**RC2: Louise Mimeau**

The identification and classification of droughts events and the consideration of the spatial dynamics of events is both a novelty and an interesting contribution to the study of droughts at the global scale. I find it interesting that the authors have taken into account two different hydrological models to analyze the model uncertainty in their method. The paper is well written and can be suited for publication in HESS after a few major remarks are taken into account, mainly concerning a more in-depth discussion of the results.

Response: We would like to thank the reviewer for his in-depth assessment of our manuscript.

General Comments

I agree with the general comments of RC1 and I would like to add the following points in addition to his:

1. The paper would gain in clarity by regrouping parts 4.1 and 4.2 (the results obtained with SoilClim and mHM could be grouped together in the same figures and analyzed at the same time), and by discussing in greater depth the identifiaction of drought events with the OPTICS method (cf general comment n°2 from RC1). I also noticed in the supplementary materials that for each continent there are often distinct events with the same start and end dates (e.g. for the SoilClim model the events ranked 193 and 373 in S. America start at the end of April 1983 and end at the beginning of January 1984, or the events ranked 463, 532, 735 and 736 in N. America which all begin in August 2017 and end in November 2017). Can these events really be considered climatically distinct? What are the reasons for these distinctions between events? I think these questions should be addressed in the discussion section.

Response: The requested extension of discussion was done and is described below within the "line by line comments".

As we mentioned in reply to RC1, the aim of the article was to present enough robust methodology of GLDEs cataloging and document its employment for two independent data sources represented for both mHM and SoilClim models to demonstrate that the classification method works well for both models. The aim of the paper is not comparison of the two different model outputs. Therefore, we prefer to keep the models comparison separated by sections, except for Section 4.3 with Figure 15, which are specifically designated for this purpose. Moreover, merging SoilClim and mHM results into the same figures would make such resulting figures too extensive and complicated, same as their descriptions, that and particularly the describing text could confuse readers, regarding what concerns SoilClim and what mHM results.

Concerning the examples of GLDEs, occurring at the same time on the same continent, we acknowledge a certain level of uncertainty withing the clustering, which is inherent to any similar method. The algorithm clusters drought grids based on their spatiotemporal position and density-based clustering generally tends to merge areas with larger density into larger clusters and delimit more smaller clusters in areas with smaller density. It is not and cannot be the role of clustering to investigate whether drought in multiple nearby regions at the same time was caused by one large-scale climatic event or not. Looking for drivers of specific GLDEs could be a subsequent analysis performed using the catalogue. However, the clustering ensures some level of spatial separation of those events. And whether the separation is too strict or too loose cannot be conclusively determined by any subjective or objective method. To make this clear to the readers, we disclosed this also in the supplementary material itself.

2. I have a few reservations about the severity and dynamics indicators, which in my opinion are too interdependent and may have an impact on the classification. I suggest to replace some of these indicators with new ones (see line by line comments below) or at least to discuss the implications of selecting these indicators further in the discussion.

Response: Based on droughts delimited by some criteria and OPTICS method of the grid-point clustering, we selected indicators, which in our opinion expressed our aim the best, i.e, to classify the drought events from the spatiotemporal and dynamic points of view by the most complex way (corresponding to model outputs). Each drought classification can be based on different criteria selected by corresponding authors (see e.g., Spinoni et al., 2019 or He et al., 2020), why other authors could select other indicators. To select really independent indicators for such events we see as extremely difficult or impossible task. Because this referee comment does not obtain proposal on such totally independent indicators, it is really difficult to add any others, because we selected the best ones according to our opinion.

Spatiotemporal characteristics of investigated GLDEs will always be to certain extent interdependent, based on basic facts concerning drought occurrence. Very short droughts generally cannot develop into continental-scale events or move to a completely different location. Therefore, a complete independence between used characteristics was not possible and not aimed for. However, we tried to include all main aspects of GLDEs within chosen characteristics. Reviewers specific comments concerning these characteristics are answered below within the “line by line comments” as well as the extension of the discussion section mentioned here.

