Referee 2
Author Response

1. General comments
   This paper assesses the impacts of different storylines of UK drought based on the 2010-2012 drought event. The results demonstrate the importance of meteorological preconditions, catchment characteristics controlling recovery time and the vulnerability of UK catchments to a 'three dry winter' scenario. Overall I enjoyed reading the paper, it is nicely written and figures are well presented. There is some interesting analysis and conclusions that will be of great benefit to those working on drought in the UK and further afield. However, I do have some major comments for the authors to consider. In particular, some of the methods need clarification and better justification, and there needs to be more critical discussion and reflection on the use of storylines in drought analysis

RESPONSE: We thank Dr. Coxon for the positive feedback and suggestions on how our manuscript can be improved. We are grateful that the reviewer agrees that our results have benefit to those working on droughts. We respond to each comment given in the text below (in bold and italics).

2. Main comments
   Plausibility. As noted in the introduction, 'Storylines are defined as physically self-consistent unfoldings of past events and the plausible evolution of these events in a future climate (Shepherd et al. 2018).'. I would like to challenge the authors and encourage more critical discussion in the manuscript on how 'plausible' the storyline scenarios are. You have implemented a number of different storylines but there is very little consideration of the plausibility of these storylines in terms of the atmospheric conditions that are needed to create them. Where is the evidence that you are implementing 'plausible' changes to this event that link to physical climate processes? What is the evidence that these are really 'physical climate storylines'? You note that the 12month precipitation-deficits from the storylines are in line with other climate scenarios but many of your scenarios are based around precipitation deficits that span more than one year (i.e. up to three dry winters). The manuscript needs more critical discussion of the plausibility of the storylines and a fuller consideration of their limitations.

RESPONSE: A similar point was also raised by reviewer 1 who was concerned about the plausibility of altering observed precipitation independently of temperature in the storylines of precondition severity. Fig. R1 shows the observed relationship between monthly precipitation and PET from 1965-2015 which shows no clear correlation apart from a slight negative correlation in spring and summer. This shows that our precipitation perturbations are plausible and do not violate any correlation structures between precipitation and temperature. This is further shown by Fig. R2 (addressing temporal correlations) which shows the equivalent values after precipitation 3- (i.e. OND 2009) and 6-months (i.e. JASOND 2009) before the 2010-12 drought reduced to match OND and JASOND precipitation at four return periods, and which are seen to fall within the historical relationships. Both figures will be included in

the revised manuscript. We will also emphasize in the revised manuscript that the creation of event-based storylines in other locations should consider potential correlation between the different variables if a strong correlation is found. The Environment Agency vulnerability framework and the high-end H++ climate change scenarios were intended as a point of comparison when discussing the implications of the storylines of precondition severity instead of a justification of their plausibility. We will make this clearer in the revised manuscript.

