Authors reply to Reviewer 2's comments

Please find our responses and actions to Reviewer 2's comments below. The reviewer's comments are in italic and labelled with a (1) and our response is in normal font, labelled with a (2). Frequently we have placed a number of the reviewer's comments under the same section if they are related and our response covers them. Figures and tables shown here only in this reply have the prefix R (e.g. Figure R1) to distinguish them from those shown in the manuscript. In our existing responses to other HESSD reviewers, we have suggested that we rename the "natural" catchment to be termed the "benchmark" catchment; therefore, the new term is used throughout in our response here.

There are a few major things that we will change in the manuscript in general, which we would like to highlight here before discussing the specific reviewer's comments. We bullet point these major changes here, which all relate to the framing of the paper and the content:

- <u>Paired catchment analysis as a complementary method</u>. We now state clearly when introducing the method that the paired catchment analysis is a complementary method to help gain insight into the human influence in a catchment.
- <u>Methodological framework flow diagram to follow.</u> As this is a methods paper, we keep the focus on the methodology aspects and not the results of the case studies. We will introduce a flow diagram to illustrate more clearly the pairing of catchments.
- <u>Demonstrating the method with case studies of different human activities, which alleviate and aggravate droughts.</u> We now display the application of the paired catchment analysis on different human activities by showing a case study in which human activities alleviate droughts (UK Blackwater) as well as the existing Australian case study, which shows an aggravation of droughts due to the human activity. This changing of the UK paired catchment case study from the Kennet to the new case study of Blackwater also addresses a number of Reviewer 2's concerns about the UK pairing, as well as showing an alleviating human activity. The other reviewers also picked up this dissimilarity of the Kennet pairing.

Reply to Reviewer's comments

(1) This is a generally well written paper on an important topic and the paper has significant potential to make a very worthwhile contribution to the 'drought in the Anthropocene' debate. It is potentially a very useful advance arising from a simple yet potentially very effective idea.

(2) We thank the reviewer for their overall view of the importance of the topic and paper.

(1) Major Reviewer's comment: pairing of UK catchment

(1) Major Comment: impact of catchment (dis)similarity on the proposed method. The method is predicated on the similarity of the donor natural catchment/target influenced catchment, but in the UK example at least, the catchments are not similar enough, very likely leading to over-estimation of the anthropogenic effect.

(1) Catchments are different in terms of runoff response/catchment function, despite their similar rainfall.

(1) Given the more limited range of the Dun flows, my guess is the Q80 of the Dun will be somewhat higher, leading to inflated values for the deviations that are used to infer aggravated drought due to human effects.

(2) We realise that a large proportion of the Reviewer 2's comments are addressing the pairing of the UK case study specifically, highlighting the issue of the (dis)similarity between the UK pair. These comments help to show the level of information needed for the proposed paired-catchment method, and the importance of a suitable pairing. Although the original catchments of the Kennet and the Dun are similar in terms of their rainfall and geology, we agree with the reviewer that the catchments are different in terms of runoff response and it is difficult to attribute this difference solely due to human influences without further information on abstraction data. We have taken on board some of the reviewer's comments to help improve the pairing.

As this is a methods paper and the case studies shown purely demonstrate the method, we found a UK pair that is more similar, which is demonstrated by other studies (e.g. Tijdeman et al., 2018) and which helps to demonstrate the use of the paired catchment analysis on a different human activity. Therefore, we have changed our UK case study from the Kennet to the Blackwater catchment (Figure R1). The Blackwater catchment has a water transfer scheme delivering excess water to the Essex area, helping to keep it wet during the summer in dry years (Robinson, 2011). Tijdeman et al. (2018) suggest the pairing with a similar catchment, Chelmer, to represent the natural situation with very similar catchment characteristics. Chelmer is identified to have minor artificial influences, with the main difference being the human activities present in Blackwater.

