Response to all reviewers
Storylines of UK drought based on the 2010-12 event

Response to Editor

We thank Prof. Jim Freer for the comments and suggestions on how to improve the manuscript. We address each point raised by the Editor and reviewers below (in red). The revised manuscript with tracked changes is attached below the responses.

1) In response to your request for clarification on the publication of the story lines data then I recommend you read our publication policy on data used in papers where it is quite clear we expect this to be part and parcel of published papers now for improvements to transparency and scrutiny. I think you will find this is what most international journals are expecting. Please see https://www.hydrology-and-earth-system-sciences.net/policies/data_policy.html

RESPONSE: Thank you. We have made the input and output data publicly available via the zenodo repository (available at: https://doi.org/10.5281/zenodo.5180494)

2) I am not yet sold on your arguments for only using delta change as the bias correction technique. In papers that I have been associated with in Flooding we specifically didn't not use delta change because we were looking at more extremes and remained unconvinced that basic delta change was valuable to reflect fully those extreme behaviour biases. I would suspect that the biases in the extremes associated with drought would be so similarly impacted and thus some hard analyses needs to be implemented in your paper to show that delta change would not unduly influence the storyline results if one scrutinized the extremes a little more. As you know I wrote to you about this in the beginning that this needs to be challenged and it has been brought up in the review process. So I don't think you can justify this for extreme event behaviour by saying it's commonly used in the UK without some further hard evidence of it's impact (or not). I'm super happy to be proved that this is not important to your methods. We all recognize there is no perfect answer to bias corrections and the deficits of downscaling and RCM precipitation quantities but that doesn't mean that should be the core justification for not exploring these issues for your own best scientific scrutiny given the core aims of your paper.

RESPONSE: The aim of the climate change component of our analysis (which we must emphasize, is only one component) is to place the 2010-12 drought in a warmer world rather than to generalize over the hydrological impacts of climate change. We sacrifice generality by focusing solely on the observed event but what we gain from this is interpretability and physical realism (further discussed in the next paragraph). We quote Reviewer 1 in this respect, who agrees that our analysis "sheds light on physical catchment properties that play a key role in the propagation of multi-year drought event". The underlying principle here is Box's famous "All models are wrong, some are useful". The delta change method allows us to do this by perturbing the observed drought sequence directly. Ensuring future plausibility is inherently difficult. While more complex statistical bias correction techniques such as quantile mapping can used to assess general changes in the impacts of climate change on hydrological variables (although we note that there are many known limitations to more complex bias adjustments: e.g. Ehret et al. 2012; Maraun et al. 2017), it is much more challenging to use bias corrected climate model data to search for similar analogues to observed events. This is particularly the case for droughts (and perhaps different from flooding applications, which tend to be shorter term weather rather than longer term climatic events). This is because of concerns and uncertainty over the realism of climate model simulations for persistent circulation extremes, and with how atmospheric circulation will respond to climate change (Shepherd 2014). Previous studies have shown that climate models tend to underestimate drought persistence (particularly important for multi-year droughts like the 2010-12 drought) and where multi-year droughts are simulated, the driving mechanisms in the climate model can vary significantly between individual drought events (Ault et al. 2014; Moon et al. 2018) hence making it difficult to validate in relation to the observed 2010-12 event. If we have perfect models and large enough ensembles, storyline approaches won't be needed. Single model initial condition large ensembles may be used to search for analogue events (e.g. van der Wiel 2021) but this is subject to on-going work and is out of the scope of this study.

We therefore believe using the change factor method for the climate change component of the paper is justified here as it has the additional advantage of being easily interpretable and comparable with the other storylines analysed in this paper, which were also created by altering the observed drought sequence. As all the event-based storylines created in this study were based on the observed drought sequence, we believe this actually increases realism compared to searching for dissimilar events in bias-corrected climate model data from models that cannot reproduce the persistent circulation anomalies that lead to the 2010-12 drought. In practice, altering the observed drought sequence in this way is valuable for water resources planning as it allows for the exploration of droughts at high return periods for which there is no historical precedent and could complement approaches following existing Environment Agency guidelines. We have added our justification in the methods section to explain how we believe the change factor method is suitable for this study (lines 231-250 in the revised manuscript). We previously raised the point that the delta method has been used widely and consistently in the UK for further context to support our choice, rather than as the core justification.

