We thank Simon Gascoin for his time in providing constructive and thoughtful comments which have certainly improved the manuscript. Responses are detailed below with reviewer comment (RC) followed by an authors response (AR), in each case. Bold text indicates text sections that have been changed in the manuscript.

RC0: The data assimilation pipeline was implemented at three different "experimental scales". I agree with the first referee that the approach is interesting and that previous work may be better acknowledged. I do not have any major comment (but many minor comments, see below), except that in my opinion the paper would have been easier to read if only one of the DA approaches was described and evaluated (including at different scales). In particular, the "coarse scale DA" was only briefly illustrated while it is in my opinion the most promising approach.
AR0: We agree that the coarse scale DA has a lot of potential which we aim to further explore in a subsequent publication. The aim of this study was to demonstrate a “proof of concept” of all three approaches which we think was important as they all have their place depending on the question being asked, e.g. if the question is site scale one would use point DA which is the most accurate but costly for large area applications. Spatially distributed DA compliments our existing efficient large area methods, whereas coarse scale DA targets only bias in the large scale forcing (arguably the most important source of error particularly if one is interested in catchment/basin scale processes such as runoff).

RC1: P1L18: "Spatial resolutions of 100 m are commonly recommended for modelling of land surface variables such as snow cover or surface temperature in complex terrain": the authors may also check Baba et al. (2019) where we specifically studied this topic (see below).
AR1: We have added this interesting reference. Nice work!

RC2:P3L1: what is hyper-efficient?
AR2: Changed to “highly efficient”

RC3: P3L6: “Earth's surface”
AR3: added the comma to “Earth’s”

RC4: P4L25: this idea was surely introduced before 2018
AR4: We have cited Martinec and Rango 1981 (https://doi.org/10.1029/WR017i005p01480) one of the first studies to use snow depletion curves together with a simple snow model.

RC5: P5L14: these parameters were obtained in Greenland. This should be explicitly stated in the method and discussed later.
AR5: We have incorrectly cited these parameters, they actually originate from the study based in colorado of De Lannoy et al 2012 which in turn are based on the approach of the global study of Reichle et al. 2007. We have edited the text and caption to clarify this as follows:

“All hyper-parameters used in generating the prior ensemble are given in Table 1 and based on values from a study in Colorado by De Lannoy et al. (2012) which in turn are based on the approach of Reichle et al. (2007), which is a global study.”

“Table 1. Hyperparameters (means, variances and correlations) defining the joint probability distribution from which the ensemble of multiplicative perturbation

parameters are drawn. These parameters were obtained from De Lannoy et al. (2012) which in turn are based on the approach of Reichle et al. (2007).”

RC6: P5L27: I do not understand why the TopoSUB approach is not compatible with an iterative approach and sequential resampling of the particles.
AR6: Poor wording on our side. TopoSUB is compatible the sentiment was that we try to build a pipeline based on efficient approaches. Changed text:
“which would be more costly and less aligned with the efficiency objectives of the clustering (TopoSUB) framework.”

RC7: P6L23: Thirel et al. (2013) do not use a threshold to convert SWE to SCA but the snow depletion curve of Zaitchik & Rodell (2009). This point should be clarified.
AR7: We mean that we just use $SWE_{SCA=1}$ values from Thirel et al to account for surface roughness. However we use a binary approach based on this threshold. We have clarified this as:

“We use a simple threshold on SWE to determine the binary (snow/no-snow) snow-cover of each modelled grid cell based on SWE values that correspond to full pixel coverage (fSCA=1) given in Thirel et al. (2013), this allows us to consider surface roughness.”

RC8: P7L7: It is odd to derive the MODIS error from a study in Svalbard while there are multiple evaluation studies of MODIS snow fraction in temperate alpine regions which are more similar to Switzerland including the original paper by Salomonson and Appel (2006).
AR8: The svalbard study uses an automatic camera which provides a high resolution error estimate, which arguably could be more important than the climate zone. Salomonson and Appel (2006) evaluate against Landsat pixels. The value we use is also in good agreement with RMSE’s found for the standard NSIDC product by Masson et al. (0.154-0.157). We have therefore strengthened this statement as follows:
“This estimate is in good agreement with those found in the Alps by other studies (e.g. Mason et al., 2018), and so we use this as the as the observation error variance (σ 2 y) in the assimilation (Section 2.3.2).”

