Added value of seasonal hindcasts to create UK hydrological drought storylines
Response to reviewers

Editor's comments

Thank you for your effort in revising the manuscript. Both reviewers are happy with the revisions. That said, one of the reviewers recommended "accepted subject to minor revisions". I read their comment, and I'd like to ask the authors to revise their manuscript one more time. This reviewer is a senior scientist and practitioner working closely with governments and decision makers. It'd a missed opportunity not to include their comments. I think the suggested minor revision put your paper in a practice-informed context that benefits you as authors as well as future readers. I'll handle the next round myself to expedite the review. Thanks for your patience and well done with the process all the way to here. Just the last notch.

Response: We thank the editor for handling the manuscript and recommending minor revisions. We appreciate the opportunity to amend the paper to enhance relevance to practitioners. Please find below our response to reviewer 2 (in red) and additions made to the revised manuscript to address the reviewer's two remaining concerns.

Reviewer #2

I appreciate the considerable efforts the authors have made to clarify that their methods are not designed as forecasts nor intended to be used as forecasts. This removes much of my original objections to the paper (with a some specific exceptions, detailed below). The strengths of the paper remain: it is very well written, the methods are clearly described, the analyses are appropriate and figures are well-presented.

Response: We would like to thank Reviewer #2 for reviewing the revised version of the manuscript. Following the valuable suggestions from the reviewer, we are glad that many of the original objections the reviewer had have been satisfactorily addressed after the first round of revisions.

I remain skeptical of the utility of this work for informing decisions. What the authors have effectively done is expand a climatological distribution by considering an ensemble of Seasonal forecasts as 'plausible' realisations. This may or may not be a a reasonable thing to do: we know, for example, that seasonal forecasts, including their dynamical representations of the world, become increasingly less accurate and less realistic at longer lead times (as they become more distant from initial conditions). One risk of the methods presented in this paper is that they artificially inflate the variance of a climatological distribution, leading to unrealistically catastrophic events (either wet or dry) in their storylines at the tails of the distributions. This can lead to overly conservative (and thus suboptimal) decision

making. I think the authors should acknowledge this possibility somewhere in their paper (can be as brief as a single sentence). As I noted in the previous review, forecasts are most useful when they are as sharp as possible without being over-confident: this narrows the range of possibilities for decision makers (usually explicitly in comparison to climatology). Widening this range of possibilities - perhaps artificially - may well make decision making less optimal. Having said all this, I accept that these are somewhat philosophical objections, and others can judge the work once it is published. As the methods are very clearly described, others can readily decide whether this approach is useful.

Response: If our understanding is correct, the reviewer seems to be addressing two distinct points here. The first is the fact that storylines are not sharp forecasts. Since it is current operational practice in water resources management to consider worst-case scenarios (which is effectively a storyline approach as we have further discussed after the first round of revisions) for the purpose of stress-testing, we believe that our approach is valuable for decision-making in that context. Sharp forecasts are useful to decision-makers only if they are also accurate. Indeed, it is precisely the *lack* of forecast accuracy in current systems that motivates the pooling of hindcasts to explore unobserved parts of the space of drought possibilities in our study. Thus, to address the point, all we can do is reiterate that storylines are not forecasts, and explain that they are instead useful from a disaster risk reduction perspective (rather than e.g. for a cost-benefit optimization exercise as one might undergo based on a probabilistic forecast). The results were presented to the Drought Management Team at Anglian Water (the internal team convened during the 2022 drought), which were able to appropriately consider these results in decision-making. We have moved a sentence from the existing text and added further explanation as follows:

"It is important to emphasize that the storylines developed here were not meant to be forecasts of winter 2022/23 but instead represented hypotheses of possible river flow and drought responses to explore plausible worst cases. The approach is advantageous from a disaster risk reduction perspective as it increases risk awareness and enables water resources planners to consider a much wider diversity of plausible river flow trajectories. The storylines form one possible source of evidence during drought planning by providing a signal of potentially wet or dry outcomes to plan for." (L387-392 in the revised manuscript).