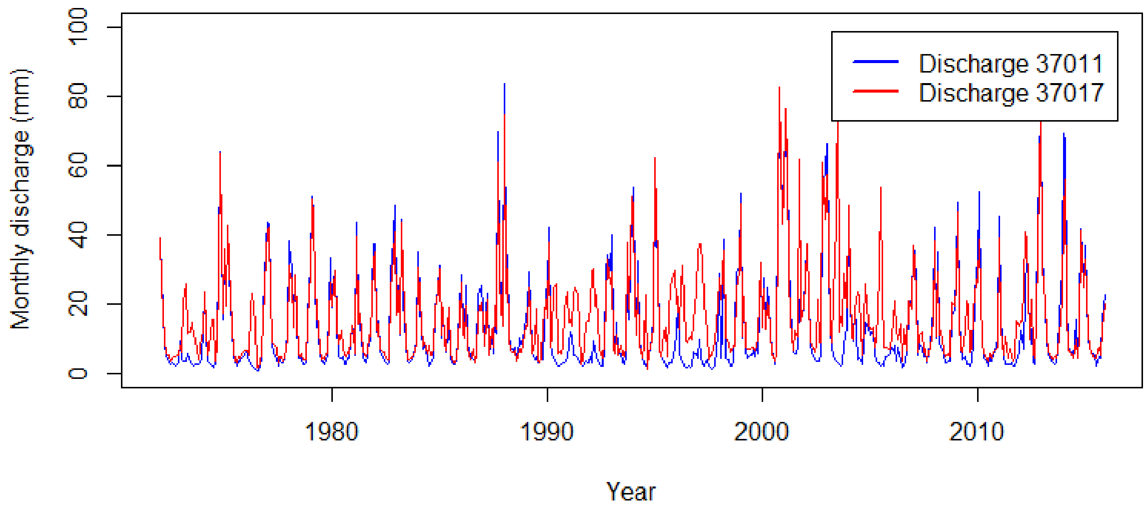

Figure R1: Discharge plotted (mm/month) for both catchments in new UK case study pair: Benchmark, Chelmer (blue) and human-influenced, Blackwater (red)

While the focus of this methods paper is to use the case studies to illustrate the method rather than attributing differences to specific causes, we think that choosing a different and less controversial pairing with more similar runoff response will better highlight the advantages of the method.

Please find below the rewritten section to replace "Section 3.2 UK paired catchments: Kennet and Dun" for the paper and the new case study and results for Table 3 and Figure 3, and updated Table 2:

"**Section 3.2 UK paired catchments: Blackwater and Chelmer**

The Blackwater catchment receives water transfers as part of the Ely Ouse water transfer scheme for the greater London area (NRFA, 2018; Tijdeman et al., 2018). The scheme was introduced in 1972 by the Environment Agency to help address anticipated water stresses due to population increase and expansion and development in the South Essex area (AEDA, 1990). Blackwater catchment was paired with a catchment nearby as its benchmark pair, Chelmer (Figure 2), due to its similarity in catchment characteristics (Table 2). Both catchments have mixed permeability superficial deposits geology (86-88%), are predominantly rural land use catchments and have similar annual rainfall totals (within 10%) (Table 2). Both catchments have very low urban extent (Chelmer 4.9% and Blackwater 5.4%) and the land uses are very similar, with arable land covering ~70-75% in both (NRFA, 2018). The observation data available for both catchments ran from 1972 to 2015 with no missing data, covering a number of important drought events in the UK.

The drought analysis shows that many droughts experienced in the natural catchment were alleviated in the human catchment due to the water transfer scheme (Figure 3; Table 3). Notably, the 1976 UK drought was not as severe in the Blackwater catchment as its benchmark pair. A number of other major drought events occurred in Chelmer in the 1990's and 2003 were not seen in Blackwater, therefore showing that they were alleviated due to the elevated flows from the water transfer scheme (Figure 3)."

Table 3: Paired catchment analysis results for UK case study, Blackwater (Human) and Chelmer (Benchmark).

