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	The Study of Sutanto and Van Lanen compares different drought identification approaches: 1) the fixed threshold level method, 2) the variable threshold level method and 3) the threshold level method applied on SSI time series, for simulated river flow at the pan-European scale. They show that (average) drought event characteristics differ based on the used drought identification method. Consequently, they show that drought event forecasts differ, depending again on the used drought identification method. Overall, the main recommendation of the paper is strong and relevant, i.e., droughts differ depending on the used method and streamflow drought forecasters and stakeholders should agree which type of drought should be forecasted. In addition, I believe that Figure 6 provides an informative message for the users and developers of hydrological drought forecasting systems.	We would like to thank the reviewer for the comments, valuable suggestions, and acknowledgement of the message in our paper that drought forecasters and stakeholders should agree at front which type of hydrological drought should be forecasted (**P18L568-570**).
2a	However, given that this paper focusses on the definitions of drought and methodology of drought identification, it sets an example which types of drought identification approaches can be used for drought forecasting applications (and how). Therefore, it should be extra "sharp" in its drought definition and identification approaches as well. At this stage, this is not the case and there are several methodological concerns that should be addressed carefully. In addition, the comparison of the results is far from straight forward. The used drought identification approaches do not only vary in overall method, but also in: 1) threshold (<10 percentile for the fixed and variable threshold approaches and around <50th percentile threshold for the SSI), 2) data accumulation period (1 month for the fixed and variable threshold based approaches vs. 6 months for the SSI), and 3) temporal resolution (daily vs. monthly).	The referee is concerned about the methodology used in our paper, i.e. in three aspects: 1) the thresholds to identify drought, 2) the data accumulation period, and 3) the temporal resolution. Our answers to these three questions are as follows: i) Our paper used the drought thresholds based on common practice in the drought community, which are in the range of 10-30th percentile of the flow duration curve (P70-90) for a Fixed Threshold (FT) or Variable Threshold (VT) and SSI below 0 (~P50). Our reason to use different thresholds (50th percentile for SSI and 10th percentile for the FT and VT) was that we would like to follow common practice for the different approaches. However, the reviewer has a point that the comparison between threshold methods (VT, FT) and SSI is not equal regarding to the use of different percentiles. Thus in the revised manuscript, we changed the thresholds from P90 into P80 for VTs and FTs (**P5L145-146**), and SSI≤-0.84 (~P80) to have a fair comparison between different drought indices (Tijdeman et al., 2020) (**P7L193-195**).

		ii) We realize that streamflow, as included SSI, comprises some catchment memory aspects (delayed flow from groundwater). Hence, in the revised manuscript, we replaced SSI-6 with SSI-1. However, we need to realize that anomalies in the accumulated flow over a longer period (e.g. SSI-6) have relevance for some purposes, such as the management of surface water reservoirs (**P6L186-P7L190**). iii) Again, we followed common practice (see item i, above) to identify drought using these methods. Many studies used daily streamflow data to analyze drought using the threshold method and monthly streamflow data to analyze drought using the standardized indices. To the author's knowledge, only Tallaksen et al. (2009) and Van Loon et al. (2019) used the monthly data to derive drought using the threshold method and these were done only for a scientific purpose (**P5L138-140**). In the revised manuscript, however, we added to the common practice approach (daily resolution), an analysis of drought characteristics using monthly streamflow data in both FT and VT drought approaches. This allows an analysis of the VT and FT threshold approach and the SSI-1 using the same temporal resolution, i.e. monthly time scale. This implies that we have two VT and FT threshold applications: daily resolution, as frequently used, and monthly resolution to allow comparison with SSI-1 (**P5L132-136**).
2b	Finally, the most novel part of this paper, which deals with the implications for drought forecasting, is rather limited and deserves more attention in my opinion.	We extended the novel part of paper to illustrate that the outcome of the forecast depends on the drought identification method. We do this by describing: (i) pan-European maps showing forecasted drought duration (Fig. 6) and timing (Fig. 7) using different drought identification methods (FT and VT with daily and monthly resolution, and SSI-1) (Section 3.2.1, **P13L405-P14L441**) (number of drought occurrence/frequency and drought deficit volume are provided in Appendix B), and (ii) a summary of forecasted drought characteristics identified using different approaches in the Rhine River using forecasts initiated from 1st January 2003 to 1st December 2003 with a lead time of 7-month. In addition we also provide information on the percentage of ensemble members showing drought for each identification method (Fig. 8, Table 3 and 4, Section 3.2.2, **P14L442-P16L519**).

