

Letter of Response to Reviewers

Dear Florent Dominé and Reviewers,

First, the authors would like to thank the reviewers for their constructive input which has greatly improved this manuscript. We have carefully considered our responses and revisions to best address the plan we mapped out initially after review. In this letter we outline the significant changes to the manuscript followed by our detailed responses to the reviewers.

We addressed the primary concerns, that the manuscript was unfocused, had a wandering discussion, and was too rich in methods and analysis throughout.

Our intent was to clarify the set-up of the manuscript which was noticed by all reviewers as muddled. Emphasis in the abstract and introduction was shifted toward how our research provides needed calibration and validation for potential air and spaceborne SWE retrievals and away from general snow remote sensing techniques. The introduction was cut by approximately 25 lines, where statements on snow density modeling and our novel aspects of the work were made concisely and impactfully.

Methods and details had been disseminated throughout the manuscript and were cleaned up in revision. Sections which we thought bogged the reader down were shifted to the appendix. We included a section in the appendix on the LiDAR point-cloud processing applied in our work. We greatly appreciated the attention reviewers gave to this aspect of the work, and we believe it will improve and educate audience understanding. Sections on the machine learning and random gaussian modelling were clarified, and a section describing SWE and uncertainty calculation was added.

It was pointed out that the results sections contained several statements of conjecture which were placed into the discussion or removed. This helped pare down the results which are now simplified and easier to read. A section on the sources of uncertainty was added to replace statements which were mistakenly placed in the methods. A supplement was created to support the various analyses presented in the text. The reviewers pointed out a significant mistake in our presentation of data validation results. We fixed this coding error and have provided scatter plots within the supplement as recommended. We also included comparisons to the ASO snow depth and SWE, as suggested.

The discussion wandered due to a lack of organization. We placed topics within specified subheadings finding many statements which were misplaced or could be improved in their explanation. We intended to cut the superfluous and condensed the discussion by about 100 lines without loss of the important information.

We appreciate the extended time offered to complete this revision, as this effort was a significant overhaul. We hope you all find this revision satisfactory, that no concern was left unaddressed. On behalf of all co-authors, we are pleased to submit this strengthened work to *The Cryosphere*.

Kind Regards,
Tate G. Meehan (corresponding author)

Responses to 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.

RC1

Line 24: How does this study represent the “next step towards broad-scale SWE retrieval”. How does broad-scale SWE retrieval depend on the methods / results shown here?

Revised to clarify the how, “Spatially continuous snow density and SWE estimated over approximately 16 km² may serve as necessary calibration and validation for stepping prospective remote sensing techniques toward broad-scale SWE retrieval.”

Introduction: There is dissonance between the initial setup of the study (which is about working toward the improvement of space-based remote sensing and talking about snowpack measurement and mapping at the local scale (e.g. which is measured by GPR, airborne lidar, drones, and field measurements).

The introduction has been tailored to speak to the methods of data acquisition and modelling, while focusing on how this is relevant for our study.

Line 34: I think this comment should be revised to “simple spatial interpolation from ground observations is difficult” - things like physiographic variables and past snowpack patterns can be useful for distributing snow measurements.

Revised as, “Snow varies over such short length scales that spatial interpolation from sparse ground observations is challenging to accurately achieve and validate.”

Line 39: While many of these technologies have, in some circumstances produced fairly accurate maps, there are many caveats and there are also many cases where they fail, or are severely biased.

Acknowledged

Line 42: I do not understand what is being said here. The first and second parts of this sentence do not seem to be related.

Revised via deletion

Line 81: This paragraph is long and disorganized. The topic sentence is about the spatial patterns of snow density being related to underlying processes, followed by a listing of (sometimes and sometimes not) spaceborne techniques that could be useful for snow density measurement.

This paragraph has been significantly revised, and broken into various focused paragraphs.

Line 89: If singling out ground temperature and roughness length, please explain why they are important in this application

Revised via deletion

Line 97: Drone based photogrammetry to get snow depth?

Yes, Clarified

What is the relationship between GPR measurements to get snow density and space based remote sensing? Can this technique be used to detect snow density using satellite data?

Revised throughout introduction to clarify our intent with statements such as, "Our research utilizes ground-penetrating radar (GPR), LiDAR, and ML to define an approach to map snow density at resolutions appropriate for air- and space-borne radar remote sensing calibration and validation."

Line 104: Rather than simply giving a high level overview of your methods, it would be nice to frame this more in terms of research questions, hypotheses, and knowledge gaps.

Revised introduction throughout to frame the knowledge gaps we address.

Line 140: Remove the word "whereas"

Accepted change.

Line 142-143: "the many forested stands in the survey domain" - this is pretty wordy. Suggest just saying "forested stands"

Accepted Change

Why wasn't the same type of instrument used in open and forested areas? How might this have affected results?

The two systems were required to reach within densely forest areas by ski, and to cover vast areas by snowmachine. I have clarified in the revision that the snowmobile could not travel into densely treed areas. I included a statement within GPR travel-time results (3.2) clarifying, "no systematic bias between the two GPR instruments was found."

Line 147: Was the same type GPS used on both GPR units (with lower accuracy in the forest)?

Clarified that a GPS receiver was used on ski, and a GNSS receiver was used on snowmachine.