	Occurrence	Duration			Deficit		
	Frequency	Total no. of months in drought	Average duration (months)	Maximum duration (months)	Total deficit (mm)	Average deficit (mm)	Maximum deficit (mm)
Benchmark	22	86	3.9	10	163.8	7.4	29.6
Human	7	16	2.3	4	39.3	5.6	16
% increase due to the human influence	-68%	-81%	-42%	-60%	-76%	-25%	-46%

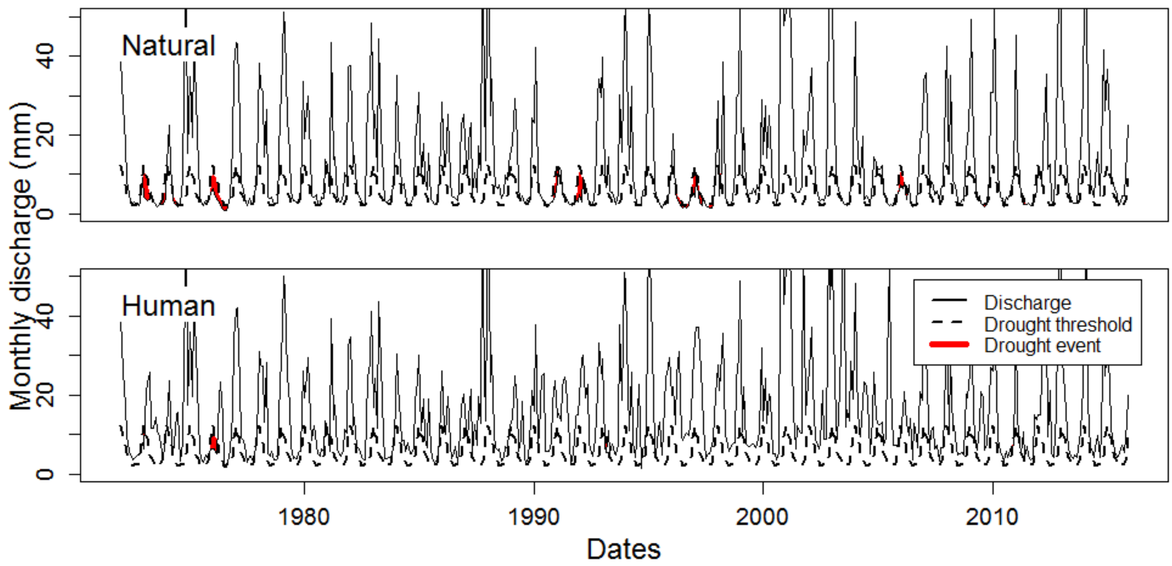

Figure 3: Drought analysis results for the UK pair, Blackwater (human) and Chelmer (benchmark) (1972 – 2015). Black solid line represents streamflow, dashed black line represents Q80 variable threshold, and red areas identified drought events.

Table 2: *Catchment data about each paired catchment case study*

Case study	Human activity	Catchment status	River/ Station	Catchment area (km^2)	Geology	Average annual precipitation (mm)	Average annual flow (mm)	BFI
UK	Water transfer scheme	Benchmark	Chelmer (37011 Churchend)	72.6	London clay and chalk, overlain with Boulder Clay	591	91	0.43
		Human	Blackwater (37017 Stisted)	139.2	London clay and chalk, overlain with Boulder Clay	579	194	0.5
Australia	Ground water abstraction for irrigation	Benchmark	Cockburn (419016 Mulla Xing station)	907	Alluvial overlying fractured rock (granite and sedimentary)	665	64	0.24
		Human	Cox (419033 Tambar Springs station)	1450	Bedrock-contained alluvial valley	732	21	0.21

By changing the UK case study, we hope to have answered the reviewer's questions. If needed, we could however also increase the number of case studies presented in the manuscript to fully demonstrate the method. In doing so, we would also cover a range of human activities, to showcase how the method can be used on different human activities and to take the emphasis away from the contentious results of only groundwater abstraction case studies. As well as a the UK Blackwater case study (water transfer) and Australia Cox case study (groundwater abstraction), we can show another UK case study (e.g. Candover, flow augmentation scheme) and a Mexican case study (e.g. Torres, urbanization). However, we have this multiple case study comparison currently as future research in preparation for publication and we would prefer to keep this paired catchments paper as a methods based paper.

(1) In terms of the results, the figures quoted in Table 3 seem very large. Given the nature of the debate around abstraction impacts, these figures could be quite contentious, as the Kennet is something of a poster child in the debate around sustainable abstraction, and the authors should do more to ensure they are meaningful.