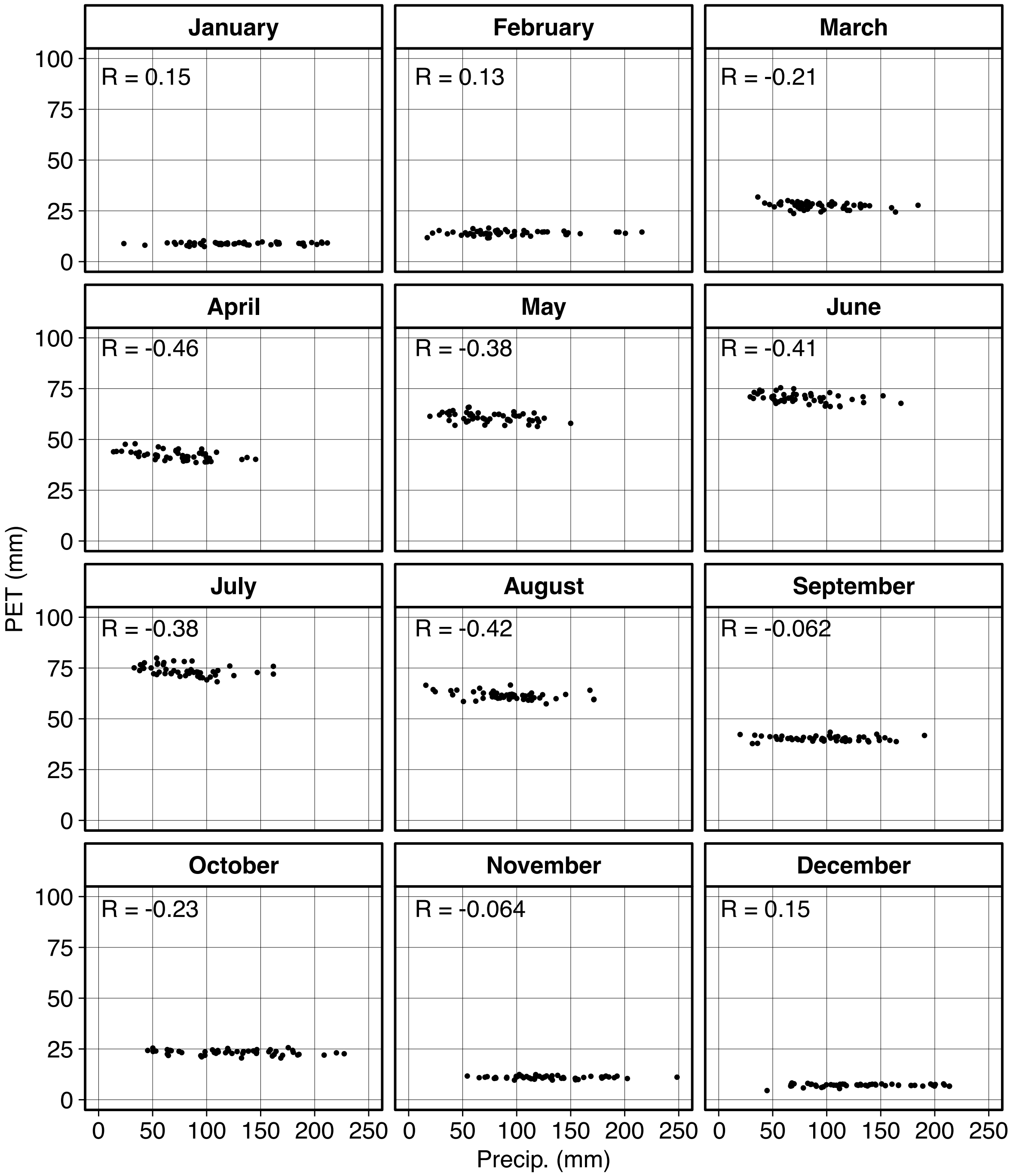

Figure R1 Observed relationship between PET and precipitation for each month for the period 1965-2015 averaged across the 100 UK catchments selected with the correlation coefficient value shown for each month.