3a	**SSI computation:** Why SSI-6? For me, it makes sense to aggregate meteorological drought indices (SPI, SPEI) to differentiate between slow and fast responding (hydrological systems), e.g., catchment with small and large storage components. However, riverflow already encompasses the accumulation and delay of the meteorological signal caused by e.g. delayed groundwater flow. From a riverflow drought perspective, it is often important to know what is currently happening in the river (SSI-1) and not what happened in the past 6 months (SSI-6). Also, the SSI-6 is not at all comparable to the 30-Day moving window used for the FT and VT approaches. This makes the interpretation of the comparison between both approaches less straight forward. Finaly, the reasoning to choose the SSI-6 over the SSI-1 because the SSI-1 results in many minor drought events does not compensate for the advantages of the SSI-1.	We do agree with the reviewer, and thus we replaced the SSI-6 results with SSI-1 (see our reply 2a, ii).
3b	Why an SSI threshold of zero to identify drought? I would not term something that happens 50% of time drought. Please note that the original SPI paper of Mckee (1993) uses a similar threshold, but has the additional requirement that the SPI should at least reach a value of -1 over the course of the drought event. In addition, an SSI threshold of zero is far from comparable to an FT or VT of Q90 used for the threshold level approaches.	The reviewer has a reasonable point here. In the revised manuscript, we changed the threshold values into P80 for the threshold methods (VT, FT) and SSI≤-0.84 (~P80) in order to have a fair comparison (see our reply 2a, i) (**P7L193-195**).
3c	Why the gamma distribution to derive the SSI? I agree that is hard to find a suitable distribution to fit to riverflow time series (line 150-151). However, that is not a good argument to simply use the Gamma distribution. There are likely to be better alternatives for your pan-European dataset (See e.g. Svensson et al., 2016, Tijdeman et al., 2020). Why no goodness of fit testing? The studies above conclude on different suitable candidate distributions for the SSI (other than the gamma distribution) that might be applicable for the current study. However, that does not mean that they can be applied on your dataset of simulated streamflow series by default, as your dataset might exhibit different properties as compared to the observed riverflow timeseries. Careful evaluation which distribution is most suitable for your set of rivers is required. Which distribution fitting method was use?	We used the gamma distribution to derive the SSI because the gamma distribution has been used for hydrological forecasting of both high and low flows (Slater and Villarini, 2018). The reviewer also recognized that it is hard to find a suitable distribution to fit all streamflow regimes in Europe (see also Vicente-Serrano et al., 2012). Moreover, no single distribution fits well with all monthly streamflow data in all river grid cells (n=+29,000), e.g., sample properties of streamflow in January might differ from those in August in all places (Tijdeman, et al., 2020) (**P6L180-184**). Our study does not focus on the selection of the best distribution for drought forecasting. We do not believe that another distribution (or other distributions) that consider differences in streamflow regime across Europe will change the main message of the study, i.e. that the outcome of the hydrological drought forecast depends on the identification method. Thus we believe it is better to simply use the widely selected gamma distribution in our

		analysis.
3d	For the forecasted SSI: Did you use the parameters of the population distribution derived from historical monthly flow values to derive the SSI for forecasted values? Or did you replace the historical values with forecasted values and than recalculated the population distribution to derive the SSI? And why, e.g., what should a forecaster do?	We used the distribution parameters derived from the observed (historic) datasets to identify the forecasted drought. Using this method, the gamma distributions were calculated from long time series of observed data, in our case 29 years, and then applied to the forecasted streamflow (Sutanto et al., 2020a, Figure A1). We did not calculate the distribution from the re-forecast datasets because the re-forecasted time series that we have are rather short (9 years) and obviously it is not the actual observed streamflow. We added this information in the revised manuscript (**P7L198-203**).
4a	**Threshold approach:** Line 123-143: Many different smoothing procedures have been applied in combination with the threshold level method. This has been done for good reason, however, sometimes resulting in an (unwanted) increase/decrease in drought occurrence, especially for the VT method. For me, a 10th percentile implies that 10% of the time series is in drought and that drought occurrence is equally distributed over the year in case of the VT method. However, by first deriving the threshold from daily streamflow data, and then smoothing both the threshold and riverflow timeseries seperately, this is not necessarily the case anymore. This might be solved relatively easily, i.e., first apply the moving average and then derive the threshold. Or you could use monthly data.	In our paper we used the moving average of the daily quantile approach (D_MA, Beyene et al., 2014) to obtain the VT thresholds. In the revised manuscript, we changed the method on how we calculate the VT thresholds. We now use monthly streamflow data to derive the monthly threshold and then we assign the monthly threshold level to each day of the month. When confronting time series of daily data (observed data, 1990-2018, and re-forecasted data 2003) with monthly threshold levels (only relevant for the VT application using a daily resolution, see our reply 2a, iii), jumps between two consecutive months might result in unrealistic drought behavior that extends around the beginning and end of each month. Therefore, we apply a 30 days centered moving average to the discrete monthly thresholds, as done, for instance, by Beyene et al. (M_MA, 2014); Van Loon et al. (2012); Van Lanen et al. (2013); Van Huijgevoort et al. (2014); Heudorfer and Stahl (2017); Van Tiel et al. (2018) (**P5L149-P6L156**).
4b	Line 366-367: You encourage using monthly streamflow data for drought forecasts but use daily streamflow in your own analyses. I would have find it logical to do this as well in this study, e.g., instead of the FT and VT approaches applied on daily data, it could be applied monthly averaged data. This also increases the comparability with the SSI. Further, is there really merit in forecasting streamflow drought duration and deficit at a daily resolution, especially for the longer lead-times? Is this being done somewhere? Can this be done with any skill? If not, wouldn't it be better to just stick to monthly data for which at least some skill might be achieved?	We added the monthly drought analysis derived from the FT and VT thresholds, as additional analysis to the daily resolution to enable comparison with the SSI-1 forecast. However, we also keep the daily analysis in our revised manuscript because the daily streamflow data is commonly used in many studies using the threshold methods (see our reply 2a, iii), incl. hydrological drought projections (Prudhomme et al., 2014; Wanders and Van Lanen, 2015; Wanders et al., 2015) (**P5L132-138**).
5a	**Results and discussion:** Section 3.2. The forecasting section, which is the most the novel part of this paper, would benefit from some more attention.	We would like to thank the reviewer for his/her valuable suggestions. We extended the forecast results with the series of 12 forecasts initiated each month (from January

