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, Fahnengbergplatz, 79098 Freiburg, Germany

Received: 31 October 2015 – Accepted: 9 November 2015 – Published: 11 December 2015

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

Published by Copernicus Publications on behalf of the European Geosciences Union.
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 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 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 hydrologic droughts are often triggered by anomalies in climatic conditions and their duration regularly depends on the persistence and magnitude of these anomalies. 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 (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 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.

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 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 related to storage (base flow index (BFI), area ($A$) and elevation ($E$)).

The following climate classification systems are also 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 evaporation is larger than the average precipitation ($\text{PET} > P$). 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 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 and the contiguous USA. 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 time step is $\leq 20$, 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 as follows:

- **KG**: according to the method of Kottek et al. (2006).
- **AI**: following the method of de Martonne (1926) 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.
- $PET > P$: sum of months with average PET (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 $> 3^\circ$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 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 interpolation of these cumulative drought duration distributions in such a way that each percentile (ranging from 1 to 100) has a value. The drought duration curves of all basins (or drought duration curve ensemble) for the USA and Europe is presented in Fig. 2a. In this study we only take into account long duration droughts; 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 (Fig. 2b, left). In a second step, these classes were used to group the drought duration curves into 5 different ensembles for each region (Fig. 2b, right). The minimum class size was set to 10 for both classes of climate classification systems and individual controls. Smaller classes 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 is presented in Table 1.

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 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 the number of percentiles with non-significant differences ($P \geq 0.05$) according to either the KS or MWU test (Eqs. 1 and 2).

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

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

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, 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 Köppen–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 average DDC...
of basins in non-seasonal snow climate with warm summers (Dfb). 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 temperate climate with cool summers (Cfc). The aridity index (AI) shows highest average DDC for basins in the lowest (most arid) AI classes for both regions. Generally, the average DDC decreases with increasing AI, 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 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, 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 (\(PET > P\)) displays an ordering of average DDC with a general pattern of higher average DDC for the basins with a high number of months \(PET > P\) and lower average DDC for basins with a low number of months \(PET > P\). Similar to the ordering of average DDC of the AI, the systematic ordering of average DDC (from high for low classes too low for high classes of \(PET > P\)) is occasionally interrupted due to an exchange between average DDC of neighboring classes.

Figure 3 (right two columns) presents the average DDC of basins grouped by individual controls. For individual control precipitation (\(P\)) of both regions, the class 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 temperature (\(T\)) in both regions show highest average DDC for both the lowest and highest temperature class. Longer drought events are thus found for basins with temperatures from the tails of the temperature distribution. However, differences in average DDC between different classes of \(T\) are not as distinct as for precipitation classes. Even smaller differences in average DDC are found for area (\(A\)). 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 ($E$) shows differences in ordering of average DDC between the two regions. For the USA, highest average DDC are displayed for the highest $E$ class whereas the highest average DDC of Europe are displayed for the lowest $E$ class. For the base flow index (BFI) in both regions; 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 USA and Europe. It reveals for the KG that basins in the Cfb climate in the USA have lower average DDC compared to Europe whereas the average DDC of basins in the Dfb and Dfc climate are higher for the USA. For the AI, 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 $> P$.

3.2 Statistical comparison

Figure 5 (left two columns) shows the results of the statistical comparison of DDC for the different climate classification systems. For the KG of the USA, DDC of basins in the Cfb climate are only similar with DDC of the Dfb climate according to both $S_{KS}$ and $S_{MWU}$. DDC of rivers in this Dfb climate show little similarity with DDC of the other climates indicating the distinction between shorter droughts for the non-seasonal warm summer Cfb and Dfb climates and longer droughts for the other climates. However, DDC of these other climates (Cfa, Dfa, Csb, Dfc, Dsb) mostly do not differ significantly among each other according to both tests. From these climates, the Dsb climate, which has the highest average DDC, is the most different and only shows similarity in DDC with the Dfc and Dsc climate (and at some percentiles with 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 according to both $S_{KS}$ and $S_{MWU}$. The differences in DDC between different AI classes is most distinct for the lowest AI classes. The higher the AI class, the more neighboring classes show similarity in DDC, whereas for these lower AI classes, only direct neighbors occasionally show similarity. The small differences in average DDC between different classes of $T < 0$ is reflected by the corresponding statistics, especially for Europe. For this region, DDC of almost all classes are similar to each other. Differences in DDC between different classes of $PET > P$ are mostly significant. $S_{KS}$ and $S_{MWU}$ indicate similarity only in DDC between neighboring classes.