3. Presenting some of the results in map form would help showing the severity and dynamics of the drought events. In particular, it would be interesting to show maps with the maximum spatial extent and the trajectory of the centroids for some of the most extreme events or for the events mentionned in the discussion section (e.g. L395, L402, L463).

Response: We understand the reviewer comment, but please consider that we are working in 10-day steps, which would mean that we should have 36 maps a year for one particular drought event. However, to illustrate the spatial nature of GLDEs, we created maps of the maximum extent and centroid movements of the three most extreme SoilClim-based GLDEs (the top 3 in Table 1) and an animation showing the development of the most severe GLDE during one year and included that all into the Supplement.

Line by line comments

L refers to line and P refers to page.

P3L74 : Please provide references for LAI, landuse and terrain inputs used for the modelling with SoilClim model.

Response: Accepted, the paragraph was extended as follows:

… SoilClim was applied to each grid with a daily input of meteorological variables that consisted of precipitation, temperature at 2 m above the ground, dew-point temperature at 2 m, wind speed at 10 m, and incoming shortwave radiation, which originates from ERA5-Land (Muñoz-Sabater et al., 2021), as well as with the leaf area index (LAI), land use and terrain inputs, also taken from the ERA5-Land dataset. The plant-available water capacity …

P3L86 : Why are the SoilClim and mHM models forced with two different datasets for the daily meteorological inputs (ERA5-Land and ERA5) ? Using the same inputs would make it easier to compare results obtained with SoilClim and mHM.

Response: The mHM is much more computationally demanding, therefore going for the higher resolution of ERA5-Land was not practical. Also it allowed slight perturbation in the modelling inputs as this particular study has not aimed at pure comparison of two different models i.e. SoilClim and mHM but to develop robust catalogue of soil drought episodes. It also needs to be noted, that the models were developed by different teams. However, both model outputs were rescaled to 0.5 ° resolution for the clustering process and as ERA5-Land is mostly just land-surface dynamic downscaling of ERA5, there is no reason to worry about huge differences in forcing which would manifest on this lower resolution.

P4L105 : Please provide the parameter values for the OPTICS clustering.

Response: In the the dbscan::optics() function in R, final parameters were: eps = 5, minPts = 20 and in dbscan::extractXi() function for cluster extraction, chosen parametr xi = 0.001.

P4L106 : How many clusters were removed from this filtering ? This information would be useful for analyzing the fraction of identified droughts that are local events versus the fraction of drought events on a broader continental scale.

Response: Accepted, following sentence was added into the manuscript:

… To eliminate regional cases with very small drought-affected areas, clusters that included fewer than 50 grids for one 10-day interval, fewer than 500 grids overall or that appeared in less than three 10-day intervals were excluded from both datasets, after which 775 clusters (further drought events) remained in the SoilClim dataset and 630 remained in the mHM dataset. In the case of both models, selected GLDEs comprise of around 15% of original clusters, however including over 80% of all grids inputted into the clustering. …

P5L148 : The relationships between the severity characteristics seem to be due to the fact that some the characteristics are inter-dependant. Characteristics b and d are directly related to c : the longer the event, the higher is the total sum of areal extents. b should perhaps be replaced with the averaged areal extent of a drought event during its duration and d with the average fraction of the drought events area with a AWR/SM value under the 2nd-percentile threshold.

Response: This problem we commented already in the response to your point 2. Here we add that severity characteristics were designed to represent spatiotemporal extent of GLDEs, i.e. their complete independence was not possible and not aimed for. Considering the reviewers suggestion, it would basically mean dividing (b) and (d) characteristics by the (c) characteristic. However, having 2 out of 4 characteristics as mean values may downplay the impact that long-term drought events have on affected areas, which is why we used basically cumulative values for characteristics (b) and (d). And since this classification is specifically focused on spatiotemporal severity, we would prefer to keep the current characteristics.

Figure 4, 5, 6, 7, 11, 12, 13, 14, 15 : Please use different color scales for dynamic classification to avoid confusion with the severity classification.

