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

Reviewer 1

Thanks for the authors for expanding on plausibility assumptions and clarifying key terms in their revisions. I think the paper can now be accepted subject to a few technical corrections.

Line 39: Missing a full stop after deficits.
Line 99: Reference missing for Doblas-Reyes et al. 2021 in the reference list, please double check references.
Line 374: I believe you meant to say "alone" rather than "along" ?

Response: We thank the reviewer for reviewing our manuscript again and for the positive comments and suggestions. The corrections have all been made.

Reviewer 2

Many thanks to the author for considering my comments and for the effort in the responses. I really appreciate the additional analyses and figures in the papers and supplementary information. I have two minor comments for the authors that relate to my original review and need further discussion in the paper:

Response: We thank the reviewer for reviewing our manuscript again and for the positive comments and suggestions. We respond to the two minor comments below (in red).

1. Model performance metrics
Thanks for the additional detail in the supplementary material and text in the methods. However, the justification for the choice of metrics still needs improving. In the paper, it is stated that 'The metrics selected are unweighted as high flows (NSE), timing of flows (logNSE), flow variability (MAPE) and overall water balance (PBIAS) should be considered equally important for river flows during the driest years.' However, there is still no discussion (as far as I can tell) of why they should be considered important and why they should be considered equally important. Why are high flows important to consider when analysing future storylines of drought?

Response: We believe that high flows, timing of low flows, flow variability and overall water balance for river flows during dry years are equally important and the parameter set used to evaluate the storylines should represent not just low river flows but the full range of flow response during dry years (including the possibility of wetter interludes during dry years). This is especially relevant for the 2010-12 drought as there were wetter periods during the drought period and the drought itself was abruptly terminated by unexpected wet conditions with high river flows (as discussed in section 3.1). The full range of flow response is also important to understand the antecedent conditions of drought events. This is important given that antecedent conditions to the 2010-12 drought is varied in the storylines of precondition severity. Using the top parameter set in the new Dry rank, Supplementary Fig.S4 shows clearly that the parameter sets are capable of simulating both the high and low flows during the 2010-12 drought period across the example catchments spread across the UK. The revised justification in the text now reads as follow (lines 239-248):

"As the LHS500 ranking was based on model performance over a long baseline period, we conduct a differential split-sample experiment to re-rank LHS500. For each catchment, the 10 driest years were selected based on mean annual precipitation (1965-2015). Model performance for each of the driest years was calculated using daily observed and simulated river flow for four of the metrics in Smith et al. (2019): NSE, logNSE, MAPE and PBIAS. The metrics selected are unweighted as high flows (NSE), timing of low flows (logNSE), flow variability (MAPE) and overall water balance (PBIAS) should be considered equally important for river flows during the driest years. This is to ensure that the full range of flow response during dry years is considered, including the potential for wetter interludes during dry years (such as that seen during the 2010-12 drought). It is also important to consider high flows during the antecedent conditions of drought events. In the context of the storyline approach, this is especially important given that antecedent conditions are varied in the storylines of precondition severity."

2. Delta change approach

I still feel that there is a lack of critical discussion related to the delta change approach in the limitations section. While I appreciate the additional text on alternative methods, there is only one sentence (as far as I can tell – apologies if I have missed this) on the limitations of a delta change approach: 'By not considering changes in the likelihood of such an event, it could underor over-estimate drought impacts from climate change.' There needs to be more critical discussion of this approach and its ability to capture how droughts might unfold in the future.

Response: The fact that the delta method retains the observed meteorological time series is often cited as the main limitation of the approach (as reflected in the limitations section and also stated in the section 2.2.3). We have restructured and expanded the limitations section to reflect on the implications of this in understanding how future droughts may unfold (lines 584-600). The revised text now reads as follow:

"Storylines in this study are based on resampling and perturbing the meteorological time series of the 2010-12 drought. The main limitation of the delta change method used to place the 2010-12 drought in a warmer world is that it retains the observed temporal variability of the observed drought. This approach is advantageous given the specific focus on the 2010-12 drought, and it avoids having to deal with potential climate mode biases in the representation of the persistent circulation anomalies that lead to drought. However, the temporal variability and sequencing of weather events may change under climate change and future changes in variability differs between GCM/RCMs. The delta method applied to the 2010-12 drought therefore means that we do not consider ways in which a drought of different nature could unfold in the future and reach similar or worse impacts as the 2010-12 drought. Future droughts where minimum river flow occurs in different seasons (e.g. summer vs winter) or driven by compound conditions (e.g multivariate heatwave drought or preconditioned drought from combination of seasonal precipitation deficits) (Zscheischler et al. 2020) thus cannot be assessed using the delta change method alone. By not considering changes in the likelihood of such an event, the delta change method could therefore under- or over-estimate drought impacts from climate change. For example, Wilby and Harris (2006) has previously shown that the direct use of statistically downscaled climate model output can lead to a smaller reduction in low flows and a wider range of projected change compared to the delta change method although overall uncertainty is dominated by differences between GCMs. However, given GCM-related uncertainty and in the absence of confident information on changes in the likelihood of multi-year circulation anomalies, using the delta change method to place a

singular event under future warming is a logically sensible approach to take, grounded in Bayesian reasoning (Shepherd 2021)."

