Reply to reviewer 1

We would like to thank the reviewer for valuable suggestions and comments. In this document, **P** refers to the page number and **L** refers to the line number in the recent paper. For example, **P3L65-70**, refers to page 3, lines 65-70.

Reviewer 1		
No	**Comment**	**Reply**
1	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!	We thank the reviewer for complements to our revised paper. This could be done because of the valuable suggestions from the reviewers.
2	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.	We thank the reviewer for these valuable suggestions. We added text to better explain the drought identification methods (see point AA, AB below). Furthermore, we added discussion about the advantages and disadvantages of the different drought identification methods in the revised manuscript (**P16L525-P17L549**). We also included thoughts about which end user could benefit from each of these methods (VTD, FTD, VTM, FTM, and SSI-1) (**P17L550-P18L566**).
3	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)	The reviewer has a point here. We changed the inconsistency in terminology used in our manuscript. We believe that the revised manuscript now has a consistent terminology throughout the text.
4	Line by line comments	
A	Line 6: "the differences of streamflow droughts using different identification approaches" -> unclear, rephrase.	We revised the sentence into ".....overview of the differences between different drought identification approaches to identify droughts in the European rivers,…… (**P1L6-7**)"
B	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	We are sorry that we confused the reviewer by using a misleading term, i.e. threshold, for a standardized approach, i.e. the SSI (whether the river is in drought or not according the SSI-1 time series). By using the suggested phrasing, we believe that the reader will be confused about the distinction between the threshold methods and the standardized methods, as introduced in the literature (Van Loon, 2015). Keeping in mind that we would like to have a clear distinction between the two methods. We will not use "threshold level" in the context of the

		standardized methods. Hence, we replaced at relevant places the "threshold level" by "limit value" (**P7L196**). So rivers are in drought according to the SSI-1 when the limit value is below -0.84.
C	Line 13,14: why define acronyms VTs and FTs in the abstract.	We believe the reviewer means L16-17. We removed the definition of the acronyms VTs and FTs in the revised manuscript (**P1L16-17).** These have been already defined earlier in the Abstract.
D	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.	We said here "Overall". If we refer to Figure 2 (drought occurrences) and Figure B2 (drought duration), then it is hard to distinguish the difference between VTM, FTM, and SSI-1. We can see a clear difference between SSI and FTM for drought timing (Figure 3). We revised the sentence into "Overall, the characteristics of SSI-1 drought are close to what is being identified by the VTM" (**P1L18-19**).
E	Line 21: "To the end" should be "in the end"	We changed the word accordingly (**P1L20**).
F	Line 39: Could remove brackets here	We removed the brackets (**P2L38**).
G	Line 46: Write out to what "these" refers.	We changed the word "these" into "the standardized drought indices" (**P2L45**).
H	Line 49: "Should be not" -> Should not be sounds more natural to my non-native ears.	We swapped the words (**P2L48**).
J	Line 55: "which is defined as ... below normal" -> add "the forecasting of " on the place of the dots.	We added the words accordingly (**P2L54**).
K	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.	We think that adding such details in the Introduction would disturb the text flow. We added "The SSI expresses the streamflow as a non-exceedance probability and …" in the method section (**P6L173**).
L	Line 74: "drought indices" -> drought indices is confusing here (refers to SPI, SPEI etc.)	We revised the sentence and specified which indices are meant (**P3L75**).
M	Line 78: "data" -> streamflow data?	Correct, we added the word "streamflow" in the revised manuscript (**P3L79**).
N	Line 82: "its" -> refers to nothing.	We revised the sentence and specified what the implications are (**P3L83**).
O	Line 88: "results" -> results and discussion sections.	We revised the sentence accordingly (**P3L89-90**).
P	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).	We added the explanation that the proxy observed streamflow hereafter is referred to as observed streamflow (**P4L100-101**) and used it throughout the rest of the manuscript.
Q	Line 107: "river streamflow" -> just streamflow	We removed the word river in the revised manuscript (**P4L109**).
R	Line 124: "threshold drought approach" -> threshold level method. Please use this (or similar) terminology consistently throughout the manuscript.	Thanks for reminding us to use consistent terminology. We use the term threshold drought approach and not the threshold level method throughout the manuscript.
S	Line 125: "Standardized drought approach"	We believe that it will confuse readers to

