Response to Henny van Lanen

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

In the study, cumulative drought duration distributions (DDCs) have been calculated using time series of observed daily streamflow from over 800 near-natural gauging stations in the USA and Europe. These DDCs were related to climate classification systems (e.g. Köppen–Geiger), climate data (e.g. long-term annual precipitation) and catchment characteristics (e.g. Base Flow Index) to investigate through a large-scale study too what extent duration of long droughts is influenced by climate and catchment controls. A visual comparison (ensemble average deviation from global average) and statistical comparison (Kolmogorov–Smirnov and Mann–Whitney U tests) have been carried out. Authors conclude that the Köppen–Geiger climate classification system does not sufficiently discriminate drought duration for climate types occurring in the USA and Europe. Individual controls have to be added. This large-scale study is important because it is based on observed flow data rather than on modelling experiments as in previous studies and it has led to interesting outcome on climate and catchment controls on drought duration. The manuscript is well written and the figures adequately support the text. It is worthwhile to be published in HESS, but improvements and some clarification are required before it can be accepted.

We thank the reviewer for his constructive feedback on the manuscript. We are grateful for his relevant and elaborate comments and suggestions on how to improve the manuscript. In this reply, we respond to each comment in order of appearance. The final implementation of the comments will be presented in the revised version of the manuscript.

Comment 1

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

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

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

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

The following differences exist between the two datasets:

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

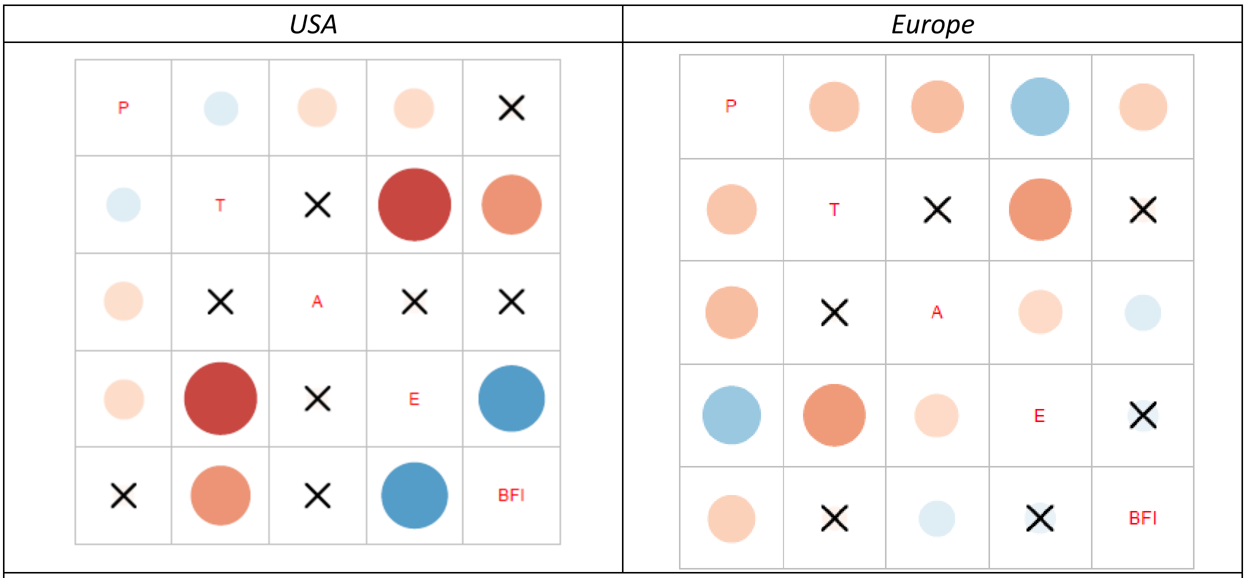

Figure 1: Correlation between individual controls for the two different regions. Size of the dots indicate the strenght of the correlation. Color indicates whether the correlation is positive (blue) or negative (red). Crosses represent non-significant correlations.