Line 160: Were these processing steps the same for both GPR units

Included a clarification on the processing steps for the skied GPR system.

Line 213: "Time series" - this is pretty vague...and perhaps not even necessary since you are only using the Feb 1st flight

Removed time-series terminology.

Line 215: What is the principle of the Štroner et al method? It is probably better, in the interest of having the work stand on its own, to be somewhat descriptive of all the methods, rather than just referring readers to other papers.

Additional Details on the point cloud processing has been provided in Appendix B

Line 230: Change “the” to “in”

Accepted change

Line 250-254 - This passage is confusing. How were you able to determine the new random error? It also seems pretty generous to call half the sample (outside the 25th-75th percentiles) outliers.

See the explanation in the revised Appendix B.2. The percentage of data used to describe outliers is dependant on the degree of misalignment between the radar and lidar sensors. Following work, not presented here, applies the same technique with refinements to the geolocation of data sources which suggests that only bottom and top 10 % of data points need to be considered as outliers going forward. This represents the progression of the technique. See the response to referee 3 for histograms defining the outliers of raw density retrievals.

Line 270 - Can you be a little more descriptive of how the threshold was determined.

Revised, “Hourly air temperature data parameterized an empirical relationship to determine the threshold for snow-transportable wind speed (Li and Pomeroy, 1997).”

Line 293-301: This passage is confusing. Can it be rephrased / reduced?

This passage has been clarified and reduced within section 2.4.4

Line 340: Does this mean that the only snow depth differences correctly captured by the lidar was the difference between canopy and open areas. Could this be commented on further in the discussion section (e.g. could it be something related to geolocation errors, differences in the scale of measurements, relatively large lidar errors relative to the variability in either of those environments individually, etc)?

A mistake was discovered in the validation code, mixing the indices between open and forest areas, which caused this spurious decorrelation. See the response to Referee 2.

Line 350: Maybe list in the text what this correlation is

Correlations have been included as suggested, along with a reference to Table 2.b. within the new section 3.8.3

Line 351-353: “The standard deviation ...” - this sentence is awkwardly phrased

Revised as, “The $\rho_{s,Pit}$ and $\rho_{s,LiDAR-GPR}$ data are both normally distributed as evidenced by Z-test (Appendix B) with overlapping standard deviations.”

Line 354: Remove “uniquely”

Accepted change.

Line 368-369: “the relative length-scales of variability for SWE closely resemble that of TWT and indicate that TWT is a better informer of SWE than either depth or density, independently” - this is confusing. Is there a way to clarify why a smaller difference in length-scale makes TWT a better informer of SWE

Clarified “... better informer of SWE spatial patterns”

Line 378: “Average of regression models” - was there just one ANN, RF, and ML model, so that the average is just the average of 3 models, or were there, for instance, an ensemble of ANN models that were averaged together?

Methods section 2.6.1 was clarified as, “Regression models were trained on the LiDAR-GPR estimated snow density using cross-validation and were applied to the surrounding terrain. Model ensembles were generated and then averaged for both RF and ANN regressions by retraining on random subsets of data. A ML snow density ensemble was composed by averaging the MLR, RF, and ANN outputs. For detail on the parameter estimation and predictor importance see Appendix A.”

Line 405: To get SWE here, do you use average snow density from all pits, as well as the random field field generated from all field density data? It doesn't seem like the SWE comparisons are fair if some methodologies have access to the validation data (e.g. average snow density) and some do not (e.g. regression maps, as far as I can tell).

Section 2.7 was added to clarify these points.

Line 408-409: “depth is primary to SWE in this environment” - what does this mean?

This conjecture has been removed from the article.

Line 409-410: “Assessed at the average value of SWE for all 96 snow pits” - this is also unclear
The language has been improved throughout this section 3.7 Spatially distributed snow water equivalent.

Line 447: 2% - are the units of this supposed to be comparable to the previous number (1 kg/m³)?

Units have been changed to kg/m³ as suggested.

Line 449-450: It is hard to reconcile how the errors in SWE found using this methodology at this field site relate to a global SWE uncertainty target.

Statement has been clarified, “Our uncertainty analysis (Fig. S8) suggests that, even using innovatory measurements from airborne and ground-based sensors, 10 % uncertainty is difficult to achieve.”

Line 539: Could you please define the acronym “FMCW”?

This paragraph has been removed from the article discussion

Line 555: Remove comma after SWE. Also, I find this statement confusing.

Revised as, “Radar travel-time informed dry-snow SWE variability better than either depth or density independently.”

Line 562: Though by comparing SWE estimates at independent snow pits, it seems like you can surmise which method probably performs better.

What is vexing about this entire process is that the true SWE remains unknown. We contend that using LiDAR & GPR, “[has] comparable uncertainty to in situ methods but with spatial continuity at resolutions practical for calibration or validation of air- or space-borne radar remote sensing retrievals of SWE.”

Line 567: Which retrieval methodologies are you referring to here?

This statement has been removed from the revised conclusions.

Line 590: I don’t think there is a regression learner toolbox. The Regression Learner app is part of the statistics and machine learning toolbox...though neural networks are not part of these tools. How did you implement the neural network models?

In recent MATLAB versions (I am using 2022a), neural networks are included in this Regression Learner application. The text now clarifies, “application” rather than toolbox.

