We thank the reviewer for the comments. We address them below point-by-point (our responses are in bold print).

1. There are two key issues with the atmospheric simulations:

a. The authors have not justified their choice of model physics. The configuration does not match any of the references on Page 6 (P6), Line 9 (L9), but in any case, those studies focused on different mountainous and climatic regions. As the authors mention, the choice of physics has a large impact on WRF results and should be optimized based on a subset of the observational data or justified in some way.

Our choice of physics schemes was informed by the previous studies on WRF that focused on SEB and glacier mass balance (Aas et al., 2015; 2016; Claremar et al 2012; Collier et al., 2013; 2015; 2018; Mölg and Kaser, 2011; Mölg et al. 2012a; 2012b). However, as the reviewer correctly noted, these studies are from different mountain regions than ours. To our knowledge, there has been no WRF application on the mountain regions where our sites lie other than Liu et al. (2016) and Wrzesien et al. (2018), which focused mainly on precipitation and SWE over a larger domain and did not analyze the WRF performance over glaciers. We are aware that WRF performance can differ due to the choice of physics schemes, as has been indicated by a large number of studies, in particular regarding the schemes used for PBL and cumulus convection (see below). We agree that more justification is needed for the schemes used in our study and we will address this issue in the revised manuscript.

To respond more specifically to this comment, we provide the justification (and plans for the revised manuscript) for each choice of the model physics used in our study:

a) Radiation: we chose RRTMG scheme as it has been the most commonly used scheme in the recent glacier studies with WRF (Aas, 2015; 2016; Collier et al 2018). We could have used CAM as in Collier et al. 2013; 2015 and Mölg and Kaser (2011), but we preferred to follow more recent studies. In our revised manuscript we plan to perform a sensitivity test that will compare RRTMG to CAM.

b) Planetary Boundary Layer and Surface Layer: our choice was Mellor–Yamada–Nakanishi–Niino (MLNN) level 3 (as a local closure scheme) as the scheme is shown to perform well across different regions (e.g Coniglio et al 2013; Gbode et al. 2018; Milovac et al. 2016; Shin and Dudhia, 2015). We chose MYNN2.5 surface layer because it is shown to be more compatible with the MYNN PBL scheme (Banks et al., 2016). Previous studies on glaciers mainly used non-local closure PBL schemes (e.g. YSU scheme; Aas et al., 2016; Collier et al., 2013; Mölg and Kaser, 2011), and Monin-Obukhov (MM5 revised) surface scheme. In our revised manuscript, we plan to perform sensitivity tests comparing the MMYN3+MMYN2.5 with the YSU+MM5 revised schemes.

c) Land surface: NOAH and NOAH-MP is the most commonly used for glacier studies (Aas et al., 1206; Collier et al 2015; 2018). Milovac et al (2016) compared the performance of NOAH-MP with NOAH and found some substantial difference in WRF performance over a domain in Germany. We used NOAH-MP, but will probably perform a sensitive test comparing this scheme with NOAH.

d) Cumulus: We chose Grell 3D, while an equally legit choice would be Kain-Frisch (as used in Collier et al. 2013; 2015; 2018). These two schemes have their pluses and minuses depending on what domain one is looking at, so it is a difficult choice to make. To address this more systematically, we plan to perform a

sensitivity test with the two schemes. There is, however, a WRF study for Korean Peninsula that suggests Grell plus Thompson microphysics scheme worked best over mountainous terrain, whereas Kain-Frisch plus Thompson microphysics worked best over the plains (Jung and Lin, 2016).

e) Microphysics: as noted above, our choice of Thompson scheme is compatible with Grell 3D cumulus scheme (Jung and Lin, 2016).

b. The land surface in the closest grid cell to the observations in the ablation zones is not classified as land ice. This inconsistency is easy to fix manually in the geo_em files before running the WRF pre-processing programs. Manual correction in the finest resolution domains (WRF 2.5 and 1.0) would appear to result in only one incorrect land-use categorization at observational points, for the southern off-glacier AWS at Castle Creek (cf. Figure 2). Reliable conclusions about the model's ability to reproduce local meteorological conditions and katabatic flows cannot be drawn when the bottom boundary conditions are incorrectly specified, and as a result, there are no glacierized grid cells neighboring the one containing the station (WRF 1.0 at Castle Creek) or there is only a single glacierized grid cell (WRF 2.5 at Nordic Glacier).

