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.
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 answer to these three questions is as follows: i) Our paper uses the drought threshold based on common practice in the drought community. Using a threshold method either a Fixed Threshold (FT) or Variable Threshold (VT), drought is identified if the streamflow falls below the threshold, which is commonly in the range of 10-30th percentile of the flow duration curve (P70-90) (Hisdal et al. 2004; Van Loon, 2015). On the other hand, the standardized indices, e.g., the Standardized Streamflow Index (SSI) identifies drought if the SSI value falls below 0, which is 50th percentile (Vicente-Serrano et al., 2012). Our reason to use different thresholds (50th percentile for SSI and 10th percentile for the FT and VT) is 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 will change the thresholds 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). ii) Our study also provides results obtained from SSI-1 (Fig. A1 and A2). The main reason we used the SSI-6 for comparison with the threshold method is that SSI-6 produces a similar number of drought events than the threshold method VT and FT (Figure 2 and Table 1). SSI-1 on the other hand produces many minor drought events (Fig. A1). This is due to the selected drought threshold (P50) we used, as mentioned in point 1 above. We realize that streamflow, as included SSI-1, comprises some catchment memory aspects (delayed flow from groundwater). Hence, in the revised manuscript, we will replace SSI-6 with SSI-1 in the main text. 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. iii) We do agree with the reviewer that our study used different temporal resolution to analyze drought, which are daily for threshold methods and monthly for SSI. 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 used the monthly data to derive drought using the threshold method and only for a scientific purpose. In the revised manuscript, however, we will add 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 will have two VT and FT threshold applications: daily resolution, as frequently used, and monthly resolution to allow comparison with SSI-1.
2b	Finally, the most novel part of this paper, which deals with the implications for	We will extend the novel part of paper to illustrate that the outcome of the forecast

	drought forecasting, is rather limited and deserves more attention in my opinion.	depends on the drought identification method. We will do this by describing: (i) pan-European maps showing forecasted drought timing and duration using different drought identification methods (FT and VT with daily and monthly resolution, and SSI-1) (number of drought occurrence/frequency and drought deficit volume will be provided in the Supplementary Material), and (ii) 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 will also provide information on the percentage of ensemble members showing drought for each identification method.
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 will switch 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 will change 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).
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	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, **P5L149-151**). The reviewer also recognized that it is hard to find a suitable

	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?	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=~10,106), e.g., sample properties of streamflow in January might differ from those in August in all places (Tijdeman, et al., 2020). 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 will add information in the revised manuscript.
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 VT thresholds. In the revised manuscript, we will change the method on how we calculate the VT thresholds. We will 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).
4b	Line 366-367: You encourage using monthly streamflow data for drought	We will add the monthly drought analysis derived from the FT and VT thresholds, as

	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?	additional analysis to the daily resolution to enable comparison with the SSI-1 forecast. However, we will 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).
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. 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?	We would like to thank the reviewer for his/her valuable suggestions. We will extend the forecast results in Figure 6 with the series of 12 forecasts initiated from January to December 2003 with a lead time 7-month ahead (see our reply number 2b) including the observed streamflow. In the revised manuscript, instead of Figure 6, we will 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 Rhine River as a table. The VT method performs better than FT (Fig. 6a) since it is in a good agreement with the observed data (Fig, 4a).
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 will extend the analysis by providing: (i) maps displaying forecasted drought timing and duration across Europe using forecast data issued in August 2003, and (ii) a table describing forecasted drought characteristics (occurrence, timing, duration, and deficit volume) 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). 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, above, point 5b) will make clear that the example of the Rhine River is sufficiently representative. In addition, we will also provide information on number of ensemble

		members for which drought was forecasted (x ensembles out of 25). We would like to stress that the evaluation of forecast skill using SSI and threshold method (VT) 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 will replace 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 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.	We thank the reviewer for the suggestions. We will add the drought duration derived from the FT, VT, and SSI approaches in the revised manuscript. However, the SSI drought deficit volume will not be added because it is impossible to derive the deficit volume using the SSI approach (major drawback of standardized approaches).
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 will provide 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).
7	**Minor comments:** Line 2: "... the term streamflow drought forecasting, rather than streamflow forecasting ..." You could briefly explain difference between the two here. We will add one sentence to describe streamflow drought forecasting in the revised manuscript.	We will add one sentence to describe streamflow drought forecasting in the revised manuscript.
	Line 5: "within" Correct?	We will replace "within" with "of".
	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 suggest to stick to the definition extreme event because we identify a drought event if the streamflow falls below the P80. Droughts are like floods called extreme events.
	Line 7, 8: "observed" might be "observations"	We will change the word accordingly.