RC9: P7L21: Masson not Mason
AR9: Corrected

RC10: P7L21: "as the as"
AR10: removed an “as the”

RC11: P8L28: cumulative distributions of what?
AR11: changed to:
“Cumulative distributions of state variables”

RC12: P10L5: It is not sure if there is a need to run a snow model at 30 m resolution especially if it does not represent wind transport and avalanches. We showed that a 250 m resolution can be sufficient to capture the main energy balance processes (Baba et al. 2019). If the resolution is set to 300 m then N_r becomes 10ˆ6.
AR12: This is a nice paper! We have since been looking at optimising the target resolution as this also reduces the TopoSUB memory requirements. We agree that wind transport and

avalanches are important processes at such high resolutions and currently investigating options to parameterise these in 1D. Note: snow is lost from steep slopes with a mass loss algorithm, however it is not redistributed as it is not clear which the downslope pixels would be.

RC13: P12L2: "which are an"
AR13: Text changed to:
“which are operational products…”

RC14: P12: 400 mm, 350 mm and 826 mm, it may be a coincidence but why not using the same precision?
AR14: We made a rounding error here values should be 401 mm, 350 mm, 826 mm
.
RC15: P14L2: the noise does not come from the NDSI-SCF relationship
AR15: We have edited this sentence to:

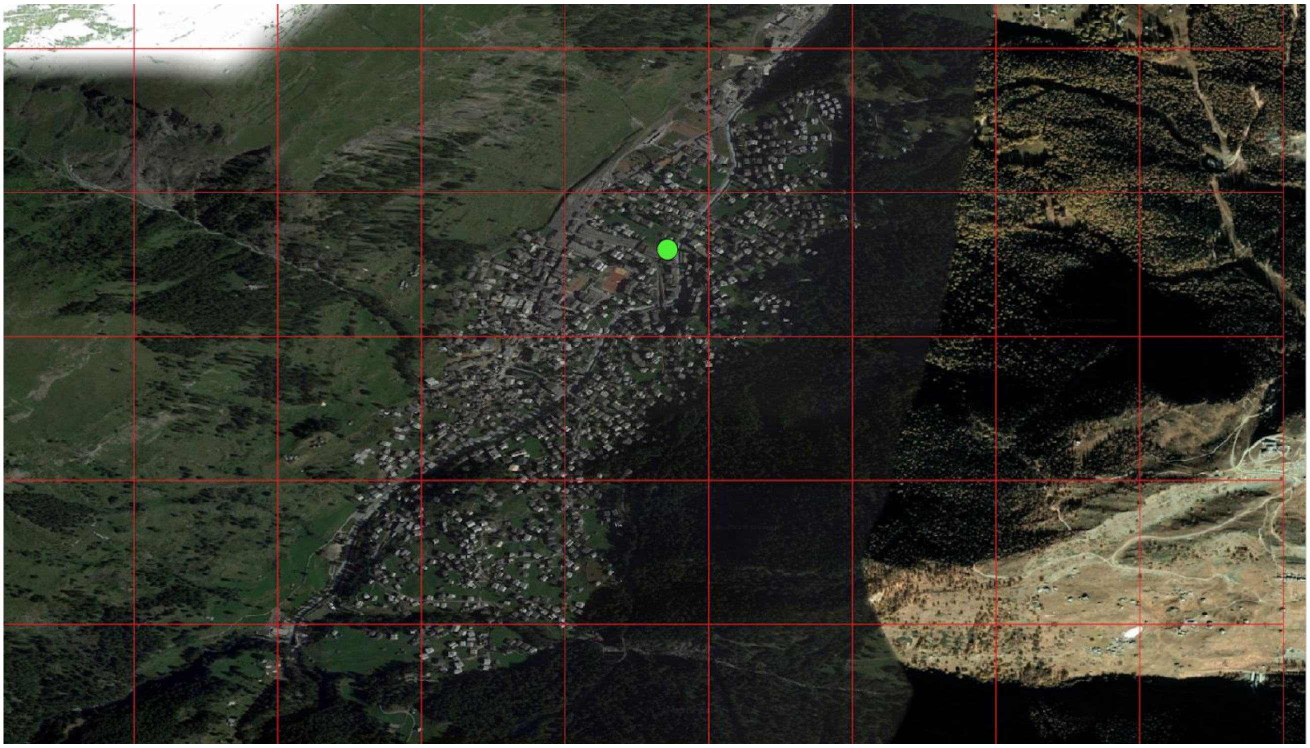
“Additionally, the MODIS products are prone to various sources of error, as discussed below in Section 6.1.5 and this adds to the difficulty in defining a robust, general algorithm that defines the start of the melt period.”

