This paper presents the application of the PEST algorithm to calibrate the WRF-Hydro hydrologic modeling system in the Beas basin in the Himalayan region. While the paper's goals could be within the scope of the journal, I found the paper hard to understand in many parts. Therefore, I recommend rejection at this stage.

Response: We appreciate the reviewer's time and feedback on our manuscript. Acknowledging the concern regarding readability, we are committed to make necessary changes in revised manuscript to enhance the clarity and overall readability of the manuscript.

The objectives are not properly explained, and little details are provided about the hydrologic and WRF atmospheric model, the datasets, and the calibration approach. The origination is also quite confusing. For example, what is the link between Sections 1.2 and 2.5?

Response: We acknowledge the reviewer for this feedback. We did not provide much details on the model itself as it is a well known model and quite popular in the community. To keep the paper concise, we kept the model description brief, however, the necessary information about model's setup and configuration is provided in section 2.3-2.4.

Regarding datasets, we have section 2.2 to discuss the observations we used and their reliability. Section 2.3 contains the information about datasets used to force WRF model with initial and boundary conditions. Table 1 contains all the adopted necessary Noah-MP parameterization options, and Table 2 contains perturbed parameters along with their range for calibration process.

Since Section 1 is a general introduction, section 1.2 provides only a brief introduction of the PEST framework used for calibration. However, Section 2 provides the actual Data and Methods, including Section 2.5 (titled WRF-Hydro calibration), which provides details on the calibration approach/methodology of WRF-Hydro. We also provided more information about PEST in a supplementary text and cited the associated technical document.

However, we will make an effort to make these sections more readable and understandable in the revised manuscript.

In many circumstances, the authors refer to other papers (largely, Dixit et al., 2023) to justify the assumptions made and the data used in this manuscript, while they should have still properly explained the implications of these assumptions.

Response: We acknowledge this concern of the reviewer. We cited Dixit et al. 2023 mostly to refer to the selection of our parameterization and observational dataset choices, with briefly indicating the highlights. For example, we used a particular set of microphysics and cumulus schemes owing to their performance reported in Dixit et al. 2023a. Since we used the same WRF simulation output as in Dixit et al. 2023a to calibrate the hydrological model, these references are relevant and help to maintain the brevity of the paper here and keep its focus on the calibration. The reason behind following this paper is the weighted ensemble approach over time that provides better results than any setup that uses a single parameterization scheme for microphysics and cumulus physics.

There are two papers in the references, as below, that were erroneously both cited as Dixit et al., 2023 in the main text. We will correct this by adding subscripts along with the citation.

Dixit, A., Sahany, S., Mishra, S. K., & Mesquita, M. D. (2023a). Seasonal dependent suitability of physical parameterizations to simulate precipitation over the Himalayan headwater. Scientific Reports, 13(1), 4756.

Dixit, A., Sahany, S., & Mishra, S. K. (2023b). Modeling the climate change impact on hydroclimate fluxes over the Beas basin using a high-resolution glacier-atmosphere-hydrology coupled setup. Journal of Hydrology, 627, 130219.

Acknowledging the concern of the reviewer, we will make an effort to discuss the implications of adopting prior studies' assumptions, while keeping the focus of this paper on the calibration process.

More generally, in my opinion, testing the effectiveness of a calibration method (here, PEST) should be done in regions where there is a wealth of data. What is learned there could then, hopefully, be translated to other areas. The authors explicitly say that the precipitation data are inaccurate in their study region. So, what is the value of this effort? This should have carefully been addressed.

Response: We understand the reviewer's concern regarding the data availability in this region and the tradeoffs it implies. However, there is a data-paucity concern in almost the entire Himalayan region, yet this region provides water supply to billions of people, necessitating efforts to predict and project possible changes despite data limitations.

Importantly, several studies using PEST in combination with WRF-Hydro or other models (predominantly groundwater models) exist, some in regions with higher data density and quality. The PEST documentation has over 40 citations as of 2024. For example, Wang et al. (2019) used a similar approach (WRF-Hydro calibration using PEST for a flood event) over the U.S. Midwest. Their study has a 3-day calibration and 12-day validation period, using sub-daily data. They show that this period and ratio of calibration to validation time is enough to explore the associated parameter sensitivity. They also reported substantial improvement in model accuracy when using PEST. There are several other studies that calibrated WRF-Hydro for specific event classes, such as Silver et al. (2017) for flood events in arid regions. Fersch et al. (2020) calibrated WRF-Hydro for 3.5 months (15 April–31 July 2015) as a compromise between the number of model runs (about 2000), which are required during hypercube sampling and PEST optimization. They reported reasonable results in terms of Nash–Sutcliffe and volumetric efficiencies. These studies used a smaller number of parameters and shorter duration for calibration than what we do here. In short, testing of PEST has been reported extensively in the literature already, warranting our application and testing in new regions such as the Himalayas. Fundamentally, careful calibration studies in data sparse regions have always been important to pave the way forward for the hydrologic modeling community. Using PEST with WRF-Hydro helps to explore more parameter search space than some traditional and more manual methods. Our objective in this study is to provide a methodological insight on WRF-Hydro calibration with limited data to study water availability for climate time-scale. We will add paragraphs in both introduction and discussion sections to highlight the existing literature around PEST applications, as described above, and better emphasize the added value of our study in that space.

