong differences between classes of ~~basins~~catchments, especially for Europe. These European ~~basins~~ catchments with most months of $T < 0$ are partly located in Scandinavia and the Alps, which have been related to

25 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~~ catchments with different drought duration characteristics.

18

More suitable for such a differentiation of ~~basins~~ catchments are the climate classification systems that take into account the dominant annual precipitation control (months with average potential evaporation larger than the precipitation ($E_{POT} > P$) and the aridity index (AI); note that the KG does not have such an annual precipitation term). $E_{POT} > P$ 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 $E_{POT} > P$ that followed the hypothesized pattern of higher DDC for ~~basins~~ catchments in the higher $E_{POT} > P$ classes and lower DDC for ~~basins~~ catchments in the lower $E_{POT} > P$ classes, which makes it a suitable climate classification system to differentiate ~~basins~~ catchments according to drought duration. The same classes for Europe and the USA show similarity in DDC for ~~basins~~ catchments located in the lower $E_{POT} > P$ classes, however, ~~basins~~ catchments located in the higher $E_{POT} > P$ 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 these higher $E_{POT} > P$ classes (Fig. 8). ~~Basins~~ Catchments located in high $E_{POT} > P$ classes of Europe mainly are from the Cfb climate whereas ~~basins~~ catchments 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 $E_{POT} > P$ 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~~ catchments 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~~ catchments in lower AI classes and lower average DDC for ~~basins~~ catchments in higher AI classes. The AI was applied
5 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 suitable to differentiate ~~basins~~ catchments according to drought duration. When comparing DDC of ~~basins~~ catchments in Europe with the USA, ~~basins~~ catchments in the lower three AI
10 classes (< 50) of the USA have higher average DDC. This difference was not explained by differences in dominant controls P (lower in Europe) and BFI (higher in Europe) for ~~basins~~ catchments in these climate classes (Fig. 7). The difference in KG climates falling into the lowest three AI classes (Fig. 8) is more likely to explain this difference in DDC. ~~Basins~~ Catchments in the lower AI classes of Europe mainly encompass the Cfb climate whereas
15 ~~basins~~ catchments in the USA 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~~ classification systems have shown to be suitable to differentiate ~~basins~~ catchments according to long duration droughts
20 in observed streamflow under a wide range of catchment properties. This suggests that, for the selected ~~basins~~catchments, 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,
25 but they do not necessarily distinguish regions with unique hydrologic drought duration characteristics. This is confirmed by differences in DDC of ~~basins~~ catchments 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~~ catchments according to drought duration within both Europe and the

USA are climate classification systems that include an absolute water balance term (AI or $E_{POT} > P$). However, both these classification systems show differences in DDC of ~~basins~~ catchments in the same classes of Europe and the USA for low AI and high $E_{POT} > P$ classes. Combining information of the different climate classification systems and individual
5   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~~ catchments of classes of ~~a variety of climate classifications systems and~~ four climate classification systems and five individual controls
10   using a dataset of near-natural streamflow records. Based solely on observations means that ~~basins~~ catchments in this 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~~ catchments with a large variety
15   of climatic and catchment properties, which allowed for a detailed comparison within and between classes of ~~basins~~catchments. Furthermore, this study only considered near-natural ~~basins~~catchments, which are potentially biased towards smaller headwater catchments. For larger ~~basins~~catchments, catchment controls such as lakes and wetlands, might have a stronger effect. However, the anthropogenic controls on streamflow drought
20   characteristics in these ~~basins~~ catchments might dominate the natural ones and therefore, these ~~basins~~ catchments were excluded in this study. For the final selection of catchments, the BFI was calculated following the approach of Gustard and Demuth, 2009 . It should be kept in mind that this approach (which uses turning points in minimum flow of a 5 day filter to define base flow) was originally designed for rainfall dominated regimes and might
25   represent base flow differently for some of the snow or glacier melt dominated catchments with long-lasting seasonal melt peaks and recessions that are thus more related to climate than to catchment controls (Gustard and Demuth, 2009) . Although out of the scope for this

research, a more catchment control-specific representation of base flow could be obtained with other calculation procedures.

