We have provided point by point responses below with author replies in bold.

Overall, I think this paper has the potential to turn into a good contribution that elaborates the influence of groundwater on the Budyko Hypothesis. The paper does not seem to have a well-described objective. I did not see a set of research questions or hypotheses to be tested. All the results presented in the paper are based on a single water year simulation in the ParFlow model, which is a fairly short time scale to convincingly report and use any groundwater related modeled variables. As I tried to figure out what the objectives of this paper might be I kept asking myself the following questions. Is the idea to:

1) develop a conceptual model for incorporating the role of groundwater (GW) to the Budyko hypotheisis (BH)?
2) parameterize the contribution of GW in the BH by relating the w parameter in the Zhang equation to a GW variable that may be obtained from observations or models, which can be used as a simple model?
3) evaluate model results to see when a Budyko type behavior is generated in systems where GW cannot be neglected (e.g. Fig 4), by modifying the source of water in the axis of the BH plotting position?

We thank the referee for their careful review of our work and their thoughtful suggestions. We agree that the intent of this work did not come through clearly enough in the original manuscript, and as a result some of our findings and our methodology choices were also not transparent. We have added the following text to the end of the introduction to clearly state the motivation for our work, our hypothesis and the goals for the paper. Additionally, we have made significant changes throughout the manuscript, especially the abstract and conclusions, to tie all of the discussion back to these goals.

> ***"…Groundwater observations sufficient to precisely characterize watershed storage changes are difficult to obtain and are not widely available. Therefore, it seems unlikely that groundwater storage calculations will be added into most Budyko analyses; more work is needed to understand the sensitivity of Budyko relationships to changes in storage and the implications of assuming of no storage changes without the ability to regularly verify this assumption.***
>
> ***We have identified three main approaches to estimate evapotranspiration (E) in Budyko analysis: First, if it's not possible to measure E directly, it is commonly estimated as the difference between precipitation and river outflow in a basin. Second, some studies measure E directly using a variety of field methods. Third, as is the case with the more recent studies that seek to account for storage changes, observed E values are augmented with measurements of groundwater surface water exchanges to estimate the 'effective precipitation' that is available for surface processes (i.e. outflow and evapotranspiration).***
>
> ***Here we hypothesize that storage changes will bias Budyko results in predictable ways, as has been indicated by previous studies, but that the direction of the bias will vary based on the way that evapotranspiration is handled within a study. We evaluate this hypothesis by comparing Budyko relationships generated following the three different approaches using the outputs of a physically based hydrologic model that directly simulates the integrated groundwater surface water system over a large spatial domain at high resolution. The three primary goals of our comparative analysis are as follows:***
>
> ***1. Evaluate the sensitivity of Budyko relationships to groundwater storage changes***
> ***2. Characterize systematic differences in the impact of storage changes on Budyko relationships***

3. ***Illustrate variability between approaches across physical settings and spatial scales"***

It was never clear to the reader why three water balance conceptualizations were used and why w were calculated for all three of them (Figs 5, 6, 7). The authors need to state what their goals were
The purpose of the three different approaches is to mimic the methods that are commonly used in the literature and to see how results would change as a function of the approach chosen. We hope that this focus comes through more clearly now with the text added to the introduction above. Also, we moved the description of these three approaches into a new sub-section in the methods section and further clarified our approach with the following summary added to the start of the methods section:

> ***"We use an integrated hydrologic model to simulate water and energy fluxes in both the surface and the subsurface. Here we apply a high resolution (1 km^2) simulation of the majority of the continental U.S. which covers more than 6 M km^2 and simulates hydrologic systems across a broad range of physical settings and storage change magnitudes. The model is driven using historical observed atmospheric forcings such as precipitation and temperature and provides gridded outputs of all water and energy fluxes throughout the system. We use simulated surface water flow, evapotranspiration and groundwater surface water exchanges to calculate Budyko relationships using three different approaches to estimate fluxes:***
>
> ***1. Calculating evapotranspiration from simulated runoff and precipitation***
> ***2. Using simulated evapotranspiration values directly***
> ***3. Using simulated evapotranspiration values directly and taking into account storage changes.***
>
> ***Differences between the approaches are compared with storage changes in each basin to evaluate the systematic impacts of these changes on Budyko relationships.***
>
> ***The numerical modeling approach used here provides several important advantages for this type of analysis. Within the numerical framework, groundwater surface water exchanges for every watershed in the system are fully characterized. This guarantees perfect closure of the water balance and means that we can mimic all three approaches within a consistent numerical framework where storage changes are directly accounted for. Furthermore, because the goal is to understand differences between approaches and not to predict local Budyko parameters the key advantage here is the ability to evaluate physically realistic behavior across a variety of physical settings and spatial scales where groundwater can be fully accounted for. Within this context, it should also be noted that the focus is on how groundwater storage changes perturb relationships. Therefore, uncertainty in local model parameters is much less important than realistic simulation of physical interactions for a range of storage changes and aridity values within a controlled numerical framework."***

