Author Response to Anonymous Reviewer #1

The authors appreciate the thoughtful review and critique of the manuscript, and the many constructive suggestions provided by the reviewer. Here we provide a general response, addressing points in the order they were presented in the review. Review comments are in bold face.

(1) Message and relevance.
The manuscript misses a substantial differentiation between meteorological and hydrological drought. More precisely, it misses the differentiation between (i) the impact of meteorological drought and (ii) the response of hydrological drought to the former. I am referring to meteorological drought as hot and dry atmospheric forcings, and to hydrological drought as e.g. water shortages and anomalously low groundwater levels, soil moisture, ET and runoff. The combination of both aspects, along with a focus on the nonlinearity of feedbacks, would make this manuscript worth publishing. The authors do, however, only address the first aspect, i.e. they evaluate the impact of meteorological drought conditions (hot, dry, hot and dry) on an annually averaged water balance without referring to the land surface/subsurface state.
In the current manuscript we chose to focus our analysis on the propagation of meteorological droughts through the hydrologic system (i.e. into a hydrologic drought). While we agree that it would also be interesting to apply the same meteorological stresses to a system that is already in a state of hydrologic drought this is outside the scope of our current analysis. Here we specifically focus on the distinguishing the impacts of various meteorological drought signatures on ET and quantify the role of lateral groundwater flow in the response, because this is where we feel our integrated modeling approach provides the most novel results to the existing body of literature. However, in Figures 4,5,7 & 8 we do also quantify all of the hydrologic drought metrics outlined by the reviewer (i.e. groundwater levels soil moisture, ET and runoff). For space reasons, we kept this analysis to an annual level. We agree with the reviewer that with this focus we may be missing an opportunity to comment more directly on other hydrologic drought impacts. We appreciate the suggestion and plan the following revisions to better highlight the connections between hydrologic and meteorological drought which are present in our simulations:

- We will expand the introduction to include this terminology and discuss existing research on pathways for drought evolution from meteorological to hydrologic drought.
- We will expand the results section by adding a new subsection at the beginning, that focuses specifically on characterizing the impacts to the water balance in more detail and quantifying the relative importance of each hydrologic drought characteristic under different meteorological drought conditions.
- We will revise the title of the manuscript to better clarify that we are evaluating the propagation from meteorological drought to hydrologic drought as follows: “Evaluating the relative importance of precipitation, temperate and land-cover

changes in the hydrologic response to extreme meteorological drought conditions in the North American High Plains".

And according to their own introduction, the fact that precipitation deficits are the main driver for (hydrological) drought, is not novel. Yet, the authors do not explicitly show that their model does simulate a hydrological drought and how the forcing perturbations impact this drought.
This is why, at the end of the manuscript, the reader is left wondering what the impacts of anomalous dry and/or hot conditions on an existing drought are, and even more simple, if a hot temperature anomaly alone can initiate or aggravate a hydrological drought.
In Figures 4,5,7&8 we directly show the impacts of a hot temperature anomaly on ET, soil moisture, groundwater depth and streamflow. In Figure 7 we compare the impacts of the hot temperature anomaly to the dry anomaly and the combined hot and dry anomaly for all of the variables mentioned above to quantify the relative importance of temperature and precipitation stresses as the reviewer suggests. We agree with the reviewer that separating out these impacts is an important step and we feel that is one of the strengths of this modeling approach. Perhaps we are misunderstanding this comment though. If the main point of the reviewer is that we are not considering the impacts of hot and dry anomalies on aggravating an existing drought, the reviewer is correct in pointing out that we chose to evaluate impacts relative to non-drought conditions rather than starting from a system that was already stressed. Still we are able to do the types of relative comparison that the reviewer is outlining here. We do appreciate the comment and we agree that we could be more explicit in our characterization of hydrologic and meteorological drought and that the manuscript would benefit from a more detailed quantification and analysis of the hydrologic drought that we simulate. Per our response above, we plan to expand our background and results sections to better highlight these points and re-frame our discussion in terms of hydrologic and meteorological drought.