Droughts were identified from near-natural streamflow records using a threshold based approach. This study focused solely on drought duration. However, there are other
5 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.

10 The drought identification method was specifically chosen to avoid artificial drought events caused by methodological choices rather than by water deficits (Beyene et al., 2014). 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
15 distributions as a function of the total number of drought events as 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 classes of DDC. However, since the used approach looses information about the frequency, it is essential to have a drought identification method that does not introduce artificial drought events and
20 thus conserves an equal fraction of time in drought for all streamflow records. Therefore, procedures that influence this fraction, e.g., 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
25 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
5 detects if one group is different from the total.

5 Conclusions

This study evaluated ~~four~~ different climate classification systems and ~~five~~ classified individual controls for their suitability to differentiate ~~basins~~ catchments according to drought duration characteristics. Results show that from the individual controls, precipitation and
10 the base flow index were most suitable differentiators. 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~~ catchments according to drought duration. The Köppen–Geiger climate classification system was able to differentiate ~~basins~~ catchments according to drought duration between
15 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~~ catchments of the same climate classes were not always comparable between Europe and the USA. For the Köppen–Geiger climate classification system, this is
20 likely related to differences in dominant controls (precipitation and base flow index) over the same Köppen–Geiger classes. The higher number of ~~basins~~ catchments 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
25 of ~~basins~~ catchments in the two regions.

Although climate classification systems that include an absolute precipitation control are most suitable to differentiate ~~basins~~ catchments according to drought duration, 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

5 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~~ catchments according to drought duration allows for more accurate stratification in comparative drought studies. However, further research is needed

10 to combine these insights into one classification system that is specifically designed to classify the sensitivity to observed hydrologic drought duration.

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.

25 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 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.: Uberarbeitete Neuausgabe von Geiger, R / Köppen–Geiger/Klima der Erde, Wandkarte 1 : 16 mill., Klett-Perthes, Gotha, 1961.

Grieser, J., Gommes, 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.:Hydrological drought explained, Wiley Interdisciplinary Reviews: Water, 2, 359–392.,doi:10.1002/wat2.1085, 2015.

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.
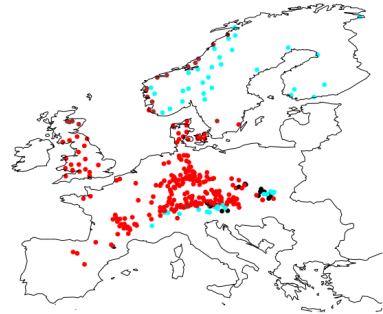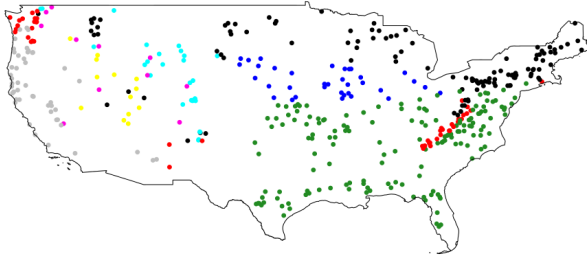
Van Loon, A. F. and Van Lanen, H. A. J.: A process-based typology of hydrological drought, Hydrol. Earth Syst. Sci., 16, 1915–1946, doi:10.5194/hess-16-1915-2012, 2012.