	Figure 6 provides a nice illustration, even though it might be a little obvious at this point in the papers that drought characteristics derived with different methods will vary, given that you apply a different threshold on the same forecast data. However: - I disagree that the drought of 2003 in the river Rhine started in August 2003. According to the SSI-1, river levels dropped to below normal anomalies much earlier. I suggest to start earlier in the year. - Why not add the observed hydrograph to the plot? - Isn't the fact that the VT method does not forecast a drought a good thing? According to this method, there was also no drought in the observed hydrograph (Fig. 4a) – how could this method have "performed better" (line 340). - Why not show the SSI-1 here?	to December 2003) with a lead time of 7-month (see our reply number 2b) including the observed streamflow (Table 3 and 4). In the revised manuscript, we present the forecasted drought characteristics (occurrence, timing, duration, and deficit volume) using different identification approaches (daily FT and VT, monthly FT and VT, and SSI-1) for the pan-European river network (Section 3.2.1) and for the Rhine River in Table 3 and 4 (Section 3.2.2) (**P13-P16**).
5b	Given the focus of the paper on river flow forecasts, I would expect more focus on the latter, and not only an exemplary timeseries river flow forecasts for one river / event. It would be interesting to include. - At least, an evaluation and discussion of the spread in streamflow forecast and especially in the spread in streamflow drought forecast, and (i.e., not only the evaluation of the median forecast). What are the ranges in drought characteristics derived from the forecast ensemble? - Consequently an evaluation or discussion of the streamflow (drought) forecasts skill, i.e., can certain "types of droughts", e.g., FT vs. VT vs. SSI, be forecasted better? The above evaluation would benefit the consideration of multiple rivers, drought events, or start months.	We would like to thank the reviewer for the suggestions. We extended the analysis by providing: (i) maps displaying forecasted drought timing and duration across Europe using forecast data issued in July 2003 (Fig. 6, 7, B1 and B2), and (ii) tables describing forecasted drought characteristics (occurrence, timing, duration, and deficit volume) for the Rhine River using a series of 12 forecasts initiated from January 2003 to December 2003 with a lead time of 7-month (median ensemble) (see also our reply number 2b) (Table 3 and 4). An analysis of the forecast using different drought identification methods for several European rivers is beyond the scope of this paper. We believe that the map showing the pan-European pattern (see item i, , point 5b) clarifies that the example of the Rhine River is sufficiently representative. In addition, we also provide information on number of ensemble members for which drought was forecasted (x ensembles out of 25) (See Table 3 and 4). We would like to stress that the evaluation of forecast skill using SSI and threshold methods (VTs and FTs) is beyond the scope of this paper. This was published in previous papers (Van Hateren et al., 2019; Sutanto et al., 2020b).
5c	Again, I would avoid the SSI-6 here, due to the strong autocorrelation of this index, which makes it relatively easy to forecast on short lead times. For example, for a forecast with a lead-time of 1 month, 5 out of 6 months are already known. Rather, I would look at the SSI 1.	As said above, we replaced the SSI-6 with SSI-1 in the main text (see our reply 2a, ii and our reply 3a).
6a	Finaly, some (non-committal) suggestions for Section 3.1 that could further improve	We thank the reviewer for the suggestions. We added the drought duration and deficit

