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

The set-up of WRF for such a downscaling analysis, the collection of related field data and the analysis of model in comparison to the measurements is a lot of work and a great effort. This is fully acknowledged but at the same time, the set-up of the study is such that the negative results (no improvement through WRF) are no surprise. The study is heavily outdated in a time when weather models run at 2 km resolution operationally (e.g. Cosmo in Switzerland) and WRF downscaling work is done at the resolution of tens of meters (Gerber et al., 2018) and not kilometers as in the submitted paper. A study that has just been published and uses a 30 m resolution to generate meteorological input for glacier energy and mass balance in the Himalayas is Stigter et al. (2018) and is showing the state of the art in the field and in fact documents that WRF is able to successfully generate useful local weather input.

From our own early work (Lehning et al. 2008; Raderschall et al.,), we know that it needs a very high resolution of below 50 m to approximate wind fields in complex terrain. And if you have correct high-resolution wind fields, you can describe boundary layer processes from snow deposition (Mott and Lehning, 2010) to snow ablation (Mott et al., 2011). I would go as far as to say that the dynamical downscaling of first flow and then full weather has been initiated with these ARPS based studies and the methodology has immediately also been applied to glaciers (Mott et al., 2008; Dadic et al., 2010) including the Mölg and Kaser study (Mölg and Kaser, 2011) mentioned in the paper. Therefore, if you do a downscaling study that will not reproduce the local wind fields correctly (in this case katabatic flows), you cannot expect to see much improvement over a re-analysis on local snow mass balance estimates. Now, if this was the first attempt and all that is currently possible, then this could still be interesting. But if much older and current studies (such as the ones mentioned above) already have pushed the limits way beyond the setup of this study and state of the art downscaling in much higher resolution shows that you can even simulate individual eddies (Gerber et al., 2017) then the results as presented in the paper do not add to our scientific body of knowledge.

Let me please emphasize that I am not trying to reject the paper because our own work

and other work as discussed above did not get cited. This may be seen as an omission

but is not even necessary or could be fixed easily. The main point is really that the

results do not allow to gain new insight and give a wrong impression on the usefulness

of dynamical downscaling because the study had the wrong/outdated design despite

all the good work that has been done in the execution. I would also fully be in favor

of publishing negative results but not if the negative results are the consequence of an

inadequate set-up such as in this study. This is unfortunate, as the paper is really well

written and nicely illustrated.

After carefully reading the reviewer's comments we realized that our objectives should have been more clearly stated. Specifically, our goal was not to use the highest possible resolution in WRF to simulate surface energy balance (SEB) at a given point on a glacier surface and to resolve turbulent eddies. Instead, our ultimate goal, as will be stated more clearly in the revised manuscript is to develop a regional glaciation modelling approach that would incorporate SEB modelling forced by coupled dynamical and statistical downscaling. The first step toward this ultimate goal, addressed in this study, is to evaluate the performance of a SEB model forced with dynamically downscaled fields at three glaciers in BC that were the subject of multi-year observations of all SEB components. With the use of WRF, we downscale meteorological variables and energy fluxes at spatial resolution that can be computationally attainable for large spatial domains (e.g. all mountain ranges in BC an Alberta) and relatively long periods (e.g. a decade). As an example of this attainability we refer to Liu et al (2016), a study that produced high-resolution downscaled climate fields at 4-km grid spacing over much of North America for 13-year period in present and future climate. In our study we downscale ERA-Interim to 2.5 km grid at all three glaciers, and further to 1 km at one glacier, for >200 days in total (four ablation seasons).

The reviewer has rejected our manuscript on the basis that our study, in particular the WRF setup and resolution, is 'heavily outdated', arguing that some recent studies (focused on snow processes) run WRF at 50 or 30 m resolution (citations of the reviewer's own work provided). While the cited studies represent indeed an impressive piece of work, they all deal with much smaller spatial scales and time periods (2 to 3 days of WRF simulations in total) than those we targeted and of relevance for our ultimate goal (regional glaciation modelling). We note that the recent downscaling work 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) and the ongoing work on climate downscaling over large complex terrain (e.g. Jung and Lin, 2016; Wrzesien et al., 2017; 2018) have been performed on the same scale (one kilometer or few kilometres) as in our study, so there is nothing outdated or wrong with our grid-spacing setup given the scale of the forcing.

We agree with the reviewer #1 and the editor that WRF needs to be re-run because there were some recognized inconsistencies in the model setup. We are confident that we can address all key criticism we received, i.e. perform the new WRF runs and sensitivity tests accordingly. In particular, we will:
1. justify our choice of model physics
2. run WRF with corrected land cover; all runs will have changing SST as boundary conditions; diffusion will be calculated in physical space
3. run WRF with consistent 2.5 km grid inner domain for all sites; and a further 850 m grid inner domain

for one site (Castle glacier)
4. 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

In addition to the general comments, I have also one additional major set-up problem,

which is the arbitrary switch-off of the cumulus convection scheme, while it is quite

clear that convection will be insufficiently resolved at a 2.5 km grid resolution. Again, it

then no surprise that precipitation simulations have a large error.