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., Haastrup, 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 the four climate classification systems (Köppen–Geiger (KG), aridity index (AI), number of months with an average temperature below zero ($T < 0$) and number of months where the average potential evaporation was larger than the average precipitation ($E_{POT} > P$)) and five individual controls (precipitation (P), temperature (T), area (A), elevation (E) and the base flow index (BFI)) and corresponding number of basins in each class (USA/Europe).
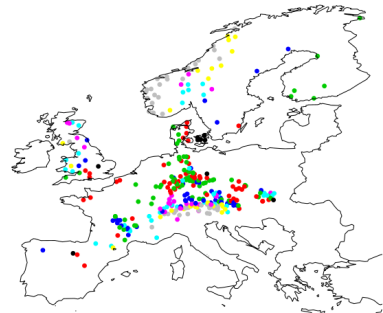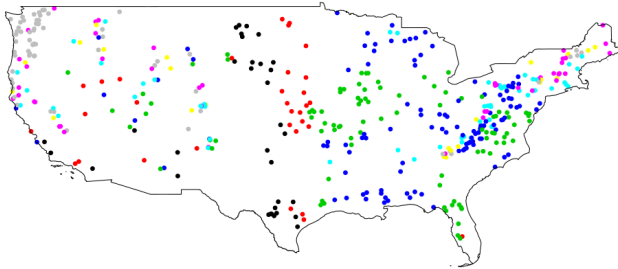
| | Climate classification systems | | | | Individual controls | | | |
KG	AI	$T < 0$	$E_{POT} > P$	P	T	A	E	BFI
Dfb(114/15)	20–30(33/11)	0(184/118)	0(20/83)	1(68/94)	1(84/78)	1(87/75)	1(100/62)	1(134/29)
Cfb(48/247)	30–40(32/59)	1(31/30)	1(27/22)	2(75/86)	2(73/88)	2(77/84)	2(101/60)	2(110/50)
Cfa(156/−)	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)	≥ 6(22/25)	≥ 6(26/17)	–	–	–	–	–
Csb(48/−)	90+(63/52)	–	–	–	–	–	–	–
Cfc(−/25)	–	–	–	–	–	–	–	–

