Review of "Streamflow drought: implication of drought definitions and its application for drought forecasting" by Sutanto and Van Lanen.

I thank the authors for their detailed reply to my comments. The authors greatly improved the manuscript. The comparison between daily and monthly resolution is a nice addition. The figures, especially the maps, are nice! However, few points remain.

Below you find a more detailed list of comments. In summary, the drought identification methods should be better explained and discussed. With the scope of your, you are setting an example for the drought forecasting community. Further, I think the discussion can focus a bit more on the implications of the results. You find many differences among methods and conclude that, based on these differences, end-users should agree upon a sharp drought definition. What would be nice is to have some more discussion on this. For example: What are the (dis-)advantages of the different methods? Which end-user would benefit from a fixed vs. variable approach. Who might be interested in the SSI over the VTM? And who would prefer daily instead of monthly data. Finally, I think that the readability of the manuscript can be increased by being more consistent in the used terminology. Use the same wording when describing the same thing (e.g. the wording you use to describe drought properties N, T, D etc., or the wording used to describe the different approaches).

Comments and suggestions:

- Line 6: "the differences of streamflow droughts using different identification approaches" → unclear, rephrase.
- Line 12: "the Standardized Streamflow Index". I do not think this is the accurate description of the approach. I would refer to it as the threshold level method applied on SSI time series
- Line 13,14: why define acronyms VTs and FTs in the abstract.
- Line 18: "Overall, the characteristics of SSI-1 drought are more or less similar to what is being identified by the monthly threshold approaches (FTM and VTM)."I am a bit surprised that SSI and FTM are the same. Especially because SSI and VTM should be very similar, and FTM and VTM show differences.
- Line 21: "To the end" should be "in the end"
- Line 39: Could remove brackets here
- Line 46: Write out to what "these" refers.
- Line 49: "Should be not" → Should not be sounds more natural to my non-native ears.
- Line 55: "which is defined as … below normal" → add "the forecasting of " on the place of the dots.
- Line 59-60: "which measures monthly normalized anomalies in streamflow and" → would at a bit more detail here, e.g., the SSI is a probabilistic index.
- Line 74: "drought indices" → drought indices is confusing here (refers to SPI, SPEI etc.)
- Line 78: "data" → streamflow data?
- Line 82: "its" → refers to nothing.
- Line 88: "results" → results and discussion sections.
- Line 98: "daily proxies for observed streamflow" → From your reply, I get why you use this terminology. However, either use it consistently throughout the manuscript, or do something like: daily proxies for observed streamflow (hereafter referred to as just streamflow for brevity reasons) ….
- Line 107: "river streamflow" → just streamflow
- Line 124: "threshold drought approach" → threshold level method. Please use this (or similar) terminology consistently throughout the manuscript.
- Line 125: "Standardized drought approach" → threshold level method applied on SSI timeseries. Please use this (or similar) terminology consistently throughout the manuscript.
- Line 128: "the water deficit in different domains of the water cycle, in our case, it is the" → redundant. Could delete.
- Line 130: ref to the original work of Zelenhasić & Salvai (1987) would fit here well.