	the manuscript: • Section3.1.1 Next to showing the amount of streamflow droughts, you could consider showing other characteristics such as the average duration, deficit volume, or the number of minor drought events. This provides valuable insights in differences between methods, and further makes the notions in 3.3.1 about regions with more minor drought quantitative. In addition, you can derive a proxy for deficit volume from standardized time series. The units are meaningless and not comparable with the deficit volumes derived with FT and VT method. However, the relative difference over Europe should pop-up.	volume derived from the FTs and VTs approaches, and only drought duration for SSI in the revised manuscript (see Fig. A2 and A3). The SSI drought deficit volume is not presented because it is impossible to derive the deficit volume using the SSI approach (major drawback of standardized approaches) (**P7L210-212**). In addition, we also added a European map showing the number of minor drought events derived using the VTD (Fig. A1).
6b	• Section 3.1.2 In addition to discussing when most drought starts, it might be interesting to see when most drought occur in difference climates. This can be presented as a series of histograms for each climate, with the month on the x-axis and the fraction of drought months that occurred in that month on the y-axis.	This is an interesting suggestion. In the revised manuscript, we provided a summary of drought characteristics (number of drought occurrence/frequency, timing, duration, and deficit volume) for 5 Köppen Geiger climate regions identified using different approaches (daily FT and VT, monthly FT and VT, and SSI-1) (Table 1 and 2).
7	**Minor comments:** Line 2: "... the term streamflow drought forecasting, rather than streamflow forecasting ..." You could briefly explain difference between the two here.	We added text to describe streamflow drought forecasting in the revised manuscript (**P1L3-4**).
	Line 5: "within" Correct?	We replaced "within" with "of" (**P1L6**).
	Line 6: Be careful with terming these extreme events. They are anomalies, but something that happens on average at least once every year, as is the case in your study, is not an extreme event.	Naming of extreme events has always a sense of subjectivity. We removed the words (**P1L7**).
	Line 7, 8: "observed" might be "observations"	We changed the word accordingly (**P1L8**).
	Line 7: "a LISFLOOD model"... are there more?	There is only one LISFLOOD model. We changed "a" in "the LISFLOOD model" (**P1L8**).
	Line 10: add method to VT and FT, e.g. variable threshold level method.	The word "method" was added in the revised manuscript (**P1L14**).
	Line 10: You also apply a threshold based approach on SSI time series. Mention this here.	An explanation about threshold to identify drought in SSI was added. However, we do this in the Methods section (**P7L193-194**). Threshold-based drought indices (called deficit characteristics in Hisdal et al., 2004) are fundamentally different from the standardized -based drought indices (Van Loon, 2015).
	Line 16: "Eliminate". Not true. You can still have 1-day droughts with these TL approaches.	We removed the sentence in the revised manuscript.
	Line 24: "IPCC" should be "The IPCC".	Thanks for the correction (**P2L29**).
	Line 34: This sentence slightly contradicts	We revised L34 to avoid possible

	with Line 1, where you state that drought forecasting is a key element of DEWS. I would expect there to be some examples. Which contemporary "DEWS" include streamflow drought forecasting, using the approaches as described in the paper (FT, VT and SSI), not just streamflow forecasting)?	contradiction, i.e. "One of the elements to be included in a NDPP is a Drought Early Warning System that in addition to real-time monitoring contains …" (**P2L37-39**). In the preceding sentence we explain the abbreviation NDPP (National Drought Policy Plan) (**P2L36**). Furthermore, streamflow drought forecasting, using all the approaches as described in the paper (FT, VT and SSI) are developed in the EU H2020 ANYWHERE project (for background, see Sutanto et al., 2020a) (**P18L570-572**).
	Line 41: "evaporation" should be potential evapotranspiration Line 47: "used" should be "be used" Line 85: "Proxy" should be "Proxies"	We revised the text accordingly (**P2L46, L53, P4L98**).
	Line 49: Mention that you specifically focus on simulated streamflow drought.	We changed "hydrological drought forecasting" into "streamflow drought forecasting" (**P2L54-55**).
	Line 75: "There" should be "There is"	We removed the word "There". Thus the sentence became: "…., which demonstrates that none of the hydrological drought forecast approaches fit all needs" (**P3L83**).
	Line 89: "proxy observed streamflow" could just be "simulated streamflow"	We would like to keep the term "proxy observed streamflow" to indicate that in principal people would like to use observed data, but these spatio-temporal streamflow observed flow data do not exist. Hence, flow data obtained from a hydrological model driven by observed weather data are used as proxy for observed (same as EFAS-WB in Arnal et al., 2018 or offline simulation in Yuan et al., 2017). This is similar to reanalysis data that are a proxy for observed weather. In some cases these simulated data are just called observed, which we think should be avoided.
	Line 112: "re-forecasted data 2003" should be "re-forecasted data of 2003" Line 119: "in" should be "for" Line 147: "median" should be "expected median". Line 179: "definitions" … "drought identification approaches" might be better. Line 221: "drought that has" should be "droughts that have"	We changed the text accordingly (**P5L128, L141, P6L179, P8L250, P16L512**).
	Line 128: "were moving averaged" rephrase	The sentence was corrected (**P6L158-159**).
	Line 134: "For the threshold" …this refers to variable threshold approach I guess? In this section, make the clear distinction between FT and VT and seperately explain how both are derived.	The threshold here refers to both FT and VT. We revised the sentence (**P5L149-P6L156**).
	Line 138-140: add here that MA introduces a significant amount of auto-correlation, which affects the skill of the river flow forecast for the first 30 days significantly.	We added an explanation about the effect of 30DMA on the forecast skill (**P6L170-172**).
	Line 155-160: Add here that it is quite easy	We replaced the SSI-6 with SSI-1 in the main

