We thank the reviewer for their encouraging and constructive comments. Their comments (in italics) are addressed below. Answers to the reviewer are given in blue.

Major comments:

Throughout the paper forest structure is parameterized from a Digital Surface Model (DSM). If I am reading this correctly this is a DSM that is normalized to terrain to give a DSM in terms of vegetation height? (Z-axiz of Figure 2)? To be consistent with literature this should rather be termed a canopy height model (CHM) rather than a DSM. https://www.earthdatascience.org/courses/earth-analytics-python/lidar-rasterdata/lidar-chm-dem-dsm/

We parameterize forest structure from a Digital Surface Model (DSM) which is the top of the surface (as defined in the short course definition in https://www.earthdatascience.org/courses/earth-analytics-python/lidar-raster-data/lidar-chm-dem-dsm/), i.e. CHM+DTM.
You are right the caption of Figure 2 was unclear. We will change that and also stress in the introduction that canopy structure metrics are derived from DSM's.

Can you clarify the type of model you are presenting in context of the existing models? There was much discussion of other existing models and a clear compare/contrast of what you are presenting would be beneficial. This is simple empirical relationship with 1 or two input variables rather than a physical/mechanistic parameterization. This is critical to clarify so that future users can determine how to use this moving forward.
We agree that it is important to characterize the type of model. We add "empirical" when we refer to model throughout the manuscript to make this more clear: e.g. in the abstract: "We present two novel empirical models …", in the discussion: "We proposed two empirical models for.." resp. "We have derived just one empirical model for the standard deviation of snow interception.-.." and in the conclusion: "The empirical models integrate forest parameters..".

A clearer description of what was being measured at the various sites is needed. It was not immediately apparent that all of this was based on snow depth difference only and ignored density. The assumption that density of new snow accumulation is the same between open and forested areas is critical. Is it reasonable to assume that the standard error of 9.31 kg/m2 of new snow estimates is greater than observed snow density differences between open and forested? Even immediately after snowfall events there will be differences in density associated with unloading/compaction on the dripline of tree crowns versus the influence of blowing snow redistribution/erosion or not in clearings? At these locations is it reasonable to assume snowfall is the same between forest and clearing locations – any preferential deposition patterns evident? Variable blowing snow deposition/erosion in clearings versus forests? In the end do you have any observations that you could demonstrated that density differences are negligible or provide these values in terms of SWE? Any errors in density differences could lead to relatively large errors in interception ratios, especially for small events, and this needs to be clarified.
We agree that snow densities can be quite different between open and forested sites, especially with compaction and redistribution of snow by wind. Therefore we chose to parameterize snow depth and not snow water equivalent (SWE), which would require

empirical parameterizations for the snow density itself (see Section 3.2 Subgrid parameterization for forest canopy interception). Until now we do not have available spatial, reliable SWE measurements comparable to the data set used in this study.

Note that the Swiss data was immediately acquired following a new snow fall event such that the influence of unloading, compaction and an eventual wind impact on the snowpack can be assumed reasonably small (see Moeser et al, 2015b).
In France, snow depth data was collected within a few days after the snow fall. The operational flat field site in France shows low wind speed and has therefore only very limited snow drift. Average hourly wind speed (at 10m height) is 1.2m/s over a period of 1993-2011 (Morin et al., 2012).
The open field site in the US may have been influenced by wind redistribution and compacting that might have created differences in the snowpack.

However, the data set from France and the US were only used as independent validation data sets, and the results are promising.

We improve the data description where we will make clear that we used observed snow depth differences, i.e. independent measurements.

The transferability of this model is tested by applying the Swiss parametrization to French and US sites. While results are promising for between these sites I would temper some of the speculation (339-353). Relative to the large range of climatic conditions of cold-regions forests globally these sites represent relatively warm locations. As expressed elsewhere there has been variability in interception model performances between maritime and continental locations not to mention more temperature cold regions versus cold arctic treeline/tundra locations. Before recommending this for universal and widespread applications this model should be tested if possible at other locations that represent more end members.
While we agree that the novel models should be tested for a broad range of climatic conditions including extreme climate conditions and at various geographic sites we believe that the three sites already cover substantial variability (as shown by mean air temperatures and precipitation sums). We therefore believe that the novel models could perform sufficiently well in other climate conditions (though of course extremes have to be investigated). At the moment we do not have more snow interception data sets available that would allow an extended evaluation.

Note however, that we also point out the limitations of the models when applied for a deciduous forest. This was discussed in line 373-385.

We rephrase line 339-353 to soften our recommendation.

The approach implemented is to parameterize an empirical relationship. This will not work perfectly for all scenarios/locations obviously. Is it possible to quantify the uncertainty of the parameters and how they may vary between sites / vegetation types? How stable are these parameters?