(2) We agree that the numbers were very large for our Kennet case study, but this might be realistic given the published concern over over-abstraction in the Kennet area, and recent changes to abstraction rates to reduce this (2014 EA issued a notice reducing the permitted abstraction from Axford and revoked the abstraction licence on the little River Og). Therefore it is legitimate that impacts seen over the previous decades could reflect this over-abstraction.

e.g. "Thames Water accused over dry River Kennet in Wiltshire", BBC, 27 September 2011 (https://www.bbc.co.uk/news/uk-england-wiltshire-15076406)

e.g.2: "For over 20 years there has been local concern that flows in the Upper Kennet above Marlborough have been affected by over-abstraction' (Kennet catchment partnership). (http://www.kennetcatchment.org/issues/over-abstraction/)

However, given that the focus of this paper is to introduce the method, there is concern that bringing in this level of information about the individual case studies will distract away from this. We will include a summary sentence demonstrating the level of required information about the human activity and reported impacts of the presented case studies, but a more detailed paragraph should be kept for Supplementary information (if needed) rather than to extend the current paper and take the focus away from the method itself.

(1) **Overall reviewer's comments: considering other approaches to bolster the method / convincingly demonstrate the method:**

(1) Given these concerns, the authors could consider some approaches to bolster the method and provide verification – e.g., how this method performs relative to other approaches or modifications, e.g. detecting deviations based on rainfall (as in Tijdeman et al. 2018) and PE.

(1) Paired catchment analysis is a staple of experimental hydrology, but is much harder to do in 'real world' examples when it is not possible to control all variables except the main intervention of interest, and even the latter may be poorly understood.

(1) I feel major revisions are needed to convincingly demonstrate the method, through modifiying it to allow some tolerance in the donor/target relationship, verifying the methods using independent abstraction data, or benchmarking it against other methods.

(2) We argue that the paired catchment approach can be extremely beneficial if you have the right level of information to justify the pairing, but not enough detailed data on the human influences

themselves (i.e. time series of abstraction rates) or to run alternative approaches (i.e. upstream-downstream approach, Rangecroft et al., 2016; observation-modelling approach, Van Loon and Van Lanen, 2013; 2015). Therefore, the paired catchment approach can help provide insight into the human influence in the catchment from the observation data available and might be a first step to obtain more data to apply alternative approaches. Exploring the use of other methods to obtain more robust results for the example case studies is not the purpose of the paper, but we will definitely reframe to explain that this method can be complementary to others.

Furthermore, we have also changed the framing of the paper in terms of the case studies presented. We now propose to show a case study which has a human activity aggravating hydrological droughts (e.g. Australia Cox – groundwater abstractions) but also a case study in which the human activity alleviates hydrological droughts (e.g. UK Blackwater – water transfer). We believe that this will help demonstrate the method and its applicability better.

To help illustrate the method further, we also propose the addition of a flow diagram for the pairing of catchments for the analysis.

(1) In general the approach could be strengthened considerably by taking a more water balance approach as done in the classic paired experimental catchment studies, and also in the study of Prosodcimi which incorporates climate variables to account for any confounding effects.

(2) In the new framing we will emphasise the use of the paired catchment approach when less detailed information is available to still make an assessment of the difference between the two catchments. The use of the same time period as a comparison between the two catchments means that overall climate is accounted for in the analysis, which is an advantage compared to methods comparing data pre- and post-disturbance.

The water balance approach requires a high level of information, including actual evapotranspiration, storage changes (e.g. soil moisture, groundwater), and does not focus on drought specifically; instead we are looking for a method which can be applied to multiple case studies to give an idea about the impact of human activity on drought only based on commonly available data.

(1) Reviewers comment: include abstraction rates information

(1) verifying the method using independent abstraction data, or benchmarking it against other methods.

(1) The Kennett is very well known to experience major abstractions, which have been non-stationary over the series. But more could be done to follow this study up – there is anecdotal information on abstractions in various grey literature sources I found online (below).

(2) As mentioned, some abstraction values are available for the UK case study through grey literature. However, it is difficult to obtain actual abstraction values, rather than licenced values, and

for the Australian case there is no information available at all. Therefore it can be very hard to give an estimate of the overall net groundwater abstraction during the time series.

