Author Response to Peter La Follette Interactive Comment:

The authors appreciate the thoughtful review and critique of the manuscript, with many constructive suggestions. Here we provide a general response to points raised in this review. The following discussion responds in the order points were presented in the review.

The reviewer notes that simulations lasted for 3 model years and suggests that we discuss the temporal evolution of nonlinearities.

- We agree that temporal evolution of drought is an important topic; however, in this paper we focus on results in the last year of model simulations in order to emphasize spatial scaling and factor interactions.
- Prior to the runs used for simulation, the model was initialized at an equilibrium state. That is, the simulation was run until annual storage change dropped below a threshold of 1% annual precipitation. Thus, the model began from a steady state configuration.
- Temporal evolution of drought is a large topic in itself and we feel that it is out of the scope of the present study; however, we will add text clarifying that we focus on the last (3rd) year and explaining the rationale.

The reviewer notes that the caption of Figure 12 does not state exactly what "nonlinearity" is plotted.

We agree this should be more explicitly stated and will revise to clarify that the nonlinearity is a comparison between the multi-factor hot/dry run versus the single factor hot run and dry run.

The reviewer notes that the paper lacks comparisons to measured data, and suggests that we should add such a comparison in the case of antecedent soil moisture.

- We agree that comparison to data is an important and challenging step of model studies. There are observations that could be used to explore the same research questions that are addressed here; however, none of them are directly comparable to the present model because the present model is not a reconstruction of any historical drought.
- The findings of this model, therefore, cannot be taken as a direct prediction for central North America. Rather, their value lies in suggesting system-scale phenomena such as the nonlinear combination of factors.
- It would be possible to design experimental studies or analyses of existing measurements to further explore these phenomena; however, we feel that is a future research topic and not possible within the scope of this study.
- We will add further discussion of this important question to the paper.

The reviewer queries whether the nonlinearities could be model artifacts, and suggests further discussion.

- We acknowledge that all models include simplifications and assumptions that in some cases produce artifacts, results that lack a physical basis. In this case, we have employed an integrated hydrologic model that minimizes these assumptions because it represents all important processes of the hydrologic cycle using physically based equations, as we discuss in the methods.

- Additionally, the perturbation experiment has the same forcing, subsurface configuration, etc for each case and alters one factor at a time, so that differing model outputs result from a difference in one input only.
- We acknowledge that the precise results of the model do not constitute a direct prediction, but we believe the system-scale results will not be artifacts of the model itself. We will add more discussion of this point.

The reviewer points out some missing citations in the reference list; we will add those and check that all cited references are listed.

The reviewer notes that PRISM climatological data and the model inputs and results have different grid resolutions, and queries how this was handled.
We agree that PRISM data has a coarser resolution than the model grid. We resampled PRISM rasters to the model grid before preparing forcing data. Additionally, lateral flow of groundwater in the model has the effect of smoothing resolution artifacts. This lateral flow of groundwater is primarily driven by topography, and the topography dataset is at the model resolution. We will add text to the methods clarifying this point.

The reviewer asks for more details about how the drought forcing data was prepared.

- We began with hourly North American Land Data Assimilation System (NLDAS) reconstructions of temperature and precipitation from a baseline water year. For our baseline we chose water year 1984, which is one of the most average water years for the United States in recent decades. We then increased temperature and decreased precipitation using anomalies drawn from a major drought in the region.
- To find the drought anomalies, we used PRISM data for water year 1934, a year of severe drought, and the 1920s, the non-drought, immediately preceding decade. We took months of water year 1934 to represent a "drought January" "drought February" etc. We averaged months of the 1920s to arrive at a baseline for that region at that time, a "non-drought January" "non-drought February" etc.
- We compared the months to create anomalies. For example, we subtracted "non-drought January" temperature from "drought January" temperature to find the January temperature anomaly. The averaging and subtraction was done for each pixel of the model grid, producing a spatial map for each month.
- As the last step, we modified the hourly baseline temperature data by adding the anomaly for the appropriate month in each cell.
- Precipitation data was processed in the same way except that we found a percent change for each month instead of an absolute difference, to avoid negative precipitation values.
- We will add a clarification of this point to the methods.

The reviewer asks for more detail regarding crop types and their representation, and suggests we could include a more detailed parameterization of the crops.

- We agree that different crops will affect the details of drought evolution in different ways, but feel it is important to note that ParFlow-CLM is not an agricultural model. Each cell is assigned 1 vegetation type and all crops are represented as the same "croplands" vegetation type. This is an approximation, as is our choice to remove all vegetation in these cells for a simulated crop failure.
- Analyzing the details of crop type and its impact is, we believe, outside the scope of the present study and not critical to address the study questions.
- We will add a brief discussion of this point to the methods.

The reviewer notes that Figure 12a shows that the median of the boxplots increases slightly in nonlinearity across scales, and queries whether this is a contradiction to our point that "nonlinearity is less at large scales."

- We agree that the median nonlinearity becomes more positive for runoff, and more negative for ET as scale increases. Our interpretation focuses on the spread of the boxplots, and on noting that extremely nonlinear responses happen more at small scales.
- As the scale increases, these responses average out to become less extreme, as shown by the decreasing spread of the boxplots.
- In a subcontinental basin as a whole, there is indeed a small positive nonlinearity in runoff: The change in runoff under both temperature and precipitation increase is slightly larger than that due to the separate effects of temperature and precipitation. However, in small basins, the change in runoff due to both variables can be much smaller or larger than expected.
- We will add further discussion to clarify this point.

The reviewer queries why Water Year 1984 was selected as our baseline; we will clarify 1984 was an average water year in the United States.

The reviewer identified a number of typos; we appreciate this and will make the corrections in revision.