Responses to Editor & Reviewer Comments

General and specific referee comments were responded to in completion, without exception.

Our responses are given in italicized text, following the posted comment.

Editor Comments

Line 59. "grain-bond characteristics" may not be clear to all readers. How about "snow type" or "snow type e.g. depth hoar vs rounded-grain snow", or any other option you would propose.

Revised as, "snow type characteristics (e.g., faceted crystals versus rounded-grain snow)"

Lines 217-220. I had to read this sentence 3 times to really understand it. How about breaking it into 2 shorter sentences? In particular, the construction ", we suggest," is far from ideal.

This sentence has been removed, as this topic is provided in the discussion.

Referee Comments

L13 for spaceborne remote sensing => « with remote sensing methods », the article has little to see with spaceborn methods.

Accepted

L28 Maybe add a more general sentence about climate change impact on snowpack at global scale before zooming on the western US ?

The authors have decided to maintain the focus of the article on the western US and how this research pertains to our study of Grand Mesa, Colorado.

L43 light detection and ranging (LiDAR; e.g., Deschamps-Berger et al., 2023; Hu et al.,2021) aboard ICESat-2 (Abdalati et al., 2010).

Revised as recommended, "light detection and ranging (LiDAR) with ICESat (Treichler and Kääb, 2017) and ICESat-2 (Hu et al., 2021; Deschamps-Berger et al., 2023; Besso et al., 2024)."

L57 empirical models spatially distribute density in SWE estimates => empirical models are used to spatially distribute density in SWE estimates

Revised as recommended.

L65 « Machine learning (ML)... » Is ML different from empirical models ? The same Broxton et al. (2019) citation is used L57. It makes this last sentence a bit confusing.

Elder et al., 1998, Wetlaufer et al., 2016, and Broxton et al., 2019 now reference the ML approaches on L66 and have been removed from L57.

L65-67 Further, how different is « verification » from « validation » ? Verification sounds vague and makes the whole sentence circular (validation is lacking requiring validation to gain confidence).

Revised as, "… are often distributed over vast areas with little validation, or consideration to the underlying physical processes, required to gain an acceptable level of model confidence."

L70 « simpler » => « simple » ? or simpler than what ?

Simpler than dynamic compaction schemes. The word simpler has been removed, as it is subjective.

L80 « require appropriate constraints » quite vague, please precise.

Revised as, "Yet, many radar remote sensing retrievals require constraints on the snow depth, density, stratigraphy, and microstructure to be presently reliable (Tsang et al., 2022)."

L90 «Our work leverages **airborne** LiDAR » ?

The word airborne has been added to the sentence as recommended.

L95 « highlights interactions between snow [...] on the densification process » interactions on ?

"On" has been changed to "in", as, "…highlights interactions between snow, terrain, vegetation, and wind in the densification process…"

L97 « to assimilate parameterizations » assimilation data is the usual term, what are you refering to here?

Revised as, "Our work addresses the need for high accuracy, distributed density measurements as assimilation data for parameterizations of snow densification…"

L155 Eqn 1. It not clear how $CHH-$HV would be calculated from this equation. Is it by setting i=1 and i=2, respectively in the first and second sum ?

The Equation has been clarified by correcting the notation of the j[th] iteration in the summation as $C(t) = \frac{1}{2}\sum_{j=t-N/2}^{t+N/2}\left\{\left[\sum_{i=1}^{M} S_{i,j}\right]^2 - \sum_{i=1}^{M} S_{i,j}^2\right\}$,

L160 Eqn 2. Does $CHH-$HV depend on *t* ? Add *max(C)*. The max is calculated within the window or all the data ? $CHH-$HV and C(t) seem interchangably used for coherence

(L155, L158) and normalized coherence (L160). Please clarify the notation.
Corrected

L190 « were resampled to the 1 m resolution » using nearest-neighbour algorithm ? Precise
Revised as, "…bare-earth and vegetation data products were resampled using the nearest-neighbour approximation…"

L197 « to co-register the LiDAR » from your answer, I understand that you do not shift or translate the LiDAR data but rather associate LiDAR points with snowpits. Could « pair », « identify », « match » be more appropriate than « co-register » ?

