We would like to thank Referee #1 for the positive feedback and helpful suggestions on this manuscript. Below we give point-by-point responses to the comments (bold and italic).

1) The paper describes a study where a large-scale, high-resolution MODFLOW-groundwater model of Germany has been used to assess a range of potential changes to groundwater and baseflow drought hazard based on three change scenarios. The scenarios are: i.) a changed recharge regime with wetter winters and drier summers (SSHIFT), ii.) changes to antecedents conditions associated with three major historic episodes of drought in Germany (SEVENT), and iii.) recovery from drought (SRECOV). These scenarios were co-designed in part with the Climate and Water Initiative of southern Germany's federal states (KLIWA) (L67-86) with the aim of stress testing the sensitivity to drought of groundwater and baseflow. Although, the geographical focus of the study is Germany, the paper addresses questions relevant to a wide readership and is clearly in the scope of HESS.

The description of the model setup (Section 2) is adequate given that more details can be found in the paper by Hellwig et al. (2020) who developed the model. However, a critical assessment by the authors of the model's suitability, including method of calibration and appropriateness of it's underlying assumptions, for the current application would be helpful. The description of the scenario design and modelling approach (Section 3) is generally clear and well-reasoned. However, the scenarios appear somewhat arbitrary. In particular, the formulation of the SRECOV scenario is less convincing than the other two scenarios. To assess the maximum duration for groundwater recovery from severe drought, the lowest simulated groundwater heads are taken as an initial condition and groundwater heads are simulated using long-term average monthly recharge as input until an arbitrary recovery has been achieved. Although adequately described, the motivation and justification for the details of this scenario are not given. The results are presented well, both graphically and in their description in Section 4.

The national-scale groundwater model is certainly limited in its local validity. We reflect on the limitations regarding several aspects in the Discussion Section. However, we agree that another reflection on the suitability of the model for the scenarios applied will be useful and will add this to the model and/or scenario description together with further details on model calibration etc.

We agree that the formulation of SRECOV might appear arbitrary. We selected the scenario to address the stakeholders request for a better understanding of the drought termination period. To account for the local differing conditions, we adopted the idea of a 'composite map' and did not select one specific drought year but rather the lowest heads of the simulated period. We agree that the recovery threshold is arbitrary, however, additional analyses with other thresholds (40 and 50-percentile groundwater head) resulted in the same patterns. As we find that the resulting T_{rec} are strongly related to T_{max}, which is an aquifer characteristic independent of meteorology/recharge or initial conditions, it is reasonable that our choices regarding initial conditions and recharge are not substantially relevant for the results. We will discuss these points in the revised manuscript and rephrase the motivation to make clear that we use this scenario to learn about the general time scale of recovery from a severe drought and exact values of T_{rec} depend on initial conditions, assumed recharge and recovery threshold.

2) The Discussion provides a number of interesting insights into the results. For example, the authors make the observation at L287-290 that: "the different responses of baseflow and groundwater are important to consider for an effective water management in a changing climate. For example, in a climate with higher annual recharge sums but more frequent summer droughts groundwater

droughts might become less severe while the baseflow drought hazard becomes more severe with potential impacts on economy and ecology". Given that the scenarios that led to this observation were shaped by stakeholders, it would be interesting to know if and how stakeholders might use such information. More generally, given the nature of the set-up of the paper (e.g. L67-74) it would be interesting to hear the author's views on any specific implications of the results of their study for drought planning and management. These could be described, however briefly, in the Discussion.

Both baseflow and groundwater deficits can have negative impacts, but the affected sectors and planning tools can be quite different. In Germany, public water supply relies to a large proportion on groundwater resources (economic/social impact). SSHIFT will help to make water supply resilient, e.g. by diversifying sources with different seasonal sensitivity; SEVENT may be used to test water tower storages at extreme event scale together with SRECOV to plan durations of measures that may be necessary. Additionally, groundwater dependent ecosystems can be vulnerable to groundwater droughts (ecological impact). Baseflow droughts also affect surface water quality and water ecology (ecological impact) as well as navigation, energy production and tourism/recreation (economic/social impacts). Therefore, different stakeholders will face different challenges and use the scenarios differently to design adaptation or to plan mitigation measures for the emergency. We will name some examples in the revised version. Also, the combination of the two variables studied may be useful. In case of increased baseflow drought hazard but less change in groundwater drought hazard one option might be to switch or add – as far as practicable - water use from surface water to groundwater. For agricultural irrigation this process has already taken place in some parts of the study region, e.g the Upper Rhine valley.
We will discuss these points in the revised manuscript.

3) Specific comments: L229 I think that the authors meant "relative" not "relevant"?

Agree

4) Section 4.1. The authors make a number of observations relating to the groundwater and baseflow changes being more pronounced under average conditions than for drought, and this is also highlighted in the Summary at L326. A brief interpretation and discussion of the implications of these observations would be helpful.

Agree. In particular it is interesting to see, that an increased variability does not necessarily lead to more pronounced extremes. We will add a brief interpretation to the Discussion Section.

5) Section 4.3 and Figure 9. The main feature of the analysis of recovery time appears to be the essentially bi-modal nature of Trec, this being most evident in the Trec vs. Tmax plot in Fig. 9. It would be interesting to hear what the authors think might be contributing to this result. Does it reflect intrinsic characteristics of the modelled system, is it an artefact of the model structure or calibration, or is it some combination of both? Perhaps such a discussion could be added to Section 4.3?

Yes, the bi-modal nature of T_{rec} is strongly related to T_{max}. The large differences in T_{max} with some regions responding within few months and others over several years are a characteristic we can also find in observations (e.g. see Figure 7 in Hellwig et al., 2020), so we don't think it is just an artefact of the modelling. Anyway, the model does not capture Tmax equally well everywhere. For example, in Hellwig et al. (2020) it was noted that T_{max} was "overestimated by the model in the

porous aquifers in the lowlands and underestimated in higher elevations". Certainly, this will also influence the pattern of Trec. We will reflect on this in the Discussion Section.

6) L348-350. The first and only mention of the application of this approach to is in the Conclusions. This seems strange. It may be appropriate to include these observations in the Discussion, but not in the conclusions?

Agree