Figure 1. ~~Basin~~ Catchment locations and two ~~corresponding~~ climate classifications (Köppen-Geiger and the aridity index). A description of these two climate classification systems is presented in Section 2.1.
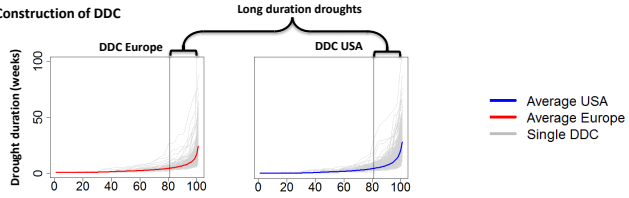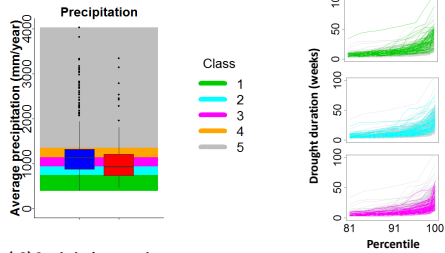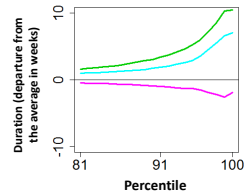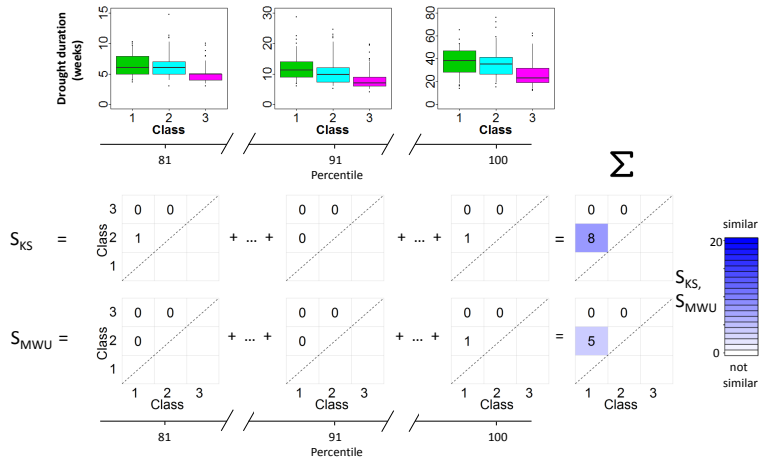
Figure 2. Conceptual approach. a: total ensemble of drought duration curves for both Europe (left) and the USA (right), as an example. b (left): example of the grouping of drought duration curves based on precipitation classes with boxplots of precipitation values for ~~basins~~ catchments in both the 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~~ catchments 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.
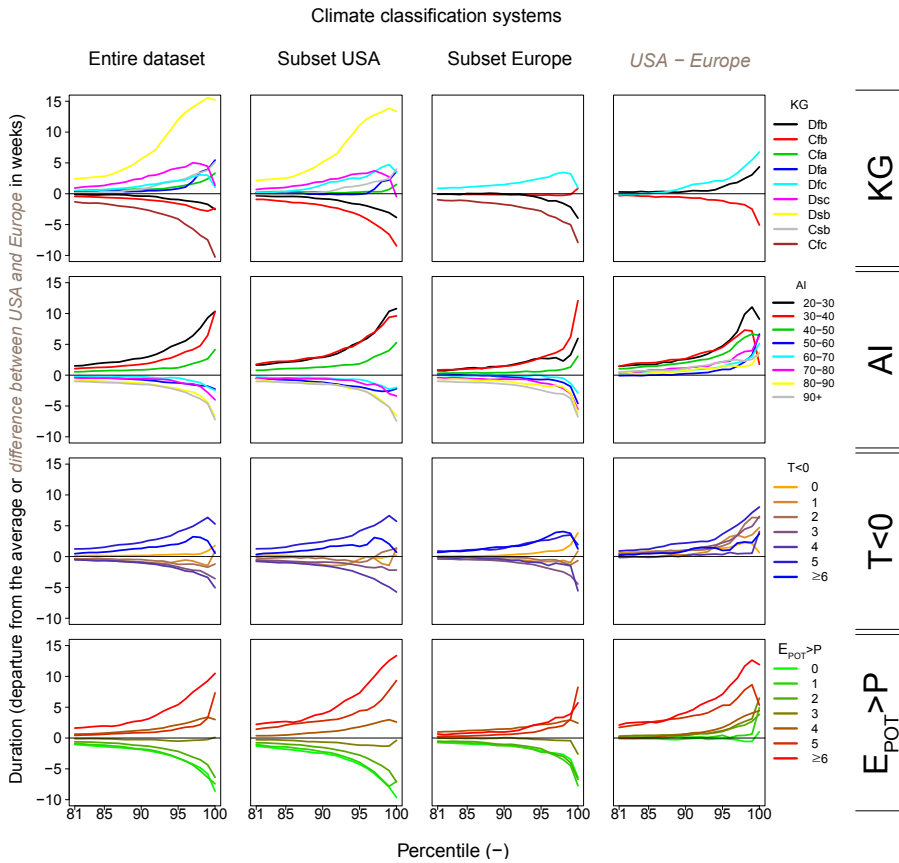
Climate classification systems

Entire dataset   Subset USA   Subset Europe   *USA − Europe*

Duration (departure from the average or *difference between USA and Europe* in weeks)

Percentile (−)