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

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

Overall, we think it is worth to present both the combined analysis and the analysis for each region. To accomplish this, we suggest a re-arrangement of figures (see example Fig.2). This involves replacing Figure 3 and 4 in the current version of the manuscript with a new Figure 3 for the different climate classification systems (KG, AI, T<0, PET>P) and a new Fig. 4 for the individual controls (P, T, A, E, BFI). These new figures will contain for each classification/control:

- *DDC for the combined dataset*
- *DDC for the USA*
- *DDC for Europe*
- *DDC USA minus DDC Europe*

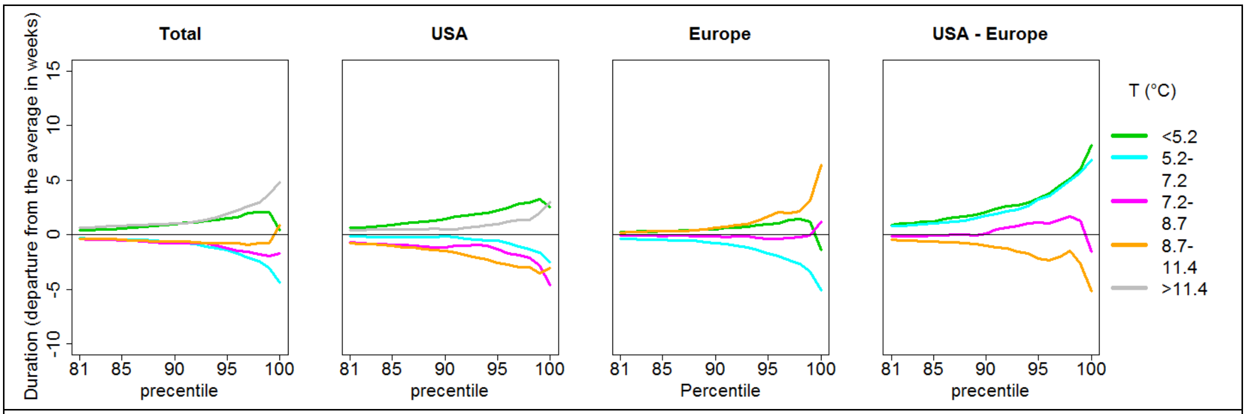

Figure 2: Exemplary row of new Fig. 3 and 4. (example for temperature)

Comment 2

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

We agree with this comment and will discuss the importance of deficit volumes in the new manuscript.

Comment 3a

Previous more limited studies (e.g. few catchments, only simulated flow) have shown that both climate and catchment properties control drought duration. This is confirmed by the current paper in the Abstract (pg. 12878, line 20), Discussion (pg. 12891, lines 5-11), Conclusions (pg. 12896, line 18). However, in the Discussion (pg. 12894, lines 6-8), it is suggested that climate classification systems only can be used to discriminate drought durations. This cannot be concluded based on the selected catchment characteristics.

This message was not intended. We agree that text is misleading and will clarify this in the revised manuscript.

Comment 3b

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

We agree that elevation and area do not directly or not necessarily describe storage related processes. Nevertheless, they are often used in regionalization approaches as proxies for controls on streamflow dynamics and indeed our study shows that they represent proxies for different controls *depending on geographical settings. In some areas, there is a relation between aridity and elevation and in others there is a relation between snow processes and elevation (Salinas, 2009). For a set of Austrian catchments, elevation was found to reflect seasonal storage in snow and glaciers (Van Loon, 2015). We also show that elevation plays a role, however, influenced by the contrasting correlation between precipitation and elevation presented in Fig. 1. Catchment area possibly provides an indication for the amount of storage within the catchment (e.g. Salinas, 2009). In the revised manuscript, we will better introduce the assumption behind the choice of control variables.*

