This paper explored how model parameter uncertainties propagate to future projections. This is an important topic and the simulation experiments are generally well designed. However, I fully concur with the first reviewer that the major contribution of this study is not well identified, given the large number of uncertainty studies already in the literature. Also, some of the key procedures and their rationale should be explained more clearly. Therefore, substantial changes are necessary before the manuscript can be considered for publication.
Reply: We thank Reviewer 2 for the thoughtful review that will improve the quality of this work. We have carefully considered all suggestions and outlined a set of proposed revisions in the following response.

Major comments:

1. It would be beneficial if the authors would highlight their additional contribution compared to Mendoza et al. (2016), who also concluded that parameter uncertainty could affect the direction and magnitude of projected changes, based on four hydrologic models (including VIC) and three US catchments.
Reply: We thank the reviewer for raising this important distinction and have made additional clarification of the additional contribution of this paper relative to Mendoza et al. (2016) (and others) in the revision. Mendoza et al. (2016) looked at the effects of model structure, objective function, multiple local optima, and forcing calibration dataset to the projections of flux and state variables in several basins at monthly time scale. In this study, we applied downscaled climate data from 18 GCM models and chose two future phases (Phase 1 is 2040-2069 and Phase 2 is 2070-2099). We then quantified the projection uncertainty from both model parameters as well as scenario choice, for the Boulder Creek Watershed. Our study is different from Mendoza et al. (2016) in several aspects: 1) our study used multiple future scenarios from GCMs—enabling a broader characterization of the relative uncertainty from scenario choice—rather than applying a single future pseudo global warming scenario (adding a mean climate perturbation to historical conditions) in the case of Mendoza, thus our study can more fully put parametric uncertainty in the context of future scenario uncertainty, 2) our study evaluates the effect of uncertainty at different time scales (annual, monthly, and daily) rather than the exclusively monthly time scale analysis of Mendoza, thus enabling our study to provide insights to the impacts of the respective uncertainty sources at shorter time scales and thus may shed light on the impacts on hydroclimatic extremes that generally occur at shorter-than-monthly timescales.

Seiller et al. (2017) also examined the effect of parameter uncertainty on future projection, based on multiple catchments, GCM and lumped hydrological models. Could the uncertainty from parameters be compared to uncertainty introduced by different GCMs?
Reply: Yes, we thank the reviewer for this suggestion, this is correct, we can directly compare the magnitude of parametric uncertainty to the uncertainty that results from considering different GCMs. Most studies to date, have ignored the role of parametric uncertainty on climate change sensitivity. Among the comparatively few studies that do consider parametric uncertainty in future projection, the majority demonstrate that the uncertainty introduced by different GCMs was higher than that from parameters. In the revision, we now directly compare the uncertainty from parameters to uncertainty introduced by different GCMs. Seiller et al. (2017) investigated the effects of calibration metrics (i.e. selecting different objective functions) on future projection. However, our study would argue that even calibration with the same objective functions that

produce largely similar historical performance, large differences can be seen in hydrologic response to future projections.

What are the minimum parameter sets the authors would recommend for decision-making purpose, considering the computation requirement?
Reply: We have added a brief discussion on this topic in the revision. This study seeks to foremost advance understanding into the contribution of parameter uncertainty to future hydrologic projections, with the most profound result being that in this watershed, different parameter sets can produce different directions in hydrologic changes. For a situation where the parametric ensemble shows both increasing and decreasing response, we might recommend to a decision-maker that the change sensitivity is not robust, since the ensemble includes positive and negative changes, although the mean or median change may be positive or negative. For this analysis, we have selected a small river basin so as to avoid the situation where compensating errors may offset each other, e.g. where increasing and decreasing responses may cancel each other out, as well as to ensure that computation capacity for calibration was not a limitation for the experiments. We did not examine whether an increase in the number of parameter sets would reduce 'uncertainty' or not. Even if the ensemble mean values would become stable when the number of parameter sets became large, it would be difficult to conclude that the ensemble mean value would be the best "decision recommendation" for future sensitivities. This challenge is within the realm of a more fundamental question: *whether we can treat all parameter sets equally or not?* While we agree (and articulate in the revision) that proposing a minimum number of parameter sets would be useful for decision makers, we also posit that this may be catchment dependent and more importantly it would depend on the timescale of interest (e.g. flooding versus drought), which is a larger question worthy of a separate study.

By focusing on one model and one catchment, the authors could also carry out more in-depth analyses such as on the plausibility/sensitivity of some parameter values.
Reply: We would like to thank the reviewer for this suggestion. The sensitivity of some parameter values may be very informative for understanding the role of each parameter. In the revision we have added a supplemental analysis into the role of individual parametric sensitivity to the overall model sensitivity, using a variance-based approach. While the focus of the manuscript remains on the parameter sensitivity to future climate change, the parametric sensitivity helps us to understand whether this change can be most easily attributed to a single parameter or whether it's more equally distributed across those parameters that were calibrated.