A few more responses on the choice of data for our study:

We selected a year with average-precipitation for the calibration process, and years with both average and below-average precipitation for the validation. Our findings indicate that a model calibrated with average precipitation also produces satisfactory accuracy with below-average precipitation, attesting to the fidelity of the calibrated model across a range of annual precipitation percentiles. As often with hydrologic calibration, more data would be better, but this anomalous precipitation year at least constitutes a useful out-of-sample test. Furthermore, we calibrated and tested a range of methods and precipitation realizations, totalling ~7000 model years, which itself is a significant computational effort, which would grow exponentially with extended time periods.

Regarding precipitation data, we emphasized the observed precipitation data uncertainty not its inaccuracy, a prevalent issue across the entire Himalayan region (*Bolch et al. 2012; Hartman et al. 2013; Hewitt 2005; Li et al. 2017*). Available gridded observations over the region are inconsistent, though APHRODITE has been reported to be the most reliable (*Andermann et al. 2011; Yatagai et al. 2012; Ménégoz et al. 2013; Palazzi et al. 2013; Li et al. 2017; Ghimire et al. 2018; Dimri et al. 2013; Mathison et al. 2013; Mishra et al. 2019*). Thus, we presented the uncertainty among observation datasets and argued why it is more reasonable to use APHRODITE as a reference dataset (see Dixit et al. 2023). Also, these observations are not used to force the hydrology model rather used to select the most reasonable set of choices for parameterization schemes.

Andermann, C., Bonnet, S. & Gloaguen, R. (2011). Evaluation of precipitation data sets along the Himalayan front. Geochem. Geophys. Geosyst. 12(7), Q07023

Bolch, T. et al. (2012). The state and fate of Himalayan glaciers. Science 336(6079), 310–314.

Dimri, A. P. et al. (2013). Application of regional climate models to the Indian winter monsoon over the western Himalayas. Sci. Total Environ. 468, S36–S47

Dixit, A., Sahany, S., Mishra, S. K., & Mesquita, M. D. (2023). Seasonal dependent suitability of physical parameterizations to simulate precipitation over the Himalayan headwater. Scientific Reports, 13(1), 4756.

Fersch, B., Senatore, A., Adler, B., Arnault, J., Mauder, M., Schneider, K., ... & Kunstmann, H. (2020). High-resolution fully coupled atmospheric–hydrological modeling: a cross-compartment regional water and energy cycle evaluation. Hydrology and Earth System Sciences, 24(5), 2457-2481.

Ghimire, S., Choudhary, A. & Dimri, A. P. (2018). Assessment of the performance of CORDEX-South Asia experiments for monsoonal precipitation over the Himalayan region during present climate: Part I. Clim. Dyn. 50, 2311–2334.

Hartmann, H. & Andresky, L. (2013). Flooding in the Indus River basin - A spatiotemporal analysis of precipitation records. Glob. Planet. Change 107, 25–35.

Hewitt, K. (2005). The Karakoram anomaly? Glacier expansion and the 'elevation effect,' Karakoram Himalaya. Mt. Res. Dev.

Li, L., Gochis, D. J., Sobolowski, S. & Mesquita, M. D. S. (2017). Evaluating the present annual water budget of a Himalayan headwater river basin using a high-resolution atmosphere-hydrology model. J. Geophys. Res. Atmos. 122(9), 4786–4807.

Mathison, C. et al. (2013). Regional projections of North Indian climate for adaptation studies. Sci. Total Environ. 468, S4–S17

Ménégoz, M., Gallée, H. & Jacobi, H. W. (2013) Precipitation and snow cover in the Himalaya: From reanalysis to regional climate simulations. Hydrol. Earth Syst. Sci. 17(10), 3921–3936.

Mishra, S. K. et al. (2019). Past and future climate change over the Himalaya-Tibetan Highland: Inferences from APHRODITE and NEX-GDDP data. Clim. Change 156, 315–322.

Palazzi, E., Von Hardenberg, J. & Provenzale, A. (2013). Precipitation in the Hindu-kush Karakoram Himalaya: Observations and future scenarios. J. Geophys. Res. Atmos. 118(1), 85–100.

Silver, M., Karnieli, A., Ginat, H., Meiri, E., & Fredj, E. (2017). An innovative method for determining hydrological calibration parameters for the WRF-Hydro model in arid regions. Environmental modelling & software, 91, 47-69.

Wang, J., Wang, C., Rao, V., Orr, A., Yan, E., & Kotamarthi, R. (2019). A parallel workflow implementation for PEST version 13.6 in high-performance computing for WRF-Hydro version 5.0: A case study over the midwestern United States. Geoscientific Model Development, 12(8), 3523-3539.

Yatagai, A. et al. (2012). APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. Bull. Am. Meteorol. Soc. 93(9), 1401–1415.

Finally, the level of English should be significantly improved.

Response: We acknowledge the reviewer's suggestion, and will put efforts to improve the language to enhance readability and clarity in conveying our research and findings effectively.