RC2

L13 A bit surprising to focus here on spaceborne measurement while they are absent of the rest of the article.

Revised to emphasize snow mass estimations rather than spaceborne remote sensing, "Estimating snow mass in the mountains remains a major challenge for spaceborne remote sensing."

L13 "enables landscape-scale snow covered area" can "enable" be grammatically used like that?

This is acceptable grammar; however this has been removed in revision.

L18 "essential snow physics" is a word missing?

Revised as, "essential snow research applications".

L23 "The root-mean-square error between the distributed estimates" A doubt remains whether this RMSE is i) between modeled density and snowpit density measurement or ii) the variability of the various models.

Clarified, "The root-mean-square error between the distributed estimates and in situ observations..."

L24 maybe keep depth and SWE with the same unit, cm?

mm are the standard unit of SWE

L25 "Wind, terrain, and vegetation interactions display corroborated controls on bulk density that show model and observation agreement." I thought that one conclusion of the article was that snowpit are too sparse to sample the terrain variability.

Snowpit measurements are too sparse to corroborate this finding. Models referred to here are ML regression densities and wind factor parameters, observations include meteorological data.

L29 "declining of"?

Revised as, "snowpacks in the western U.S. declined ~ 20 %..."

L35 "from ground observations is not possible" this should be tempered.

Text Has been revised as, "...challenging to accurately achieve and validate"

L38 "instruments, which" delete the ",,"?

Accepted

L40 Replace WorldView with "high-resolution satellite stereo images" and maybe keep Marti et al., 2016 and McGrath et al., 2019.

Accepted

L42 LiDAR and photogrammetry

Accepted

L43 "whereas" I do not see opposition between the two parts of the sentence.

Replaced with "furthermore"

L51 "in space more significantly" => "more in space"

Accepted

L54: give the spatial variability scale of snow depth as well

The spatial variability of snow depth is given in the results of this manuscript.

L65 Similar, other works looked at converting snow depth to SWE, see for instance and the citations therein:

Thank you for this information.

L71 "more simple" => "simpler"?

Accepted

L116 "from mean density measured in snow pits with the airborne LiDAR snow depths"? not clear, please rephrase.

Revised, "We evaluated the degree of improvement in SWE at the sub-catchment scale by comparing SWE distributed using airborne LiDAR snow depths and density estimated from regression techniques, random-gaussian field synthesis, and the mean density measured in snow pits."

L131 Cite some of these works.

Included the following references.

*Boyd, D. R., Alam, A. M., Kurum, M., Gurbuz, A. C., & Osmanoglu, B. (2022). Preliminary Snow Water Equivalent Retrieval of SnowEX20 Swesarr Data. International Geoscience and Remote Sensing Symposium (IGARSS), 2022-July, 3927–3930.
<https://doi.org/10.1109/IGARSS46834.2022.9883412>*

*Singh, S., Durand, M., Kim, E., Pan, J., Kang, D. H., & Barros, A. P. (2023). A Physical-Statistical Retrieval Framework to Estimate SWE from X and Ku-Band SAR Observations. International Geoscience and Remote Sensing Symposium (IGARSS), 2023-July, 17–20.
<https://doi.org/10.1109/IGARSS52108.2023.10281838>*

L202 Please repeat here the IOP date.

The dates for the IOP have been reiterated at the beginning of the paragraph, "Snow pit observations and manual depth probe measurements were collected throughout the 27 January – 12 February 2020 IOP..."

L210 "the local distance between two point clouds" This sounds like a significant difference with most snow depth studies (e.g. ASO Painter et al., 2016) in which the snow depth is calculated along the vertical direction as the difference between gridded products. I think it should introduce a discrepancy as well with the snow depth measured in snowpits.

This technique has been clarified within Appendix B.1

L219 "lidar" written LiDAR elsewhere
Corrected

L219 Could the statistics of the elevation difference over stable terrain (snow free, unchanged) be used to estimate the uncertainty?

This technique is typically accomplished over plowed roads, where snow depth change is expected to be zero. This result has been included in Section 3.8.1.

L224 "We applied a k-d tree searcher (Bentley, 1975) to co-register the LiDAR coordinates within a 1 m radius of the GPR TWTs" I understand co-registration as moving one the elevation dataset for instance with a translation vector (X,Y,Z). Is it the case here? If so, can you cite an article using this k-d tree searcher algorithm? It does not seem very common.

The kd-tree is simply a fast algorithm for finding coordinate points within a defined radius of a set of query points.

L226 "By calculating the maximum cross-correlation lag on continuous segments of transect data" not clear what this operation is.

The statement has been clarified as, "By calculating the lag distance of the maximum cross-correlation between GPR TWT and LiDAR snow depth on continuous segments of transect data"

L239 2.4.3 Error and uncertainty are described at various places of the article (L247, L416- 417, L443). This is confusing. Please reorganize it.

Uncertainties for the various measurements and results have been placed within a new Section 3.8.

L249 "are spatially uncorrelated suggesting that the errors are random and can be treated with filtering." =>"are spatially uncorrelated and can be treated with filtering." "filtering" is quite vague, not any filter would be adapted.

Revised as, "errors are spatially uncorrelated and can be treated with statistical filtering"

L252 "we reduced the random error to ± 30 kg/m³ by filtering outliers." Not clear whether you optimized filtering until this error level was reached or if it is a fortunate consequence.