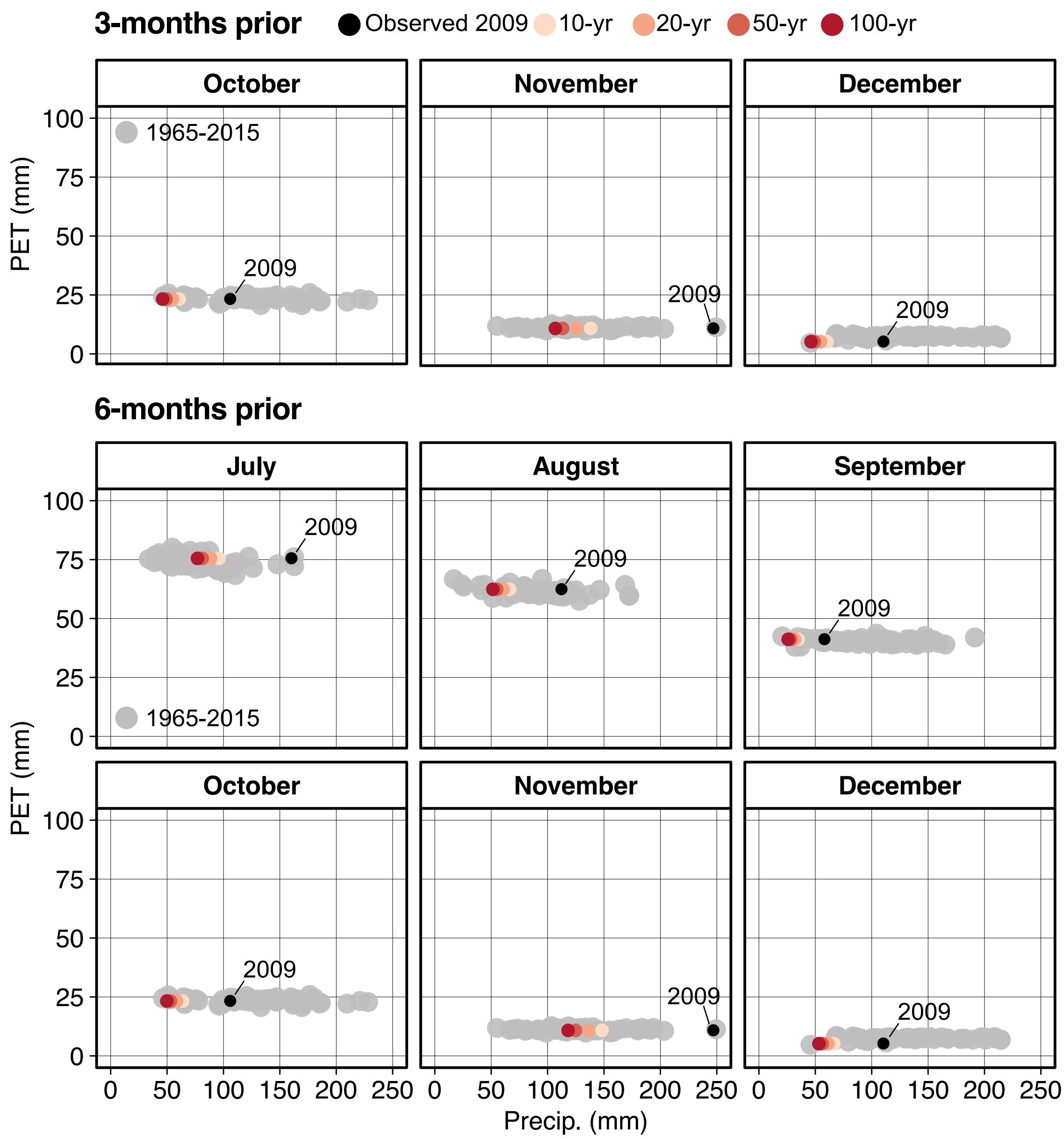

Figure R2 October to December monthly precipitation and PET (1965-2015) (top) and July to December monthly precipitation and PET (1965-2015) (bottom) The black circle indicates observed value in 2009 while the colored circles indicate the value after the precipitation 3- (top) and 6-months (bottom) prior to the 2010-12 drought is reduced at four return periods.

With regard to precipitation perturbations over a longer time period, a large number of previous studies have investigated the occurrence and likelihood of sequences of dry winters and their implications for UK water resources. Past multi-year droughts have been shown to include at least one dry winter (Environment Agency 2009; Watts et al. 2012; Folland et al. 2015). The effects of successive dry winters have been shown through multiple reconstructions of precipitation, river flows and groundwater levels across the UK (e.g. Spraggs et al. 2015; Barker et al. 2019; Watts et al. 2012; Bloomfield et al. 2019). In river flow reconstructions, a "third dry winter" scenario was shown to cause significant reduction in storage at key reservoirs in East Anglia during 1943-46 (Spraggs et al. 2015). Similarly, precipitation reconstructions showed that even longer dry spells are plausible with "the long drought" between 1890-1910 characterized

