Review of Hellwig et al (2020), Stress-testing groundwater and baseflow drought responses to synthetic climate change-informed recharge scenarios

This paper tackles an important topic of how groundwater and baseflow will respond to changes in recharge. To test this, the study uses MODFLOW to explore how groundwater and baseflow change in response to three different recharge scenarios across Germany. The recharge scenarios are informed from stakeholder interactions and the combination of the scenarios targets different characteristics of groundwater and baseflow drought responses. The study concludes that a shift in rainfall to wetter winters and drier summers will not cause decreases in groundwater resources in general, but water managers need to consider the potential for more severe groundwater droughts following prolonged dry spells. The figures are well presented and the paper is generally well written.

The results could be of significant interest to the scientific community. However, my overall assessment is that major changes to the paper with additional simulations are required before the paper is suitable for publication. Currently the paper explores a very limited set of scenarios and thus does not robustly "stress test" or truly assess the sensitivity of groundwater and baseflow drought responses to different scenarios. It is difficult to have confidence in the conclusions that are presented in the paper when they are based on a single change for each scenario. This becomes particularly important given the significant non-linearities between changes in groundwater head and baseflow, as highlighted by the authors. A critical assessment of the model's suitability to simulate groundwater and baseflow drought responses is also needed.

These comments are discussed in more detail below, which I hope the authors find useful.

Major Comments

Scenarios – The scenarios are very limited. If the aim of the paper is to test and attribute specific sensitivities as noted in the introduction then a larger number of simulations should have been undertaken. Conclusions such as "a seasonal shift of recharge (i.e. less summer recharge and more winter recharge) will therefore have low effects on groundwater and baseflow drought severity" need to be based on more than a single scenario of +/-15% to be robust. Specific comments are:

(1) *SShift* – This scenario applies a 15% increase in recharge for winter months and a 15% decrease in recharge for summer months to the whole time series. Running a single set of percentage changes applied to the whole timeseries provides a very limited view of the question posed of "How will a changed recharge regime with wetter winters and drier summers change the inter-annual variability and water availability during droughts?". The authors should explore this in more depth by running additional scenarios that vary the percentage increases.

(2) *Srecov* – The justification for this scenario is quite weak compared to the other two scenarios and again is very limited in that it only explores the response under the assumption of long term average recharge.

(3) *Comparison between scenarios* – In the discussion and conclusions, comparisons between the scenarios are made. However, it is difficult to be confident in these comparisons as only a single scenario is assessed. For these comparisons to be robust additional simulations need to be performed to assess the sensitivity of the drought response to each scenario.

Model Evaluation – I agree with reviewer 1 that a critical assessment of the model's suitability for this application is required in Section 2. The authors need to demonstrate that the model can effectively reproduce the metrics that are used in this paper to assess groundwater and baseflow drought responses (e.g. the recovery time T_{rec}, inter-annual variability, percentile thresholds, performance during "benchmark droughts") and how this varies spatially and temporally for Germany. Currently, the discussion in Section 2 centres on model performance for T_{max} which is

based on correlations and not focused on the (likely) non-linear drought responses that are being assessed here.

Minor Comments and Technical Corrections

Abstract L7. Please change to "depend on **the** systems' sensitivity"

Introduction L25-28. I would move (or remove) the two sentences starting with "Contrary to surface water, groundwater is hard to…" to L44 where you discuss the absence of observational data and use of groundwater models in more detail.

Introduction L49. Replace "more and more" with "increasingly"

Introduction L55. "Climate models (often) lack alterations in the sequencing of future wet and dry spells". This sentence needs to be supported by some references.

Equation 1 L114. What does the '*f*' denote?

Section 3 L164. It is not entirely clear to me how you calculate inter-annual variability – can you clarify and provide the equation?

Discussion. It might be worth adding some sub-section headings to the discussion to break up the text a little for the reader.

Discussion L267-268 "Also, the recovery time Trec from a severe drought varied accordingly (SRECOV)." I am not sure what you mean here – can you clarify?

Supplementary Information. Figures S1-S4 are very difficult to interpret and the figure quality is poor (i.e. they are quite blurry). Can you make the maps bigger and ensure the figures are incorporated at high resolution so that they are clear to the reader.