Revised as, "Via median filtering and interpolating through outliers error estimates were reduced to 30 kg/m³ (Appendix B.2)."

L256 When was the last snowfall prior the IOP?

Certainly, new snowfall would introduce greater discrepancy. Regardless of any new precipitation, it is necessary to consider changes in snowpack height due to densification and wind redistribution, which we have attempted to account for.

L312-318 This paragraph does not match the title of this section. It could be move in 2.6, in a section gathering all the spatialisation methods used.

This paragraph has been moved to a new subsection 2.6.2 Random Gaussian Field Model

L327 What is the spatial resolution of the density maps? All this smoothing must reduce the actual resolution of the maps.

The density maps are posted with a 1 m pixel size.

L340 "The in situ snow depth observations (hs,Pit and hs,Probe) compare well with the LiDAR snow depths throughout the entire domain (R² = 0.61, RMSE = 11 cm, ME = 0 cm), however within the open and forested domains individually, LiDAR and GPR estimated snow depths are uncorrelated with in situ snow depths (Table 1)." This is puzzling. Can a set of points be correlated while two subsets including all the points are not correlated? A simple scatter plot, even in supplement, would help get a better feeling of the data agreement.

This is an important comment, thank you. A mistake was found in matching the indices of raster pixels to those at probed locations when splitting the data into forested and open areas for validation. Scatter plots have been provided in the supplement, and the tables have been updated with the corrected values.

L346 "The GPR data were acquired within a few metres of, but not directly beside the faces of snow pits, which necessitated a radius for pairing observations. The choice of 12.5 m matches that of the filtering during the sensor integration step (Section 2.4.3)" This should be in method, not results.

This information has been placed in methods Section 2.2

L348 "Measurements accumulated over 12.5 m distance introduce inherent variability on the order of 10 % (Section 3.3), which, along with differences between representative observation scales may explain the weak correlation between estimated density and the in situ measurements." This should go in discussion.

Information regarding the spatial variability has been placed in a new section 3.8.3. Speculation regarding the correlation has been placed in the discussion as recommended.

L349 "weak correlation" sounds contradictory with the 13% RMSE mentioned which sounds good.

Moving deliberation on weak correlation to the discussion has removed the seemingly contradictory connotation of this statement.

L353 (and L383) Isn't there strong correlations between S_x , the Wind Factor and the topographic predictors used to estimate the densities? In that case, isn't it expected, by construction, that the densities and the wind proxies will be correlated?

While true that topographic predictors are used to estimate the density, we see this correlation as well in the GPR-LiDAR measured density. The appearance of snow density features in the machine learning models resembles the upwind slope patterns most closely in the direction of the prevailing wind. This suggests that the terrain features in a particular aspect, here 200-220, are the most causal of the GPR-LiDAR measured patterns and those predicted by the machine learning model ensemble.

L355 "**suggests** that the method retrieves density patterns which are related to the degree of exposure and shelter due to topography and vegetation" => to discussion

Revised, as recommended within the discussion.

L364 "the expected variability among co-located ρ is approximately 2 %," Based on which semi-variogram? Not clear. Again, this sounds more like discussion to me.

This will be clarified, as either variogram presents a similar result.

L367 "to resolve spatial patterns" => "to resolve spatial patterns at finer scale." There might be spatial pattern of larger scale than 100 m.

Revised as recommended

L368 "depth and density **formulate** TWT and SWE," formulate is not clear. By definition, the SWE is the product of depth and density.

It is too nebulous for the context to clearly describe how TWT is dependent on snow depth and density. The comprehension that both SWE and TWT are related by depth and density is sufficient. The word choice "relate" has replaced "formulate".

L370 3.4. Please provide metrics comparing the density modeled and measured (mean, median bias, standard deviation, RMSE...).

Statistical metrics comparing the measured and modeled densities are provided in Table 2.

L376 "This coincided" what is "this"?

"This" has been replaced with "The appropriate hyperparametrisation"

L378 "as the random field contains little meaningful spatial information" This is quite expected (idem L388). By construction it is spatially random. I am not sure what is the point of this randomly distributed density. Is it used in other studies?

The randomly distributed density is presented as an informed benchmark model which we hoped to outperform with the heavy lifting of machine learning. It provides in a visual sense how randomness appears as a snow field, which in turn bolsters the appearance of the spatial patterns supplied by regression modeling.

L379 "spreads the strengths" formulation unclear.

"Spreads" has been changed to "combines."

L391 "The lack of a large-scale topographic trend in density, such as one driven by elevation or aspect, evinces the role of forest vegetation on density." is this supported by a figure showing density against elevation or aspect? I don't understand the link between the two parts of the sentence, the elevation/aspect and the forest vegetation.

Revised, "The role of forest vegetation on snow density is evinced in this topographically simple environment because a large-scale topographic trend, such as one driven by elevation or aspect, is not saturating the signal in density."

L407 "with an appropriate correlation length and prior mean and spread but maintain a larger bias" too many "and", hard to understand.

Revised, "The ensemble-modelled densities explain more variation in the distributed SWE estimates than those distributed by random gaussian processes."

L409 "covariance exists" it always does, doesn't it? Do you mean that it is large or small?