by three or more successive dry winters punctuated by wet interludes (Marsh et al. 2007). Quantifying transition probabilities of consecutive dry seasons, Wilby et al. (2015) found that the longest spell of consecutive dry winter or summer half-years spanned 4 years (including 4 dry winters) in the 1870s in the England and Wales Precipitation (EWP) time series. The same study also found the longest observed sequence of consecutive river flow deficit reached 5.5 years during the 1988-93 drought in southern England.

Motivated by similar aims as the "three dry winters" storyline in this study, water companies have previously considered the hypothetical situation of a third dry winter following the 2004-06 drought which was characterized by two consecutive dry winters (Environment Agency 2009). Similarly, there was widespread concern and expectation in early 2012 that dry conditions would persist based on the prevailing atmospheric conditions at the time (Bell et al. 2013; Spraggs et al. 2015). The repetition of a dry year to represent continued dry conditions is therefore a reasonable and plausible case to investigate given concerns at the time. Figure R3 shows more clearly the differences in atmospheric circulation between the repeated year and the year replaced. Average geopotential height at 500hPa (Z500) anomalies from ERA5 for 2010 (repeated year) compared with 2012 (replaced year) show that summer and winter 2010 was characterized by high pressure over parts of the UK throughout the year. Conversely, 2012 was characterized by low pressure over the UK and high rainfall totals across 2012 which terminated the drought. We will expand and cite the studies above to better justify the plausibility of the three dry winter storyline in the revised manuscript.