	to forecast the SSI-6 for short lead times, given the strong autocorrelation of the timeseries. E.g., for 1-month lead-times, you already know five months and only have to forecast one.	text, thus the explanation of preceding observed data is not necessary there.
	Line 162-164: Please explain how you classify an event with varying SSI values into one category.	In our study we only focused on the median ensemble and not the whole ensemble (25 members). Thus if the median value of SSI is in between -1 and -1.5, we classify the event as moderate drought.
	Line 162-177: Did you derive the climate classification yourself using the approach described in Peel et al (2007)? Or did you use their dataset?	We used their dataset (**P8L235-236**).
	Line 188: "Lower than median streamflow" ... Not necessarily true. Technically, above median streamflow can still be a negative SSI and vice versa. Depends on the sample and (goodness of fit) population distribution to derive the SSI.	We use the threshold SSI<-0.84 to identify SSI drought in the revised manuscript (**P7L193-195**).
	Line 189: Figure 3 does not show that streamflow droughts occur every year.	Figure 3 shows the drought timing i.e. the month in which commonly start, and not drought occurrences. The latter we show in Fig. 2.
	Line 200: This is comparing apples and pears, as the thresholds are completely different.	We changed the threshold values, i.e. special application of VT and FT thresholds, for better comparison in the revised manuscript (see our reply 2a, i).
	Line 203-206: Could this not be compensated by a higher number of drought in winter for the VT?	Sorry, we have to disagree. The VT method takes into account the seasonality.
	Line 228. "(Coincides with hydrologic years in most of Europe)" remove: unneeded repetition. Line 264-266. Is the last part, i.e., about the lowest and n-day minimum flow, needed? Interrupts flow.	We removed the sentences.
	Line 266-267. Looking at Fig. 5a, I find the SSI-1 timeseries much more informative about drought in the river Rhine. Rhine drought reaches is maximum in summer 2003, and recovers in winter 2004. For me, this make much more sense than the SSI-6 timeseries. Was the drought in the river Rhine really a multiyear event? Were there impact directly related to Rhine river flows over the course of 2004?	In the revised manuscript, we use only the SSI-1 forecasts rather than the SSI-6. Drought 2003 in Europe was one of the severe drought events and this was applied to the Rhine River as well. The impact of 2003 drought on the Rhine flow was apparent but we are not aware if there was an impact in 2004 meaning that there would have been a multi year drought.
	Line 270. For me, this description of drought in the river Rhine makes much more sense. It would make even more sense if you would use a more appropriate drought threshold (maybe SSI-1 < -0.84, corresponding to the 20th percentile). I don't see the problem of having 2003 split up in different events and question why it is better to use an SSI-6 and thereby inflate the event to a multiyear drought.	We do agree with the reviewer and therefore we revised the manuscript by using the drought threshold SSI<-0.84 and only using SSI-1 (see our reply 2a, i) (**P7L193-195**).
	Line 285: "C" should be century. Line 361: "rare extreme drought events" ...	Typo was corrected (**P12L365**) and we removed sentence (L361).

	extreme events are by definition rare. Rephrase.	
	Line 295-302. Why limit yourself here to the four case study Rivers and the limited time window? You could directly compare the number of drought events & their deficit volumes over a longer time period and for all the catchments (starting by deriving the difference between Fig. 2a and b).	We do agree with the reviewer and, as stated above, we extended our analysis by providing drought duration using different approaches (Fig. A2, see also Table 1 and 2). The limited time window in Figure 4 was made to increase the readability. This was done by showing 2003 drought events in north, central, and east Europe and 2005-2006 droughts in south Europe. In the Supplementary Material we have given the drought characteristics for all four selected rivers (Tabel S1 and S2) derived from data of the entire period (1990-2018). This allows a comparison with Table 1 and 2, but please note that drought characteristics obtained for individual rivers over the period 1990-2018 may deviate from the general pattern, as reported in Table 1 and 2 (Section 3.1.1), because the drought analysis of a specific river only involves streamflow generation upstream of the river grid cell that has been selected to represent the river.
	Line 312-329. According the definition of drought according to VT and the SSI, droughts are expected to occur for an equal amount of time over the year. Please provide an explanation for the distinct temporal differences in drought occurrences. Or is this still referring to the start month of the drought?	Yes, we refer to the drought timing, which is identified as the month when drought mostly started. In the revised manuscript we use in a specific application of the FT and VT methods monthly streamflow data, thus in that case there is no discrepancy in the temporal resolution between threshold methods and SSI.
	Line 309: "(except for the Rhine River)" this contradicts with the discussion in the paragraph above.	For the selected river basins (Section 3.1.4), we did the analysis only for the selected river grid cells. The discussion in the paragraph before this section was for the whole of Europe in general. We moved the detailed analysis of four selected river basins to the Supplementary Material.
	Line 337: Not only meteorological drought, also streamflow drought according to the SSI-1 (Fig. 5a).	We do agree with the reviewer. However, this sentence has been deleted in the revised manuscript.
	Line 354. Which is good, because there was no drought according to the VT, or?	Here the VT method did not forecast drought in August 2003 using the 30DMA. The 30DMA, however, is very useful in reducing minor drought events and it is also recommended to increase the forecast skill (previous version: **P12L354-360**). In the revised manuscript, we changed the forecast initiation months to April and July.
	Line 382: "eliminate" ... not correct as minor droughts can still occur.	We removed the paragraph since our study does not discuss the forecast skill.
	Line 372-373: "the FT method produces higher drought deficit volumes and duration than VT" not shown for the pan-European dataset.	We added drought duration and deficit volume in the revised manuscript (see Figure A2, A3, Table 1, 2, Figure 6, B2, Table 3, and 4).
	Line 375: "occurred" should be "started".	We changed the word accordingly (e.g. **P17L542**).
	Line 377: "what being identified by"	We revised the whole paragraph.