Comment 3c

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

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

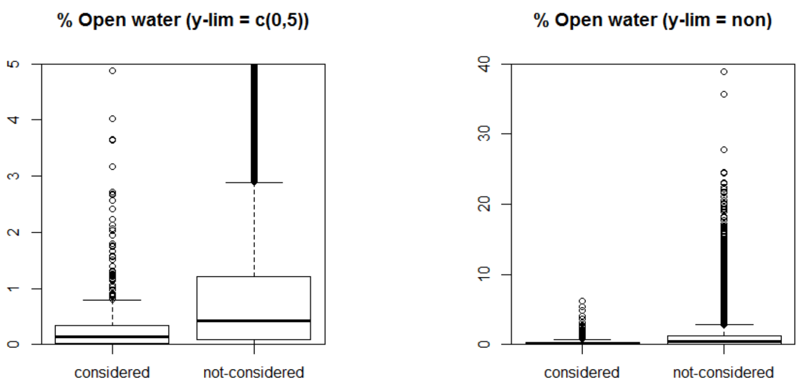

Figure 3. Fraction of open water for the considered basins with near-natural records and the not-considered basins (Left: y-axis limited by 5, right no y-axis limit).

Comment 3d

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

Good point. We will briefly add this issue to the discussion section.

Comment 3e

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

We share this opinion of the referee. In the revised version of the manuscript we will address this by:

- *the clarifications of our objectives and new figures that include the combined dataset*
- *improved explanation of the complementary role of individual control variables in the climate classification systems and the choice of the variables elevation and basin area*
- *discussing the possible effects of lakes and soils and representativeness of the basins with near-natural streamflow conditions.*

Comment 4

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

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

Comment 5

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

We will remove or revise these statements.

Minor comments:

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

Will be changed in the revised manuscript.

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

Will be removed.

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

Ok

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

Ok, see response to comment 3b.

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

461 for the USA, 347 for Europe. We will add this in the revised manuscript

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

Always 40 years of data. We will clarify this in the revised manuscript.

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

Since we compared two separated regions, we did not deem it necessary to reflect similar time periods. We see more value in including a larger number of stations while reflecting recent times. We tested if occurrence of long duration droughts was higher in the non-overlapping time periods (Fig. 4). This figure shows that 2006-2008 had a relatively large proportion of long duration droughts in the USA whereas 1965-1969 did not have notably more long duration droughts in Europe. We could briefly reflect on this in the discussion by adding references to the major drought events in the common and different time periods of the two regions.