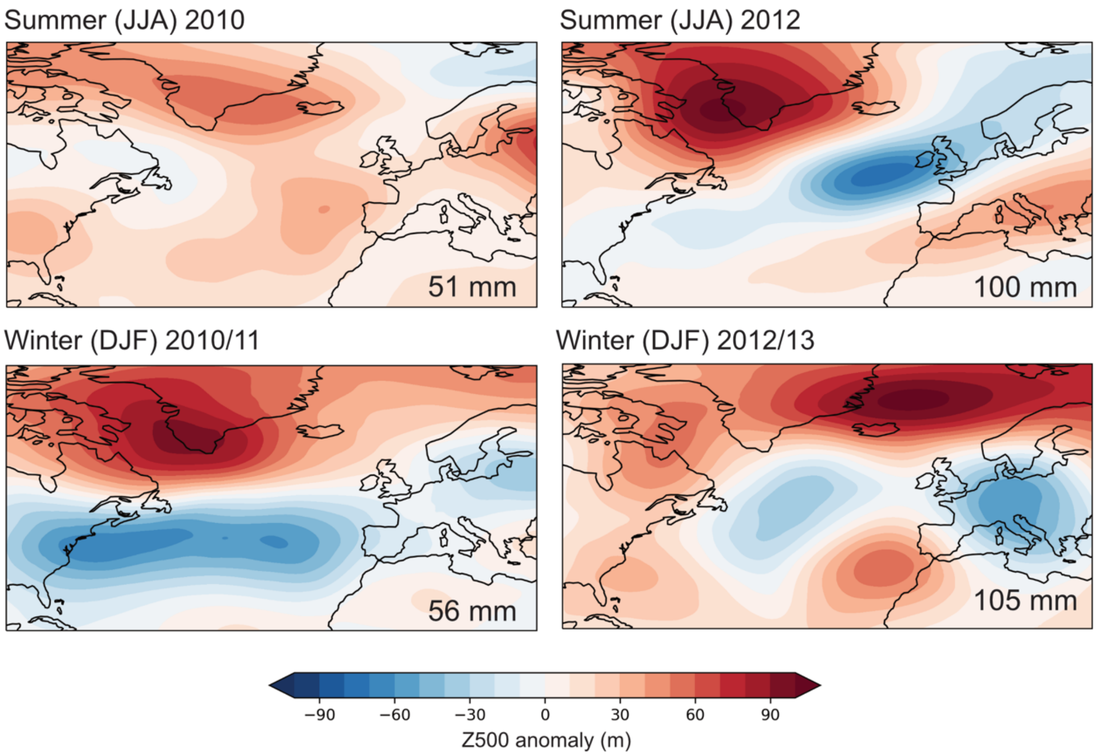

Figure R3 Summer and winter geopotential height (m) at 500 hPa (Z500) anomalies relative to 1978-2015 from ERA5 for 2010 (left) and 2012 (right). The average monthly rainfall totals (mm) are shown in the corner for the respective year.

3. Delta change approach. Aligned with the comment above is the use of the delta change approach to represent changes in climate. There are a whole host of problems with delta change approaches (see Fowler et al, 2007 https://doi.org/10.1002/joc.1556) and again, in terms of plausibility, I think it is difficult to argue that applying mean monthly factors to a past drought event gives you a realistic picture of the 'hydrological impacts of climate change'. Again, there is no critical discussion of this in the paper.

RESPONSE There are a whole host of problems associated with bias correction and downscaling (Maraun et al. 2017), with the realism of climate model simulations (especially for persistent circulation extremes), and with knowing how atmospheric circulation will respond to climate change (Shepherd 2014). There is no easy answer here, and nobody can claim to predict the future under such conditions. The storyline approach sacrifices generality for physical plausibility. In particular, our aim here is to place the 2010-12 drought in a future climate instead of generalizing the hydrological impacts of climate change across dissimilar drought events. We believe the delta change method is suitable for this as retaining the baseline temporal sequence of the 2010-12 drought increases realism and enables quick comparison with the other storylines which were also created based on altering the observed time series of the 2010-12 drought.

Although the delta method omits the influence of changes in wet/dry sequences in the storylines of climate change, the other storylines created in this study consider changes to the wet/dry sequence of the observed drought. Despite the limitations of the delta change approach, it remains widely used in hydrological climate change impact assessments globally. It also remains the most widely used method for UK catchments and has been used consistently since the 1990s to reach important conclusions on the potential impacts of climate change on UK water resources. We will expand on this justification of the delta change method in the methods section in the revised manuscript. There are alternative emerging methods available to place historical events under future warming. This includes searching for analogue events (e.g. Cattiaux et al. 2010), the use of large ensemble climate model data (e.g. van der Wiel et al. 2020) or atmospheric nudging of climate models (e.g. van Garderen et al. 2021). We will expand on these alternative approaches in the revised manuscript.

4. Estimating return periods. In Section 2.2.1 you use annual average three month rainfall from 1965 – 2015 to estimate 10, 20, 50 and 100-year return periods. Firstly it is not clear what the source of this rainfall data is (I assume CEH-GEAR as this is referenced below?). Secondly, if it is CEH-GEAR (or Had-UK) then the rainfall data are available for much longer time periods (1890- 2017). So why choose a shorter time period which could make your estimates less robust, particularly when you are trying to estimate a 1 in 100 year return period of rainfall?

RESPONSE: Apologies for the confusion, the precipitation data we used was CEH-GEAR. We will amend the typo in the data availability section in the revised manuscript. We chose the time period of 1965-2015 as the baseline period as we

did not have temperature (and PET) data for the longer time period for hydrological modelling. However, we agree that estimates could be more robust with the full dataset. As suggested by the reviewer's comments, we can revise our estimates of rainfall return periods using the full CEH-GEAR dataset and amend Figures 5 and 6 accordingly.

It should be noted that the aims for the storylines of precondition severity are not to improve the estimates of rainfall totals at a particular return period but rather to investigate sensitivity of different catchments to various magnitudes of rainfall perturbations. We believe this aim was satisfied with the estimates of return periods using the shorter period (1965-2015) and do not anticipate changes to the overall conclusions with return periods calculated from the full dataset.