Yes, our approach is to parameterize an empirical relationship. This model was then validated with independent data sets at different locations and for one different vegetation type, suggesting that it is also applicable in different climates/forests. Nevertheless, it is clear that more data is required to confirm this. To specify the robustness of our fitted parameters we will include the confidence interval of the estimates.

<u>*Specific comments:*</u>

First sentence on abstract and line 19-21 are a little contradictory.

This will be modified.

26-33: Transition to discussing surface albedo is abrupt. While snow interception/albedo is a critical feedback it is not extensively discussed hereafter? Can this section be simplified?

We agree and we will rephrase the transition.

41-42: Awkward sentence

Agreed, we will rephrase this sentence.

102: define more clearly what indirect interception measurements are.

Indirect interception measurements were introduced in line 37-39 in the introduction. We will expand the explanation in the data section as well.

151-154: what may the influence of different point cloud densities be upon the CHM derivation? Are there any recommendations you could make on what should be collected in future for proper a parameterization of your models?

In general the higher the point cloud, the greater the potential detail of the model. Specifically, if multi-return LiDAR is being integrated, then the higher the density of the last returns, the higher the potential detail of the DSM. This translates into a higher resolution CHM as well, since this data is subtracted from the raw data to create canopy heights from elevations.

1-m resolution DSMs computed from points clouds above 5 returns/m^2 are usually quite consistent, and are generally considered as suitable from automated tree detection. Local artifacts (NA or low pixels) can be expected due to heterogeneous scanning pattern on the ground, and to canopy penetration variability depending on forest type and beam intensity and divergence. But the description of the canopy at 1 m resolution is quite robust for such densities. Higher densities are probably required in the case of deciduous species with LiDAR data acquired in leaf-off conditions.

Ten years ago, country-wide acquisitions would be typically between 0.5 to 2 returns/m^2. Current available or scheduled country-wide datasets are now around 1-5 returns/m^2 (e.g., Denmark 5 returns/m^2, North-Rhine Westphalia in Germany 4 returns/m^2, Spain 1 return/m^2, France 2 to 5 returns/m^2).

We can expect that thanks to technical improvements in LiDAR sensors, the density of 5 returns/m^2 will be exceeded in most countrywide campaigns in the next decade. Besides, acquisitions over smaller areas (municipalities…) have usually higher densities.

205: how are you getting from SP at a point to mean sky view factor. How is the space being discretized? Are you computing SP on a fine scale grid and averaging values over a coarser scale of interest?

Here, we derive the sky view factor *F*sky using Eq. (1) for each fine-scale grid cell in a DSM. A spatial mean is obtained by averaging all fine-scale grid cells within a coarse-scale grid cell. The spacing thus depends on the available fine-scale grid cell resolution of the DSM. We do not use *F*sky derived from SP. We will make this clear in line 199-205.

248-250: can you clarify this reversed response?

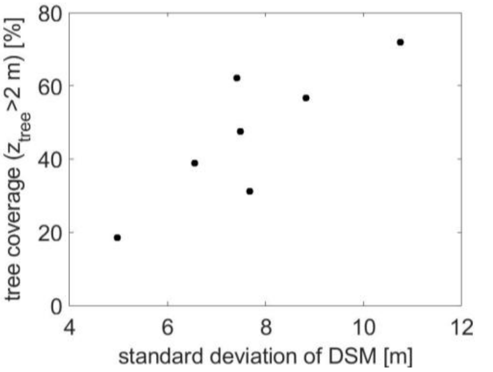
Deriving the sky view factor *F*sky using Eq. (1) for a fine-scale grid cell in a DSM implies a calculation on the DSM. Deriving *F*sky from HP or SP allows a view from below canopy. Since we do not use HP derived *F*sky and to avoid confusion we remove the second part of the explanation.

279: Why do we want to know the standard deviation of snow interception. Can you articulate a broader reason to calculate this?

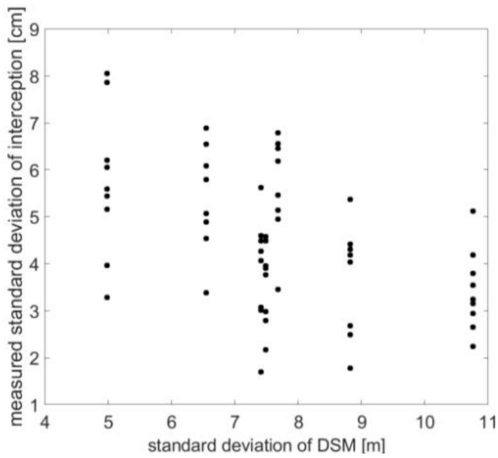
While we had some explanations in line 86-88 we will add some at the beginning of Section 3.2.