I personally like the analyses of the nonlinearity of feedbacks, which should, in my opinion, be the main topic of the manuscript and could help to increase relevance. Specifically, what would be really interesting, is the combination of both drought aspects along with the nonlinearity analysis. This would intuitively lead to interesting questions, such as: Does the nonlinearity of the land surface feedbacks to meteorological drought forcing aggravate or dampen the hydrological drought? Does it change extremes (as indicates in title!)? How does it impact severity and extent of hydrological drought? How does land use buffer the impact of (the nonlinear feedbacks of) meteorological drought on hydrological drought? The simulation experiments seem to be designed to address exactly these questions, but the manuscript does not.
We appreciate the suggestions and agree that the nonlinearity analysis is a novel component of our work which we should perhaps highlight more strongly. Because we are not coupled to an atmospheric model for these simulations we cannot directly evaluate some of the reviewer's suggested research questions here that refer to

feedbacks from the hydrologic system up to the atmosphere. We like the ideas suggested here though and in response to this comment we will do the following:

- Currently, nonlinearity is addressed under our research question 2, in results section 3.2. We will expand the discussion and directly evaluate the effects of vegetation change/land use on hydrologic response (i.e. the final question suggested by the reviewer.
  In the title we used the term 'extreme' drought because we have selected a historically extreme drought perturbation to apply to our system not because we were evaluating the role of hydrologic drought in changing this extreme meteorological condition. In response to this comment and the previous comments we will revise the title as to: "Evaluating the relative importance of precipitation, temperate and land-cover changes in the hydrologic response to extreme meteorological drought conditions in the North American High Plains".

(2) Methodology and presentation.
In addition, both, methodology and analyses would benefit from a more precise description. In the following, I will list a couple of (important) issues that remain unclearto me and hampered my understanding:
- Model selection. I understand the advantage of ParFlow as a numerical, physics-based model which simulates lateral flow over other models, such as VIC and SWAT, and the advantage is clear from the description. I do, however, not see the need to "badmouth" other models if they are neither being used and compared, nor validated against observations.
We completely agree that there is no need to 'badmouth' different modeling approaches and that was not our intent here. We added discussion of VIC and SWAT approaches because they are frequently used in similar studies of drought. Our intent was to be clear about the key differences in the physical approaches of these models and the specific advantages that ParFlow for the questions we wanted to ask in this study. We also noted in this section that ParFlow is the most computationally demanding and that this is a limitation for our approach. Indeed, our results suggest that less complex models may be adequate for large scale questions, in some circumstances. We will revise this section to ensure that we are not doing this in a way that disparages any other models, but simply weighs advantages and disadvantages.

Moreover, I do not understand the comparison of lateral flow/ no lateral flow influence on ET in Fig. 13. This is not connected to the research questions proposed and setup and results are not well explained. Do you apply a constant water table in the free drainage runs? If so, did you perform a separate spinup for those runs? Or might the differences in Fig. 13 simply arise because you have different water table depths?
We included the free draining runs to address our third research question, on spatial scaling and complexity. Many of the results in Section 3.3 suggest that model responses at large scales are linear and predictable, which in turn raises the question whether big-picture questions, at subcontinental scales, could be addressed without integrated hydrologic modeling and its km-scale complexity. The free-draining run tests

this by removing interactions between cells. Without lateral flow, the grid cells can be considered as a package of single column models, run across various soil types, slopes and land cover.

- We did not apply a constant water table or a separate spinup for the free draining run, and it does have lower water tables than the other runs. We accounted for this by using a free draining baseline run in calculating free draining ET impacts; however, it is still possible that a generally lower water table resulted in a water limited system and decreased ET once plant transpiration stopped. We will revise the methods to clarify our setup of the free draining run, and add details about the effect of a lower water table on the differences in Fig 13.
- We tested the same big-picture question (what is the average effect of crop disturbance on ET?) with and without lateral flow. The results in Fig 13 showed that crop disturbance increased ET in the normal case, but decreased ET in the free draining case. This is an example where even a large scale prediction depends on the representation of lateral flows and interactions within the model. We will revise to include more discussion of how the free draining runs relate to the research questions and overall point of the paper.