RC16: P14L14: another important limitation is the poor accuracy of the MODIS product in dense forest areas. In particular, I wonder if it could be the cause of the DA failure observed near Zermatt rather than the “urban effect”. In any case the consequence of the lack of reliable snow detection in dense forest areas must be discussed since the DA scheme is presented as applicable at a global scale.
AR16: This is a really good point, MODIS pixels even partially containing forest would definitely have an additional source of error. We have added to the discussion issues related to forest cover as a final sentence to 6.1.5 Observational errors:

“A final important limitation of the scheme is the lack of reliable fSCA retrievals in forested areas, which applies to any optical sensor (e.g. as mentioned in the description of the ADS data).”

However, we think our original hypothesis is still most likely as the Zermatt observer station is right in the middle of Zermatt town (see green dot on figure below) and the MODIS pixel footprint (red grid) is not contaminated by forest in this case. We have also edited the text to make it clear that this is point-scale DA which is important as the fSCA signal then comes from a single pixel and there is no risk of “forest contamination”:

“Figure 12 gives an example of point-scale DA..”



Author edit: P14L30: Language correction: changed “So” to “Therefore”.
Author edit**:** P14L16: Grammar correction: “As you can see” to “It can be seen”.

RC17: P15L23: The reference for the Theia snow products is Gascoin et al. (2018).
AR17: Now corrected.

RC18: P15L30: "1 km not ideal" but the results show that 500 m is useful.
AR18: This refers to the MODIS LST product here, we do not show results for that. Surface temperatures are expected to be very heterogeneous in a 1 km footprint in mountain regions therefore it is not clear how useful assimilation would be. We have reformulated this sentence as:

“For additional datasets (other than fSCA) land surface temperature (LST) can be retrieved from both MODIS and Landsat and provide a means to constrain uncertainty in the surface energy balance. However, the current MODIS LST products are coarse at 1 km with respect to the expected heterogeneity of LST in mountain regions (Gubler et al. 2011).”

RC19: P16L23: "data was obtained from We"
AR19: Removed erroneous text “data was obtained from”.

RC20: P17: I tried to explore the code in the Github repository but it contains tens of R and Python files from multiple projects; it would be a great addition to the paper if the code was a bit more documented to allow reproducing the results of this paper or even better to allow other interested people using the DA scheme in another study area (just a suggestion!).

AR20: We are working on python packages for this project - but as scientists first and software developers second this takes additional time!

RC21: P18 Endrizzi et al. not a discussion paper
AR21: Corrected

RC22: Figure 3: figure labels are too small.
AR22: We have increased label size.

RC23: Figure 8: what does represent the spread? (full ensemble?)
AR23: we have added the following text to the caption:
"The shading and solid lines show the 90th percentile range and median of the prior (red) and posterior (blue) estimates."

RC24: Figure 11: top panels are HS not fSCA
AR24: We have corrected labels and axis scaling.

References

De Lannoy, Gabriëlle J. M., Rolf H. Reichle, Kristi R. Arsenault, Paul R. Houser, Sujay Kumar, Niko E. C. Verhoest, and Valentijn R. N. Pauwels. 2012. "Multiscale Assimilation of Advanced Microwave Scanning Radiometer–EOS Snow Water Equivalent and Moderate Resolution Imaging Spectroradiometer Snow Cover Fraction Observations in Northern Colorado." *Water Resources Research* 48 (1): W01522.de lannoy

Gubler, S., J. Fiddes, S. Gruber, and M. Keller. 2011. "Scale-Dependent Measurement and Analysis of Ground Surface Temperature Variability in Alpine Terrain." *The Cryosphere* 5: 431–43.

Reichle, Rolf H., Randal D. Koster, Ping Liu, Sarith P. P. Mahanama, Eni G. Njoku, and Manfred Owe. 2007. "Comparison and Assimilation of Global Soil Moisture Retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR)." *Journal of Geophysical Research* 112 (D9): 1697.