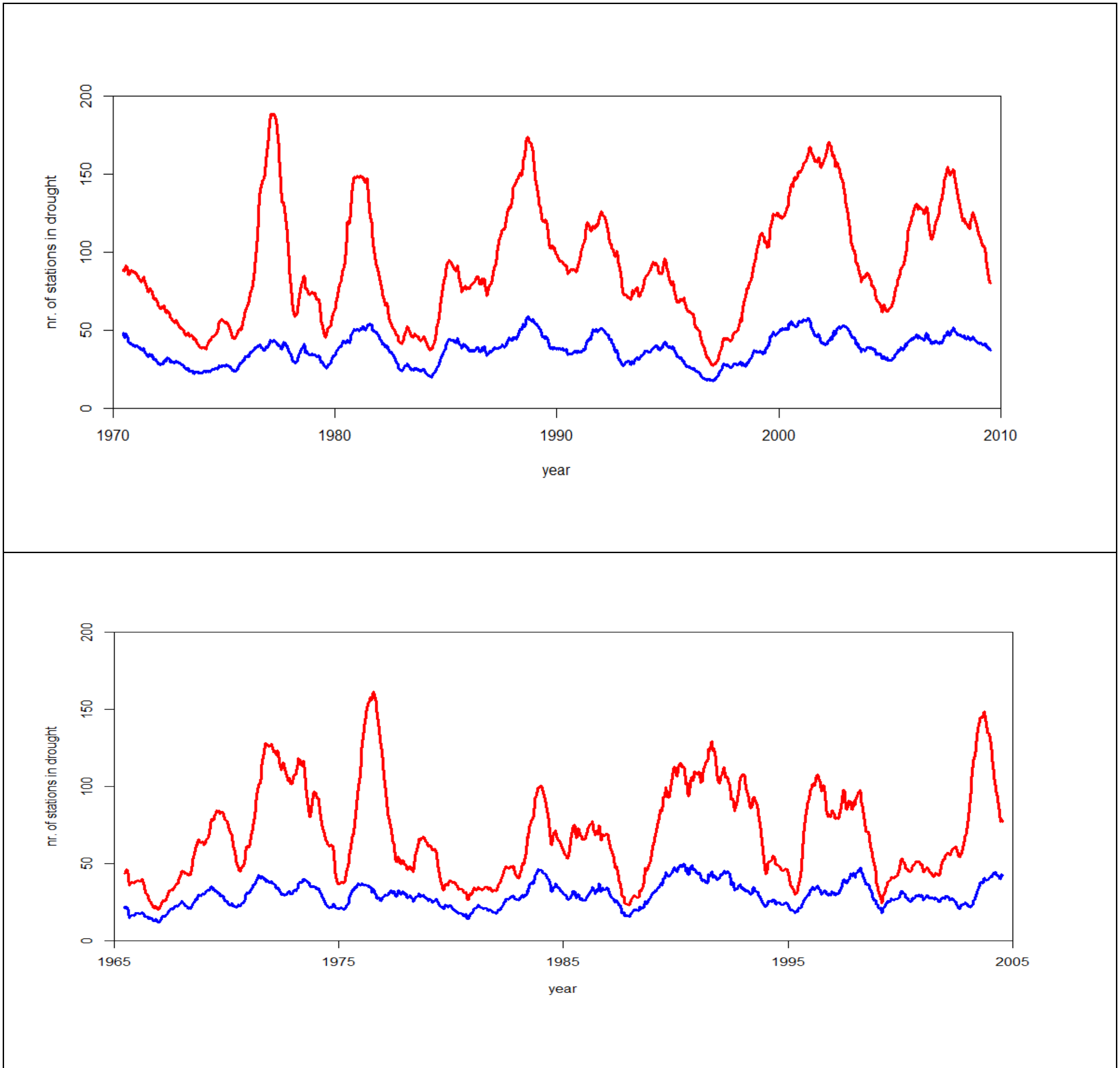

Figure 4: Number of station in drought for each calendar week (2080 week in total) for the USA (upper graph) and Europe (lower graph) after applying a moving average with a centered window of 52 weeks. Blue line = all droughts with a duration shorter than the Q81 drought duration. Red line = All droughts. Note that the difference between the blue and red line reflects the amount of stations in drought for the long duration droughts.

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

Will be defined.

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

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

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

Linear interpolation. Will be added to the revised manuscript.

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

We will add a sentence that describes this in the revised manuscript.

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

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

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

Yes, as suggested by reviewer 2, we will replace "equal class size" with "equal number of basins"

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

Yes, we will replace class size with number of basins.

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

Yes, class boundaries are based on the entire dataset. We will clarify this in the revised manuscript.

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

We will use this suggestion in the revised manuscript

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

We acknowledge the subjective decision to use the average DDC over the median DDC. However, differences in ranks between groups are covered by the statistical analysis.

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

We will take this suggestion into account for the revision of the manuscript.

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

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

- *remove Fig. 5 in the current version of the manuscript*
- *create two new figures to present the statistics (new figures 5 and 6)*
    - *one for the different climate classification systems (KG, AI, T<0, PET>P)*
    - *one for the individual controls (P, T, A, E, BFI)*
- *these new figures will contain for each control:*
    - *statistical similarity measures for the combined dataset*
    - *statistical similarity measures for the two regions*
    - *statistical similarity measures between the same classes of the USA versus Europe*
- *each control will be represented by one row of 4 figures (example in Fig. 5).*

