Review of "The impact of elevation and flow dynamics on hydrological drought and wet spell characteristics in semi-arid southeast Arizona" by Lu et al.

The authors modify the combined threshold level method and consecutive dry period method, originally proposed by Van Huijgevoort (2012), to be more applicable for regions with strong precipitation seasonality (e.g., monsoon season). In the modified version of the method, consecutive dry / drought day values were compared with the same variable for all other years for each calendar day seperately (and not with events as was the case for the old method). By doing so, the drought identification method allows droughts to persist from the wet to the dry season. Additionally, the drought identification method results in a uniform percentile distribution for each calendar day. The authors applied the method to (tributaries of) a case study river in Arizona (US), and reveal more satisfying identification of streamflow drought events, as well as clear differences in drought occurrence and characteristic between the main river branch (without base flow) and upstream tributaries for which base flow plays a more important role.

In general, I like the modification of the methods as it allows for a more consistent comparison of drought over space and time, e.g., the 20th percentile threshold is exceeded 20% of time for each DOY and (sub-)catchments. On the other hand, there is one aspect of the methods that I did not find very satisfying, and I am curious to the opinion of the authors. In addition, there are some other major and quite some minor comments that should be addressed before the manuscript is suitable for publication.

Major comments.

1. The proposed method and modifications are definitely an improvement. However, there might be one more point of consideration. With the suggested method, you kind of force droughts to start in the flow season. Droughts starting in the dry season often have to "catch-up" with droughts starting in the flow season, as their consecutive number of drought/dry days is lacking behind. However, even when their consecutive dry day number is lacking behind, their total period of zero flow might last way longer then the drought that started in the wet/flow season. The official start of the lacking-behind dry season drought can only occur when some earlier starting droughts (in other years) are terminated. In other words, the official start point of a drought starting in the dry season might be after a prolonged period of zero discharge. I believe that a more optimal definition would take into account the total period of zero flow, and:
    - Encourage the authors to come up with another modification of the method that somehow take into account the total length of the dry spell or provide an argument why the do not think this is needed.
    - *An option might be to assign each zero flow day the value of the total length of the dry period. So, if you have a dry period of 10 days, you give each day a value of 10, rather than a value from 1:10. You can do a similar percentile ranking based on these values. This might work….*
    - In any case, the authors should definitely investigate how long a period of zero flow can occur before a "drought" starts and discuss this.
2. I engorge the authors to be more careful with the use of the wording "hydrological extremes" throughout the manuscript. They mostly look at **anomalies**, not extremes. This is an important distinction to make upfront (especially with regard to the communication of the results to a broader audience), especially because some of the "extremes" are not "as extreme" (see next comment)
3. Figure 6 (upper panel) reveals that only a very small discharge event (only visible when completely zoomed-in) triggers the identification of a wet period. This period then last for a

30 days as the authors apply a 30-day moving average. I agree that this one-day <0.1 mm discharge event can be abnormal, but to classify this entire period as wet might not be the most suitable.

   a. The authors should discuss this, especially in the context of the term "hydrological extremes" used throughout the manuscript.
   b. In addition, droughts might develop more gradually then wet periods. Could the authors comment on whether the use of a 30-day moving average window is equally suitable to both drought and wet period?

4. I do not agree that one should definitely correct for human induced step changes and trends for drought analyses (Section 3.3). From an instream perspective, human induced changes and modifications in drought characteristics can be of critical importance as well. In addition, the proposed method of recalculating drought characteristics for different reference periods (before and after change points, where the change point might vary depending on the river), might hinder a fair comparison of drought characteristics over space and time, as each time a different reference period is used. So, in case the authors want to correct for trends and step changes, it should be better justified why (in Section 3.3), and the consequences should be discussed.
5. I like the consistency of the proposed method, i.e., the 20th percentile is exceeded 20% of the time. However, later on, the authors apply different pooling approaches. Did the authors test how these approaches modify the amount of time anomalies occur for each catchment and DOY?

Minor comments:

The manuscript requires some more thorough editing. Suggestions (and some other points) in the minor comments below.