	-> threshold level method applied on SSI timeseries. Please use this (or similar) terminology consistently throughout the manuscript.	change the SSI into the threshold level method applied on SSI. Please see our reason in point B above, and we revised text to respond to the comment made by the reviewer (**P7L196**).
T	Line 128: "the water deficit in different domains of the water cycle, in our case, it is the" -> redundant. Could delete.	We revised the sentence into "….to calculate the water deficit in streamflow" (**P5L130**).
U	Line 130: ref to the original work of Zelenhasić & Salvai (1987) would fit here well.	Reference (Zelenhasić and Salvai, 1987) was added (**P5L131**).
V	Line 144: How was the data aggregated: Sum or mean?	The data was averaged. We added this information in the revised manuscript for clarity (**P5L146**).
W	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.	We removed the sentence as suggested.
X	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.	We agree with the reviewer the statement we made about Q80 and Q70 is not necessarily true. We decided to remove the sentence in the revised manuscript.
Y	Line 150: "be straightforwardly be" -> remove one be.	We thank the reviewer for pointing out typo. We removed the first "be" (**P5L152**).
Z	Section 2.2.1. It is still not clearly described how the daily threshold is arrived.	Below (point AA) we clarify the calculation of the thresholds.
AA	"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.	The reviewer is correct. First, we averaged daily data into monthly data (**P5L146**). Second, we calculated the threshold level for each month. Third, we assigned the monthly threshold to each day of the respective months to obtain a first estimate of the daily thresholds (**P5L152-3154**). Lastly, we applied the 30DMA to these daily threshold to obtain the final daily thresholds. The smoothing is done to avoid jumps in the threshold (**P5L155-157**). The adopted approach in this study has been widely used in the scientific literature, e.g. Van Loon et al., 2012; Van Lanen et al., 2013; and Van Huijgevoort et al., 2014; Beyene et al., 2014. This method is called M_MA in Beyene et al. (2014). The reviewer is correct that the VTD can also be derived from daily data of the flow duration curve (called D_MA in Beyene et al., 2014). Our study analyzes the streamflow drought across Europe and not only for a specific region e.g. a mountainous region (snow region) or a semi-arid region. We decided to use the M_MA instead of D_MA or other methods because this method has been widely applied in many drought studies.

		We added a discussion about the use of M_MA instead of D_MA or others in the revised manuscript (**P6L160-167**).
AB	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.	The alpha and beta parameters of the gamma probability density function are estimated for each grid cell and for each month of the year. We calculated the alpha and beta by using the method of moments. We added this information in the revised version (**P6L177-178**).
AC	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 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.	We expanded the background of SSI in the revised manuscript (**P6L173-174**). We believe that providing results on the testing of the suitability of the gamma distribution to derive the SSI for all river grid cells in Europe and each month (in total > 348,000 parameter sets) is beyond the scope of this study. The main message in our study is that the different approaches produce different drought characteristics, which we think is not substantially impacted by the choice of another probability distribution. We added text on the use of different probability distributions that might be more suitable for streamflow drought in some cases than the gamma distribution including references for studies that performed this analysis (**P6L178-182**). We also added a remark on the choice of the probability distribution at the end of Section 3.2 Implication of different drought identification approaches to forecast streamflow (**P17L546-549**).
AD	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.	We revised the sentence (**P8L226-228**).
AE	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.	The sentence was deleted.
AF	Line 305. Also negatively correlated for the threshold level method applied on the SSI.	We removed part of the sentence, because the negative correlation applies to all drought identification approaches (**P10L306**).
AG	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.	We agree with the reviewer and removed the sentence accordingly.
AH	Line 348: "the SSI-1 ..." -> and VTM	We revised the sentence accordingly (**P11L350**).
AI	Line 360-361: "(Tallaksen and Van Lanen,	Reference was removed.