The authors attempt to address the second issue by manually changing the land-surface type in the grid point containing the AWS at Nordic Glacier. However, this is the smallest glacier studied and may represent an underestimate of the impact of atmosphere-glacier feedbacks on the presented results. In addition, the horizontal resolution of the finest domain (WRF 2.5) is not well suited for this study site, as the authors acknowledge on P27, L15. For these reasons, I think the simulations should be repeated with accurate bottom boundary conditions (see minor comment 2 about SST). This change would also help to streamline and simplify the manuscript (i.e., the authors could remove P6, L29-34; P13, L17-26; P14, L24-29; Section 3.4).

Initially, we did not want to do this (manual) correction because we trusted the ESA 300 m input data used in WRF. ESA data has the latest glacier inventory (RGI) incorporated for this region and, when plotted on 300 m grid it has all the ice/snow grid cells correctly spaced (see Figures 2 and 3 in the manuscript). However, the same data when extrapolated to 2.5 km and 1 km (ourput from geogrid) appears to have a westward shift of one grid cell for 2.5 km case and two grid cells for 1 km case (Figures 2 and 3). We recently detected the cause of this problem in the index file for ESA data, prior to running geogrid script. Thus the final output data was incorrectly extrapolated despite using the correct ESA input data. We now corrected this issue and, as a result, all AWS sites are correctly recognized as ice/snow land category. We thank the reviewer for insisting on getting this done correctly as this prompted us to identify and fix the

problem.

2. The approach to the glacier simulations underutilizes the information provided by WRF, perhaps due to the incorrect land-surface categorization. For example, for most of the SEB simulations presented in the paper, daily mean albedo is specified from observations for both AWS- and WRF- forced runs (e.g., P20, L1; P21, L8) rather than using the simulated WRF value (which should be optimized in the source code, see minor comment 9). They use a calibrated value for fresh snow density and apply a temperature threshold for determining the frozen fraction of precipitation, however both of these fields are available from the microphysics scheme. This approach, in particular the albedo treatment, makes the comparison between the two forcing datasets less informative than it could be.

We agree with the reviewer and, especially in the light of corrected problem with the land-surface categorization, we think it is possible to better utilize the information from WRF model output. We are confident that we can address all key criticism received, i.e. perform the new WRF runs and sensitivity tests accordingly. In particular, we will output frozen fraction of precipitation, albedo, sensible and latent heat fluxes directly from WRF to be compared with our SEB model results and simple accumulation model.

Minor comments

1. P6, L7: Why did the authors create a new set of domains for the simulation down to 1-km grid spacing at Castle Creek? Other options would be to nest a fourth domain or use the WRF program ndown to force two separate D3s of 2.5-km and 1-km grid spacing. The latter would be the most consistent in terms of lateral boundary information, and both would be more numerically efficient.

We had some computation problems with the 1-km run setup, so we separated the two domain setups in order to resolve this problem. In the revised manuscript we plan to do the domain setup consistently as the reviewer suggested.

2. P6, L15: Why was SST kept constant and how does this impact the simulations, some of which exceed two months in length? This time-varying field is provided by ERA-Interim and is easily incorporated into the simulations using the wrflowinp bottom boundary updates.

Similarly to the above comment, we had some computation problems initially with the changing SST boundary conditions, which prompted us to perform the runs with the SST kept constant. We are, however, aware that this is far from ideal. In the revised manuscript we plan to use changing SST.

3. P6, L24: Please provide the exact spin-up time and a reason for this choice.

We will do so in the revised manuscript.

4. P7, Table 3:

a. Are the timesteps correct? If yes, is the model solution stable and physical with a timestep of 20 times the grid spacing at Castle Creek?