We agree that human activities are non-stationary over the investigated time period. This is one of the reasons why we calculate overall drought characteristics for the whole time series, rather than an event-by-event analysis.

(1) Reviewers comment: Impacts and abstraction data

(1) Finally, to really demonstrate the success of the method, it would be nice to have some independent verification of the suggested impacts. I appreciate access to abstraction data is not straightforward for the UK, but might be possible for one catchment, at least for derived data on impacts rather than particular abstractions.

(2) We would like to be careful not to allocate too much of the paper to the two case studies as we feel that this will detract away from the focus of the paper, which is the method itself. However, we have included more information here to give a background on the new UK pairing (see new Section 3.2). It is also important to note that gaining the level of information we have for the UK is possible, but it is difficult to get abstraction data for many other regions of the world, including Australia.

(1) I would suggest some dialogue with the EA would be worthwhile, as there seem to be naturalised data (be decomposition and/or modelling) available for the Kennet for various part studies.

(2) We fully agree that a comparison between observation data and naturalised data for the human influenced catchment would complement this method and paper, to see if the results are similar. However, this is not the focus of the paper and we do not fully agree with the use of the EA naturalisation data as we do not have information about which processes and activities are included in the modelling and/or decomposition. The naturalisation data has large uncertainties as well and cannot be used as a benchmark or independent verification.

(1) Reviewers comment: using threshold from benchmark catchment

(1) However, transferring a threshold directly from one catchment (reading off the Q80 flow value from the natural catchment and applying it to the influenced one) to another seems like a potentially dangerous business. This might not be a problem if one is just trying to estimate flows at an ungauged site, and can report uncertainties; but in the present method any biases arising from the data transfer could be very misleading.

(2) The transferring of thresholds from the donor catchment to the influenced catchment is a basic principle of the paired catchment approach. Furthermore, the transferring of thresholds from the benchmark situation to the human-influenced situation has been used in other existing literature, with regards to the comparison of naturalised data and observed flow (observation-modelling framework, Van Loon & Van Lanen, 2013; Van Loon & Van Lanen, 2015). It is also used by the large-

scale hydrological modelling community when analysing future droughts – use of pristine threshold for both the pristine and human scenarios to calculate the modelled human impact on hydrological droughts (e.g. Wanders & Wada, 2015).

The main issue with using a threshold established on the human-influenced catchment discharge is that the effect of the human activities is then included in the threshold used to calculate droughts. This makes it harder to compare the observed situation with the expected normal. The use of the benchmark catchment for the threshold allows a better representation of the expected normal without the human activity.

We currently choose not to explore this avenue of analysis because it would complicate the paper too much. However, we show the results here for one case study, UK Blackwater, to example the difference that can be observed for the reviewer (Figure R2 & Table R1). Because the human influence is now included in the threshold (Figure R2), using the own threshold results in lower numbers of the human influence on drought (Table R1). We feel that this underestimates the human influence on drought, but if required, we can include both the application of own thresholds and the application of the benchmark threshold in the manuscript.