Revised as, "The errors among snow depth and density are uncorrelated ($R^2 = 0.03$) with negligible covariance."

L410 "has a negative bias of approximately 7 %," hard to compare with the other bias mentioned above which is in kg m⁻².

Revised to include units of SWE, "bias of -20 mm (7%)"

L414 "R2 = 0.16" seems like weak correlation. What is the R2 of "the errors among snow depth and density are uncorrelated"?

R2 values included, "uncorrelated ($R^2 = 0.03$)"

L417 "Following from the propagation of errors for relative errors in snow depth and density, we estimated the SWE uncertainty to first order (Raleigh & Small, 2017)." this should be in methods and better explain.

This has been described in a new methods section 2.7

L418 "The distributed relative SWE uncertainty is presented in Fig. 10 and is negatively correlated with the distributed SWE ($R^2 = 0.44$)." I understand this as: the larger the SWE, the smaller the error. But I understand that the inverse was true for snow depth and density (L414-415). Sounds inconsistent.

The attention to statistical detail in your review has been very helpful. You are correct that these statistics were inconsistent. After doubling back into the analysis code, it was found that both depth and density were negatively correlated with the measurements of 96 the snow pits within the study domain. LiDAR depth errors calculated among the ~30,000 validation depth measurements revealed that the correlation among snow depths was spurious. Snow depth was found to be uncorrelated with the errors ($R^2 = 0.04$). Using simple linear regression, we modelled the snow density errors as a function the machine learning regression ensemble density. Using the RMSE (11 cm) for relative errors in snow depth, linear errors in density, and linear error propagation we estimated the SWE uncertainty to first order (Raleigh & Small, 2017). These errors are presented within the supplement (Fig. S9).

L435 "Sensitivity analysis" how does this articulate with the SWE uncertainty in 3.7. I understand that the uncertainty is estimated in two ways: by comparison with the snowpits values and with the sensitivity analysis. This should be clarified in the relevant parts (methods, results) and evaluated.

Section 3.8 has been dedicated to uncertainties, and their relevance in our analysis.

L440 "the leading source of error in our density measurement is spatial misalignments" but further, the impact is estimated to be of 1 kg/m², that is way less than the 15 kg/m² reported or, I guess, the difference with the snow pits.

Thank you for the insight. The takeaway is that the roughly 30 kg/m³ error is baked into this analysis, due to the data misalignments. Further perturbations of data alignment led to only a small increase in error.

L441 It is not clear what this cross-correlation is exactly but if it provides the misalignment value, could it not be used to correct it?

The value of misalignment is not spatially uniform, attempts were made to correct this, however, I determined it best practice to accept this error and inform the readers of the impact of such misalignments.

L445 What does integration mean?
Combining GPR and LiDAR sensor data.

L447 "up to 2 %" use consistent units, kg m⁻² in the same sentence.
Corrected 2 % to 5 kg/m³

L455 "*the choice of uncertainties propagated*" it makes it sound like these choices are purely subjectives and not well informed.

Revised as, "Uncertainty propagated"

L456 "*uncertainty in midwinter tends to reduce at peak SWE*" cite a relevant source for this statement.
This conjecture has been removed from the article.

L457 "*remarkably difficult*" does the mean density of several snowpits work as well (L475)? If so, the remarkable difficulty is rather to spatialized the density.

Subjective language has been removed, "Our findings (Fig. 9) are within the 10 % goal"

L465 "*improved ...importance*" increased?

Revised, "increased LiDAR feature predictive power."

L470 "*can be treated as random normal*" Say clearly if it has a normal distribution or if it is assumed.
Revised," appears as random normal variable."

Fig 3, 4, 5, 7: Divergent color map (red, yellow, blue) is not really appropriated for variables which are always positive. It is traditionnaly used for variables centered on zero (yellow).
Our usage of the divergent colormap is to show anomalous areas, centered around the mean value.

Fig 3. Could be moved to supplement.
Placed within Appendix B

It should be considered to put in a single figure (with adapted colorbar) the snow depth,

the density and the swe map. It is hard to identify the patterns similarity as long as they are spreaded over different figures.

Thank you for the suggestion, a figure showing depth, density, and SWE side-by-side has been included in the supplement (Fig. S7).

Fig 7. This could be moved to supplement and rather show the density map used most in the article (mean of the models).

Accepted, Fig. 6 is now the mean of ML ensemble density.

Figure B1 what are the grey bars? I would move this figure to main manuscript.

Bars are the histogram which was then fit with a kernel smoothed density estimator.

Figure C1. Could be merged with Fig 8.

Figure C1 has been moved into the Supplement.

RC3

Scope: This manuscript covers a lot of ground, what are the specific questions that are being addressed and how does the analysis presented speak to these questions?

Our manuscript presents a large portion of material to develop the technology which allows us to ask questions regarding the snow cover. The narrative of the document, perceived as wandering, shows the natural progression of my years of research.

First, Large data must be processed and curated. We ask how this can be done most effectively, and we answer that question by developing an automated radar processing technology and utilizing open-source software to reduce an accurate lidar snow depth raster from point cloud data. Then, we must show that LiDAR and GPR can be successfully fused to retrieve snow density. The question we ask, “can this be accomplished,” as this technique had not previously been completed at the time I began the research and manuscript.