3) I also agree with comments made about the calibration of the model metrics not being very well justified or seemingly related to the core issue that the models produce historical drought behaviour well. The reasons and justification for these equally weighted metrics (which seem ad hoc at this moment) must be improved and that may need some additional analyses. I'm sure you have your reasons and analysed this in more detail than the paper currently shows, so we need that more intelligent calibration approach better identified.

RESPONSE: The Editor and Reviewer 2 are concerned about how the four selected metrics were selected and whether the fact that they are equally weighted would affect the choice of the parameters to simulate droughts. The ability of the top 500 parameter sets (LHS500) to reproduce periods of historic drought in both timing and magnitude has already been demonstrated in Smith et al. (2019). As the original LHS500 was ranked based on model performance over a long baseline period, a differential split-sample test is conducted by re-ranking LHS500 based on performance during the driest years using four of the six metrics in Smith et al. (2019). As we're calculating the metrics for river flow during the driest years, we don't believe any of the four metrics can be considered more important than the others as high flows (NSE), timing of flows (logNSE), variability (MAPE) and overall water balance (PBIAS) during dry years are all equally important. Model performance for the top parameter set in the original LHS500 rank and the Dry rank are comparable but, in some catchments, the Dry rank show better performance during the driest years. The top ranked parameter set in the original LHS500 ranking remains unchanged in the Dry rank for 17 out of the 100 catchments. For the majority of catchments (54 out of 100), the top parameter set in the new Dry rank is within the top 10 of the original LHS500 rankings. For the remaining catchments, the top parameter set in the new Dry rank are all found in the top 100 of the original LHS500 rankings. We have added an additional figure (Supplementary Figure S4c) comparing the top parameter set in the Dry rank and its corresponding position in the original LHS500 ranking.

More generally, however, the details of the model parameter calibration process are not critical to the findings of this paper. We are seeking to apply a model which produces plausible hydrological simulations corresponding to the range in catchment and climate conditions across the UK. The key point is whether the model is informative and helps us evaluate the storyline concept – and we believe it does. As Reviewer 2 recommended, we have also provided two additional figures in the supplementary materials. Fig S4 shows high NSE and logNSE values for catchments across the UK using the top parameter set of the Dry rank. Fig S5 shows simulated river flows across 2010 and 2012 and clearly shows that the model is able to reproduce low river flows and drought conditions.

4) Finally the reviewers have asked for improvements to the justification of the recovery time metrics. This was another issue I need in my editorial review of your initial paper. I am still confused as to how a simple on any one day threshold metric has value to look at drought recovery which is a longer term process. Is there really no more intelligent way to approach this that deals with that longer term process? I'd like to see more critical discussion of that please.

RESPONSE: In our initial response to the Reviewer's comments on this, we have clarified that the catchment recovery time as calculated in the study is not indicative of the time taken for catchments to fully recover from drought to non-drought conditions. Instead, it is meant as an indicator for how long the influence of the precondition perturbations were felt for each catchment. This is consistent with similar indices proposed to investigate the impacts of changes in initial conditions and we have added those references in the revised text. To avoid further misunderstanding, catchment recovery time is renamed precondition persistence time which more accurately reflects what is shown in the results. The Editor is correct that other metrics would be needed if the aims are to consider full recovery from drought to non-drought conditions. As this was not the aim of the storylines of precondition severity, we point the reader to Parry et al. (2016) for calculating drought termination metrics to achieve this.

Response to Reviewer 1

General comments
The submitted manuscript addresses a very relevant topic for water risk management, (i.e. low likelihood/high impact events) and does so using storylines, a novel approach that allows the investigation of plausible but unrealized high impact events. The selected storylines are based on the 2010-2012 UK drought event and explore imposed changes to 1) Precondition severity, 2) Temporal drought sequence, and 3) Climate change. The implications of such changes are assessed by quantifying changes to streamflow maximum intensity, mean deficit, and duration. The results do not only facilitate the realization that it could have been worse/it possibly will be worse but also sheds light on physical catchment properties that play a key role in the propagation of a multi-year drought event. In general, the manuscript is well written and structured and the results are relevant to a broad community interested in novel approaches that tackle environmental risk management and future climate change impacts. I have few minor concerns that I share in what follows:

RESPONSE: We thank the reviewer for the positive feedback on our manuscript. We are grateful for the comments and suggestions on how our manuscript can be improved. We respond to each comment below (in red).