Reviewer 3

The paper (from my point of view) is now way too long (more than 700lines and 13 Fig.). I recommend that a part of the analyses is removed to streamline the paper (e.g. focus on one instead of two storylines?). However, the topic and the idea of the paper is really important for the community and I lean towards a 2nd round of revisions and not towards a rejection. A lot of confusing terms have been removed in the first round of reviews, this is nice to see and there is some progress in the paper. Unfortunately, the paper is now somehow inflated and should be narrowed down again to one (or two) key messages.

Response: We thank the reviewer for reviewing our manuscript. We are grateful the reviewer agrees that the paper is useful for the community and that there is progress after revisions made after the first round of reviews. We have reviewed and have shortened the overall length of the paper. We have removed one figure from the main text (Fig 10, now Fig.S10) and the length of the paper is now close to the length of the original submitted manuscript and contains the same number of figures.

We do not think that any of the storylines should be removed as all the storylines considered explore different aspects of the 2010-12 drought. The storylines of seasonal contribution show the importance of autumn conditions in the 2010-12 drought and should be considered in assessment of future multi-year droughts. The storylines of precondition severity show the sensitivity of the drought to plausible drier preconditions that are linked to both the spatial dynamics of the drought and also catchment characteristics (e.g. hydrogeology). The "Dry year before" and "Dry year after" storylines show distinctive drought responses between the fast- and slow-responding catchments and represent plausible alternatives of the observed drought. The storylines of climate change using the UKCP18 projections place the drought at different global warming levels and showed worsening of conditions with temperature rise. We showed that all storylines are capable of matching or exceeding conditions observed in past droughts and are thus important elements to consider in the unfolding of future high-impact drought events.

Fig.1: This figure is added to the manuscript after first round of reviews. The explanation given (L190-204) is helpful to understand how the severity of preconditions is assessed. However, I am not totally convinced that PET (as response variable on the y-axis) is the best choice (variability is rather small), perhaps AET would be better (i.e. possibility to identify situations where the systems are water-limited).

Response: The figure was added after reviewers 1 and 2 both raised concerns over whether the perturbations made for the storylines of precondition severity would violate any correlation structures between PET and temperature (hence the choice of PET on the y-axis). We believe this to be appropriate given the fact that PET and precipitation are both inputs to the GR4J hydrological model and this comparison directly shows that the perturbations made to the input data to the hydrological model are physically plausible.

Fig. 4: Please use another encoding than colour for the points on the map (e.g. varying point shapes).

Response: Done.

Fig.7: Same color gradient for latitude and persistence time is confusing. Perstence time could be binned and log-scaled. Please give a metric for the panels in Fig. 7 that quantifies the strength of the relationship between x and y.

Response: Done.

Regarding the conclusions: Please state clearly here what is found not what is done (e.g. sentences L693-699 are too vague). Paragraph L 711-715 is not really helpful, what is the value/the message here? Please consider that storyline approaches are relatively new for the community and the benefits compared to climate change scenario modelling approaches should be a) clearly mentioned and b) justified by some values.

Response: Lines 711-715 places this study in relation to practical water resources planning and the potential to bridge probabilistic (return period) estimates with the storyline approach to understand extreme droughts with no historical precedent. We agree it might be better for this to be placed elsewhere and a shortened version of this has been included in the discussion (lines 564-567).

We have rewritten the conclusions to show 1) the extent to which the 2010-12 drought could have been worse given the various storylines and 2) that the storyline approach adds value and complements traditional climate change impact assessments (lines 621-644). To further clarify the added value of the storyline approach compared to existing approaches, we have also added reference to a recently accepted review paper (Chan et al. in press) (lines 56-61 and lines 561-564). The review paper categorizes studies on the hydrological impacts of climate change in the UK by their methodological approaches and outlines outstanding research gaps that can be tackled by emerging approaches such as the storyline approach.

References

Chan, W.C.H., Shepherd T.G., Facer-Childs, K., Darch, G. and Arnell, N.: Tracking the methodological development of climate change projections for UK river flows, Progress in Physical Geography: Earth and Environment. in press.

Shepherd, T. G.: Bringing physical reasoning into statistical practice in climate-change science, Climatic Change, 169, 2, https://doi.org/10.1007/s10584-021-03226-6, 2021.

Wilby, R. L. and Harris, I.: A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, Water Resources Research, 42(2), https://doi.org/10.1029/2005WR004065, 2006.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., Maraun, D., Ramos, A. M., Ridder, N. N., Thiery, W., and Vignotto, E.: A typology of compound weather and climate events, Nat Rev Earth Environ, 1, 333–347, https://doi.org/10.1038/s43017-020-0060-z, 2020.