If one needs to improve the use of the BH for regions where GW can not be neglected, one could work with the original model inputs of observed P and Q, and calculated Ep, and parameterize w = f(GW, E/P, Ep/P) and use this w in the original model and test it .. In your case E would come from PArFLow. Apparently this does not seemed to be the objective of this paper, but I felt that Fig 6b came close to this idea but stopped there.. Finding w value for the indirect method (Fig 6a) did not make sense to me as E=P-Q won't give the "correct" ET and therefore why would you calculate w using this ET/P. Please better state what you objectives are.

Again we hope that the modifications to the text have made our goals clearer. Our purpose is not to predict shape parameters but to provide a demonstration of the sensitivity of these parameters to both groundwater storage changes and the methodology chosen for a study. We agree that inferring evapotranspiration as P-Q won't give the 'correct' ET value. However, as this is a very common method used in Budyko analyses we feel it is important to understand how groundwater storage changes influence Budyko behavior when this approach is used. The scatter in Fig. 6 shows that it's not possible to predict shape parameters as a function of groundwater contributions alone. This is to be expected given the other watershed characteristics that have been shown by previous research to impact Budyko curves. However, the point of Fig. 7 is to show that for a given watershed the shift in shape parameter resulting from a non-zero groundwater contribution will follow regular patterns. It may not be feasible for most studies to incorporate groundwater surface water interactions into their Budyko analyses (especially for those studies that do not even have evapotranspiration measurements to use); we seek to demonstrate how shape parameters will shift results across a broad range of settings and groundwater contribution levels so that other studies can take this account when interpreting their results even if it's not possible to directly verify the no storage assumption with observations. We have added the following text to the conclusions emphasize this point:

> ***"These results also have implications for the myriad of studies that seek to relate shape parameters for Budyko curves to other watershed characteristics. The conceptual models shown here illustrate that groundwater contributions will shift points in consistent and predictable ways when other variables are held constant (i.e. if you apply a consistent groundwater contribution across the entire range of aridity values or consider the shift of a single point with a given aridity value). However, we use the results from our integrated hydrologic model to demonstrate that, within complex domains, groundwater surface water exchanges are spatially heterogeneous and depend on watershed characteristics such as aridity values that can also influence Budyko relationships. The scatter in Figs. 6 and 7 demonstrates that groundwater contributions cannot easily serve as an independent predictor of the shape of Budyko relationships. This also shows that in large comparative studies, the bias caused by groundwater surface water interactions may not be readily apparent because it will vary from watershed to watershed."***

Interpretation of Figure 4,5,6,7 need help. The paper does not sufficiently discuss the processes that lead to patterns in these figures.

We appreciate the comment and upon review we agree that the discussion of these figures could be improved. We have significantly expanded the discussion around these figures (as well as the rest of the results) to more explicitly step through the relationships shown and what they mean.

Abstracts lines 25, 26 what do you mean by best results? Best of what? Is the "best" ˇ represent better predicted water balance by the BH, modified in this study, against modeled water balance? Or did you develop a simpler model of water balance that gives consistent predictions with ParFlow?

We agree that this language was vague. The abstract has been significantly revised in response to other comments and this text is no longer included.

Modeling methods: In this study modeled data comes from ParFlow, which was used for only a single water year (1985), starting from a steady-state groundwater configuration. Obviously the question is – why would you use a single water year. I wonder if the system won't respond to this steady-state assumption when you start running the model with the actual climate forcing of 1985. The paper