	rephrase	
	Figure 1: Nice. What is the difference between light and dark grey in e.g., the Alps?	We only have grey color for ET region (Alps).
	Figure 2: You could add the upper boundary, e.g. 30-xx instead of >30.	We changed the whole figure in the revised manuscript.
	Figure 3: "The timing for drought was determined based on the first month of each drought event." This is the same as what is said in the beginning of the caption.	We removed the last sentence to avoid duplication.
	Figure 4: Some droughts are hardly visible (e.g. in Figure 4a). It might work to use a log-scale Figure 4: Axis lables: m3 sec-1 or m3 / sec instead of m3/sec	We changed the axis into m^3 sec^{-1}.
	Figure 4: Are the grey vertical lines the hydrological years?	We added an explanation for the grey vertical lines.
	Figure 4. You might consider using a different color when VT and FT overlap.	We revised the color as suggested (orange color).
	Figure 5. Add grey vertical lines here as well.	We added the grey vertical lines.
	Figure 6. Same comments as for Figure 4 and 5.	We revised the figure accordingly.
	Table 1. Would be interesting to also compare average deficit volume and timing.	We added drought duration and deficit volume in Table 1, 2, 3, and 4.

References:

1. Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B., and Pappenberger, F.: Skilful seasonal forecasts of streamflow over europe?, Hydrol. Earth Syst. Sci., 22, 2057–2072, https://doi.org/10.5194/hess-22-2057-2018, 2018.
2. Beyene, B.S., Van Loon, A.F., Van Lanen, H.A.J., and Torfs, P.J.J.: Investigation of variable threshold level approaches for hydrological drought identification, Hydrol. Earth Syst. Sci. Discuss., 11, 12765–12797, 2014
3. Heudorfer, B., and Stahl, K.: Comparison of different threshold level methods for drought propagation analysis in Germany, Hydrology Research, 2017, 48 (5): 1311–1326. https://doi.org/10.2166/nh.2016.258.
4. Hisdal, H., Tallaksen, L. M., Clausen, B., Peters, E., and Gustard, A.: Hydrological Drought Characteristics. In: Tallaksen, L. M. & Van Lanen, H. A. J.(Eds.) Hydrological Drought, Processes and Estimation Methods for Streamflow and Groundwater. Development in Water Science 48, Elsevier Science B.V., pg. 139-198, 2004.
5. Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, Proc. Natl. Acad. Sci., 111, 3262–3267, doi:10.1073/pnas.1222473110, 2014.
6. Slater, L.J., and Villarini, G.: Enhancing the predictability of seasonal streamflow with a statistical-dynamical approach, Geophysical Research Letters, 45(13), pp.6504-6513, 2018.
7. Sutanto, S.J., Van Lanen, H.A.J., Wetterhall, F., Llort, X. (2020a). Potential of pan-European seasonal hydro-meteorological drought forecasts obtained

from a Multi-Hazard Early Warning System. Bulletin of the American Meteorological Society. https://doi.org/10.1175/BAMS-D-18-0196.1.

8. Sutanto, S.J., Wetterhall, F., Van Lanen, H.A.J. (2020b): Hydrological drought forecasts outperform meteorological drought forecasts. Environment Research Letters, 15: https://doi.org/10.1088/1748-9326/ab8b13.
9. Tallaksen, L.M., Hisdal, H., and Van Lanen, H.A.J.: Space-time modeling of catchment scale drought characteristics, Journal of Hydrology, 375, 363–372, 2009.
10. Tijdeman, E., Stahl, K., and Tallaksen, L. M.: Drought characteristics derived based on the Standardized Streamflow Index: A large sample comparison for parametric and nonparametric methods, Water Resources Research, 56, e2019WR026315, https://doi.org/ 10.1029/2019WR026315, 2020.
11. Van Hateren, T., Sutanto, S. J., and Van Lanen, H. A. J.: Evaluating uncertainty and robustness of seasonal meteorological and hydrological drought forecasts at the catchment scale-case Catalonia (spain), Env. Int., 133, 105206, https://doi.org/10.1016/j.envint.2019.105206, 2019.
12. Van Huijgevoort, M. H. J., Van Lanen, H. A. J., Teuling, A. J., and Uijlenhoet, R.: Identification of changes in hydrological drought characteristics from a multi-GCM driven ensemble constrained by observed discharge, Journal of Hydrology, 512, 421-434, doi:10.1016/j.hydrol.2014.02.060, 2014.
13. Van Lanen, H. A. J., Wanders, N., Tallaksen, L. M., and Van Loon, A. F.: Hydrological drought across the world: impact of climate and physical catchment structure, Hydrol. Earth Syst. Sci., 17, 1715–1732, https://doi.org/10.5194/hess-17-1715-2013, 2013.
14. Van Loon, A. F., Rangecroft, S., Coxon, G., Naranjo, J. A. B., Ogtrop, F. V., and Van Lanen, H. A. J.: Using paired catchments to quantify the human influence on hydrological droughts, Hydrol. Earth Syst. Sci., 23, 1725-1739, https://doi.org/10.5194/hess-23-1725-2019, 2019.
15. Van Loon, A. F.: Hydrological drought explained, WIREs Water, https://doi.org/10.1002/wat2.1085, 2015.
16. Van Loon, A. F., Van Huijgevoort, M. H. J., and Van Lanen, H. A. J.: Evaluation of drought propagation in an ensemble mean of large-scale hydrological models, Hydrol. Earth Syst. Sci., 16, 4057-4078, doi:10.5194/hess-16-4057-2012, 2012.
17. Van Tiel, M., Teuling, A. J., Wanders, N., Vis, M. J. P., Stahl, K., and Van Loon, A. F.: The role of glacier changes and threshold definition in the characterisation of future streamflow droughts in glacierised catchments, Hydrol. Earth Syst. Sci., 22, 463–485, https://doi.org/10.5194/hess-22-463-2018, 2018.
18. Vicente-Serrano, S. M., López-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Azorin-Molina, C., and Morán-Tejeda, E.: Accurate computation of a streamflow drought index, J. Hydrol. Eng., 17(2), 318-332, doi:10.1061/(ASCE)HE.1943-5584.0000433, 2012.
19. Wanders, N. and Van Lanen, H. A. J.: Future hydrological drought across climate regions around the world modelled with a synthetic hydrological modelling approach forced by three General Circulation Models, Nat. Hazards Earth Syst. Sci, 15, 487–504, doi:10.5194/nhess-15-487-2015.

