Replies to Chang Liao

We would like to thank Chang Liao for his interest in our paper and for uploading his comments.

Below, comments by Chang Liao are in italic font and our replies are in normal font.

This manuscript tries to untangle one of the most challenging problems in hydrology, and it has implications to more than hydrology models: why even a calibrated hydrology model is not reliable for future simulations? While the authors lay down quite great effort to test and examine some hypothesis, its vision and credibility may be shorten by some major limitations. It is great to see authors went through input driving data (precipitation, temperature, etc.) to all the way up to discharge. The whole analytical process was very convincing. Regardless of the model details, I only have a couple of concerns and comments.

First, various spatially distributed hydrology models were used across scales. The authors need to justify why HBV is representative here. There are models considering vegetation dynamics for example.

> We are not claiming that the applied model is representative for all hydrological models and acknowledge that there are models that consider vegetation dynamics. However, conceptual HBV-type models are often used in the context of national scale climate change impact assessments. The fact that in this study, HBV did not result in reliable discharge simulations in a transient climate is thus concerning and very relevant for studies that apply HBV-type models (or similar models that neglect changes in vegetation dynamics).

Second, as authors pointed out many sources may contribute to model low performance, I suggest there should be at least more evaluations of various hydrological processes. For example, the spatial maps of snow cover, SWE, canopy interception, runoff, snowmelt, soil moisture, etc. A cost function only focus on discharge will likely miss a lot of information. We all know a combination of different parameters can produce the similar results but only one of them is the correct set. The only way to reduce this uncertainty is to examine every single step.

> We agree that including more data on other variables than discharge in the objective function is a good idea. However, for most of the suggested fluxes or state variables there are no observations to compare to (or, available observations are not directly comparable to the modelled variable, as for example for remotely sensed soil moisture). Since many of the study catchments are in a mountainous region, snow data are a relevant data source and we will add a comparison to observed snow data to strengthen the manuscript.