- **Title:** a major part of the paper deals with modifying the combined TLM and CDPM. I think this should somehow be reflected in the title.
- Line 3: regularly occurring no flow conditions instead of "no flow conditions"
- Line 5: The wording "hydrological extremes". I would be very careful with this wording throughout the manuscript. You mainly identify anomalies, not extremes.
- Line 16: the abstract does not mention anything specific about the impact of elevation although it is in the title. Maybe it can be added here.
- Line 17: correct the spelling of "drought and wet spells patterns"
- Line 18: wet spell**s** and drought**s** instead of "wet spell and drought are rare".
- Introduction: I strongly encourage the authors to review some more recent literature, especially about the application of the combined method between 2012 and now.
- Introduction: be more careful in your referencing, e.g.: Line 25: The sentence about the costs of drought with a reference to Dai (2011) is also in the Van Huijgenvoort paper. However, they refer to a different contribution of Dai (2011).
- Line 38: "severe dry spells impacting the land surface". You could be clearer here.
- Line 38: Where does "This" refer to?
- Line 41: if you focus on both dry and wet spells, it would be good to give them equal amount of attention. Now most focus on droughts.
- Line 43: remove "(" after e.g.
- Line 50: You could add to this argument that some studies kick-out stations with regularly, occurring zero flows. (I am guilty as well; Tijdeman et al., 2016).
- Line 51: "incorrect higher threshold". Why is a higher threshold incorrect? What is a correct threshold?

- Line 54: if true, the authors could add that the CDPM was developed for precipitation.
- Line 56: The main problem of the method was the difficulty to characterize the transition from flow to low-flow periods, or? Not the identification of the start.
- Line 61: elevation is not a direct control. In this study, elevation encompasses a variety of different controls such related to snow, the occurrence of groundwater, and the occurrence of local rainfall events.
- Line 68: "To limit the impact of human influences" I would remove this from the sentence as it reads as one of the goals from the study.
- Line 82: suggest to remove "detailed"
- Line 83: suggest to remove "detailed"
- Line 86: "as explained in the introduction" If you explain something in the introduction, you do not have to repeat.
- Line 92: "300-350 mm of precipitation" → per year?
- Line 110: Be consistent. You use both stream flow and streamflow.
- Line 118: spell instead of "spells"
- Line 134: you could also refer to the threshold level method here.
- Line 135: you could explain why these methods do not work well in these domains.
- Line 138: More careful phrasing: I doubt that this is the most widely used method, especially not for wet "hydrological extremes".
- Line 139-141: More careful phrasing needed as this is not nescecairly true. Some (of the mentioned) studies methods use daily flow values to define a monthly or annual threshold (e.g., fixed threshold).
- Line 143: All referenced studies are drought studies. Any example of the use of the TLM to identify wet spells?
- Line 143-144: the threshold level method can never distinguish between a dry period and drought. The problem with regularly occurring zero flows:
    - Droughts are indicated for all zero flows or for none of the zero flows, depending how you define your quantiles.
    - It is quite complicated to rank zero flow events, i.e., how do you rank five zeros (see Stagge et al. (2015) how they deal with zero precipitation in the SPI).
    - These problems should be presented more clearly (either here or in the introduction), as you are going to solve them later.
- Line 162: I would describe the rescaling method more clearly.
- Line 164: ">20%" this is not a wet spell.
- Line 166: "Major advancement" → Any references to studies that applied this method would be nice her. Here, or in the introduction.
- Line 204: why 25 and 75? Why not 20 and 80 (your drought and wet spell thresholds).
- Line 182-187: Maybe use bullet points here as well. This would highlight the modifications you made.
- In general, I would strongly encourage the authors to use some variables and equations in both the new and modified method description (as is done in the prev. paper, where it increased the clarity).
- Section 3.3. A reference to Sadri et al., 2016 (HESS) would fit nice here (uses same kind of tests to identify human modifications of low flow).
- Line 200: "as indicated in the introduction …" -> repetitive and not needed.
- Line 202: "is necessary to identify and correct for these impacts" do not agree (see major comment (3). Please elaborate why.
- Line 224: "The previous section presented …" → suggest to delete.