	Line 7: "a LISFLOOD model"... are there more?	There is only one LISFLOOD model. We will change "a" in "the LISFLOOD model".
	Line 10: add method to VT and FT, e.g. variable threshold level method.	The word "method" will be added in the revised manuscript.
	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 will be added. However, we will do this in the Methods section. 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 will change the word "eliminate" into "minimize".
	Line 24: "IPCC" should be "The IPCC".	Thanks for the correction.
	Line 34: This sentence slightly contradicts 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)?	We will revise L34 to avoid possible 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 …". In the preceding sentence we will explain the abbreviation NDMP (National Drought Policy Plan). 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).
	Line 41: "evaporation" should be potential evapotranspiration Line 47: "used" should be "be used" Line 85: "Proxy" should be "Proxies"	We will revise the text accordingly.
	Line 49: Mention that you specifically focus on simulated streamflow drought.	We will change "hydrological drought forecasting" into "streamflow drought forecasting".
	Line 75: "There" should be "There is"	We will remove the word "There". Thus the sentence becomes: "...., which demonstrates that no one fits all.....".
	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	We will change the text accordingly.

	median". Line 179: "definitions" ... "drought identification approaches" might be better. Line 221: "drought that has" should be "droughts that have"	
	Line 128: "were moving averaged" rephrase	The sentence will be corrected.
	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 will revise the sentence.
	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 will add an explanation about the effect of 30DMA on the forecast skill.
	Line 155-160: Add here that it is quite easy 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.	We will replace the SSI-6 with SSI-1 in the main 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.
	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 will use the threshold SSI<-0.84 to identify SSI drought in the revise manuscript.
	Line 189: Figure 3 does not show that streamflow droughts occur every year.	Figure 3 shows the drought timing 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 will change 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 will remove the sentence.
	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	In the revised manuscript, we will 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

	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?	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 will revise the manuscript by using the drought threshold SSI<-0.84 and only using SSI-1 (see our reply 2a, i).
	Line 285: "C" should be century. Line 361: "rare extreme drought events" ... extreme events are by definition rare. Rephrase.	Typo will be corrected.
	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 will extend our analysis by providing drought duration using different approaches. 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.
	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 when drought mostly started. In the revised manuscript we will 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.
	Line 337: Not only meteorological drought, also streamflow drought according to the SSI-1 (Fig. 5a).	We do agree with the reviewer, we will revise the manuscript accordingly.
	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 (**P12L354-360**).
	Line 382: "eliminate" ... not correct as minor droughts can still occur.	We will change the word into "minimize".
	Line 372-373: "the FT method produces higher drought deficit volumes and duration than VT" not shown for the pan-European dataset.	We will add drought duration and deficit volume in the revised manuscript.
	Line 375: "occurred" should be "started".	We will change the word accordingly.
	Line 377: "what being identified by" rephrase	We will revise the sentence into: what is being identified.

	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 will revise the legend accordingly.
	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 will remove 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 will change the axis into m^3 sec^{-1}.
	Figure 4: Are the grey vertical lines the hydrological years?	We will add an explanation for the grey vertical lines.
	Figure 4. You might consider using a different color when VT and FT overlap.	We will revise the color as suggested.
	Figure 5. Add grey vertical lines here as well.	We will add the grey vertical lines.
	Figure 6. Same comments as for Figure 4 and 5.	We will revise the figure accordingly.
	Table 1. Would be interesting to also compare average deficit volume and timing.	We will add drought duration and deficit volume in Table 1.

References:

1. 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.
2. Van Loon, A. F.: Hydrological drought explained, WIREs Water, https://doi.org/10.1002/wat2.1085, 2015.
3. 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.
4. 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.
5. 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.
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. 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

9. 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.
10. 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.
11. 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.
12. B Heudorfer, K Stahl: 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.