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

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

T<0
- 0
- 1
- 2
- 3
- 4
- 5
- ≥6

E_{POT}>P
- 0
- 1
- 2
- 3
- 4
- 5
- ≥6

KG

AI

T<0

E_{POT}>P

Figure 3. Averages of the ensembles of subsets of drought duration curves between the 81st and 100th percentile (average DDC) for ~~basins~~ catchments 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~~catchments, i.e., average of: all ~~basins~~ catchments (first column), all ~~basins~~ catchments in the USA (second column) and all ~~basins~~ catchments in Europe (third column). The fourth (right) column shows the difference in average DDC of ~~basins~~ catchments in the same climate classes for the USA and Europe (average DDC USA minus average DDC Europe).
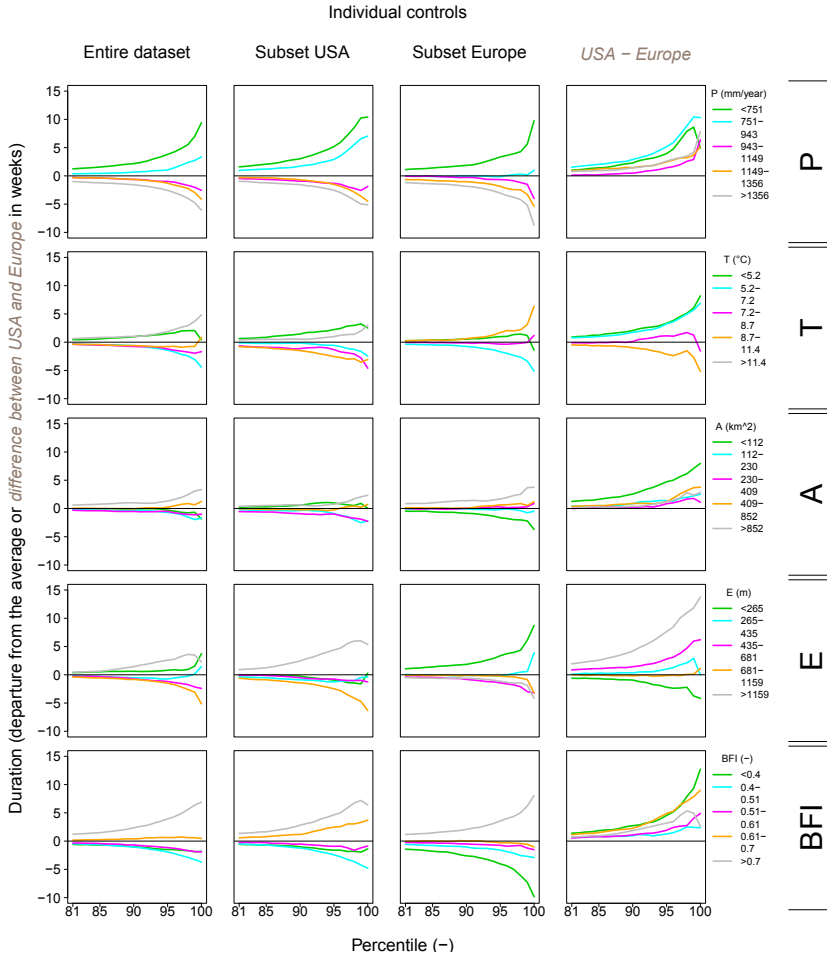
Figure 4. Averages of the ensembles of subsets of drought duration curves between the 81st and 100th percentile (average DDC) for ~~basins~~ catchments in different classes of individual controls (rows) for: the entire dataset (first column), the USA (second column) and Europe (third