- One missing first result section would be the evaluation of the newly proposed method / comparison between the new and old method. (4.1). Now, this is more or less done in the methods section. However, I believe this better fits the results.
- Line 231: The fact that you apply a moving average on your data should be presented earlier (before the introduction of the combined method).
- Line 233-234: Did the authors recalculate the threshold after applying a different moving average interval?
- Line 235: Introduce abbreviation "PDSI"
- Define duration in the methods and be consistent. Is this the total time in drought for a certain year (as x-axis Figure 4c) or the length of the drought / dry spell (e.g. Figure 8).
- There are many unnecessary repetitions at the start of each new section (e.g., the first five lines of 4.3). I would suggest deleting these to improve readability.
- How are probabilities in Figure 8, 9, 10 etc. calculated. And can the authors confirm that the dips in probabilities at the left and right side of Figure 8b,e are actually there? And not an artefact of any kind of smoothing that does not take into account observation before and after the plotted period?
- Line 251-255: This is repetition and could be deleted.
- Line 257: As indicated in Section 2.2 … (repetition).
- Line 258: spells instead of "spell"
- Line 258: I could not find a reference to Figure 5 before (and after) you introduce Figure 6.
- Line 268: Discharge instead of "Moisture"
- Line 285: For both stations?
- Line 301: For the representative station or for all stations?
- Line 304: I would not call a 100 days "very short"
- Line 346-350: This is all repetition, which the authors could consider to remove.
- Line 355-356: to much brackets, especially to much ")" …
- Line 368: Use of wording "hydrological extremes" (see earlier comment)
- Results general: there is some discussion in the results (e.g., 295-300). I would suggest the authors to more strictly separate.
- Discussion general: too much repletion of the results and too less interpretation. For example, various other studies show the effect of pooling and smoothing, and the authors could relate their findings to these works (+ a recommendation which smoothing method to use). In addition, there are also quite some new results in the discussion. I suggest separating.
- Line 408: "impact of elevation and flow conditions." Elevation does not have a direct impact (and "flow conditions" is a bit of a vague term).
- Line 426: "dominant driver" of what?
- Line 432: "short duration hydrological extremes" … again not extreme. Anomaly.
- Figure 2: You might add the trend from 1960-onward to the legend to show that it is less or absent.
- Figure 2: Location nr. 2 has only data from 1967 –onward (according to table 1). Location 3 or error in table 1?
- Figure 3: I have difficulties to compare the timeseries in the panel plots. It might be better to separate them? And maybe use a log-scale, so you can see higher flows as well?
- Figure 2,3,4,5: be consistent. You now use discharge (mm) vs. runoff (mm) vs. discharge (m3/sec) vs. moving average discharged(?) in Figure 5.
- Figure 4c, x-axis: duration or days in drought?
- Figure 4, caption: remove ")"
- Figure 7: I would not use green for wet spells (subjective opinion). Maybe light blue as is done in the USGS WaterWatch? https://waterwatch.usgs.gov/

- Figure 8: I would separate the probability plots from the main plots, as these are not really related.
- Figure 8: do the lower probabilities on the right and left side of the probability plot indicate a lower probability of occurrence (e.g., left of Fig. e)? Or is this some kind of a smoothing effect (due to a possible absence of points at the start and the end of the year).
- Figure 8: modify the plotting area to fit all observations.
- Figure 8-11: What is the difference between the probability (8-9) and density (10-11)?
- Figure 11: Why a MA period of 1 year? If you apply a one-year MA, you probably remove all zero values and are comparing results for the TLM.
- Figure 12: Any pooling applied here?
- Figure 13: For similar reasons: why a 1-year MA average window.

Stagge, James H., et al. "Candidate distributions for climatological drought indices (SPI and SPEI)." International Journal of Climatology 35.13 (2015): 4027-4040.

Sadri, S., Kam, J., and Sheffield, J.: Nonstationarity of low flows and their timing in the eastern United States, Hydrol. Earth Syst. Sci., 20, 633–649, https://doi.org/10.5194/hess-20-633-2016, 2016.

Tijdeman, E., Bachmair, S., and Stahl, K.: Controls on hydrologic drought duration in near-natural streamflow in Europe and the USA, Hydrol. Earth Syst. Sci., 20, 4043–4059, https://doi.org/10.5194/hess-20-4043-2016, 2016.