- Numerical experiments. The use of different experiment names, e.g. "Hot and Dry", "hot/dry", "Hot and Dry" (are they all the same "Hot and Dry" run (6) from Tab. 2?) is really confusing and makes it hard to follow. Please unify.
We will review for consistent terminology in the revised manuscript.

- Simulation period and dependency on land surface/subsurface state. I am totally confused about the simulation period and the setup of the numerical experiments: Which years are simulated? Why is the model set up with data from 1984?
We have summarized our explanation in response to these questions below. We will also expand the methodology to include these points in the revised manuscript.

- We use synthetic drought scenarios that are not an exact reconstruction of any historical drought, but rather an example of severe meteorological drought. We start with a baseline water year, then add perturbations singly and in combination, as shown in the manuscript Table 1.
- We used Water Year 1984 as the baseline because it is one of the most average water years in the United States in recent decades.
- The forcing data was prepared as described below:
    - 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.

Do you simulate a hydrological drought period, and if so, why do you not evaluate the impact of your forcing perturbations on evolution and extent of drought?

- We do not reconstruct any specific historical drought; however, the forcing is derived from the 1930s, which was one of the most extreme meteorological and hydrological droughts recorded in the study area. We feel it is reasonable to expect significant hydrological impacts from this meteorological forcing, and we characterize those impacts in the results section as discussed below.
- We agree that temporal evolution of drought is an important topic; however, in this paper we focus on annually averaged results in the last year of model simulations in order to emphasize spatial scaling and factor interactions. 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 (3^{rd}) year and explaining the rationale.
- We currently characterize the spatially distributed impacts in Section 3.1 with Figure 5. We agree that the spatial extent of impacts is an important topic and we will expand the discussion to include more characterization of the spatial extent/patterns of impacts including various thresholds to characterize the severity of drought impact.

Even if you do not simulate a hydrological drought, it is still important to evaluate the differences in relation to the land surface and subsurface states. What is the relative importance compared to the natural (modelled) variability? E.g. the reader does not know if a water table difference of ~1 m (Fig. 5) is on the order of natural, (inter-)annual variability and if it occurs in a region of shallow or deep water tables.

Here too we are not entirely sure what the reviewer means when they say we are not simulating a hydrologic drought. It is true that our simulation starts from a baseline non-drought condition, but we apply severe impacts to precipitation and temperature (i.e. forcing the model with a meteorological drought) in order to simulate a hydrologic drought. That point aside though we appreciate the suggestion and we agree that some additional discussion of our simulated impacts in the context of model variability would help put results into context.

- We will use our added results section (noted in our first response) to place impacts in context of seasonal and spatial variability within the model. We think this will help us better characterize the severity of our simulated impacts and improve the discussion of hydrologic drought per the reviewer's previous comments.

- Time scales. The authors only show annually averaged differences, which do not allow to address drought and extreme impacts (as indicated by the title!). The limitation of presenting annual averages becomes evident in Fig. 10, which shows the relation between "antecedent soil moisture" on ET. First of all, I do not really see a "clear break" (p. 17, l. 9).

To keep our discussion and figure count manageable we chose to focus on spatial variability more than temporal variability (in this case seasonal oscillations). We highlight the impact of spatial averaging on muting the simulated impacts similar to what the reviewer is noting for temporal averaging. As noted in earlier responses, the term 'extreme' refers to the drought perturbation we chose to apply. In Figure 10, approximately 350 mm of soil moisture marks a transition between energy limited and water limited behavior in both the hot and the dry run. We will revise the discussion to use a different term to describe this instead of "clear break."

Secondly, do you use an annual average as antecedent soil moisture? Soil moisture varies at much shorter time scales; and a grid cell (region) might move from an energy limited towards a soil moisture limited state within a year, and especially during a drought. Maybe it makes sense to look at shorter time scales...otherwise I do not see the merit of Fig. 10.