column). Average DDC are displayed as departures from the overall average of DDC for the specific selection of ~~basins~~catchments, i.e., overall average of: all ~~basins~~ catchments (first column), all ~~basins~~ catchments in the USA (second column) and all ~~basins~~ catchments in Europe (third column). The fourth (right) column shows the difference in average DDC of ~~basins~~ catchments in the same classes of individual controls for the USA and Europe (average DDC USA minus average DDC Europe).
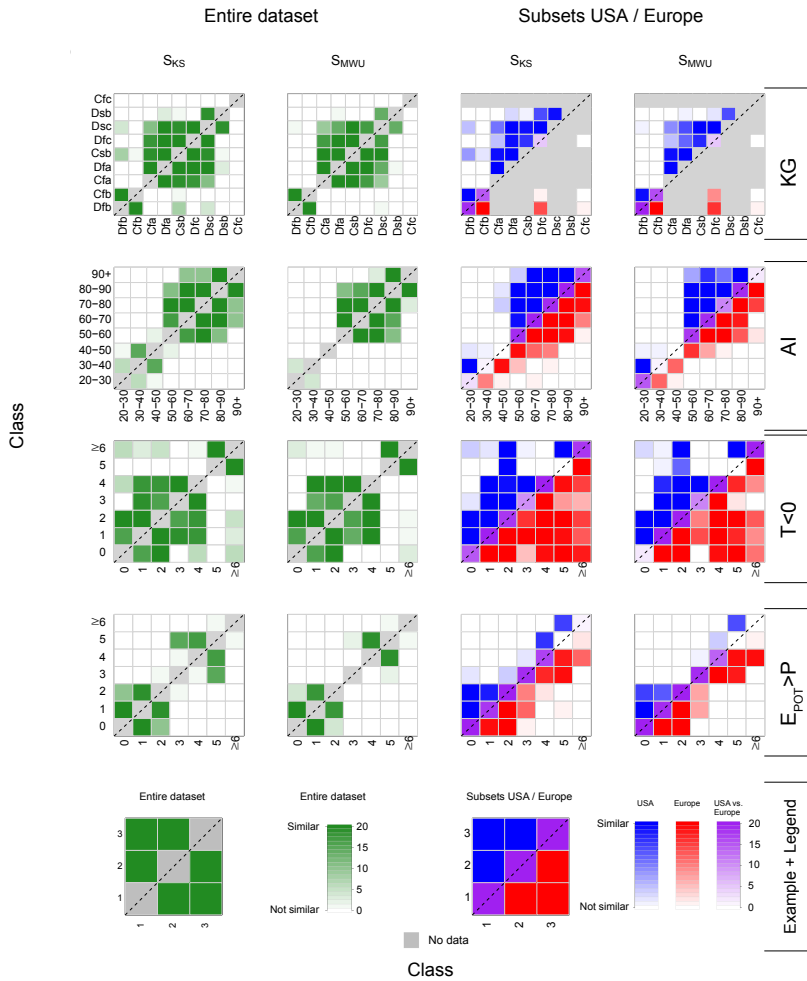
Figure 5. Number of percentiles with similar DDC values of ~~basins~~ catchments in different classes of climate classification systems according to the KS and the MWU test, reflected by two measures of statistical similarity (S_{KS} and S_{MWU}). Left two columns show these measures of similarity for 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~~ catchments 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~~ catchments was smaller than 10 in one of the two regions).
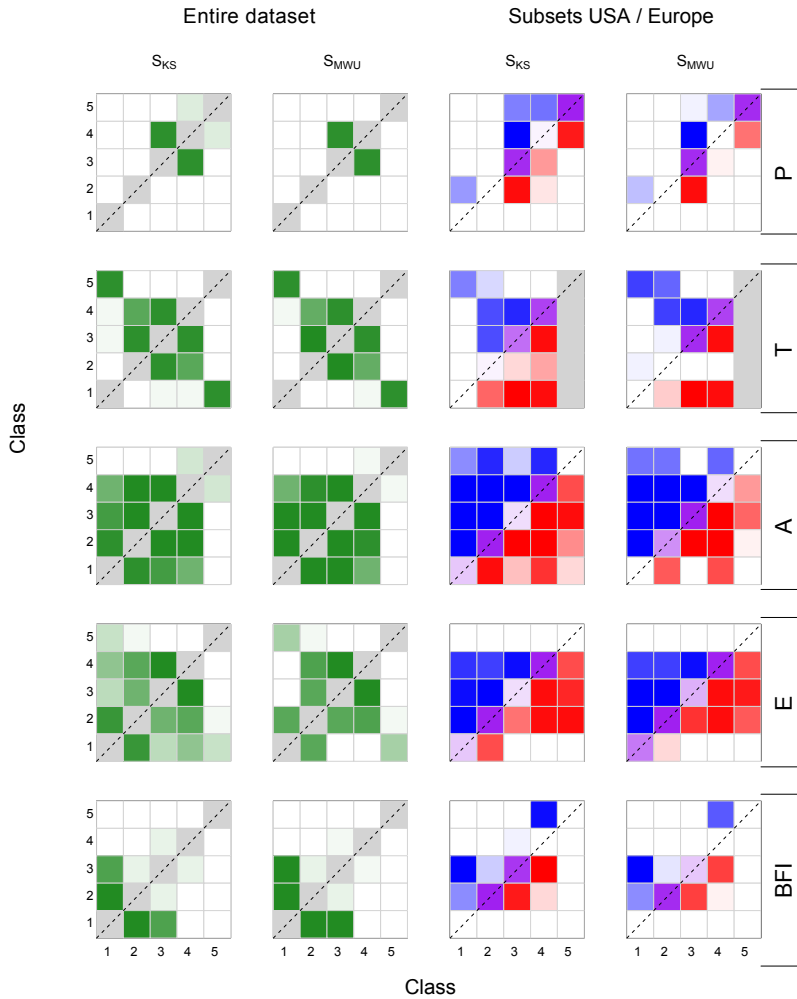