Figure 5 (right two columns) displays the statistical comparison of DDC grouped by individual controls. Average DDC of 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, 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, confirming that long duration droughts in this region are significantly longer in both colder and warmer basins. Area ($A$) hardly shows significant differences in DDC between classes for both Europe and the USA. Basins of the highest $E$ class for the USA have significantly higher DDC compared to DDC of basins in the other $E$ classes of this region, whereas for Europe, basins of the lowest $E$ class have significantly higher DDC. For the BFI, DDC of different classes are often significantly different from each other 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 controls. Our result 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 controls, however, are not the same between the classes of different climate classification systems (Fig. 6), which in the end affects their overall suitability to differentiate basins according to drought duration.

For the KG climate classification system in the USA, the only climate that was not influenced by seasonality in precipitation nor the occurrence of a cold or hot season, Cfb, show the lowest average DDC (shortest droughts) and was only comparable with DDC in the Dfb climate. This Dfb climate was 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, it is not statistically significant when comparing the percentiles of the DDC.

The hot summer climates without seasonality in precipitation (Cfa, Dfa) have higher average DDC than their warm summer variations (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 the Dfa and Cfa in the USA. The differences in $P$ between the hot and warm
summer climates (Fig. 6) 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 the hot summer climates show above average DDC. However, their DDC show similarity with the 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.

KG climates that showed highest average DDC were 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 (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, 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 are the climate classification systems that take into account the dominant annual precipitation control (months with average potential evaporation larger than the precipitation (PET > P) and the aridity index (AI); note that the KG does not have such an annual precipitation term). PET > 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 > P that followed
the hypothesized pattern of higher DDC for the higher PET > P classes and lower DDC for the lower PET > 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 > P classes, however, basins located in the higher PET > P classes show significantly higher DDC for the USA. One possible explanation could be the difference in distribution of KG climates between these regions for the higher AI classes (Fig. 7). Basins located in high PET > P classes of Europe mainly are from the Cfb climate whereas basins of 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 > 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 the Thornthwaite formula 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 lower AI classes and lower average DDC for basins of the 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 Europe with the USA, 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 classes (Fig. 6). Differences in KG climates falling into the lowest three AI classes (Fig. 7) is more likely to explain this difference. The lower AI classes of Europe mainly encompass the Cfb climate whereas in the USA these classes mainly consist of the hot summer climates (Cfa, Dfa) that have showed longer drought durations.

Overall, different climate classifications systems have shown to be suitable to differentiate basins according to drought duration for 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 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. 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 > $P$). However, both these classification systems show differences in DDC between the same classes for Europe and the USA for low AI and high PET > $P$ classes. Combining information of the different climate classification systems and individual 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 of near-natural streamflow records. Based solely on observations means that basins in this data set 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.

Droughts were identified from the near-natural streamflow records using a drought identification method that 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 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 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 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., Fleig et al., 2006) or 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 influenced by seasonality 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.

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 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


de Martonne, E.: L’indice d’aridité, Bulletin de l’Association de géographes français, 3, 3–5, 1926. 12884


12900
Table 1. Classes of climate classification systems and individual controls and corresponding class sizes (USA/Europe).