Questions arise in these “engineering” steps, such as is there a signal buried in the noise? We answer that question, yes and yield error bars. When an appropriate filter design is applied, a pattern appears along the radar transects. We make a significant discovery! What is causing the observed spatial patterns of snow density? What are the correlation lengths of these snow features? And, how can we best distribute this new snow density information? We answered the latter, by providing empirical methods using statistical models and machine learning, to best represent, what we believe are meaningful spatial patterns in the data.

Now, we may begin to ask ourselves scientific questions. What physical processes might be driving the patterns we see? How can we infer such physics processes from empirical data? From literature, we pose the wind, terrain, and vegetation hypotheses, and we answer thoroughly, by determining which of our hypothesized terrain and vegetation parameters are most responsible for these patterns and their model dependencies. And we supplement this result to ascertain the impact of wind on snow density patterns, independently. We confirm that wind, in a particular direction, interacts with the terrain and trees of Grand Mesa to explain a significant portion of the pattern variance, both observed by LiDAR—GPR and modeled using terrain features.

We learned a lot in this process, it took a lot of heart, and we are wishing to share this entire story with the cryosphere. We intended to have made this story more clear and concise through this revision, despite how long and winding this research has been, and hope this effort shows.

Comparison with existing techniques and demonstration of improvement: the authors miss a fruitful opportunity to compare their SWE map with other spatially distributed estimates of SWE (such as those prepared explicitly for the NASA SnowEx campaign by the Airborne Snow Observatory(ies), and potentially other SWE maps from instrumentation deployed as a part of the SnowEx 2020 effort. While I understand the spatial resolution of 1m is quite different from the 50m ASO products, the sum of averages should allow a meaningful comparison and contrast and is an obvious omission from the work (in my opinion).

The authors believe such a data comparison is an important contribution, that had been stifled by data availability. This healthy competition allowed the opportunity to contribute our own lidar snow depth data. The comparison between our data and ASO data has been analyzed in this revision. The results provide insights to the degree of information lost as a function of resolution and lends credence to the accuracy of point cloud processing.

Snow depth map: I am wondering why the authors chose to generate their own snow depth map from the point cloud data using Multiscale Model to Model Cloud Compare (M3C2) rather than just use one of the official SnowEx products from the same lidar data produced by Airborne Snow Observatory(ies). The M3C2 method is really targeted toward benefits over rasterized data in rough and complex terrain, which is far from the flat plateau of the Grand Mesa. This also seems like a missed opportunity to connect the current research/work to existing frameworks within the snow community – for instance, it would be useful to know how these products compare and if M3C2 provides clear benefits (as raster-based subtractions for snow depth are widely used by the community).

Thank you for your suggestion. At the project's initiation, there was no existing snow depth map for February 1, 2020, and the ASO (Airborne Snow Observatory) data had not been published. Within the revised appendix, I have provided additional details about our methodology in response to the previous comment. It's worth noting that the M3C2 method could prove beneficial for flat areas, given its effectiveness on both rough and smooth surfaces.

Snow density inconsistency: Line 340 "The LiDAR–GPR inferred average snow density shows ... and greater variability in the open terrain than areas sheltered by forest canopies." I would have expected the opposite, the snow pit measurements do not support this statement (from what I can tell) and, given the MLR/RF/ANN models were trained using the Lidar-GPR densities (with high variability in the exposed areas, it is no surprise that the MLR/RF/ANN models are also disagreeing with the snow pit measurements w.r.t spatial variability of bulk snow density).

Perhaps a brief explanation on this inconsistency is required.

Snow pit data suggest that the standard deviation of snow density is approximately equal between open ($\pm 19 \text{ kg/m}^3$) and forested areas ($\pm 20 \text{ kg/m}^3$). As we have stated, the LiDAR—GPR derived snow density shows higher variability in open areas ($\pm 32 \text{ kg/m}^3$) than in forested areas

(± 19 kg/m³). Modeled results, however variable on model archetype, also tend to show greater variability in the open yet within the same rough order of magnitude as measured in snow pits. We cannot rule out the possibility that snow pit locations incurred a bias due to sampling strategy, nor that the opposite isn't true of the GPR sample locations. We can confidently state, by the amount N fold data observations, that the LiDAR—GPR retrieved density over a greater variety of terrain, and possible variability, within the study domain on Grand Mesa. We also cannot rule out that data inconsistency isn't driven by inherent spatial variability on length scales beneath our threshold of observation due to spatial filtering (10 % variability at 25 m) and that of snow pit measurements (2.5 % nugget variability).

Broader appeal: I am wondering if there were other SnowEx locations that had GPR data to replicate this study?

Bonnell et al. (2023) uses similar methods for deriving snow density along GPR transects.

Line edits

Line 225 - "We found that errors in the co-registration of these data are the leading source of error in the estimated densities." What is the magnitude of this error?

Errors in retrieved density are initially on the order 150 kg/m³ this error was filtered appropriately which reduced to the random error to approximately 30 kg/m³. This information has been re-presented in Appendix B.2.

Line 245 - "At the 1-sigma level, errors of approximately ± 150 kg/m³ can be expected from this sensor integration method". This magnitude of deviation in bulk snow is of course large. I am wondering how the filtering on this would go without "SnowEx level" (high quality and many of) in-situ measurements?