	2004)" what is this reference doing here? They give me a definition of multi-year drought? Might remove.	
AJ	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.	We thank for the suggestion. We added the reference (**P12L377**).
AK	"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).	We thank the reviewer for the suggestion. We revised the sentence into "As expected, the SSI-1 droughts follow the pattern of VTM droughts because both metrics consider seasonality" (**P12L374-375**).
AL	Line 396-399: Nested sentence – difficult to follow.	We divided the sentence into two: "Our generic finding that the streamflow drought characteristics (frequency, duration, timing) derived using different identification methods differ is in line with the observations made by Vidal et al. (2010). Their study in France also concluded that different identification methods (only standardized-based indices at multiple time scales) generate different drought characteristics" (**P13L400-403**).
AM	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.	The discussion about why drought identification approaches differ, as well as advantages and disadvantages of different approaches was added in the revised manuscript (**P16L525-P17L549**). Moreover, which end user would benefit from each of these methods (VTD, FTD, VTM, FTM, and SSI-1) are also discussed (**P17L550-P18L566**).
AN	Line 538: "fixed" -> variable	We thank the reviewer. The typo was corrected (**P18L585**).
AO	Line 543: "The start of SSI-1 droughts is closest to VTM droughts" as expected (see above).	We revised the sentence into "The start of SSI-1 droughts is closest to VTM droughts because both methods use a monthly resolution and consider seasonality" (**P18L590-591**).
AP	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.	We added some examples of end-users that might benefit from different drought identification methods in the revised manuscript (**P17L550-P18L566**).
AQ	"forecast both standardized-based and threshold-based drought indices." -> Nice! Do they also forecast using fixed and variable thresholds?	Thanks for asking this question. The ADEWS forecasts droughts in streamflow, groundwater, runoff, and precipitation using the VTD approach (Sutanto et al., 2020) while the EDO only forecasts a combined indicator consisted of the SPI, SPEI, a Soil Moisture Index (SMI, soil moisture anomaly), the fAPAR (vegetation) anomaly, and low flow

		index (Cammalleri et al., 2020; 2021). The low flow index also uses VTD. The fixed threshold methods (FTD and FTM) are not used in any of these DEWS. We added discussion about the use of drought identification approaches in both DEWSs in the revised manuscript (**P19L609-616**).
AR	"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.	We thank the reviewer for his/her suggestion. We added a discussion about advantages and disadvantages of different drought identification methods in the end of Section 3.2. (**P16L525-P17L549**), as well as a discussion about the end user that would possibly benefit from each of these methods (VTD, FTD, VTM, FTM, and SSI-1) (**P17L550-P18L566**). See, also point AM.
5	Figures	
A	Fig 1. Mention the four basins in the caption.	The name of four river basins was added in the figure's caption (**P29**)
B	Fig 2. (caption) "Drought occurrences" -> number of drought occurrence (consistent with 2.3)	We revised the caption accordingly (**P30**).
C	Fig 3. "Months when drought mostly started" -> drought initiation time	We revised the caption into "Drought timing (onset)…" (**P31**)
D	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).	Purple color is used to identify SSI drought. We believe it is better to stick to the color choice as it is.
E	Fig 7: "Section 2.3 explains how the drought timing is determined using forecast data." -> Not needed.	We removed the sentence accordingly.
F	Fig 8. Why not connect all ensemble members to last observed month?	When we designed the figure, we thought that the figure would read best, when the x-axis would cover the same period Jan 2003 – Jan 2004 for both initiation dates (April, July). The observed streamflow covers the whole period, whereas the forecast ensemble covers 7 months. Hence, we decided not to revise the figure.
G	Fig 8. Add "threshold" to the legend of Figure e-f.	The legends for Figure 8e and 8f were updated. In Fig. 8e and 8f "limit value -0.84" is added (**P36**)
H	Table 1: explain why no timing (T) in table 1 for Europe.	Timing (onset) from each region differs and it is not a quantitative measure. We could calculate the median timing or average but we believe this does not make sense. For example, if the timing in one climate region is in spring and in another climate region in autumn, then the median or average timing will be in summer, which is not correct. We explained this in the revised manuscript (**P37**).
I	Figure A2: color scales are not matching, i.e., red color 90 days and 5 months	We changed the colors for monthly threshold in the revised manuscript (Now Figure B2, **P42**).