5. Catchment recovery time. I don't really understand why you choose the baseline simulation as your threshold for the catchment recovery time. This isn't necessarily an indication of the catchment having 'recovered' – the baseline simulation may still be very low flows. Is the time calculated from the very beginning of the simulation? This metric needs to be better clarified and justified.

RESPONSE: The aim of quantifying the "catchment recovery time" was to investigate how long the influence of different precondition perturbations persists for each catchment and how that might relate to physical catchment characteristics. The metric was calculated from the start of the perturbation until the influence of the perturbation is no longer detected (<1% compared with baseline). Based on the reviewer's comments, we realize the use of the term might be confusing. We will define this better in the revised manuscript to clarify that the catchment recovery time as calculated in this study is not indicative of how long the catchment took to recover from drought conditions but is instead indicative of how long the influence of precondition perturbations lasts for each catchment. A similar experimental set-up was also proposed in Staudinger and Seibert (2014) which calculated a catchment "relaxation" and "persistence" time from perturbations to initial conditions. This is also consistent with Stoelzle et al. (2020) which proposed the evaluation of drought stress tests using what they termed "catchment recovery duration" which is defined in the same way as in this study.

6. Model Performance metrics. Better justification for this choice of metrics is needed – what do they represent and why are they appropriate for this analysis? Should NSE (a metric focused on high flows) really be given equal weighting? Some maps of model performance (where dots are coloured by their best NSE/logNSE value for example) would be useful so we can see the spatial differences in model performance. I would expect more detailed analysis of how the model performs for the 2010-2012 event given the focus of the paper.

RESPONSE: We agree that there should be more clarification and discussion of model performance during past droughts and the 2010-12 drought. The model parameters were taken from Smith et al. (2019). In that study, the authors used a Latin hypercube sampling calibration approach across the selected metrics and demonstrated the model performance of the top 500 parameter sets. The

authors also showed that periods of drought identified from simulated river flow match observed occurrence of past droughts well. This study takes the top 500 parameter sets from Smith et al. (2019) and re-ranks them based on performance during dry years. We will clarify this in the methods section to justify how the parameter sets and calibration strategy are appropriate for drought analysis. As suggested, we will include maps of NSE and logNSE in the supplementary material. We also propose to include Figure R4 which shows simulated river flow during the 2010-12 drought for nine example catchments spread across the UK, showing the ability of the model to reproduce low river flows across the catchments during this period.