In-situ measurements were used to validate the LiDAR snow depths and derive appropriate error bars of 10 cm therefrom. We also estimated the uncertainty in LiDAR snow depth by maximum standard deviation within the point cloud to raster processing. Uncertainties in GPR travel-times were also derived inherently. Filter design was approached using the range in snow pit densities, however other, more intuitive, choices could have been made. This analysis could be run without the use of in-situ measurements.

Line 250 – The Grand Mesa is really flat, I'm wondering what magnitude of "geolocation errors in LiDAR-derived snow depths" you are expecting here?

We expect negligible geolocation errors in the LiDAR product, the intent of this sentence is to convey the fusion of these two data sets, which has been revised. "The combined measurement errors in LiDAR-derived snow depths and GPR TWTs and errors in geolocation upon integrating these data may translate to errors in the retrieved density."

Line 289 - "pixels within a 3 m buffer of LiDAR vegetation greater than 0.5 m height". With snow depths of ~ 1m across the Grand Mesa, are you also including vegetation that was buried?

Yes, we include vegetation that is buried by snow cover. We are using the vegetation height segmentation from the September 26, 2016 ASO snow free product. Using a threshold of these

data to classify vegetated areas, however, ignores the effect snow cover has on vegetation height.

Line 210: For clarification, the February 2020 flight was actually collected 1-2 Feb (over 2 days)
Thank you for the clarification, this has been revised where appropriate throughout.

Line 215: I'm curious about the M3C2 method – were you seeing resulting snow depths of zero depth over the plowed roads?

We observed minimal scattered snow depth, ranging between 0 and 5 cm, along the main roads, while higher snow depth (above 1 m) was detected at the edges of the roads (Figure 1). These values were excluded from the final map through masking. This information has been provided in the revision within Section 3.8.1.

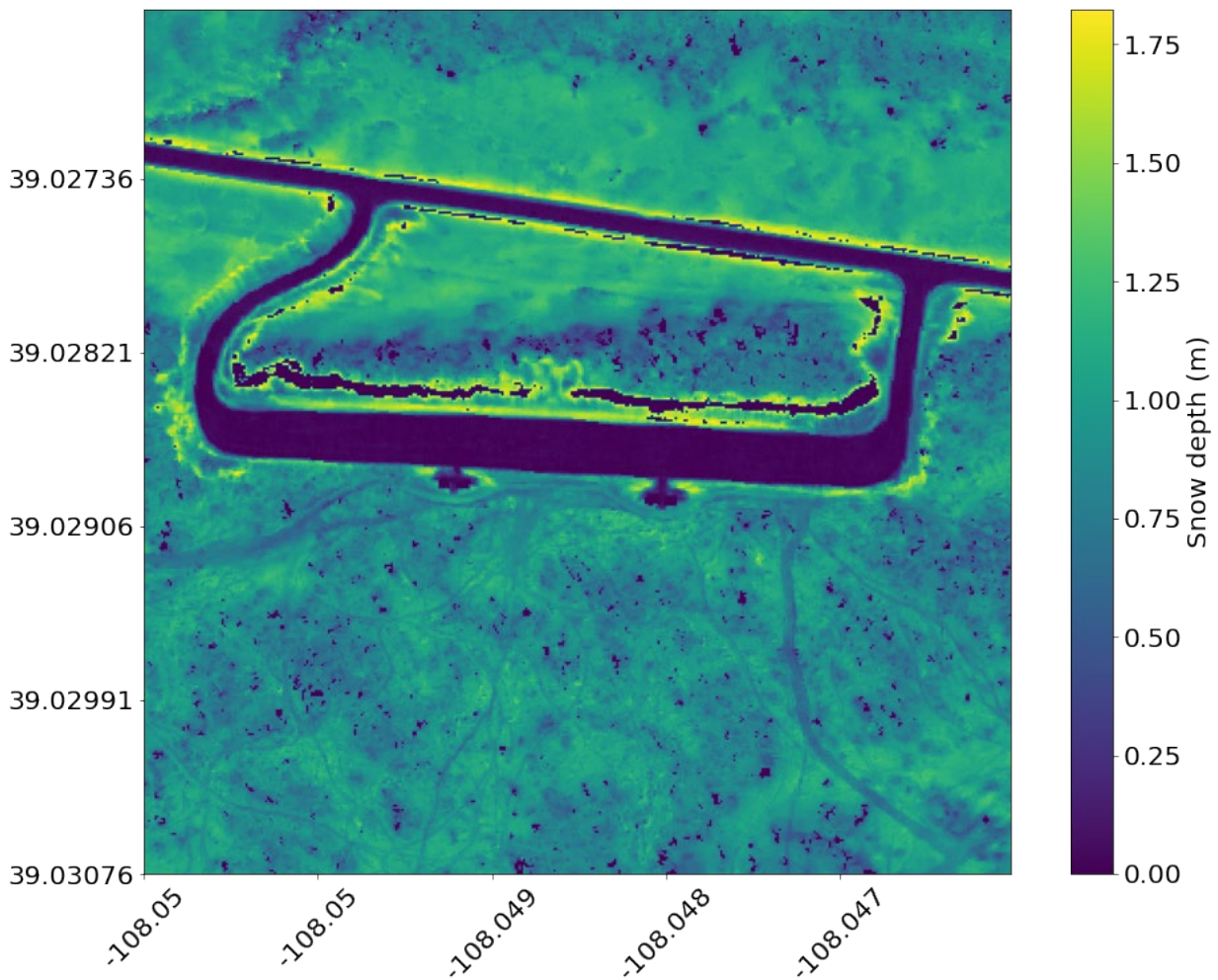


Figure 1. Snow map of a random road across Grand Mesa. The snow depth values on the road are between 0 and 5 cm.

