We thank the reviewer for his comments and suggestions.

We agree with the reviewer that we are ultimately testing a model with a model. However, these models are independent from each other, and the models used are so simple that they are surely not overfitted. The starting point for our virtual experiments were hydrological models calibrated on observations. As stated in the study, we aimed at exploring and targeting the role of one specific characteristic (soil storage) for explaining significant offsets from the Budyko curve. We tried to isolate this influence on the water balance, and understand a bit better its qualitative and quantitative behavior and relationship to the Budyko curve.

Validation datasets for catchments including root zone depths are indeed difficult to obtain. It would furthermore also be challenging, if not impossible, to disentangle the other second-order influences (confounding variables) on the mean water balance. Such an approach would again require the discussion about the deviations from the soil storage perspective in this hypothetical validation dataset – which would probably end up being a whole study on its own. This motivated us to look at it in a different manner. We decided to test the role of controls related to soil storage volume in the Budyko framework in virtual experiments that allow masking out other influences.

The fact that we are focusing on soil characteristics in this paper does not mean that we consider these other factors being minor or absent. We do not want to conclude that soil characteristics are “the” first-order determinant of the deviations, but rather one among several controls, which we also mention explicitly. We focused on soil storage because, from a physical point of view, this factor should determine the available storage volume required to buffer the slower evaporation process. Testing this hypothesis in our virtual experiments showed that varying the soil parameters within reasonable ranges was sufficient to make the systems reach the Budyko curve in most cases. In our opinion, these results are a strong indication that soil storage characteristics can indeed have an important role in explaining deviations from the Budyko curve, and should not be neglected when studying these effects. We will carefully revise discussion and conclusions in this direction.

Thanks for pointing out the importance of the synchronicity of P and ETp. The importance
of the phase lag did not appear that striking to us in this study, since there is no phase lag in the Peruvian basin (one of the two seasonal ones). For the other seasonal catchment (in the US) it does occur though and might well play a role. We will complement the discussion in that regard. In general, we tried to discuss potential explanations for deviating behavior from our soil storage-based reasoning; however, this part can of course never be all-encompassing.

We are also grateful for the other, more technical comments, and will gladly incorporate them in a revised version.