mentions that PArFlow simulations were done for historical climate in CONUS. I wonder why the authors did not use the full length of simulations and evaluate the –mean annual water balanceâA˘Twith ˘ GW contribution in the BH hypothesis, instead of just using a single year which I presume creates some rapid transient conditions in the beginning of the model run as the water table would respond to the 1985 forcing. Running the model with a historical climate forcing data and evaluating the long-term water balance with long-term-average estimated flux variables, including groundwater seems to be the logical way to go. I'm having a hard time accepting the justification of the use of a single water year. BH is ideal for long-term-aver water balance conditions as well. So logic tells me to use longer simulations.
It's clear from this comment that some of the description of our modeling approach and decisions were unclear. When we said that we were using 'historical climate', we meant that we were using the observed historical climate data for water year 1985 not that we have completed a long term transient simulation over a longer historical climate period. We have added text to clarify that the historical climate forcings we used are only for the one-year simulation that was completed. Also, the referee is correct that there is some initial instability when we transition from the steady state-groundwater configuration to the transient ParFlow CLM simulation. We did take this into account and discarded some initialization before starting the one-year simulation presented here. We have also added text to clarify this point.

We also agree that the one-year simulation period is not ideal if our goal is to achieve a long-term water balance. However, as we hope has been clarified in this response and in the revised manuscript, our focus is on understanding the impact of groundwater surface water exchanges when long term equilibrium is either not the case or cannot be verified. We have added the following text to the methods section of the revised manuscript to be more explicit about this choice:

> ***"The 1-year simulation presented here intentionally violates the steady state assumption. The purpose of our analysis is to evaluate the impact of net storage changes on Budyko relationships, therefore a steady-state simulation is not the goal. It can also be argued that storage changes will vary from year to year or depending on the multi-year period analyzed. The 1985 simulation year is not presented as a prediction of long-term storage variability, it is simply used to sample a range of groundwater surface water exchange across variable climates and physical settings. We present a general framework for understanding the impacts of storage changes in various Budyko formulations using water year 1985 as a representative example."***

In the methods the G term need to be more clearly explained in my opinion. In reading the paper I went back and forth a few times to make sense of what authors might have meant by G but I'm still not clear.. My intuition tells me that groundwater contribution would be the net volume of groundwater staying in the basin at the end of a water year.. I imagine G is not always a contribution as in some cases G may flow out of a basin in which case G will be a sink term. Your groundwater surface water exchange can practically be infiltration or saturation excess overland flow.. I'm not following what this definition means in the context of eq.(6). Your sign convention in the G/P plots should be explained.
The groundwater contribution term is not a measure of groundwater storage or lateral groundwater fluxes; it represents the net exchange between the surface and the subsurface. In response to this comment we have added the following discussion to better clarify what is and isn't included in the groundwater contribution term (new text is underlined). Also, throughout the results section we have added text to remind the reader that a positive groundwater contribution means a net flux of water from the surface to the surface.

"There are multiple ways to estimate groundwater contributions within the model. Using gridded model outputs, the exchanges across the boundaries of every river cell could be summed to determine net contribution of groundwater to overland flow. Similarly, we could aggregate hourly changes in groundwater storage for every sub basin to determine total storage exchanges. Because we are interested in the net contribution of groundwater to streamflow and evapotranspiration for this analysis, we can take a simpler approach. Within our numeral framework we have guaranteed closure of the water balance for every watershed and therefore the net change in groundwater storage that contributes to the surface water budget is simply P - Q_{out} – E, based on Eq. (6). When calculated this way, G encompasses the total groundwater surface water exchanges (i.e. changes in storage) required to support the simulated outflow and evapotranspiration. It should be noted that in this formulation G encompasses both exchanges between groundwater and surface water, which can be either positive fluxes from the surface to the subsurface or negative fluxes from subsurface to the surface, as well as changes in surface water storage. The assumption is that, over the annual simulation, changes in ponded water are small relative to groundwater surface water exchanges and so we refer to G as simply groundwater storage changes or groundwater contributions. We follow the convention that a positive groundwater contribution denotes water that is infiltrating from the land surface to the subsurface whereas a negative value indicates groundwater discharge which can either occur from groundwater supported E or baseflow contributions to streams.

This approach is focused solely on the net contribution of groundwater to the surface water budget. Nested systems of local and regional lateral groundwater flow are simulated within the model and previous work has evaluated spatial patterns and physical drivers of lateral groundwater imports and exports across the domain[Condon and Maxwell, 2015; Condon et al., 2015], as well as groundwater residence times [Maxwell et al., 2016]. Here we focus only on net exchanges with the surface that are relevant to the Budyko formulation. We do not need to quantify lateral exchanges in the subsurface directly for these purposes; however, it should be noted that the lateral redistribution of groundwater that occurs within the model is still vital to generating realistic groundwater configurations and supporting groundwater surface water exchanges."