We disagree that this qualifies as a problem. The cumulus parametrization can be turned off below 4 km model resolution (see Weisman et al. 1997). The rule of thumb (see WRF current documentation and guidelines from NCAR) is to switch it off below 5 km resolution, but certainly at 2.5 km and below the parametrization is not needed as the model is eddy 'permitting' (not resolved, but permitted; meaning that the eddies are there but not greatly resolved). Please note the previous studies on WRF application on glaciers also switched off the cumulus convection scheme for the inner-most domain with resolution below 5 km (Aas et al., 2015; 2016; Claremar et al 2012; Collier et al., 2013; 2015; 2018; Mölg and Kaser, 2011; Mölg et al. 2012A; 2012b)

One final major point, which can either be a typo or a serious misconception is the

statement on p. 27 l. 10 when the authors talk about adiabatic cooling in the katabatic

wind on the glacier. Of course, descending air masses warm by an adiabatic process.

This is indeed a typo; we thank the reviewer for spotting it. We meant 'advective cooling' as stated on page 22, line 7. The working assumption here is that the cold air from glacier accumulation area drains non-adiabatically downslope. The air is expected to warm up adiabatically as it descents downslope, but because the shallow jet of air exchanges heat in contact with the glacier surface (which cannot exceed 0ºC), the air warms up less than it would if the process was adiabatic -> this results in advective cooling at the lower station (in the ablation area).

Some detailed comments:

p. 5 l. 6: Roughness lengths should be consistent with model resolution as sub-grid

topography needs to be represented by the roughness

Please note that we did not alter the default roughness lengths in the WRF model, i.e. in the Noah-MP land surface model. Thus there is no inconsistency with WRF model resolution in our model runs. As explained in the text, the roughness lengths (calculated from the eddy-covariance measurement at each site) are only used in our SEB model, which is run off-grid with WRF output data.

p. 5 l.8: Why did you not use a precipitation lapse rate?

The point was to make a direct comparison of WRF output with observations that are not altered or corrected in any way (bias corrections or application of assumed lapse rates).

p. 9 Eq. 2: I don't know why there is p/p_0 in this equation. This does not come from

the original derivation of the bulk formula (see e.g. textbooks by Brutsaert or Stull) and

the influence of air pressure is already there via the air density.

This is a typo; we thank the reviewer for spotting it. The air density (ρ_a) at each site is derived as the air density at standard sea-level pressure (ρ_0 = 1.29 kg/m3 at 0ºC) multiplied by the ratio between the air pressure at each site (p) and the standard sea-level pressure (p_0 ; 1013 hPa). So it should be ρ_0 instead of ρ_a in these equations.

p. 19 l.11: There is some physical argument why roughness for momentum could be

different from roughness of scalars (but empirical evidence is missing, see e.g. Schlögl

et al., 2017). However, assuming two different roughness lengths for moisture and

temperature has no theoretical justification

We agree that there is no theoretical justification. However, we use the roughness lengths (seasonal mean) as assessed from our eddy-covariance data from each site. Thus we use the empirical values for all three roughness lengths and this empirical values show difference between roughness for humidity and temperature.

p. 13 l.3: You could diagnose local stability (at least over a melting surface – not

sure you had a surface temperature measurement) and then use an adequate stability

correction.

Please note that we do diagnose local stability (z/L) which is used in the bulk method with the stability corrections to derive turbulent heat fluxes (page 10, line 6-10). It is true that we could use the same parameter (z/L) to diagnose whether a log-linear or log profile of wind applies, and then perform the wind corrections accordingly. However, due to the poor resemblance between diagnosed and observed (eddy-covariance derived) z/L (see Fitzpatrick et al, 2017; Radic et al., 2017) we chose not to introduce any additional uncertainty in the wind profiles (either observed or WRF-derived) by implementing this potential wind corrections.

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

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

Clarke, G. K., Jarosch, A. H., Anslow, F. S., Radic, V., and Menounos, B. (2015): Projected deglaciation of western Canada in the twenty-first century, Nature Geoscience, 8, 372–377, https://doi.org/10.1038/NGEO2407.

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

Fitzpatrick, N., Radic, V., and Menounos, B. (2017): Surface Energy Balance Closure and Turbulent Flux Parameterization on a Mid-Latitude Mountain Glacier Purcell Mountains, Canada, Frontiers in Earth Science, 5, https://doi.org/10.3389/feart.2017.00067.

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

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,

Radic, V., Menounos, B., Shea, J., Fitzpatrick, N., Tessema, M. A., and Dery, S. J. (2017): Evaluation of different methods to model near-surface turbulent fluxes for a mountain glacier in the Cariboo Mountains, BC, Canada, Cryosphere, 11, 2897–2918, https://doi.org/10.5194/tc-11- 2897-2017

Weisman, M.L., W.C. Skamarock, and J.B. Klemp (1997): The Resolution Dependence of Explicitly Modeled Convective Systems. Mon. Wea. Rev., 125, 527–548, https://doi.org/10.1175/1520-0493(1997)125

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

Wrzesien, M. L., Durand, M. T., Pavelsky, T. M., Howat, I. M., Margulis, S. A., & Huning, L. S. (2017). Comparison of methods to estimate snow water equivalent at the mountain range scale: A case study of the California Sierra Nevada. Journal of Hydrometeorology, 18(4), 1101–1119. https://doi.org/10.1175/JHM-D-16-0246.1