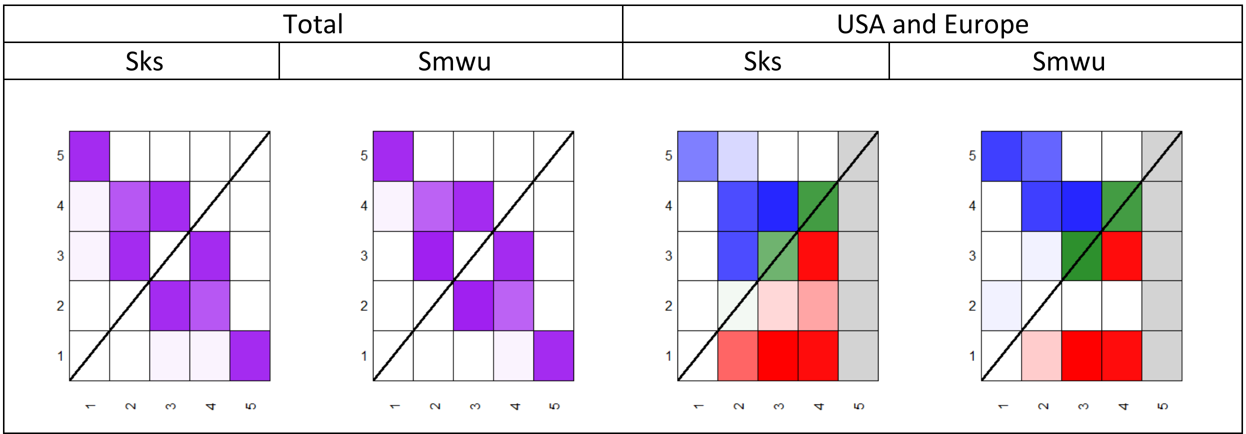

Figure 5: Exemplary row of the newly proposed Fig. 5 and 6 (example for temperature). Left two matrices are for the analyses of the entire dataset. Right two matrices are for the analysis of the datasets separated by region. The darker the color, the more percentiles are similar. Purple for the entire dataset, Blue for the USA, Red for Europe, Green for the comparison between Europe and the USA, grey indicates when a class is not considered (in this case, Temperature class 5 for Europe).

We realize that the right two figures are more difficult to interpret now and will therefore provide an exemplary figure in the legend of the newly proposed Fig. 5 of the manuscript (Fig. 6).

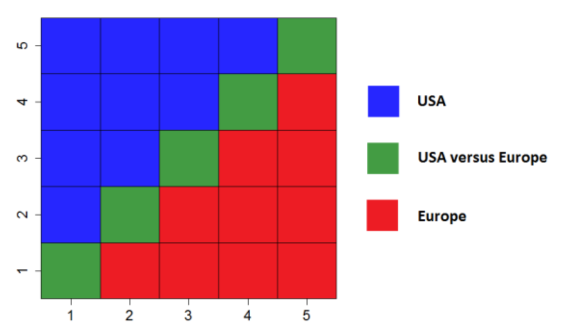

Figure 6: Exemplary figure to be included in the legend of the newly proposed Fig. 5 of the revised manuscript

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

Thanks for pointing out this mistake. We will correct this in the new version of the manuscript.

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

Ok

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

Ok

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

Ok

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

We will enlarge the smaller text of this figure.

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

We will leave it out in the revised manuscript

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

We will duplicate these labels and place them below or above the box plots.

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

Ok

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

We will change this in the revised manuscript.

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

We will replace this in the revised manuscript

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

See suggested changes to the figures in comment 1 and example in Fig. 2. Caption text will be improved.

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

See Fig. 5 and 6 of this reply and the corresponding reply to the related minor comment. We will improve the figure captions.

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

See Fig. 5 and 6 of this reply and the corresponding reply to the related minor comment.

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

We will apply the suggested changes to the new figure in the manuscript.

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

We will apply the suggested changes to the caption of this figure.

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

Ok

***References**:*

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

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