I understand plausibility to be a key property of the designed storylines. The first storyline proposes varying 3- and 6- months prior precipitation conditions to the 2010-2012 drought event independently of other climatic variables used in the model simulation. Such manipulations do not consider correlation structures in the data. I find that not completely justified and slightly weakening the plausibility assumption. For example, the potential presence of autocorrelation among successive monthly precipitation values or the correlation between precipitation and temperature are not considered. The authors can potentially mention these concerns in their discussion to further strengthen the plausibility argument.

RESPONSE: We agree that further information is needed to discuss the implications of the precondition storylines on the correlation between potential evapotranspiration (PET) and precipitation. We have added two new figures to address this comment (Figure 1 in the revised manuscript and Figure S2 in the supplementary materials). Figure S2 in the supplementary materials shows monthly precipitation and PET from 1965-2015. Apart from a slight negative correlation between precipitation and PET in spring and summer, there appears to be no clear correlation between the variables in the remaining months from 1965-2015 data. Figure 1 in the revised manuscript shows the equivalent values after precipitation 3- (i.e. OND 2009) and 6-months (i.e. JASOND 2009) before the 2010-12 drought was increased/reduced to match mean OND or JASOND precipitation at four return periods. The changes in precipitation prior to the drought does not appear to be outliers compared to the observed relationship between precipitation and PET from 1965-2015. We also emphasize that the creation of event storylines in other locations outside the UK should consider potential correlation between the different variables if a strong correlation is found.

Autocorrelation among successive monthly rainfall values is mostly not statistically significant (within the 95th confidence interval) apart from the short-term and decays rapidly after the first 1-2 months. For some stations, there are statistically significant but low autocorrelation values highlighting rainfall seasonality. Low monthly autocorrelation for rainfall is also seen in previous studies when considering the performance of stochastic weather generators (e.g. Kilsby et al. 2007; Serinaldi and Kilsby 2012; Chun et al. 2013). Given the low autocorrelation found in the observed data, the 3- and 6-months perturbations in the storyline of precondition severity do not violate existing autocorrelation structures and are valid and plausible. We have amended the text to reflect this.

I see that some consideration is given in the paragraph starting at Line 516, nevertheless, I find that rather short and in itself not fully convincing. If I understand correctly, the authors address plausibility for the precondition storylines by comparing the resultant 12-month precipitation deficits to outputs of high-end climate change scenarios. They argue that the preconditioning storylines are plausible as these are contained within the range of outputs from high-end climate change scenarios. Nevertheless, I expected that plausibility concerning these particular storylines should address whether such conditions are possible in the current climate.

RESPONSE: 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, rather than as a justification of their plausibility. We have amended and moved this text to Section 4.2 in our discussion of the value of the storyline approach to highlight these storylines as alternatives to existing projections. As discussed in the previous response, we have expanded our justification of the plausibility of the precondition storylines.

The authors state that they apply the delta approach in its standard form (line 189) where historical variability is retained. This formulation confuses me a bit as I am not sure what a non-standard form for the delta approach is.

RESPONSE: The standard form of the change factor approach, as applied in this study, retains historical variability with monthly change factors. There have been different modifications or variations to the delta approach proposed in the literature. They mostly consist of ways to calculate percentile- or quantile-based change factors for relative changes in wet and dry days and rainfall intensity (e.g. Anandhi et al. 2011; Willems and Vrac 2011; Ntegeka et al. 2014). Anandhi et al. (2011) also reviews and presents a classification of different variants of the change factor method. Although there are several modifications, the standard delta method as used in this paper remains the most widely used. We have clarified this by referencing these studies in the revised manuscript.

Can the authors expand on this in their discussion to address limitations associated with the method they chose and possibly elaborate on other potential methods that can be used to answer questions such as: How would that particular event look like in a warmer world? (e.g. Wehrli et al. 2020).