There is a typo. Timesteps should be corrected to: 75, 15, 3 s.

b. In complex terrain, diffusion should be computed in physical space (diff_opt set to 2) for more accurate results where coordinate surfaces are sloped.

We thank the reviewer for this comment. We will follow this advise in the revised WRF setup.

5. P8, Figure 2: The authors state that they updated the land-use data using ESA CCI, however certain areas that appear to be at least 506. P9/10: The paragraph explaining the bulk aerodynamic method could be removed, since it is well established and the reader is referred to Fitzpatrick et al. (2017).

As stated above we will be using corrected ESA data (corrected index file) so these results will most likely be revised. We will consider removing the paragraph explaining the bulk method.

7. P10, L15-16: What ice/snow albedos are supported by the measurements?

The chosen albedo values (in the SEB model) agree with those observed at our sites (see Figure 4).

8. P16: I suggest showing the comparison for precipitation, as it plays a role later in the modelled surface height changes.

This will be addressed in the revised manuscript.

9. P24, L8: The albedo for glacierized grid cells is a prognostic variable in WRF. The default lower bound (variable ALBICE in phys/module_sf_noahmp_glacier.F) is set unrealistically high at 0.67 and should be changed to a value consistent with bare ice for more realistic simulations and atmospheric feedbacks.

This will be addressed in the revised manuscript. Note that, due to corrected land-surface categorizations, out sensitivity tests will not be the same.

10. P26, L2: Please rephrase to match what is described in the methods.

This will be addressed in the revised manuscript.

11. P28, L17: Further optimization may be possible. For comparison, we run a three-domain configuration down to sub-kilometer grid spacing with dimensions exceeding 300x300x50 and are able to complete more than 20 simulation days in one

day of wall-time on 500 processors.

We thank the reviewer for pointing this out. This computational efficiency with 500 processors is indeed impressive. As a comparison, we operate with 30-50 processors and the given example would take us 10 days.

Technical comments

1. I suggest changing “near-surface” to “surface” in the title, as only surface mass and energy fluxes are considered.

2. P1, L5: for clarity, I suggest changing “nested within the ERA-Interim” to “forced at its lateral boundaries by ERA-Interim.”

3. P1, L6: change “spatial resolution” to “grid spacing.”

4. Section 2.1: I suggest referencing Figures 2 and 3 where applicable to make it easier for the reader to follow the station locations.

5. P6, L2: “advanced research version of the WRF model.”

6. Throughout the paper, please change “(see text)” to refer to the relevant section.

7. P14, L28 and elsewhere: I think the phrase “(not shown)” is overused in the manuscript. I suggest removing some of the statements if they are not important or introducing supplementary material.

8. P27, L4-5: This sentence appears nearly word-for-word in Collier et al. (2013).

Tables Figures

Figure 1: I suggest adding shaded model topography.

Figures 2 and 3: Please label the axes or provide a scale. For Figure 3, is there a pink triangle?

Table 3: Please provide the grid dimensions.

All technical comments above will be addressed in the revised manuscript.

References:

Aas, K. S., T. Dunse, E. Collier, T. V. Schuler, T. K. Berntsen, J. Kohler, and B. Luks (2016): The climatic mass balance of Svalbard glaciers: a 10-year simulation with a coupled atmosphere-glacier mass balance model. Cryosphere, 10(3):1089–1104.

Aas, K. S. , Berntsen, T. K. , Boike, J. , Etzelmüller, B. , Kristjánsson, J. E. , Maturilli, M. , Schuler, T. V. , Stordal, F. and Westermann, S. (2015): A Comparison between Simulated and Observed Surface Energy Balance at the Svalbard Archipelago , Journal of Applied Meteorology and Climatology, 54 (5), pp. 1102-1119 . doi: 10.1175/JAMC-D-14-0080.1

Banks R.F., Jordi Tiana-Alsina, José María Baldasano, Francesc Rocadenbosch, Alexandros Papayannis, Stavros Solomos, Chris G. Tzanis, (2016) Sensitivity of boundary-layer variables to PBL schemes in the WRF model based on surface meteorological observations, lidar, and radiosondes during the HygrA-CD campaign, Atmospheric Research, 176–177, 185-201.