20. Wanders, N., Wada, Y. and Van Lanen, H.A.J.: Global hydrological droughts in the 21st century under achanging hydrological regime, Earth Syst. Dynam., 6, 1–15, doi:10.5194/esd-6-1-2015, 2015.
21. Yuan X., Zhang, M., Wang, L., and Zhou, T.: Understanding and seasonal forecasting of hydrological drought in the Anthropocene, Hydrol. Earth Syst. Sci., 21, 5477–5492, doi:10.5194/hess-21-5477-2017, 2017.

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 1		
No	**Comment**	**Reply**
1	The authors performed an intercomparison of three different streamflow drought indicators, with the goal to highlight the differences in the drought characteristics associated to each index and to detail the implication on drought forecast. I found the overall goal of the study meaningful, given the confusion that still arise among scientists and operational users on the topic, but I also found the paper and its structure generally out of focus. The key message of the paper "....developers of DEWS and end-users should clearly agree among themselves upon a sharp definition on which type of streamflow drought is required to be forecasted for a specific application." is in my opinion, even if relevant, better suited for a short communication or letter paper rather then a research paper.	We would like to thank the reviewer for the acknowledgement of the goal of our study. We appreciate the suggestion from the reviewer that our manuscript is better suited for a short communication paper rather than a research paper. However, a short communication paper would only be an option, if a systematic intercomparison of threshold and standardized streamflow drought indices across Europe obtained from commonly used identification approaches would exist. Such intercomparison, however, does not exist and consequently a technical paper is needed, which describes and discusses the use of different drought identification approaches (in this paper five approaches) to derive streamflow drought across Europe (Section 3.1). This has to precede the section that deals with the implication on drought forecasting, which was extended to show the differences in a more comprehensive way (Section 5.2). Hence this paper should be a technical research paper instead of a short communication paper.
2	The research results that should support this conclusion as reported in this paper are somewhat lacking in both clarity and rigorousness.	We believe that the conclusions of our study support the results that different drought indices generate different number of drought occurrences/frequency and timing, which are also related to climate regions (**P17L537**). We believe that we improved clarity and rigorousness of our results in the revised manuscript through making the drought identification methods more consistent in terms of: (i) thresholds, (ii) data accumulation period, and (iii) temporal resolution (see our reply number 3 below for more details).
3	The main drawback of the analysis is the fact that the authors uses three drought indicators that rely on quite different input data and basis hypotheses to conclude that they provide a different picture of drought. This result is quite obvious after an attentive read, given the background premises: - daily data for threshold methods vs. monthly data for SSI. - 90th percentile for threshold methods vs. median for SSI (SSI=0). - Event-based approach for threshold vs. single monthly value for SSI All these discrepancies in the	We would like to thank the referee for the comments and valuable suggestions. Our paper used the drought threshold based on common practice in the drought community, which are in the range of 10-30th percentile of the flow duration curve (P70-90) for a Fixed Threshold (FT) or Variable Threshold (VT) and SSI below 0 (~P50). Our reason that we used different thresholds (50th percentile for SSI and 10th percentile for threshold for the VT and FT) was that we would like to follow common practice for the different approaches. However, the reviewer has a