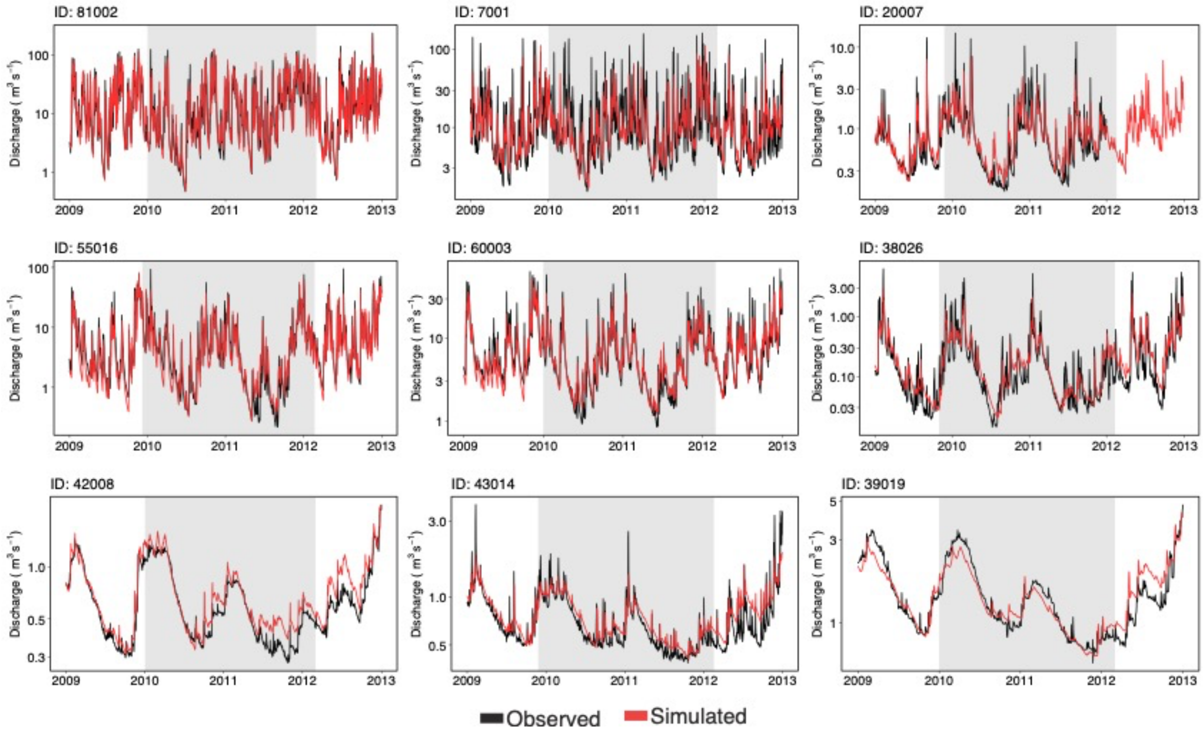

Figure R4 Daily observed (black) and simulated (red) river flow across nine example catchments from the top parameter set in re-ranked parameter ensemble from Smith et al. (2019). The y axis is presented in log scale.

7. Data Availability. The data availability section needs to cover all the data used and produced in the paper. Will you be making the storyline input data available (i.e. the modified rainfall and temperature timeseries) for others to use? Will you be making the outputs available? This is important for reproducibility, transparency etc.

RESPONSE: As the aim of our study was not to create a new dataset, we did not describe the input and output data in the data availability section. However, if the editor and reviewer believe that the input and output data could potentially be of interest to the community, we can make the modified rainfall and temperature time series and the simulation outputs for each storyline available in the interest of transparency (possibly via zenodo or similar repository?).

Technical comments

1. L14. 'highly conditioned by its meteorological preconditions'. Not entirely sure what you mean here, can you clarify?
   RESPONSE: We will rephrase. What we meant was that the spatial and temporal characteristics of the 2010-12 drought were highly influenced by the meteorological conditions 3- and 6-months prior to drought inception.

2. L55. You might also consider citing Dobson et al (https://doi.org/10.1029/2020WR027187) which considers the future spatial dynamics of droughts and water scarcity across England and Wales.
   RESPONSE: Thanks for pointing us to Dobson et al. (2020). We will cite this in the revised manuscript as suggested.

3. L116. It would be useful to add a map of the catchments (with the catchment boundaries) into the supplementary information. This would help highlight their size and spatial coverage across GB.
   RESPONSE: Thanks for the useful suggestion. We will add this to the supplementary information.

4. L150. 'The temporal variability of the reduced preconditions precipitation'. This doesn't make sense to me and should be reworded.
   RESPONSE: We will rephrase this.

5. Figure 9 – how much variation is there in the percentage/absolute changes between the different clusters? i.e. are the projected changes in rainfall very different for cluster 1 compared to cluster 5? Might be worth adding these plots to the supplementary information for context as most of the subsequent analysis is focused on the changes for each cluster.
   RESPONSE: We will add a supplementary figure on projected change in rainfall across the different clusters as suggested. We anticipate this will broadly reflect differences between the clusters in changes in drought characteristics in the storylines of climate change section (i.e. Fig.11).

6. Figure 12 is quite blurry – can you increase the resolution?
   RESPONSE: We will modify this.

References

Barker, L. J., Hannaford, J., Chiverton, A. and Svensson, C.: From meteorological to hydrological drought using standardised indicators, Hydrology and Earth System Sciences, 20(6), 2483–2505, doi:https://doi.org/10.5194/hess-20-2483-2016, 2016.