RESPONSE: We have expanded on the limitations relating to the use of the delta method. We have added reference to Wehrli et al. (2020) and van Garderen et al. (2021) as examples of spectral nudging. We have also expanded on other methods to construct event storylines under climate change. Alternative approaches to investigate extreme events in a warmer world would be to search for analogues or events similar to the 2010-12 drought (for example, analysis of weather types or circulation patterns – e.g. Cattiaux et al. 2010, or through the use of large ensemble climate model data – e.g. van der Wiel et al. 2021).

The validity of the change factor method was also raised by Reviewer 2 and the Editor. We have expanded on the reasons for choosing the delta method in the Methods section in the revised manuscript (see earlier response to Editor; lines 231-250 in revised manuscript).

It is clear to me why storylines are relevant as complementary information to already existing approaches that rely on GCM projections to quantify the hydrological impacts of climate change. I also do understand how these two approaches are very much different in scope. Nevertheless, the authors use the terms "scenario-driven approach" as a particular feature of GCM driven assessments in an attempt to contrast their approach and I find that slightly misleading. Storylines are still very much scenarios to my understanding, event-based in that case, and with a focus on plausibility rather than probability. I don't see why they wouldn't qualify as scenario-driven. The author themselves state that (i.e. line 143): "storylines follow similar methodologies employed in previous studies to create scenarios". I, therefore, recommend revisiting specifically this phrasing to reduce confusion and facilitate the understanding of what is meant by storylines.

RESPONSE: It is true that in lay usage, "storyline" and "scenario" are somewhat interchangeable terms. However, in climate change science, the word "scenario" is firmly established as corresponding to a specified socio-economic pathway associated with a particular climate forcing, and to use it in any other sense would cause unnecessary confusion. Shepherd et al. (2018) discussed this issue as part of their rationale for the use of the term "storyline" in the context of physical climate, and their definition of physical climate storyline has now been adopted by the IPCC Glossary. Here we specifically use the concept of a physical climate "event storyline", as illustrated in Box 10.2, Figure 1 of the AR6 WG1 report (available from https://www.ipcc.ch/report/ar6/wg1/#FullReport). We have now clarified our language and made reference to the IPCC usage of these words. We would argue that from this perspective, it is not misleading to distinguish between scenario-driven and storyline approaches. Physical climate storylines can be created independently of scenario-driven GCM ensemble projections to represent situations or conditions that could lead to significant impacts, and can complement results from scenario-driven GCM ensemble projections. To avoid misunderstanding, we have checked all our uses of "scenario" and "storyline" and provided clarifications in each case.

We agree that line 143 may be confusing to readers. We consider the similar methodologies in the previous studies as cited to also be storylines that could be used to complement climate change projections and stress test hydrological systems. We have removed any mention of "scenario" in this case.

Another point related to terminology: Can the authors explain their use of the term "counterfactual" when discussing future impacts of climate change. As the climate change storyline refers to a hypothetical event in the future, I find it a bit unclear why that would qualify as a counterfactual.

RESPONSE: We agree with the reviewer that the use of the term "counterfactual" when discussing the storylines of climate change is potentially confusing. All use of the term when discussing the UKCP18 climate projections have been removed.

Technical comments
I am slightly confused by this sentence: Line 373, "The drought is estimated to worsen for the "Dry year before" storyline for all clusters except for mean drought deficit for Cluster 4 for SSI-6". I believe something along the lines of: "The drought defined by SSI-6 is estimated to worsen for the "Dry year before" storyline for all clusters except for mean drought deficit for Cluster 4 " is a bit more clear.
RESPONSE: We have rephrased as suggested.

Line 378: is -> are
RESPONSE: Done.

Line 381-382: I believe something in the punctuation of the phrase is incorrect. Please check that.
RESPONSE: We have modified this.

Response to Reviewer 2

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 the reviewer agrees that our results have benefit to those working on droughts. We respond to each comment given in the text below (in red).

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 raised by Reviewer 1 who was concerned about the plausibility of altering observed precipitation independently of temperature in the storylines of precondition severity. In response to that comment, we have added Figure 1 in the revised manuscript and Figure S1 in the supplementary materials to illustrate that our perturbation of precipitation 3- and 6-months prior to the observed drought does not violate any correlation structures between precipitation and potential evapotranspiration (PET). We have also emphasized in the revised manuscript in the methods section that the creation of event-based storylines in other locations should consider potential correlation between different variables if a strong corelation is found. The discussion of the H++ high end climate change scenarios as intended as a point of comparison rather than as a justification of our storylines of precondition severity. We have moved this to the discussion.