<table>
<thead>
<tr>
<th>KG</th>
<th>AI</th>
<th>T &lt; 0</th>
<th>PET &gt; P</th>
<th>P</th>
<th>T</th>
<th>A</th>
<th>E</th>
<th>BFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dfb(114/15)</td>
<td>20–30(33/11)</td>
<td>0(184/118)</td>
<td>0(20/83)</td>
<td>1(68/94)</td>
<td>1(84/78)</td>
<td>1(87/75)</td>
<td>1(100/62)</td>
<td>1(134/29)</td>
</tr>
<tr>
<td>Cfb(48/247)</td>
<td>30–40(32/59)</td>
<td>1(31/30)</td>
<td>1(27/22)</td>
<td>2(75/86)</td>
<td>2(73/88)</td>
<td>2(77/84)</td>
<td>2(101/60)</td>
<td>2(110/50)</td>
</tr>
<tr>
<td>Cfa(156/--)</td>
<td>40–50(92/78)</td>
<td>2(14/33)</td>
<td>2(83/33)</td>
<td>3(98/64)</td>
<td>3(47/115)</td>
<td>3(77/85)</td>
<td>3(84/78)</td>
<td>3(67/95)</td>
</tr>
<tr>
<td>Dfc(29/49)</td>
<td>60–70(56/45)</td>
<td>4(46/18)</td>
<td>4(128/61)</td>
<td>5(105/57)</td>
<td>5(161/--)</td>
<td>5(115/47)</td>
<td>5(105/58)</td>
<td>5(80/83)</td>
</tr>
<tr>
<td>Dfa(35/--)</td>
<td>80–90(24/28)</td>
<td>≥ 6(22/25)</td>
<td>≥ 6(26/17)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Dsb(13/--)</td>
<td>90(63/52)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Ctc(--)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
Figure 1. Basin locations and two corresponding classifications (Köppen–Geiger and the aridity index), used in this study. A description of these two climate classification systems and their classes is presented in Sect. 2.1.
Controls on hydrologic drought duration in Europe and the USA

E. Tijdeman et al.

C1) Visual comparison

A) Construction of DDC

B) Grouping of DDC

C1) Visual comparison

C2) Statistical comparison

$S_{KS}$

$S_{MWU}$

Percentile

Duration (departure from the average in weeks)

Drought duration (weeks)

Class

similar

not similar

$S_{KS}$, $S_{MWU}$

KS, MWU

0

20

DDC Europe

DDC USA

Percentile
**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 Europe (red) and the USA (blue) and background colors indicating the class ranges for each equal size class. **(b, right):** corresponding exemplary ensembles of DDC groups for classes 1, 2 and 3 for the USA. **(c1):** Visualization of average DDC of the three exemplary classes displayed as departures from the total average of DDC of the USA. **(c2):** Statistical comparison of distributions of DDC per percentile at each percentile between 81 and 100 (displayed for percentile 81, 91 and 100 in the boxplots). Significance of differences in DDC 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.** Average DDC (displayed as departures from the total average of DDC of each region) of all classes of different climate classification systems and individual controls.
Figure 4. Upper graphs: Difference in average DDC for the same climate classification system classes in Europe and the USA (average DDC USA minus average DDC Europe). Colors correspond to the legend of Fig. 3. Lower rows: measures of statistical similarity $S_{KS}$ and $S_{MWU}$ between DDC per climate classification system class in the USA and Europe.
Figure 5. Measures of statistical similarity $S_{KS}$ and $S_{MWU}$ between DDC of all climate classification system classes and classes of individual controls for the USA (blue, above the diagonal of each matrix) and Europe (red, below the diagonal of each matrix).
**Figure 6.** Distribution of individual controls $P$ (upper row) and BFI (lower row) for the classes of different climate classification systems for the USA (blue) and Europe (red). Background colors indicate the ranges of classes of the individual controls (see Fig. 3). Box: percentile 25, 50 and 75. End of lines: percentile 5 and 95. Points: outliers.
Figure 7. Distribution of different KG climates for all basins with an AI smaller than 50 (left column) or PET > P of 5 or more months (right column) for both the USA and Europe.