References

Van Loon, A. F. Hydrological drought explained. WIREs Water, doi:10.1002/wat2.1085 (2015).

Zelenhasic, E. & Salvai, A. A mthod of streamflow drought analysis. Water Resources Research, 23(1), 156-168, https://doi.org/10.1029/WR023i001p00156 (1987).

Reply to reviewer 2

We would like to thank the reviewer for valuable suggestions and comments. In this document, **P** refers to the page number and **L** refers to the line number in the recent paper. For example, **P3L65-70**, refers to page 3, lines 65-70.

Reviewer 2		
No	**Comment**	**Reply**
1	I would like to thank the authors for carefully considering my suggestions and comments. Overall, I found the revised version much improved, with the additional analyses adding very interesting insight on the topic.	We thank the reviewer for complements to our revised paper. This could be done because of the suggestions from the reviewers.
2	I still have only one major concern with the presented paper. The authors stress the difference between monthly and daily methods in term of number of event, average duration and deficit, while also highlighting how many of the reported events (especially in the VTD) are basically only minor events. Under such circumstances, average statistics may be not good proxy of the performance of the index, and in my opinion this is an issue that needs to be better addressed, with specific analyses. As an example, if the VTD reports 3 events, 1 very big (2 months) plus 2 minor (of 1 day each), whereas the VTM reports only the major event (2 months), are the two versions so different at the end (3 events vs. 1)? I think that a more "fair" comparison would be in terms of "total" quantities rather than average. Like: total deficit and total number of day under drought. I suggest to add to the mix those metrics as well, in order to better explain the differences between the different indices.	We thank the reviewer for his/her concern about the minor drought events when using the VTD. As you noted, the number of drought occurrences is a total quantity. However, we do not believe that it is good idea to change the average statistic of the other drought characteristics into total quantities. For example, total drought duration will end up in 20% of length of the time series because we used the Q80 threshold, i.e. 28 years of observations will result in 67 months drought duration (28 x 12 months x 20%). Thus, we believe it is fair to use average statistics of drought characteristics instead of total quantities. We addressed the topic of minor droughts in Section 3.1.1 (**P9L279-287**). As described there in a vast area (e.g. Cfb and Dfb climates) more than half of the drought events are shorter than 30 days. This implies that the number of VTD droughts longer than 1 month in these regions is somewhat lower than VTM droughts. We added some text to let the reader realize this aspect of minor drought (**P9L284-285**). Furthermore, we would like to add that we implemented the VTD approach as commonly done in literature, where specific methods are used to exclude minor droughts. The purpose of the paper is to compare the outcome of drought identification approach as commonly implemented. In our study, we applied the 30DMA to avoid minor drought events e.g. drought that has a duration of 1 day or 2 days. This 30DMA method, in general, removes the short drought event but it cannot completely remove the minor drought event, e.g. drought that has duration in between a few days to a month. Additional methods such as the inter-event time method (IT-method), the moving average procedure (used in this study), and the sequent peak algorithm (SPA) are also used to exclude minor droughts when using daily data, as we discuss in our manuscript