- We acknowledge that annual averages will mask the most extreme values of any given year, and that soil moisture has large variation within a year. We chose to use annual averages in order to condense a large amount of data (hourly values for each cell of the model grid) from each run into a summary that would allow us to focus on the project research questions: comparisons between runs that show interactions between variables, nonlinearity and spatial scaling.
- The merit of Figure 10 is in providing a summary of water limited versus energy limited behavior. We agree that it does not address annual variability; however, it provides a general illustration of the mechanism of energy limitation at work in the model, which is used in later discussion to explain some of the nonlinear responses.
- We will change the antecedent soil moisture to be the value at the beginning of the year, rather than the annual average.
- We will improve and expand the discussion to better connect Figure 10 and its mechanism to later sections, to clarify its place in the analysis.

Whether a grid cell is soil moisture or energy limited also depends on the soil texture, doesn't it? I am not sure that this can be as simplified as the text does it.

- Cells become moisture limited when the demand for evaporation exceeds the available water. Soil texture changes water retention, so two soils with the same water content but different textures might have different amounts of water available for evaporation. Thus, in a few cases, the same water content might lead one cell to be energy limited but another to be water limited.
- We agree that plotting all points together regardless of texture is a simplification. However, we argue that this simplification does not change the overall message of the figure; some cells are water limited, some are energy limited, and the pre-existing soil moisture affects the response of model cells to forcing changes.

- Spatial scales. Sec. 3.3 remains unclear to me, though it could be potentially very interesting. This may be mainly due to my lack of understanding what is shown in Fig. 12. Could you please clarify? Does Fig. 12 show the same comparison as in Fig. 9 but the sum? But then, which scenario is shown?

- We agree this should be more explicitly stated and will revise the caption of Figure 12, to clarify that the nonlinearity is a comparison between the multi-factor hot/dry run versus the single factor hot run and dry run.
- As the reviewer surmised, the boxplots show the spread of the data that is plotted in Figure 9 a), c), e) and g), at a variety of scales. We will clarify that connection in the text.

What are HUC6 and HUC8 basins? And which are the major basins? Which basins do you actually show in Fig. 5?

- HUC-6 and HUC-8 stand for Hydrologic Unit Code levels, a system used to classify drainages in the United States. HUC-8 basins are smaller basins that nest within HUC-6 drainages.
- We use the term major basins to mean the Arkansas, Red, and Missouri subcontinental basins.
- Figure 5 and other spatial maps show HUC-6 basins such as the North Platte, South Platte, Upper Cimarron, Republican, etc.
- We appreciate the reminder that not all readers are familiar with this terminology, and will add a clarification.

(3) Minor points.

- If I understand correctly, “anomalies” are neither climatological anomalies, e.g. of soil moisture or runoff to determine drought extent, nor are they the anomalies of forcing during the 1934 (?) drought, which are used to perturb the forcing. If I understand correctly, anomalies in the manuscript are simply the difference between a scenario and the baseline simulation. Please clarify and consider rephrasing.

This is correct; we use the term “anomalies” to describe the difference between a scenario and the baseline simulation. [We also use the term “impacts” and will review to ensure that the language is clear throughout the discussion.

- A lot of references are missing (e.g., Loon 2015; Eltahir 1998; Seneviratne 2010;

Betts 1996; Koster 2004; McEvoy 2016;...)!
We will review the manuscript to ensure that all citations appear in the reference list.

- p. 11. l. 7-9: The description of Fig. 5 in the text is misleading. If I see this correctly
(and as is later on in the manuscript mentioned), Hot (and Crops) have higher ET! Also, please be precise what "lower" WTD means...it's deeper?
We appreciate being alerted to this oversight. The Hot and Crops runs have slightly higher ET and we will revise accordingly. Lower WTD means a water table depth further below the land surface.