Response: The 7-color scale was carefully prepared in accordance with the journal's policy to use CVD-friendly scales for all figures. We would prefer to keep it as it is, since the classifications are clearly distinguished by the category names (7s / 7d) and axes names, however if the reviewer or the editor insist, we will try to prepare a second 7-color scale that also sufficiently CVD-friendly.

P13L244 : Can the authors explain this relationship between categories S and D: why events with a wide spatial and temporal coverage are also the most dynamic? I believe this might due to the method used to calculate the dynamics characteristics, which are dependent on the duration of the event (especially for the indicator a: the longer the event, the greater the sum of the distances) and the spatial extent of the event (an event can be very extensive spatially and relatively static, but because of its spatial extent, a small shift in the centroid can give a large absolute distance). The dynamics characteristics should rather be computed as ratios between distances between the centroids and the width of the spatial extent (or number of grid cells) of the events and averaged per time interval.

Response: Characteristic (c) in the dynamic classification is "The mean geographic distance between centroid positions", so it is already averaged by time, as the reviewer suggests.

Characteristic (b) is also not cumulated over time, since it is (one) maximum distance between any couple of all centroid positions. So, if the centroid is e.g. just oscillating back and forth in a small area, this value remains low and will not cumulate over time. Of course, long-term droughts have a higher potential to actually move to a completely different location, but that is exactly one of the behavior patterns (maybe even the most interesting one) we wanted to comprehend by the dynamic classification.

So, this leaves only the characteristic (a) cumulating over time as the reviewer pointed out. However, even droughts that might experience some kind of "pulsing", or moving back and forth in a small area for a longer time period may be considered dynamic from a certain standpoint, compared to actually completely static droughts. Therefore, we would prefer to keep this characteristic in the dynamic classification.

Figure 7 and 14 : Please clarify axis labels (e.g. Severity category or Dynamic category)

Response: Accepted and corrected

Figure 15 : A clear separation between the 5 continents (a and b) and categories (c) would make the figure easier to read.

Response: Accepted and corrected.

P24L385 : Could changes in land cover or irrigation, which are not taken into account in the modelling, also be sources of uncertainty and have an impact on the classification of drought events?

Response: Yes, in some areas the landuse changes would be significant but rarely on the scale considered in the project. However, we added these factors into the list of uncertainties.

P24L390 : The authors should discuss in more detail the sensitivity of drought event identification to OPTICS clustering parameters.

Response: Accepted, following text was added into the paragraph:

… However, specific settings of the algorithm parameters must be partially derived empirically to fit the characteristics of a given dataset (particularly concerning its density) and prevent either the connection of all objects (in our case, grids in individual 10-day intervals) into one large cluster, including the whole dataset, or the failure of the algorithm to create clusters at all if the clustering parameters are too strict. Aside of these extreme cases, when the clustering failed, smaller changes of the parameters lead only to minor changes in the event delimitation. Still, there is no objective method for defining "perfect" parameters; hence, clustering uncertainty is inherent and affects the length of existence of individual clusters in our dataset. …

P26 Table 5 : To make the table easier to read, please replace A and B in the table respectively with 1980-2000 and 2000-2020, Response: Accepted and corrected

and perhaps just show the relative frequencies instead of showing both absolute and relative frequencies. Response: If in both reviews appeared several times request on comparison of both mHM and SoilClim models, then we see as important to preserve both absolute and relative frequencies in this table in order to demonstrate differences or agreement between these two datasets.

P26L455 : It should be pointed out in the discussion that this statistical analysis over two 20-year periods is a little short-sighted for identifying trends (especially when some drought events can last several years).

Response: Accepted, following text was added into the paragraph:

To demonstrate this situation, Table 5 shows a comparison of the absolute and relative numbers of GLDEs during 1980–2000 and 2001–2022. As these periods are from a climatological point of view relatively short, the comparation should be taken with caution. Except for category 1d in the dynamic classification, …

L553, L590, L651, L663, L699 : Some doi or url are missing in the references.

Response: Accepted and corrected.

L964 : A line break is missing before Vincente-Serrano et al, 2022

Response: Accepted and corrected.