	Similarly, when FTD and VTD are compared, you need to find a way to distinguish between the cases when VTD detects more actual events (i.e. events in different seasons) vs. the cases when FTD detects a single event while VTD "splits" the same in multiple smaller events (but close in time). Also in this case, total quantities may alleviate the problem.	(**P16L507-514**). The FTD and VTD are conceptually different; the VTD considers seasonality whereas the FTD does not, as described in the manuscript e.g. **P5L143-145**). This means that the major difference is that the VTD may detect droughts both in the low flow season and in the high flow season when the flow is below normal. The FTD only detects event in the low flow season. This is the major reason that the VTD approach identifies more drought events than the FTD. Hence, we think, in general, we should not focus too much on the low flow season where some events may be split in more subevents according to one of the methods as a main reason for the differences in number of drought occurrences. We believe that the detailed comparison of different drought identification approaches performed for 4 selected river basins is sufficient to illustrate also the difference between VTD and FTD (Figure 4 and 5). For example, Figure 4a clearly shows minor VTD drought in spring 2003 and minor FTD drought in autumn 2004.
3	As a final comment, I found the revised text a little unpolished. I report some minor issues and comments regarding the first pages of the manuscript, but I suggest a careful revision of the full text.	We thank for the reviewer's feedback on the first pages. We read carefully the rest of the manuscript and we revised it at several places.
4	Line by line comments	
A	P1 L5. "The way, how…" is repetitive.	We corrected the text (**P1L5**).
B	P1 L14. Here it should read "more than…"	We revised the word accordingly (**P1L14-15**).
C	P1 L21. "To the end…" sounds out of place.	We revised the sentence into "In the end" (**P1L20**).
D	P2 L29. "from among others drought…" is not clear. please rephrase.	We revised the sentence into "….that impacts of drought on society….." (**P2L28**).
E	P2 L31. Given the focus of the paper on streamflow drought over Europe, it may be worth to mention this recent study (HESS 24, 5919-5945).	Suggested literature was added (**P2L30**).
F	P3 L82. Please spell our incl.	The word was fully written (**P3L83**).
G	P3 L82. "Europa" should read "Europe".	We corrected the typo (**P3L85**).
H	P4 L95. Looking at the maps, I'm assuming that only cells with a minimum contributing area are considered, especially because the threshold method may not work as intended in rivers with streamflow close to zero. This is a good place to mention that.	For the map, we only plotted the major European rivers. This indicated that small rivers are excluded from our plot. We thank the reviewer for mentioning that the threshold method used in our study does not work for rivers with flow close to zero. Indeed, the reviewer is correct. We added this information in the Section 2.2.1 paragraph 2 where we discuss the use of threshold Q80 (**P5L149-151**).
I	P4 L100. "Center" should read "Centre".	We thank the reviewer for spotting the typo

		because of the US English auto correction. The word was revised (**P4L102**).
J	P5 149-174. I suggest to summarize this discussion a bit, since is taking most of the methodology section even if this is not the key point on the analysis.	We moved the details about how the 30DMA has been implemented to avoid minor droughts in the historic and forecasted daily streamflow data to a new Appendix A (**P20L631-649)**. The other appendices have been renamed.
K	P7 L193. "… falls below -0.84…"Looking at the next section, it seems that also for SSI an event-based approach is adopted. Does an event start when SSI falls below the threshold and ends when it returns above (as for the threshold methods) or is each monthly value treated separately? Please clarify here.	We derived drought characteristics from the drought events when the streamflow falls below the threshold level (Q80) and the SSI-1 values fall below the limit value of -0.84 for both historical analysis and reforecasts. We clarified this in the revised manuscript (**P5L133-134** for threshold and **P7L195-196** for SSI)
L	P8 L252. Again, the criteria adopted to selected those 29,000 cells (i.e. minimum contributing area) need to be highlighted in the methodology. I'm sure that Europe is covered by much more than 29,000 5x5 km2 cells.	We thank the reviewer for raising this topic. In this study, we only selected major European rivers, indicated by river cells that have average discharge above 10 m^3/sec (n=~29,000). We added information about selecting river grid cells in the data section (Section 2.1.) (**P4L105-106**).
M	Authors: The suggested information about the VT method applied in EDO was added (P18L560-562). The added reference is to the CDI index, and not to the streamflow drought index (Hydrol. Sci. J. 62(3), 346-358). In my opinion, a better place to refer to this operational index would be in the Lisflood section (since is based on these data).	We added the suggested reference (**P19L610**) and moved the Sepulcre-Canto et al., 2012 reference to the method section (**P4L115**).