Line 319– I'm not following this para.. shouldn't a positive G mean that the watershed receives flux across its groundwater boundaries and a negative G indicates net export of GW to surrounding basins.. Water that infiltrates to subsurface would just increase the storage of GW wouldn't it.. This water may stay in the watershed or exported out.. seems like concepts are a bit miss-use here or not explained clearly.. Perhaps you use a Delta Storage term in 6 and 7 and explain these referring to the storage change etc.. hard to follow here..
We refer the referee to the response to the previous comment. A positive G value indicates a net flux of water from the surface to the subsurface (i.e. a positive contribution to groundwater). G is not a measure of groundwater storage or lateral flow, it is simply reflecting the extent to which the surface water budget is perturbed by exchanges with the subsurface. We hope that the referee will find the modified discussion copied above to be easier to be more transparent.

Line 353- I'm not clear how G was calculated.. Above you said you used eq (6).. here G/P>0 is interpreted as storage gain.. Headwater of Missouri should be recharging the system and therefore they are not GW exporters.. but the region below in ND and Nebraska area should be net exporters right..? so I was expecting to so G/P0 toward the middle of Missouri where it connects to Mississippi.. Please better explain conceptual model. Please clarify– if G>0, there should be net input to the watershed and Effective precip should be P+G, and if G

Again we refer the referee to the previous response and to the studies cited there that evaluate lateral flow patterns directly. As stated above, we have done work to evaluate groundwater imports and exports with this model, but the groundwater fraction presented here is not a direct measure of lateral groundwater flow between basins (although clearly the net exchanges we are using here will be supported by lateral groundwater flow in the model).

Line 419.. I would not cite Istanbulluoglu et al., 2012 here. Istanbulluoglu et al showed the limitation of assuming ET=P-Q in the Budyko curve, and proposed to use ET= P-QDeltaS, where DeltaS is change in groundwater storage assuming no net export/import of GW.

We agree; although Istanbulluoglu et al 2012 did use the inferred approach, their intent was to evaluate the limitations of this approach when storage changes occur. We agree that they should not be grouped with other studies that were not considering groundwater changes and the citation has been removed.

Line 453..don't cite Istanbulluoglu et al., 2012 here

This has also been corrected in the revised manuscript.

Line 525.. Istanbulluoglu et al., 2012 used the inferred ET approach to show its limitations– not as the proposed method to calculate ET from P and Q. Incorporating the contribution of groundwater in the water balance equation to calculate ET led to a more consistent trend in the evapotranspiration ratio and aridity index.. See Figs 6a,b and Fig 7a,b,e,f. I think this paragraph should better summarize their results.

We have expanded the summary of Istanbulluoglu et al. as follows (new text is underlined):

> ***"This point is also made by Istanbulluoglu et al. [2012] who <u>evaluated the impact of groundwater storage changes on</u> Budyko relationships using the inferred evapotranspiration approach <u>and adjusting for storage changes using estimates from groundwater observations</u>. They provide a similar conceptual model to Fig. 6d <u>describing consistent shift within the Budyko space as a function of groundwater contribution</u>. However, for the four basins in Nebraska that they evaluated they found a negative relationship between inferred evapotranspiration ratios and aridity. This was attributed to a strong negative correlation between groundwater contribution fraction and aridity index<u>. In other words, for this subset of basins, they show that the resulting trend is controlled by the dependence of groundwater contribution on other watershed characteristics.</u>"***

Works Cited:

Condon, L. E., and R. M. Maxwell (2015), Evaluating the relationship between topography and groundwater using outputs from a continental-scale integrated hydrology model, Water Resources Research, n/a-n/a, doi: 10.1002/2014WR016774.

Condon, L. E., A. S. Hering, and R. M. Maxwell (2015), Quantitative assessment of groundwater controls across major US river basins using a multi-model regression algorithm, Advances in Water Resources, 82, 106-123, doi: http://dx.doi.org/10.1016/j.advwatres.2015.04.008.

Istanbulluoglu, E., T. Wang, O. M. Wright, and J. D. Lenters (2012), Interpretation of hydrologic trends from a water balance perspective: The role of groundwater storage in the Budyko hypothesis, Water Resources Research, 48(3), n/a-n/a, doi: 10.1029/2010WR010100.

Maxwell, R. M., L. E. Condon, S. J. Kollet, K. Maher, R. Haggerty, and M. M. Forrester (2016), The imprint of climate and geology on the residence times of groundwater, Geophysical Research Letters, 43(2), 701-708, doi: 10.1002/2015GL066916.