We would like to thank Dr. Farmer for his insightful comments. Here we present our response to the comments and plans of revision.

Reviewer comment

Firstly, I would ask that the authors provide a little more support for their selection of climate scenarios. They state that the scenarios selected represent a medium and a high emissions scenario. While I am supportive of this approach given the current state of emissions, I feel the authors should provide objective reasoning for not selecting a lower emissions scenario. Providing objective reasoning will more strongly avoid accusations of bias selection.

Author reply

(1) We did not analyze any RCP2.6 or RCP6.0 experiments simply because of data availability. There are only RCP4.5 and RCP8.5 outputs in the MACA climate datasets. We agree that a lower emission scenario (RCP2.6) would provide some additional information. However, we believe that the four scenarios (near- and far-future under RCP4.5 and RCP8.5) can provide a sufficient variety of future climates for our goals.

Reviewer comment

I was happy to see that authors' discussion of precipitation and model uncertainty. However, I feel this is a very important issue and could use additional discussion. As the previous reviewer observed, it can have a dramatic effect on the interpretation of the results as well as the application of statistical hypothesis testing. Do the current results and understanding of uncertainty allow you to posit anything about the informal significance of these changes? How might decreased uncertainty improve accuracy?

Author reply

(2) We agree that the uncertainty propagated from GCMs, especially the projection of precipitation, is one of the most important issues in climate change impact studies. A comparison of the GCMs' capability of reproducing climate variables might be beneficial, but there could be lots of issues in this exercise. First of all, accuracy of the state-of-the-art GCMs is still quite limited when it comes to long-term precipitation predictions. Second, different models show advantages and disadvantages in capturing different characteristics (mean, distribution, extreme events, et al.) or at different spatiotemporal scales. Regretfully, the accuracy of climate modeling is out of the scope of this study. A lot more work is needed to have realistic long-term prediction. Thus, our current studies of climate change impact are not just about predicting the future, but also about demonstrating what will happen if certain assumptions or tendencies hold true in the future.

Reviewer comment

Finally, related to the question of uncertainty, I would like the authors to discuss the partitioning of attribution a little more fully. As the previous reviewer has observed, P and T are not treated as independent forcing in this work. Could you provide some additional discussion on this point? More importantly, I found myself looking for a justification on how these could be separated. For example, what changes in P are not derived from changes in T, whether directly or indirectly? It seems that any change in runoff due to precipitation may be indirectly interpreted as a change induced by T. While I do not disagree with the authors premise, I feel it should be more clearly articulated.

Author reply

(3) We have added more description in Section "2.3.2 Detecting of potential changes in future" to clarify the method and the use of the climate data. P and T are indeed related to each other in the climate system, but the interactions between them is hard to interpret or quantify without a rigorous climate modeling study. We want to clarify that the effects of P and T were only separated from a statistical perspective. For example, when we ran the model with historical temperature and future precipitation, we were not saying it is a possible future. The only possible future we showed in this study was the simulations with coupled P and T series derived from GCMs. Besides, our objective is not to predict future runoff in the actual years up to 2099, but to examine the long-term characteristics based on 30-yr averages. We believe that our method is more robust than traditional sensitivity test approaches because the inner-annual and inter-annual variations of P and T series were implicitly included. Traditional approaches usually simply assume that precipitation or temperature changes by a certain amount [*Karl and Riebsame*, 1989] or remains constant over time [*McCabe and Wolock*, 2011], but our continuous simulations based on a large dataset (three 30-yr periods and 20 GCMs) makes it more reliable to examine the long-term patterns in a highly uncertain future.

References

Karl, T. R., and W. E. Riebsame (1989), The impact of decadal fluctuations in mean precipitation and temperature on runoff: a sensitivity study over the United States, *Climatic Change*, *15*(3), 423-447.

McCabe, G. J., and D. M. Wolock (2011), Independent effects of temperature and precipitation on modeled runoff in the conterminous United States, *Water Resources Research*, *47*.

Reviewer comment

Again, not retreading the ground of the previous reviewer, I will leave my review here. I have not conducted an editorial review of the work, though I did notice, as did the previous reviewer, that some editorial work is needed before publication; I am sure that HESS will provide this resource. In addition to the comments of the previous review, I would ask that the authors pay particular attention to statistical testing. For example, the null hypothesis of a test should always be stated with an alternative hypothesis. Additionally, the results of a test should never be reported exclusively as "significant" or "not significant". Given the wide availability of computational software, this determination should always be coupled with a reported p-value, allowing a true assessment of the strength of the evidence. Binary significance is no longer appropriate in scientific literature. In a similar vein, it may behoove the authors to refer to climate change rather than global warming, the former being a more accurate representation of the situation.

Author reply

(4) We fully agree that p-value is very important for statistical tests. We showed the results of p-value in Table 2 in an earlier version. However, too many numbers in the table makes it difficult for readers to see the key results. We therefore decided to simplify it and replace the p-values with signs. We will rephrase this part considering both reviewers' comments.

We replaced 'global warming' with 'climate change' at Page3 Line17.