	drought definition make the intercomparison a mere exercise, and its outcomes are hard to translate into actual general considerations.	point that the comparison between threshold methods (VT and FT) and SSI is not equal regarding to the use of different percentiles. Thus, in the revised manuscript, we changed the threshold from P90 into P80 for VT and FT, and SSI≤-0.84 (~P80) to have a fair comparison between different drought indices (Tijdeman et al., 2020) (**P7L193-195**). We also agree with the reviewer that our study used different temporal resolutions to analyze drought, which were daily for the threshold methods and monthly for SSI. Again, we followed the common practice to identify drought using these methods. Many studies used daily streamflow data to analyze drought using the threshold method and monthly streamflow data to analyze drought using the standardized indices. To the author's knowledge, only Tallaksen et al. (2009) and Van Loon et al. (2019) used the monthly data to derive drought using the threshold method and these were done only for a scientific purpose (**P5L136-140**). In the revised manuscript, we added analyses of drought characteristics using monthly streamflow data in both FT and VT drought approaches.
4	An additional drawback is the general lack of details on the implementation of the three approaches, which severely limits the possibility for the readers to extrapolate meaningful information from the research outputs.	We elaborated more the method section and added more drought characteristics, such as drought duration and deficit volume using different methods in the revised manuscript (Section 2.3, **P7-8**). By adding more results especially in the forecasting section (Section 3.2) (see our reply 5 below and 6c), the reader hopefully will clearly see the differences in drought characteristics because of different drought identification methods (Fig. 6, 7, 8, B1, B2, Table 3 and 4).
5	Finally, the analysis on the implications on drought forecast, which should be the main focus of the paper according to the title, is very limited in scope, and it needs to be significantly expanded in order to keep it as the focus of the paper.	We thank the reviewer for the suggestion. We expanded the analysis using the series of 12 forecasts initiated in each month (from January 2003 to December 2003) with 7 months lead time for each initiation. We do this by describing: (i) pan-European maps showing forecasted drought duration (Fig. 6) and timing (Fig. 7) using different drought identification methods (FT and VT with daily and monthly resolution, and SSI-1) (Section 3.2.1, **P13L405-P14L441**), (number of drought occurrence/frequency and drought deficit volume are provided in the Appendix B), and (ii) summary of forecasted drought characteristics identified using different approaches (FT and VT with daily and monthly resolution, and SSI-1) in the Rhine River using the series of forecasts initiated from 1st January 2003 to 1st December 2003 with a lead time of 7-month (Fig. 8, Table 3 and 4, Section 3.2.2, **P14L442-P16L519**).

6a	**Specific Comments** **Introduction** The authors should better highlight how different definitions of streamflow drought in DEWS exists also for two reasons: 1) different users have different needs that can be accommodate by different indicators (e.g. river navigation may be affected more by FT droughts that VT droughts), 2) different available input data lead to different definitions (e.g. threshold methods may not be suitable for monthly data, and daily data may not be available in near-real time).	We describe the reason why different definitions of streamflow drought exist in DEWS in the Conclusions (**P18L559-562**). However, we would like to leave the decision of using which drought identification approach to the users (**P18L568-573**).
6b	**Data and Methods** The description of the different drought indices need to be more explicit. How the drought events are defined for each index? How is the onset computed? Severity? Duration? Any event definition in the SSI? Etc. . . Also, more consistency on the adopted thresholds need to be enforced (why SSI=0 is used as threshold when 90th percentile is used for VT and FT?). It is also worth to mention that a VT method based on the same LISFLOOD data is currently operationally implemented as part of EDO (https://edo.jrc.ec.europa.eu/).	We expanded the method section in the revised manuscript, as suggested (Sections 2.2.1 and 2.2.2, **P5-7**). We added some information on drought characteristics in the method section, such as drought timing or onset (month when drought starts), number of drought occurrences/frequency, duration, and deficit volume (Section 2.3, **P7-8**). As mentioned above, we changed the drought threshold into P80 for FT and VT and SSI≤-0.84 (~P80) for the standardized index (our reply number 3) (**P7L193-195**). The suggested information about the VT method applied in EDO was added (**P18L560-562**).
6c	**Results and discussion** There is a clear unbalance between the historical analysis and the forecast. Give the title of the paper, I would aspect much more emphasis on the latter.	We extended our forecast analysis by providing: 1) maps displaying forecasted drought timing and duration across Europe using forecast data issued in July 2003 (Fig. 6, 7, B1 and B2), and 2) tables describing forecasted drought characteristics (occurrence, timing, duration, and deficit volume) using the series of 12 forecasts initiated from January 2003 to December 2003 with a lead time of 7-month (median ensemble) (see our reply number 5) (Table 3 and 4). The latter, however, can only be performed only for one river. In addition, we also provide information of number of ensemble members indicating drought in percent (x ensembles out of 25) (See Table 3 and 4).

References:

1. Tallaksen, L.M., Hisdal, H., and Van Lanen, H.A.J.: Space-time modeling of catchment scale drought characteristics, Journal of Hydrology, 375, 363–372, 2009.
2. Tijdeman, E., Stahl, K., and Tallaksen, L. M.: Drought characteristics derived based on the Standardized Streamflow Index: A large sample comparison for parametric and nonparametric methods, Water Resources Research, 56, e2019WR026315, https://doi.org/ 10.1029/2019WR026315, 2020.

3. Van Loon, A. F., Rangecroft, S., Coxon, G., Naranjo, J. A. B., Ogtrop, F. V., and Van Lanen, H. A. J.: Using paired catchments to quantify the human influence on hydrological droughts, Hydrol. Earth Syst. Sci., 23, 1725-1739, https://doi.org/10.5194/hess-23-1725-2019, 2019.