Employing a simple glacier mass balance model forced with dynamically refined ERA-Interim at 20km with irrigation, the study mainly focuses on reproducing the observed mass balance gradients of the WKSK glaciers, as the role of irrigation on summer snowfall increase as well as the impact of increasing westerly solid precipitation both resulting in positive or balanced mass balances have already been reported either by the authors or other studies. I very much like the idea of the study.

Quite common among the most of climate modelling studies focusing the WKSK region is the lack of their validation against (high-altitude representative) observations within the WKSK region itself, and especially against those that actually include snow, although such observations are quite a few.

Lacking such validation initially, revised study now includes more stations, adjust WRF-20km temperatures prior to their comparison with the stations, and also, validates the WRF-20km against the ERA5, ERA5-Land and HAR fine scaled reanalysis datasets.

Besides above additions and revisions in the study, I am of the view that although observed mass balances of the WKSK glaciers are reproduced - at least to some degree of satisfaction - however path to such outcome seriously lacks realism. There are large and unquantified uncertainties at each step of reproducing the delicate mass balance gradient of the WKSK glaciers, which need to be first reduced to the extent possible, quantified and then assessed for their possible impacts on the results and conclusions. My particular concerns are given below:

MAJOR CONCERNS:

To establish the robustness of the WRF-20km simulations for further use in mass balance modelling of the WKSK glaciers, its scale should be refined and its validation must be performed over the complex terrain of the focused regions against few available high-altitude observations from the WKSK region, instead of only against the stations from non-glaciated surrounding areas. Because, relative to the less complex terrain of the surrounding areas, coarse grid size dynamical refinement (as in the study) perform poor over highly concentrated topography of the western Tibetan Plateau, Karakoram and adjacent regions, irrespective of the forcing datasets, featuring substantial cold (6-10deg) and wet biases. These biases are spatially heterogeneous and are difficult to adjust statistically, particularly for precipitation. Not surprising that in the revised study, the WRF-20km temperatures have been adjusted only for comparison against the station observations but not for calculating melt season temperatures for mass balance modelling, whereas, precipitation was not adjusted at all. Substantial cold biases at 20km grid size over complex terrain may result in overall shorter melt season, reduced energy available for melt and anomalous snowfall amounts. Reducing these biases and then quantifying the effect of remaining biases on the results are therefore fundamental to establish the robustness of climatic simulation, and in turn, of modelling delicate mass balance gradients of the WKSK glaciers. Establishing the robustness of climatic simulation requires at least introducing a convection-permitting scale and resolving valley scale physical processes explicitly, by introducing a further model nest spanning over the WKSK region and then extensively validating it against available high-altitude stations from SIHP WAPDA, PMD, Agha Khan Agency for Habitat, CMA, EVK2-CNR, and others from the WKSK region at least within the 2000-2010 period, which has been used for mass balance comparison.

Further, precipitation in the only used model realization is highly sensitive to the chosen cumulus parameterization, microphysics and related physics in the climate model. Any change in physics leads to significantly different magnitudes of precipitation and signs of change, beside other climatic conditions. Effort to quantify the sensitivity of results to chosen physics in the model is missing. Recommendation from the literature could have been useful too. Chosen model physics is actually based on de Kok et al., 2018, who state that their 7-year simulation was not aimed at accurately reproducing the reality but only to show the effect of irrigation, unlike this study.

The implementation of irrigation in 20 km WRF simulation lacks realism as it perturbs precipitation at each timestep based on monthly crop water demand rate calculated by PCR_GLOBWB forced by different dataset (may be CRU TS2.1 and ERA-40), and most importantly, ignoring the on-ground facts, such as, deficit conditions and the irrigation efficiency. This can lead to anomalously higher moisture availability that yields increased snowfall in the neighboring regions and subsequently can positively affects the mass balance results. Hence, it is important to realistically implement the irrigation in the model to avoid introducing spurious atmospheric moisture amounts that are favorable to the conclusions of the study. Qualitative validation against non-validated proxy evapotranspiration observations does not add to the robustness of the WRF-20km irrigation and requires to be replaced with quantitative validation against reliable datasets including quantification of their uncertainty and subsequent impacts on the conclusions.

Other comments:

- What is the statistical significance of presented trends in temperature and snow fall? Do presented slopes actually carry any physical meanings?
- Line 180-185: how it is established that GLEAM provides unrealistically low evapotranspiration in heavily irrigated arid regions in July. Validation of WRF-20km against the employed evapotranspiration proxy observations is completely subjective and unreliable. In fact, landing in the middle of multiple non-validated unreliable datasets does not establish robustness of the WRF-20km simulations.
- I think, unlike west Kunlun Shan, Karakoram region feature accumulation during winter and spring. Validation of winter and spring climatic conditions seems important here.
- Is the water demand calculated based on ERA-Interim? I guess that the water demand calculated by the PCR-GLOBWB was based on CRU/ERA-40 datasets, which are different than those used here. If yes, any explanation on the effects on results should be added.
- How precipitation at each time step was perturbed in the model is not clear. How it has been achieved?
- Implementing irrigation through continuous light rain in the study completely ignores the significant impact of irrigation timing on the climate and seasonality and the state of vegetation. I think direct perturbation of soil moisture is a better approach that imitates irrigation via a flooding of the surface and disregards other reservoirs such as the canopy layer. Hence, it is important to know what effect introducing of continuous precipitation had on realistically reproducing surface parameters? I hope gridded observations at least over the plain areas of south Asia are representative and can be used for validation.
- Describe negative trend of regional glacier mass balance for the WKSK in Figure 7(d).

- Line 156: Within upper Indus basin, observations from a number of automated weather stations are available from SIHP, Pakistan since 1994/1995 up to 4440 m asl and from the long-term PMD stations up to 2200 masl since 1960s or earlier. For example in Norris et al., 2018. Importantly, SIHP automated weather stations include both snow and rain. Additionally, snow heights and snow fall amounts from the Weather Monitoring Posts from the Agha Khan Agency for Habitat and a few observations from EVK2-CNR are a valuable database for validation. For this, station selection criteria can be further relaxed to the available observations as the validation of the whole length of simulation does not seem to be mandatory in the data scarce region.
- Lines 220-225: Unlike Waqas and Athar (2018), several studies suggest statistically significant strong cooling at least in July and September months over both low and high-altitude stations within the HKH region.
- The study mainly focuses on reproducing the annual mass balances of the WKSK glaciers featuring delicate changes using a highly simplified lumped mass balance model. It would have been better to model the mass balance of these glaciers using more sophisticated model and on an intra-annual scale as measurements/variables from the model are not an issue.