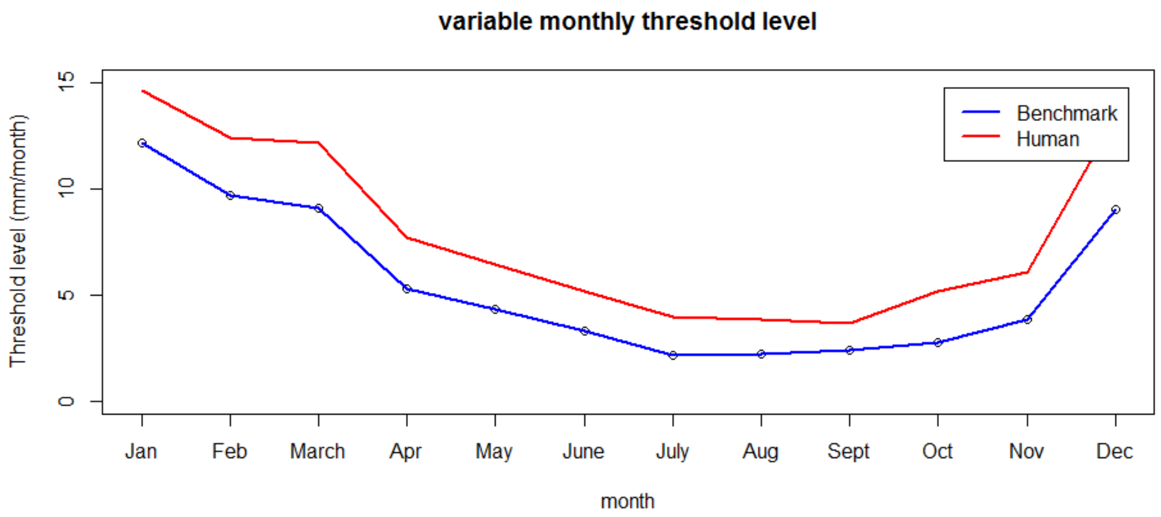

Figure R2: Variable Q80 thresholds for the benchmark catchment (blue) and human catchment (red)

Table R1: UK Blackwater case study: Percentage change in hydrological drought characteristics due to human influence calculated with either own station thresholds or benchmark threshold.

Hydrological droughts % change detected in human catchment compared to benchmark	**Freq.**	**Duration**			**Deficit**		
		Average	**Max**	**Total**	**Average**	**Max**	**Total**
Own threshold	+14	-32	-40	-23	-24	+1	-14
Benchmark threshold	-68	-42	-60	-81	-25	-46	-76

(1) Reviewers comment: Impact of catchment (dis)similarity on the proposed method

(1) Put simply, the method applied in this paper can only work if the donor's natural flow regimes is near-identical to the 'theoretical natural' flow regime of the target site. Any deviation between these regimes will be interpreted as anthropogenic; when it could just be due to the variations between two catchments that appear quite similar but are in fact different.

(1) It is difficult to find suitable pairs. Even when catchments are in principle very similar (geology, rainfall etc), the concept of 'uniqueness of place' is a major obstacle.

(2) We agree that 100% proof of similarity cannot be given when datasets are used of catchments already disturbed by the human activity. Experimental hydrology (control and treated catchments/plots) and models are also unable to provide this. We can only reliably check the similarity of both catchments if we have long time series (preferably 30 years to calculate the thresholds) for the donor and the human-influenced catchment prior to human disturbance. In practice, this will be very rare or none-existing for catchment pairs.

(1) As a result, I do not think the authors can claim 'attribution', and the claims of the paper need to be reconsidered.

Note that this an important difference in the urbanisation paper (Prosdocimi et al., 2015) or in the classic experimental catchments, which all incorporate some data on the intervention in question into the analysis (e.g. the land cover data used by Prosdocimi et al.).

(2) We agree that we can rephrase our results from attribution to likely cause. We will cover this in our revised discussion section, stating that is it about all human influences (and some minor differences between catchments under natural conditions).

We can include abstraction rate information, even if from grey literature sources online, when available, and more detailed information on land use. As previously stated, we are cautious to not include too much detail about the two case studies (Cox, Blackwater) as it is a methods paper and the focus should remain on demonstrating the method, not on the discussing the case studies themselves.

(1) Specific reviewer's comments

(1) A technical matter: In Figure 3, I'm surprised to see so little of the flows being below the threshold. It does not look like 20% of the flows are below the threshold to me – can the authors please check?

(2) We can confirm that it is correct. The total number of months in drought (Table 3) is close to 20%: 93 months in drought out of 540 months (17%). It is not the full 20% because we have dropped minor droughts that were only 1 month in duration.

(1) P2, L21. Another approach is using deviations in the P-Q relationship, e.g. Tijdeman et al., 2018.

(3) This will be added in.

(1) P2, Intro. The paper would do well to refer to the expansive literature in hydroecology which also tackles a similar problem of estimating 'natural' flows for sites, against which impacted flows can be compared. The classic papers of Brian Richter are a good start, and I'm fairly sure methods have been proposed to transfer natural flow percentiles (but using a whole FDC approach; try the DHRAM work by Andrew Black, Dundee as a start). Another area where this is done routinely is through the LowFlow software produce, a regionalisation product which estimates natural and disturbed FDCs at any site. Its not drought specific, but definitely has a very similar aim.