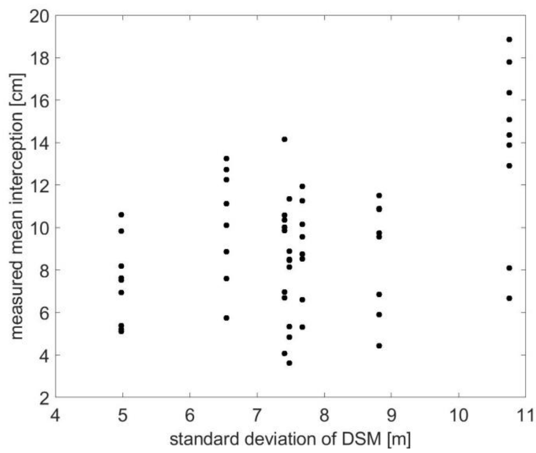
286-287: As canopy gets to be more homogeneous the spatial variability of interception increases? How?

The larger σ_z the more trees are in a field area (50x50m^2 plot), i.e. the denser the coverage:



Based on our data set larger σ_z implies larger spatial mean interception, but lower spatial variability of interception. Lower σ_z implies lower spatial mean interception, but larger spatial variability of interception:





We will rephrase the explanation in line 286-287.

To avoid infinity for an extreme value of σ_z=0 but keeping the functional form we introduce σ_z=*f*(1/(1+σ_z)). This new form changes modeled interception values and performances only slightly. We will change this in the revised manuscript.

322-324: Could a global product (between 51.6_ N and 51.6_ S at least) for these metrics
be derived from the GEDI platform? https://gedi.umd.edu/
Thanks a lot for pointing us towards the GEDI platform. This is a very promising mission. A GEDI product exists with 25 m grid resolution for ground elevation and canopy top height (L1A-2A). We assume a product like thus could be used to compute *F*sky spatially (Eq. (1)).

However, a 25 m fine-scale DSM is much coarser than the resolutions used here for developing a snow interception model. This might need a scale-dependent investigation.

365: Typo "ASifferences"
Corrected.

381-383: Deciduous will have very different behavior than coniferous vegetation. Could you reoptimise your model for deciduous specific sites? Would be interesting to know if the same scaling laws were applicable to know if a separate deciduous scaling parameterization is needed or not.
We agree that snow interception models should be verified for different vegetation species. As discussed in line 371-375 Huerta et al. (2019) showed very recently that current interception models developed for coniferous vegetation required recalibrating of fit parameters to be applicable in deciduous forests. However, the same scaling laws were applicable.
Our models were developed and validated for coniferous vegetation. We further validated the models with indirect interception measurements from two measurement campaigns conducted in a deciduous forest in the US. Larger biases resulted. However, we could not perform a solid validation of our models with this data set since the LiDAR point cloud was acquired during leaves-on conditions, which led to overestimations in modeled interception.
To develop empirical interception models for deciduous forests measurement campaigns and a LiDAR acquired during leaves-off conditions are required.

386-388. Why? Can you justify this a bit more?
Unfortunately, it is not fully clear to us what your question is.
The novel interception models presented here, use forest structure metrics which can be derived spatially on a DSM without tedious field measurements. The accuracy of the derived metrics is dictated by the resolution of the DSM. In contrast a more accurate presentation of forest structure metrics might be achieved using field measurements (e.g. HP) but then a spatial coverage is not feasible.

412-416: The full summary of the various interception efficiencies would be better presented in the results rather than in the conclusion for the first time.
We prefer having this summary in the conclusions. It is not really a result, but rather characterizes observed snow interception in general compared to our datasets. It further confirms previously observed annual snow interception fractions which were mentioned at the beginning of the introduction.

423-425: long challenging sentence
We will rephrase this sentence.

Figure 1: Can scale bars and north arrows be consistently sized and located on edge of orthophotos? Where snow depth measurement setups the same for each point in the respective sites?
Yes, the snow depth measurement setups were the same for each point in the respective sites. We will make Figure 1 more consistent.

Figure 2: what is grid resolution of DSM (aka CHM)? What UTM zone is applicable for the respective easting/northing? Correct sig figs on the easting northing?
The coordinates of Figure 2 are displayed in the Swiss reference system CH1903+. It is metric similar to UTM. Grid resolution of the CHM's is 1 m as for the DSM's. This information will be added to the caption.

Figure 4-6: "Parametrized" or "Modelled" interception on y-axis label
We change the labels to "Modeled interception".