- Line 144: How was the data aggregated: Sum or mean?
- Line 146-147: "The Q80 was considered as the drought threshold because most of the rivers across Europe are classified as perennial rivers." → I would remove this sentence. Why would one use a different threshold for Intermittent Rivers? And not a different drought identification approach (e.g. Van Huijgevoort et al. 2012)? The next sentence provides enough justification of why Q80.
- Line 149: "fewer drought events" → Nitpicking here, but not necessarily true: Q70 could also mean that few minor Q80 droughts are pooled together in one larger Q70 event.
- Line 150: "be straightforwardly be" → remove one be.
- Section 2.1. It is still not clearly described how the daily threshold is arrived.
- "whereas for the VTD method, the calculated monthly thresholds were firstly assigned as the threshold levels for each day of the respective months" → Is this correct? Isn't the VTD usually derived from daily data of the flow duration curve within a certain month?
    - If it is correct (I guess so after reading 4c). please discuss that a threshold derived from monthly data might be different from a threshold derived from daily data.
    - If it is not correct: Please clarify.
    - Didn't the cited study of Beyene et al. (2014) find that other threshold smoothing procedures were more suitable for e.g., highly seasonal (snow) regimes? Please discuss.
- Line 178: How did you estimate the parameters of the gamma distribution? L-moments, Maximum likelihood estimation, or a combination of the both. Please add.
- Section 2.2.2. You study sets an example for a broad community. This is obviously a good thing! However, I feel certain topics should be more carefully explained and discussed.
    - Please provide a bit more background about the SSI, e.g. it is a probability index, it has certain assumptions, it has uncertainties etc.
    - Ok – you use the gamma distribution, fine. However, what I would highly encourage is to include one more map to the supplementary material that shows the suitability of this distribution across all rivers. For example, you can derive a goodness of fit metric (Shapiro-Wilk, KS or something else) and show for each river how many months pass this this goodness of fit metric. Also, please discuss that other distributions might be more suitable for streamflow. Testing goodness of fit is a regularly ignored, but essential step, before using any standardized drought index.
- Line 225-226: "Obviously, the average deficit volume in a river grid cell, which we use in the historic analysis, equals the total deficit divided by the number of droughts." → Suggest to delete this sentence as it is indeed obvious.
- Line261-262: "This happens when the streamflow falls below the threshold, which is Q80 (VTs and FTs) or equal to SSI<-0.84 in our study." → consider deleting.
- Line 305. Also negatively correlated for the threshold level method applied on the SSI.
- Line 346: "This precipitation had a more marked effect on the SSI-1 drought than on the VTM and FTM droughts." → Visually, yes. But you are kind of comparing apples and pears, i.e., changes in absolute flow versus changes in SSI (standard normal distribution). Would remove or rephrase.
- Line 348: "the SSI-1 …" → and VTM
- Line 360-361: "(Tallaksen and Van Lanen, 2004)" what is this reference doing here? They give me a definition of multi-year drought? Might remove.
- Line 365-375; You might have a look for the work of Vicente-Serrano, as he did a lot of research to both the Ebro Basin and SSI.
- "The SSI-1 droughts follow the pattern of VTM droughts" this comparison is done a few times. You might at somewhere that these results are expected, given that the SSI and VTM are very similar metric (only difference is the probability distribution fitting step).
- Line 396-399: Nested sentence – difficult to follow.
- Section 3.1: general – What I miss a bit in this section is the interpretation of the results. For example: there are earlier / longer / more minor droughts with this method as compared to the other method. So what? Why and for who is this important? And which drought monitoring and early warning application benefits more from a VT as compared to a FT method and vice versa.
- Line 538: "fixed" → variable

- Line 543: “The start of SSI-1 droughts is closest to VTM droughts” as expected (see above).
- Line 555: “The differences in drought frequency, average duration, timing, and deficit volumes
- Between VT droughts (incl. SSI-1) and FT droughts highlight the importance of whether end-users of drought forecasts should take seasonality into account or not.” Good point! But what would strengthen it are some examples of end-users that might benefit from a fixed versus variable drought definition.
- “forecast both standardized-based and threshold-based drought indices.” → Nice! Do they also forecast using fixed and variable thresholds?
- “based upon the provided description of the identification method and product.” → Nice point again – but could pick-up on this point a bit more in the discussion. I think it is even more crucial to provide accurate guidance with interpretation than a bunch of different products.

Figures:

- General: very nice maps!
- Fig 1. Mention the four basins in the caption.
- Fig 2. (caption) “Drought occurrences” → number of drought occurrence (consistent with 2.3)
- Fig 3. “Months when drought mostly started” → drought initiation time
- Fig 4-5. Please ignore if it does not make sense – but just wanted to note that I found red a more logic color for overlapping deficits (not orange). Or maybe you can do purple instead of orange for overlap? (red + blue = purple).
- Fig 7: “Section 2.3 explains how the drought timing is determined using forecast data.” → Not needed.
- Fig 7. Why not connect all ensemble members to last observed month?
- Fig 7. Add “threshold” to the legend of Figure e-f.
- Table 1: explain why no timing (T) in table 1 for Europe.
- Figure S1: color scales are not matching, i.e., red color 90 days and 5 months