The second point raised by the reviewer is whether the storylines are physically realistic since at a long lead time, model solutions may drift from the real world into a physically unrealistic model world. We agree that if the hindcasts contained extreme storylines that were outside those that could be produced in the real world, suboptimal decision-making could result. Thus, establishing the physical plausibility of these storylines is crucial. Yet doing so is inherently challenging since by definition, extreme storylines will lie in the tails of the distributions and will not be well sampled (if at all) in the observations. That is why we conducted model evaluation for our stated purpose, for precipitation (Figures 1 and 2), the circulation drivers (Figure 3), and the atmospheric patterns themselves (Figure 5). Our approach was also compared with the existing ESP method (Figure 7). Further, we have made the following addition to

the text in the methods section to address the reviewer's concern about the potential of exaggerating the likelihood of unrealistically low rainfall:

"The SEAS5 hindcasts also do not seem to be exaggerating the likelihood of low rainfall as 4 out of the 38 observed winters fall within the lowest 10% of the standardised modelled rainfall distribution. There is a 75% chance of this occurring according to the binomial probability formula." (L105-108 in the revised manuscript)

We have also added the following text in the discussion:

"In such an approach it is crucial to establish that the storylines are physically plausible, otherwise decision-making could be suboptimal. Yet doing so is inherently challenging since by definition, extreme storylines will lie in the tails of the distributions and will not be well sampled (if at all) in the observations. That is why we conducted model evaluation for our stated purpose, for precipitation (Figures 1 and 2), the circulation drivers (Figure 3), and the atmospheric patterns themselves (Figure 5). The reliability of the storyline approach ultimately rests on the physical plausibility of the storylines that are produced. However, this is the case for any exploration of extreme outcomes, and is not particular to the storyline approach." (L401-406 in the revised manuscript)

One section that I did still have a minor objection to is the following (L345-354):

"Although this study did not consider the likelihood of a particular storyline for winter 2022/23, further subsets to the hindcast winters can be made to provide weights for particular storylines that are considered more likely than others over time (e.g. based on prevailing atmospheric circulation patterns). Given the large sample size of the hindcast winters, future work could also condition storylines based on their preconditions. For example, for the 2022 drought, storylines can be created by selecting only winters in the hindcasts with a wetter than average preceding November (as was observed in November 2022). This approach also takes advantage of forecasts of winter circulation characteristics (or weather regimes) which may be more reliable than forecasts of winter precipitation; these circulation forecasts can help inform plausible weightings assigned to particular storylines (Richardson et al. 2020). When employed during an ongoing event, this approach may also shed light on the conditions required for drought termination, for example by calculating drought termination metrics in Parry et al. (2016) for each storyline."

I realise this section is speculating on future improvements, but to me it too strongly retains the sense that the methods presented here are useful in forecasting. I suggest removing this section: as already noted, in my view the methods presented here are not useful for forecasting, and I think the method is unlikely to be useful in prediction, even with the suggestions of future work. This is because the method relies upon ensembles from seasonal climate prediction models, and these models are already more sophisticated alternatives to the inevitable prediction selection etc. required to condition climatological distributions on initial conditions.

Response: Although we do see potential in hybrid storyline/forecasting methods, in order to avoid possible confusion we have removed all references to forecasts in this

text and made clear that each of the possible extensions we suggest can be considered as a hypothetical counterfactual, rather than a forecast. The text (now separated off into a separate paragraph) now reads:

“Although this study did not consider the likelihood of a particular storyline for winter 2022/23, further subsets to the storylines can be made over time or retrospectively. For example, storylines of winter 2022/23 could have been created by selecting only winters in the hindcasts with a wetter than average preceding November (as was observed in November 2022). When employed during an ongoing event, this approach may also shed light on the conditions required for drought termination, for example by calculating drought termination metrics in Parry et al. (2016) for each storyline. The same approach can be used after an event to explore downward counterfactuals and the hydrological impacts should the event have turned out worse. For example, subsets of the storylines can be made to explore consequences should winter 2022/23 have turned out even drier than observed or if the preceding November had been drier than average.” (L361-369 in revised manuscript)