Bloomfield, J. P., Marchant, B. P., and McKenzie, A. A.: Changes in groundwater drought associated with anthropogenic warming, 23, 1393–1408, https://doi.org/10.5194/hess-23-1393-2019, 2019.

Environment Agency: Impact of long droughts on water resources: https://www.gov.uk/government/publications/impacts-of-long-droughts-on-water-resources (last access: 21 June 2021), 2009.

Folland, C. K., Hannaford, J., Bloomfield, J. P., Kendon, M., Svensson, C., Marchant, B. P., Prior, J., and Wallace, E.: Multi-annual droughts in the English Lowlands: a

review of their characteristics and climate drivers in the winter half-year, 19, 2353–2375, https://doi.org/10.5194/hess-19-2353-2015, 2015.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A., and Mearns, L. O.: Towards process-informed bias correction of climate change simulations, 7, 764–773, https://doi.org/10.1038/nclimate3418, 2017.

Marsh, T., Cole, G. and Wilby, R.: Major droughts in England and Wales, 1800–2006, Weather, 62(4), 87–93, doi:10.1002/wea.67, 2007.

Shepherd, T. G.: Atmospheric circulation as a source of uncertainty in climate change projections, 7, 703–708, https://doi.org/10.1038/ngeo2253, 2014.

Smith, K. A., Barker, L. J., Tanguy, M., Parry, S., Harrigan, S., Legg, T. P., Prudhomme, C., and Hannaford, J.: A multi-objective ensemble approach to hydrological modelling in the UK: an application to historic drought reconstruction, 23, 3247–3268, https://doi.org/10.5194/hess-23-3247-2019, 2019.

Spraggs, G., Peaver, L., Jones, P., and Ede, P.: Re-construction of historic drought in the Anglian Region (UK) over the period 1798–2010 and the implications for water resources and drought management, Journal of Hydrology, 526, 231–252, https://doi.org/10.1016/j.jhydrol.2015.01.015, 2015.

Staudinger, M. and Seibert, J.: Predictability of low flow – An assessment with simulation experiments, Journal of Hydrology, 519, 1383–1393, https://doi.org/10.1016/j.jhydrol.2014.08.061, 2014.

Stoelzle, M., Staudinger, M., Stahl, K., and Weiler, M.: Stress testing as complement to climate scenarios: recharge scenarios to quantify streamflow drought sensitivity, in: Proceedings of the International Association of Hydrological Sciences, Hydrological processes and water security in a changing world - Hydrological Processes and Water Security in a Changing World, Beijing, China, 6–9 November 2018, 43–50, https://doi.org/10.5194/piahs-383-43-2020, 2020.

van der Wiel, K., Selten, F. M., Bintanja, R., Blackport, R. and Screen, J. A.: Ensemble climate-impact modelling: extreme impacts from moderate meteorological conditions, Environ. Res. Lett., 15(3), 034050, doi:10.1088/1748-9326/ab7668, 2020.

van Garderen, L., Feser, F., and Shepherd, T. G.: A methodology for attributing the role of climate change in extreme events: a global spectrally nudged storyline, Natural Hazards and Earth System Sciences., 21, 171–186, https://doi.org/10.5194/nhess-21-171-2021, 2021.

Watts, G., Christierson, B. von, Hannaford, J. and Lonsdale, K.: Testing the resilience of water supply systems to long droughts, Journal of Hydrology, 414–415, 255–267, https://doi.org/10.1016/j.jhydrol.2011.10.038, 2012.

Wilby, R. L., Prudhomme, C., Parry, S., and Muchan, K. G. L.: Persistence of Hydrometeorological Droughts in the United Kingdom: A Regional Analysis of Multi-Season Rainfall and River Flow Anomalies, J. of Extr. Even., 02, 1550006, https://doi.org/10.1142/S2345737615500062, 2015.