Thank you for your valuable comments and your suggestions for completing our literature digest. In the following, we want to address some of those works, describe our motivation and our view on how our study fits in.

Donohue (2012) and Yang (2016) both build on Choudhury's parametric Budyko equation, even though Donohue's approach to relate the parameter \( n \) with physically meaningful characteristics from Porporato 2004's water balance model is very interesting to tackle the issue of lack of physical meaning. As stated in our introduction and as shown extensively by Reaver et al. 2020 (cited in the intro), we have doubts that the \( n \)-parameter can be linked to catchment in a clear and transparent manner, it is in fact a largely empirical exponent. We think that our straightforward approach is more transparent to explain deviations from the non-parametric Budyko curve.

We agree with the reviewer, that Daly et al. (2019b) indeed present a very interesting approach to introduce a new hydrological space and a storage limitation by combining the physical storage capacity with the temporal variability of \( \text{ET}_p \) and \( P \), and suggest that soil storage is a key parameter in terms of the Budyko offsets. In fact, our findings are similar, though we also characterize the influence “finer pores” exert on water limitation of \( \text{ET} \). We will acknowledge their work in a revised version, which will also help to discuss the interaction of soil storage with climatic variability in more detail.

Gentine et al. (2012) also investigated the relationship between the Budyko curve and (amongst other) soil storage. Our findings fit into what they reported for the Budyko-optimal rooting depths, while their results are based on a more complex soil water balance model. Our analysis, in turn, also includes the sensitivity of the Budyko position to the development of soils or its corresponding parameters. Plus, we found that the water balance is also sensitive to capillarity-controlled transport limitation, within reasonable ranges of field capacities.

The Budyko curve provides estimates of the mean hydrological partitioning as a function of climatic dryness. The fact that so many catchments cluster around the Budyko curve shows that climate is the first-order control for this partitioning. At the same time, it is an important factor for the development of soils and vegetation, which are dependent and interdependent elements of the hydrological system (Troch 2014). A potential evolution towards the Budyko curve or towards the maximization of its resources like water
Berghuijs 2020), should then be related to the potential to store water from temporally varying climatic forcing.

In light of these thoughts, we consider it insightful to explore the relationship of soil storage and the non-parametric Budyko curve in a virtual experiment approach. We wanted to test with realistic boundary conditions (meteorological forcing), and different parametrizations (catchments), how systems position themselves in the Budyko space with evolving soil volume and retention characteristics. The computational burden of this simplified HBV model is not so significant that we could not add more catchments to the study. However, we do not see that it will yield different results. This would be the case if we analyzed calibrated root zone storage parameters of the model in relation to the Budyko curve offsets. Here, we use the beta store of the HBV model as a learning tool to explore Budyko curve offsets by first-order sensitivities to soil parameters, after the model was calibrated. The sensitivity pattern will of course gradually change, when using a different catchment at the same dryness index, but we do not expect anything fundamentally new or different. The asset of using a model is to learn exemplarily, and we do assume that sensitivities to total storage and field capacity are meaningful and interpretable (see also work of Gharari and Hrachowitz using the FLEX Topo).

Our study was motivated by the various attempts of using the Budyko framework for deriving expected values for constraining hydrological models or estimating the water balance in data-scarce and data-uncertain catchments for practical applications, and the question what will cause the considerable offsets from Budyko that are observed in various cases. Several ideas are discussed the literature. We wanted to focus on the soil in detail, and tried to analyze the sensitivities and offsets that can be expected through the storage volume and capillary transport limitations of the evaporation flux in a straightforward modeling approach. We investigated parameter spaces spanned by both the soil storage characteristics, which neither are independent of each other, nor are not static characteristics of a catchment. We therefore also analyzed 2D parameter spaces and tried to look at them from the perspective of soil evolution. We are not aware of a similar study into these issues. Most Budyko offset studies focus on explaining current catchment water balances and most of them use parametric Budyko frameworks. We will clarify this point in the revised version.

A very extensive debate about the spatial scale and “applicability” issues of the Budyko framework is probably beyond the scope of this discussion. On the one hand, also larger catchments ≥ 10,000 km$^2$ (Budyko 1974) show clear offsets from the Budyko curve that must be related to second-order controls such as for instance soil storage or climate variability, even though the influence of specific catchment idiosyncrasies certainly increases at smaller scales – however, some of the mesoscale catchments we analyzed didn’t show a larger offset than other larger catchments. On the other hand, mesoscale catchments often present the scale of interest, and in our opinion, this corresponds exactly to the spatial scale where such second-order controls like storage can play a crucial role in the mean water partitioning (e.g., Daly 2019b). Our study explicitly investigates soil-related deviations at this mesoscale. The MOPEX dataset, which is used for many of the Budyko studies, shows a distribution where most catchment areas are below 3000 km$^2$, thus technically not in the range of “applicability” of Budyko. The data, however, still show clustering of the catchments around the Budyko curve – with smaller or greater deviations. We will try to add a comment that acknowledges the discussion about spatial scales and the Budyko curve in the revised version.

We would also consider your minor suggestions in a revised version of the manuscript. Thank you again for your helpful input!