2. Based on the results, it is also not quite clear if the authors' general conclusion is fully supported ("multiple optimal parameter sets are needed in order to make meaningful projections of water resource availability into the future"). What projections would be meaningful for this basin, for example, would a few percent increase or decrease in annual flow play a significant role in water supply here, or would the timing be more important?
Reply: We would like to thank the reviewer for this suggestion. We agree that the general conclusion is not well written (or overstated), and we have modified it in the revision. Our primary result here is that different parameter sets can result in large difference in future projections. The results demonstrate the need to consider multiple optimal parameter sets in order to provide a more complete range of future projections of water resource availability into the future. Boulder Creek is an important water source as it provides drinking water, agricultural irrigation, aquatic habitat, recreation, and so forth. Both the timing and magnitude of flows matter in this watershed (e.g. Murphy et al., 2003). Accordingly, we revise the manuscript to

more thoroughly contextualize how the changes shown by this work would affect water resources and decisions here and in similar basins.

Additionally, the importance of parameter values would be better understood in the context of other uncertainty sources, and previous study showed the relative importance of model structure and parameter values was catchment dependent (Kay et al., 2009). Therefore, it might be worthwhile for the authors to rethink the conclusions from their results.
Reply: We agree that the importance of parameter values would be better understood in the context of other uncertainty sources. We acknowledge that uncertainties due to model structure, objective functions, and hydrological indicators can be catchment dependent. Here, we chose a topographically complex watershed to investigate the contribution of model parameter uncertainty relative to GCM uncertainty. Both the type of watershed, as well as the (only) two sources of uncertainty examined have implications for our conclusions, e.g. are most relevant for snowmelt dominated montane domains. Furthermore, we evaluate the role of uncertainty at different time scales (annual, monthly, and daily) and for different hydrological variables. We now interpret our results within the context of Kay et al. (2009) and have revised the manuscript to avoid overstatements in conclusions.

3. I also have some questions regarding the methodology. In P4L19 "eight parameters were calibrated at 1/8 degree spatial resolution": are there different values on each 1/8 degree grid, or is each parameter the same for the whole basin? Another question is on the climate forcing data. What is the bias correction method used in USBR, and why do the data need to be corrected again by the delta-change method? Would the multiplicative method the authors use lead to some unrealistically high daily precipitation values in future, if the RatioPm is large?
Reply: We thank the reviewer for raising these issues. To clarify, each parameter set is applied over the entire basin. The bias-correction and spatial disaggregation (BCSD) climate data from USBR used a quantile mapping method on location-specific datasets, between gridded observations (Maurer et al., 2002) and the GCM historical data. For value in each grid cell and variable, cumulative distribution functions (CDFs) of conditions from both observed and GCM historical datasets were constructed. At each percentile rank in the GCM historical CDFs, observed values in the same rank were identified and applied so that the historical GCM and observed CDFs match. Quantile mapping was performed (by USBR) by populating the sample distribution using a 15-day moving window centered on each calendar day. Quantile mapping adjustments from historical GCM and observed datasets are then transferred to future time slices (USBR, 2013; Abatzoglou and Brown, 2012). The RatioPm varies from 0.76 to 1.39, with an average value of 1.02. The values have seasonal fluctuations, but generally center around 1.02. Using this adjusted dataset from USBR, we calculated and applied the delta change method, e.g. the difference in future climate relative to historical, in order to impose regional climate variability on the model. We have clarified the methodology in the revised manuscript.

As substantial changes are needed, I would only raise two minor comments at this stage:

- There are numerus grammatical errors in the manuscript such that a careful proofreading is mandatary. For example in the abstract alone: L16 "parameter sets to"- >"parameter sets for"; L23 "result to"->"lead to". - Figure 4. Which period is validation period, which is calibration?
Reply: Thank you for your suggestions, we will thoroughly proofread the revised manuscript. In Figure 4, the validation period is 1991-2010 and the calibration period is 1981-1990, which are

specified in L12-14 on Page 5. The different lengths of the calibration and validation time periods refer to the similar approach in Bastola et al. (2011). We will include this information in the updated figure caption.

References Kay, A.L., Davies, H.N., Bell, V.A., Jones, R.G. (2009) Comparison of uncertainty sources for climate change impacts: flood frequency in England. Clim. Change 92, 41–63. doi:10.1007/s10584-008-9471-4.

Seiller, G., Roy, R. & Anctil, F. (2017) Influence of three common calibration metrics on the diagnosis of climate change impacts on water resources. J. Hydrol. 547, 280–295. doi:10.1016/j.jhydrol.2017.02.004

Additional references:
Abatzoglou, J. T., and Brown, T. J. (2012) A comparison of statistically downscaling methods suited for wildfire applications. International Journal of Climatology, 32:772-780, doi: 10.1002/joc.2312.

Bastola, S., Murphy, C., and Sweeney, J. (2011) The role of hydrological modelling uncertainties in climate change impact assessments of Irish river catchments. Advances in Water Resources, 34:562-576, doi:10.1016/j.advwatres.2011.01.008.

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., and Nijssen, B. (2002) A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States. Journal of Climate, 15:3237-3251, https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.

Murphy, S. F., Barber, L. B., Verplanck, P. L., and Kinner, D. A. (2003) Environmental setting and hydrology of the Boulder Creek Watershed, Colorado. In S.F. Murphy et al. (ed.) Comprehensive water quality of the Boulder Creek Watershed, Colorado during high-flow and low-flow conditions, 2000: U.S. Geological Survey Water-Resources Investigations Report 03-4045, pp. 5-26.

USBR (2013) Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections: Release of Downscaled CMIP5 Climate Projections, Comparison with Preceding Information, and Summary of User Needs. U.S. Department of the Interior, Bureau of Reclamation, Technical Services Center, Denver, Colorado, 47pp.