- Fig. 9: The differences in Fig. 9 (c-d-e-f) are not percentages, are they? 0.4 % difference in Fig. 9 seems rather small and not significant (or do you mean 40% as
mentioned in the text on e.g. p. 17?).
This is correct; c-d-e-f show fractional differences. We will revise the legend.
What do grey colors in Fig. 9 mean?
Grey colors show a limited amount of extreme values that fall outside the color scale. We set the color scale to show the most relevant variability, which required excluding some outliers.
Do I understand correctly, that Fig. 9 shows (a,c,e,g): Hot (3) + Dry (4) - "Hot and Dry" (6) ? (b,d,f,h): Hot (3) + Dry (4) + Crops (5) - "Worst case" (7) ? (with numbers referring to runs from Table 2)

- Subpanels a,c,e,g compare the hot and dry run (6) to the individual hot (3) and dry (4) runs and subpanels b,d,f, and h compare the worst case run (7) to the hot and dry (6) and crops (5) runs.
- The nonlinearity is calculated as described on page 9, lines 13-24. First, individual impacts were calculated for each run relative to the baseline, by simply subtracting baseline values. Then a fractional difference was found between the impacts in a multifactor run, and the expected impacts found by adding together individual runs.
- Red colors in Figure 9 mean that the multifactor impact was smaller than expected, and blue colors mean that it was larger.
- We will expand the caption and/or discussion to clarify.

- There are no tables representing results; hence the paragraphs on p. 13 l. 8-11 and p. 16 - l. 6-9 are confusing. Please remove/rephrase.
This discussion is supposed to refer to the data in Figures 7 and 8 (which was presented in table form in an earlier draft); we will revise and review the text for similar oversights.

- p. 14 l. 8-10: this relates to the forcing perturbations, does it? So, if I sum up the precipitation changes from Fig. X over the domain, they correspond to a total change of 40% ?
This is correct; we will clarify the sentence.

- p. 17 l. 17-18: Is this related to the results presented in Fig. 9c? (Here, radiation probably has a stronger impact than temperature).
This is correct; Figure 10 illustrates a mechanism that relates to Figure 9c. We also note that radiation did not change between the baseline and drought scenarios.

- Conclusions. (i) contradiction of "real world scenario" to what you describe in the introduction and the methods.
We agree that the model is not reconstructing a real world scenario and will rephrase the sentence. Our intent is to convey that real world scenarios include multiple factors, not to claim that this model represents a usable reconstruction of some known drought.

(ii) "ranges of variation typical of major droughts" (p. 24 l. 7) - I cannot find this classification in the presentation of the results. Could you please expand?
This phrase refers to the development of the drought scenarios, which use temperature, precipitation and vegetation changes derived from a historical drought scenario. It is meant to acknowledge that the relative importance of the factors holds for the ranges tested in the study, but not necessarily for ranges outside those simulated. For example, if temperature was increased far enough, it could have a larger impact than a small precipitation change. We will clarify this point in the methods and discussion.

- Might the perturbation of radiation be more important than temperature perturbations? (because this really indicates a limitation of energy)
Radiation was not perturbed in the study; temperature was the only change between the baseline and the hot scenario. We will clarify this point in the methods.

- p. 11 l. 11: Figure 3c?
Correct; figure 3c. We will revise the sentence.

- p. 13 l. 10: there is no Table 4.
This refers to a table that was removed in an earlier draft; we will delete this passage.

- p. 13 l. 16 and 19: Figure 6?
Figure 5 refers to the spatial map of anomalies; figure 6 plots some of these anomalies against the forcing anomalies that produced them. There is a typo in the current draft which should read: "variability in the sensitivity...is illustrated in **Figure 6** which plots...Variability in **Figure 6** shows that the same perturbation can result in a wide range of responses"

- p. 16 l. 11: How do you end up with 93% with 102 and 105 mm?
There is a typo in the current draft; it should read 105 mm, which is 93% of **114** mm. (The percentage calculation is also done with unrounded values)

- p. 16 l. 6-9: there is no table.

As noted above, this should refer to Figures 7 and 8.

- p. 16 l. 16: do you mean Figure 8?
Correct, this is supposed to say Figure 8.
- p. 17 l.8: observed is "modeled" ?
Correct, we will revise the sentence.

- p. 5 l. 9: 102 m?
Correct, we will revise the typo.

- p. 4 l.3: 1930s ?
Correct, we will revise the typo.