Figure 6. Number of percentiles with similar DDC values of ~~basins~~ catchments in different classes of individual controls according to the KS and the MWU test, reflected by two measures of statistical similarity (S_{KS} and S_{MWU}). The darker the color, the more similar percentiles (legend is presented 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~~ catchments 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~~ catchments was smaller than 10 in one of the two regions).
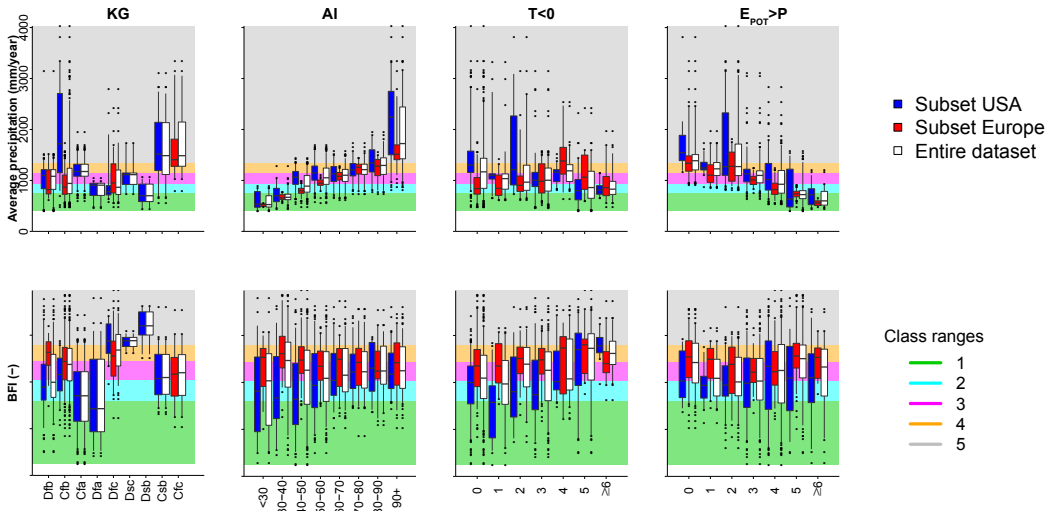
Figure 7. Distribution of individual controls P (upper row) and BFI (lower row) over classes of different climate classification systems for the USA (blue), Europe (red) and the entire dataset (white). Background colors indicate the ranges of classes of the individual controls (see Figure 4 for class ranges). Box: percentile 25, 50 and 75. End of whiskers: percentiles 5 and 95. Points: outliers.
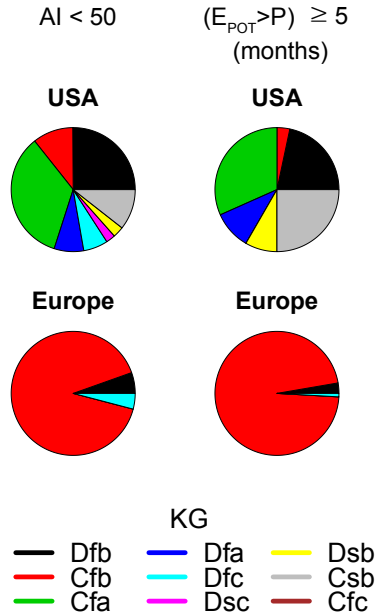
Figure 8. Distribution of different KG climates for all ~~basins~~ catchments with an AI smaller than 50 (left) or $E_{POT} > P$ of 5 or more months (right) for both the USA and Europe.