The word "find" has replaced "co-register"

L243 : « by retraining on random subsets of data. » not clear at all.

Revised as, "By retraining the model architectures on random subsets of data, 50 model ensembles were generated and then averaged for both RF and ANN regressions."

L260 « upscaled » how ? Using what algorithm ?

Revised as, "We upscaled $b_{s,LiDAR-\overline{Ens}}$ *to 50 m resolution using nearest-neighbour approximation for comparison with the 50 m ASO SWE."*

L262 «the RMSE (11 cm) was used to estimate the random error » RMSE is impacted by systematic errors (i.e. bias), thus it should not be used to estimate the random error.

Please note that our LiDAR snow depth product was validated and expresses no systematic biases.

L302 « Using supervised ML regression**,** » add coma ?

Accepted

L320 « drives **SWE** spatial patterns » ?

Accepted

L388-390 « that are on the scale of the 1 m resolution data products. » I am a bit doubtfull of that. The density models were trained on raster variables smoothed at 5 m and 25 m resolution. Thus a 1 m shift is small compared to the actual resolution of the densities and should have little impact. Especially taking into account the little variability of density at this scale (Fig 5.). Following your answer to my previous similar comment : how do you understand the fact that « f*urther perturbations of data alignment led to »* a much smaller error (1 kg m-3) than the error mostly attributed to misalignment (30 kg m-3)?

Errors caused by further perturbations of data alignment are insensitive to additional perturbation, as you have alluded to, by our processes of smoothing data and outlier filtering. In practice, we find that within a 1 m radius of probed depth validation measurements approximately 10 cm of error in snow depth and approximately 1 ns of TWT error from cross-over analysis is expected. We show how these measurement errors contribute to errors of up to 150 kg/m^3 in the density retrieval. A small perturbation in space may in fact produce a large error. Without errors existing between the registration of these data, and if you comprehend that the measurements of both LiDAR snow depth and GPR travel-time are repeatable in space, our method of retrieval is straightforward and depth and TWT will agree with snow density.

Retrieval error, as just described, is challenging to disentangle from validation errors and inherent spatial variability. Within a snow pit measurement of density, it is very common to make adjacent measurements which vary by more than 25 kg/m^3. We provide, that on average across all snow pits the sample variability among adjacent data is 2.5 %. We also find this to be the inherent, nugget variability of our retrieval algorithm. Which suggests that our retrieval is sensitive to real variability and at the appropriate length scales.

L403. Maybe comment on the fact that Yildiz et al. (2021) had a smaller study site which limited the lag considered to a maximum of ~50 m ?

Revised as, "These findings differ from a previous variogram analysis that found correlation lengths for snow density of less than 10 m at a smaller study site which limited the maximum lag separation to approximately 50 m (Yildiz et al., 2021)."

L435 « To capture the range of processes (i.e., elevation, slope, aspect, and forest attributes) » these are not processes.

"processes" has been changed to "terrain features"

L468 « is…was » unify tenses ?

Corrected

L483 « a SnowEx pit in two hours » give an estimate of the depth of the pit as it is has a major effect on required time.

Revised to clarify a 1 m deep snow pit. "For example, a team of two can fully sample a one-metre-deep SnowEx pit in two hours…"

L501 « that snow pit observations are independent and unable to resolve spatial patterns < 150 m scale » this a result of the sampling strategy, not of the snow pit approach in itself.

Revised for clarity as, "…snow pit observations following the SnowEx 2020 Grand Mesa IOP sampling strategy are independent and unable to resolve spatial patterns < 150 m scale."

Fig 2. State in the legend that the colorbar is centered on the mean value. Idem in similar figures.

Corrected

Fig 5. Interesting, more variability in depth and SWE in forest compared to open, but less in density. Could be worth commenting in 3.4 ? Could it result from relative importance of wind transport, canopy interception… ?

Fig. B2. The yellow histogramm is hardly visible. Maybe make it darker ?

Corrected

Fig S8. Did you filter out negative values ? If so, state it in the methods, if not use a colorscale allowing negative values.

Revised the Text in Section 2.4.2, "Then, we applied the point cloud differencing method to estimate snow depth on a 1 m grid (Appendix B.1). Negative snow depth values were filtered as no data values."