(2) We agree that literature from the hydroecology field can be brought in to show how they address the similar problem. We thank the reviewer for the suggestion and the references, and we will update the introduction.

(1) P4, Sect. 2.3 Given the concerns raised about the UK catchments, this section needs to be reconsidered.

(2) This section will be changed in the revised manuscript to introduce the new case study, Blackwater, and update all associated sections, tables (Table 2; Table 3) and figures (Figure 3).

(1) P6, L2. The 80th percentile is not what is being used here. This paper uses the 20th percentile, or, as is most commonly referred to in hydrology, Q80: the 80% non exceedance threshold from the flow duration curve.

(2) The authors agree that this needs rewording, and will be for the revised version throughout.

Existing phrasing of: "Here the 80th percentile was used as the threshold, meaning that 80% of the time discharge is above the threshold. The 80th percentile is a commonly used threshold to identify drought events (Hisdal & Tallaksen, 2000; Fleig et al., 2006; Heudorfer & Stahl, 2016)."

Has now been changed to:

"Here the 80% non-exceedance threshold (Q80) from the flow duration curve was used. This means that 80% of the time discharge is above this threshold. The Q80 is a commonly used threshold to identify drought events (Hisdal & Tallaksen, 2000; Tallaksen and Van Lanen, 2004; Fleig et al., 2006; Heudorfer & Stahl, 2016)."

(1) Discussion: is generally very insightful but definitely needs reconsidering in light of catchment selection issues, and claims about attribution needed to be moderated.

(2) We thank the review for their comment and the suggestion to highlight the limitations and issues further. We will change the discussion to include more about these aspects.

We thank the reviewers for all their comments and suggestions, and we hope that a number of them have been satisfied by the change of UK catchment, and that we have strengthened the paper with the suggestions enhanced analysis and provided a more clear aim (i.e. a methodological paper rather than a case-study type of paper).

References

AEDA, 1990. The Ely Ouse Essex water transfer scheme, Agricultural & Environmental Data Achieve. http://www.environmentdata.org/archive/ealit:4162

BBC, 2011: "Thames Water accused over dry River Kennet in Wiltshire", BBC, 27 September 2011, https://www.bbc.co.uk/news/uk-england-wiltshire-15076406

Kennet catchment partnership: http://www.kennetcatchment.org/issues/over-abstraction/

Rangecroft, S., Van Loon, A. F., Maureira, H., Verbist, K., Hannah, D. M. 2016. Multi-method assessment of reservoir effects on hydrological droughts in an arid region, ESDD, doi: 10.5194

Robinson, W. 2011. Essex & Suffold Water: Ely Ouse to Essex Transfer Scheme. https://www.raeng.org.uk/RAE/media/General/News/Documents/20111027-William-Robinson.pdf

Tallaksen, L.M. and Van Lanen, H.A.J. (Eds.) (2004) Hydrological Drought. Processes and Estimation Methods for Streamflow and Groundwater. Developments in Water Science, 48, Elsevier Science B.V., 579 pg.

Tallaksen, L.M., Hisdal, H. and Van Lanen, H.A.J. (2009) Space-time modeling of catchment scale drought characteristics. J. of Hydrol. 375:363-372 (doi:10.1016/j.jhydrol.2009.06.032).

Tijdeman, E., Hannaford, J. & Stahl, K. 2018. Human influences on streamflow drought characteristics in England and Wales, HESS. Vol 22(2): 1051-1064.

Van Loon & Van Lanen, 2013. Making the distinction between water scarcity and drought using an observation-modeling framework. Water Resources Research, vol 49(3): 1483-1502.

Van Loon & Van Lanen, 2015. Testing the observation-modelling framework to distinguish between hydrological drought and water scarcity in case studies around Europe. European Water, 49: 65-75.

Wanders, N. & Wada, Y. 2015. Human and climate impacts on the 21st century hydrological drought. Journal of Hydrology, vol 526: 208-220.