Claremar, B., F. Obleitner, C. Reijmer, V. Pohjola, A. Waxegard, F. Karner, and A. Rutgersson (2012): Applying a Mesoscale Atmospheric Model to Svalbard Glaciers. Advances in Meteorology, 2012:1–22

Collier, E., T. Mölg, F. Maussion, D. Scherer, C. Mayer, and A. Bush (2013): High-resolution interactive modelling of the mountain glacier-atmospheric interface: an application over the Karakoram. The Cryosphere, 7:779–795

Collier, E., F. Maussion, L. Nicholson, T. Molg, W. Immerzeel, and A. Bush (2015): Impact of debris cover on glacier ablation and atmospheric-glacier feedbacks in the Karakoram. The Cryosphere, 9:1617–1632.

Collier, E., Mölg, T., and Sauter, T. (2018): Recent atmospheric variability at Kibo Summit, Kilimanjaro, and its relation to climate mode activity, J. Clim, 31, 3875–3891, doi:10.1175/JCLI-D-17-0551.1

Coniglio C., Michael & Correia, Jr, James & Marsh, Patrick & Kong, Fanyou (2013): Verification of Convection-Allowing WRF Model Forecasts of the Planetary Boundary Layer Using Sounding Observations. Weather and Forecasting. 28. 10.1175/WAF-D-12-00103.1.

Gbode, I.E., Dudhia, J., Ogunjobi, K.O. and Ajayi V. O. (2018) Sensitivity of different physics schemes in the WRF model during a West African monsoon regime, Theor Appl Climatol., https://doi.org/10.1007/s00704-018-2538-x

Jung, Yong & Lin, Yuh-Lang. (2016): Assessment of a Regional-Scale Weather Model for Hydrological Applications in South Korea. Environment and Natural Resources Research. 6. 28. 10.5539/enrr.v6n2p28.

Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., et al. (2016): Continental-scale convection-permitting modeling of the current and future climate of North America. Climate Dynamics, 49(1-2), 71–95. https://doi.org/10.1007/s00382-016-3327-9

Milovac, J., K. Warrach-Sagi, A. Behrendt, F. Späth, J. Ingwersen, and V. Wulfmeyer (2016), Investigation of PBL schemes combining the WRF model simulations with scanning water vapor differential absorption lidar measurements,J. Geophys. Res. Atmos., 121, 624–649, doi:10.1002/2015JD023927.

Mölg, T. and Kaser, G. (2011): A new approach to resolving climate-cryosphere relations: Downscaling slimate dynamics to glacier-scale mass and energy balance without statistical scale linking, Journal ofGeophysical research, 116, 779–795, https://doi.org/10.1029/2011JD015669

Mölg, T., Großhauser, M., Hemp, A., Hofer, M., and Marzeion, B. (2012a): Limited forcing of glacier loss through land-cover change on Kilimanjaro, Nature Climate Change, 2(4), 254–258, https://doi.org/10.1038/NCLIMATE1390

Mölg, T., Maussion, F., Yang, W., and D., S. (2012b): The footprint of Asian monsoon dynamics in the mass and energy balance of a Tibetan glacier, The Cryosphere, 6, 1445–1461, https://doi.org/10.5194/tc-6-1445-2012

Shin, H. H., and J. Dudhia (2016): Evaluation of PBL parameterizations in WRF at subkilometer grid spacings: Turbulence statistics in the dry convective boundary layer. Monthly Weather Review, 144, 1161-1177, doi:10.1175/MWR-D-15-0208.1.

Wrzesien, M. L., Durand, M. T., Pavelsky, T. M., Kapnick, S. B., Zhang, Y., Guo, J., & Shum, C. K. (2018). A new estimate of North American mountain snow accumulation from regional climate model simulations. Geophysical Research Letters, 45, 1423–1432. https:// doi.org/10.1002/2017GL076664