With regard to precipitation perturbations over a longer time period with the "three dry winter" storylines, we consider this to be plausible as a large number of previous studies have investigated the occurrence and likelihood of sequences of dry winters and their implications for UK water resources. We have added additional background on the "three dry winters" situation in the methods section. In addition to this, we also added reference to a previous Environment Agency study which looked at a similar storyline with a third dry winter following the 2004-06 drought. There was widespread concern and expectation in early 2012 that dry conditions would persist based on the prevailing atmospheric circulation before abrupt record-breaking rainfall terminated the drought. An additional figure has been added to the supplementary material (Figure S2) which shows more clearly the differences in atmospheric circulation between the repeated year and the year replaced. Average Z500 anomalies for 2010 (repeated year) show were characterized by high pressure over parts of the UK throughout summer and winter. In contrast, 2012 (replaced year) was characterized by low pressure over the UK and high rainfall totals throughout the year. 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. We have added this justification in the methods section. Greater consideration of atmospheric conditions will be the subject of future work.

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: This relates to the limitations of the delta change method also raised by the Editor and Reviewer 1. The reviewer is concerned about the plausibility of applying the delta change method. However, we would argue that it actually increases realism with the additional advantage of increasing familiarity to stakeholders by retaining the observed drought sequence. A key characteristic of event storylines like the ones created in this study is that they are familiar and link directly to experiences of stakeholders (as highlighted in Box 10.2 in the IPCC AR6 report). Please also see earlier response to Editor for an expanded justification on this.

We think that characterizing Fowler et al.'s paragraph on delta-change methods as identifying "a whole host of problems" is overstating what is said in that paper. A number of caveats are mentioned (not specifically about droughts), but they are only caveats, and every method has its caveats. These caveats are also all discussed in more detail now in the revised manuscript. The assumption that GCMs simulate relative changes better than absolute values is the cornerstone of the analysis presented in IPCC reports, so it is not a radical assumption. We have expanded on our discussion in the methods section to explain why we believe the change factor method is suitable for this study. We have also expanded on this and other alternative approaches in the limitations section in the Discussion.

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: The dataset used was CEH-GEAR. We have amended the typo in the data availability section. We agree that our estimates of precipitation return periods could be most robust with the full CEH-GEAR dataset from 1900. 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. As suggested by the reviewer's comments, we have repeated the simulations for this section based on revised return periods calculated using the full dataset. This resulted in updates to Figures 5 and 6 in the original manuscript (Figures 6 and 7 in the revised manuscript) and corresponding changes to the discussion of the results.

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: In response to comments from the reviewers and the Editor, we have clarified that the catchment recovery time as calculated in the study is not indicative of the time taken for catchments to recover from drought to non-drought conditions. Instead, it is meant as an indicator of how long the influence of the precondition perturbations were felt for each catchment. This is consistent with similar indices proposed to investigate the impacts of changes in initial conditions. To avoid further misunderstanding, catchment recovery time is renamed precondition persistence time, which more accurately reflects what is shown in the results. The metric is calculated from the start of the perturbation until the influence of the perturbation is no longer detected (<1% compared with baseline). We have clarified this in the revised manuscript with references to Stoelze et al. (2020) and Staudinger and Seibert (2014) which used similar metrics to understand the influence of changes to initial conditions. Although our aim was not to investigate the time taken for the catchment to fully recover from drought to non-drought conditions, we have added reference to Parry et al. (2016) which outlines a framework to calculate drought termination metrics.

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: This point was also raised by the Editor. Please see prior response to Editor for expanded justification. As recommended, we have added supplementary Figures S4 and S5 NSE/logNSE values for each catchment for the top parameter set in the Dry rank and simulated river flow across the 2010-12 drought. Simulated river flow matches well with the observations in both timing and magnitude of low flows across 2010 and 2012. The NSE and logNSE values are high for the top-ranked parameter set in the Dry rank (>0.5) across most of the catchments. logNSE values are generally higher than NSE values. As the Dry rank is based on ranking parameter sets based on the driest years, NSE values (which as the reviewer noted is focused on high flows) are lower as a result. Catchments with relatively poorer performance are highlighted in the original text as fast-responding catchments with a "flashy" river regime in northern Scotland as identified in Smith et al. (2019).

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: We have made the input and output data available via the zenodo repository (available at: https://doi.org/10.5281/zenodo.5180494).

Technical comments
L14. 'highly conditioned by its meteorological preconditions'. Not entirely sure what you mean here, can you clarify?

RESPONSE: We have rephrased. 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.

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 have cited this in the revised manuscript as suggested.

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 have added this as Figure S1 in the supplementary materials.

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

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 have added a Figure S11 in the supplementary materials on projected change in mean annual precipitation across the different clusters as suggested. This broadly reflect differences between the clusters in changes in drought characteristics in the storylines of climate change section (i.e. Fig.12 in the revised manuscript).

Figure 12 is quite blurry – can you increase the resolution?
RESPONSE: Done.

Combined references

Anandhi, A., Frei, A., Pierson, D. C., Schneiderman, E. M., Zion, M. S., Lounsbury, D. and Matonse, A. H.: Examination of change factor methodologies for climate change impact assessment, Water Resources Research, 47(3), doi:10.1029/2010WR009104, 2011.

Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T., and Meko, D. M.: Assessing the Risk of Persistent Drought Using Climate Model Simulations and Paleoclimate Data, 27, 7529–7549, https://doi.org/10.1175/JCLI-D-12-00282.1, 2014.

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V. and Codron, F.: Winter 2010 in Europe: A cold extreme in a warming climate, Geophysical Research Letters, 37(20), doi:10.1029/2010GL044613, 2010.

Ehret, U. et al. (2012). Should we apply bias correction to global and regional climate model data? *Hydrol. Earth Syst. Sci.*, 16: 3391–3404. doi:10.5194/hess-16-3391-2012.

Maraun, D. et al. (2017) Towards process-informed bias correction of climate change simulations. *Nature Climate Change* 7, 764-773

Ntegeka, V., Baguis, P., Roulin, E., and Willems, P.: Developing tailored climate change scenarios for hydrological impact assessments, 508, 307–321, https://doi.org/10.1016/j.jhydrol.2013.11.001, 2014.

Parry, S., Wilby, R. L., Prudhomme, C., and Wood, P. J.: A systematic assessment of drought termination in the United Kingdom, 20, 4265–4281, https://doi.org/10.5194/hess-20-4265-2016, 2016.

Shepherd, T. G.: Atmospheric circulation as a source of uncertainty in climate change projections, Nature Geoscience, 7(10), 703–708, doi:10.1038/ngeo2253, 2014.

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.

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, Hydrology and Earth System Sciences, 23(8), 3247–3268, doi:https://doi.org/10.5194/hess-23-3247-2019, 2019.

Stoelzle, M., Stahl, K., Morhard, A. and Weiler, M.: Streamflow sensitivity to drought scenarios in catchments with different geology, Geophysical Research Letters, 41(17), 6174–6183, doi:10.1002/2014GL061344, 2014.

Moon, H., Gudmundsson, L., and Seneviratne, S. I.: Drought Persistence Errors in Global Climate Models, 123, 3483–3496, https://doi.org/10.1002/2017JD027577, 2018.

van der Wiel, K., Lenderink, G., and de Vries, H.: Physical storylines of future European drought events like 2018 based on ensemble climate modelling, 33, 100350, https://doi.org/10.1016/j.wace.2021.100350, 2021.

Willems, P. and Vrac, M.: Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change, 402, 193–205, https://doi.org/10.1016/j.jhydrol.2011.02.030, 2011.

Relevant changes to the manuscript:

- Further justification of the storyline approach in relation to existing approaches with reference to “physical climate event storylines” recently adopted in the IPCC AR6 Glossary
- Simulations for the storylines of precondition severity were re-ran based on estimations of rainfall return periods using the entire CEH-GEAR dataset (1900-2015) including changes to Figures 6 and 7 in the revised manuscript
- Further consideration and justification for the plausibility of the storylines of precondition severity (including Figure 1 in revised manuscript) and the “three dry winter” storylines
- Expanded discussion and justification of the delta change method used to construct storylines of climate change
- Expanded discussion on alternative approaches to construct event storylines
- Minor changes to other sections and figures