Line 225: given "spatial registration between the LiDAR and GPR varies on the order of a few metres", and the high spatial variability of snow depth, does it make sense to use 1m snow depths to invert the GPR information? In Section 2.4.3.1 you do an error analysis and perturb

TWT and snow depth - of the two perturbations which one injected the most uncertainty into resulting snow density estimate? On this, I really like Figure 3.

For the range in uncertainties expected, we find that snow depth and travel-time both contribute approximately equally to the resulting density. Using a window of 1 m when determining GPR cross-over location errors, substantiates the errors expected by to be encountered during data integration. The use of a 3 m LiDAR snow depth results in a similar density retrieval (Figure 2). The 1 m snow density has a mean and standard deviation of $276 \pm 179 \text{ kg/m}^3$ where the 3 m snow density is $268 \pm 175 \text{ kg/m}^3$.

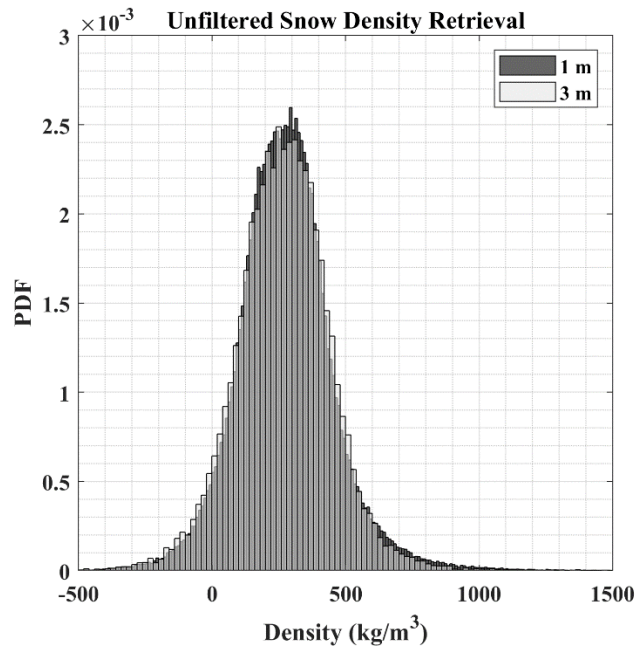


Figure 2: The unfiltered retrieved snow density using a 1 m LiDAR raster and a 3 m LiDAR raster.

Line 255: what was the magnitude of the trends here? "Separate linear trends were identified in the forested and open regions traversed with the GPR"

The trend has a range of 51 kg/m^3 with the minimum value -36 kg/m^3 and maximum 15 kg/m^3 .

Line 265: "1 October 2019 through the end of the SnowEx IOP on 12 February 2020" I'm a bit confused as to why you would include 3-12 Feb in your mean wind direction and speed calculations if you are trying to get at snow drifting ON 1-2 Feb. Perhaps I mis-read here.

It seemed more useful for the authors to present the wind data through the duration of the Grand Mesa IOP to the community. GPR data acquired on Feb. 5 was used in this effort. The Median Wind Speed Oct. 1 – Feb.1 is 202 degrees. The median wind speed Oct.1 – Feb. 12 is 200 degrees. The addition of these data does not impact the analysis.

Line 270: I thought Winstral et al., 2002 was based on the baseline snow off DEM. Why use snow on terrain? "we computed two parameters (maximum upwind slope and wind factor parameters; Winstral et al., 2002) from the 1 February lidar-derived snow surface elevations."

It seemed intuitive to me to use a snow surface, rather than bare earth, because we are using this information as validation of retrieved snow properties. The time history of snow density patterns we hypothesized to be driven by snow surface expression after a period of snow accumulation. The difference in results is negligible.

Line 285: Wouldn't some of that 0.5m high vegetation be buried in 1m of snow, and therefore not assert a drift influence? "3 m buffer of LiDAR vegetation greater than 0.5 m height"

This is good reasoning to use the snow surface elevation model in these calculations.

Line 330: The Regression models were used to "fill-the-gaps" in the density field provided by the GPR. I'm curious as to what these maps look like.

It is unclear what is being asked by this comment. The regression modelled density is presented in Figure 6, while the Lidar—GPR derived densities are presented in Figure 5.

Line 340: "however within the open and forested domains individually, LiDAR and GPR estimated snow depths are uncorrelated with in situ snow depths (Table 1)." Shouldn't they be correlated somewhat? Why are they uncorrelated with R² of 0? what do the scatter plots look like?

This was a helpful comment which drew our attention to a mistake in the tracking of indices between validation data in forest and open areas. The text has been corrected in Section 3.2 and Table 1 which overviews these data.

Line 350: "may explain the weak correlation between estimated density and the in situ measurements" Is the magnitude of the un-correlation around 10%?

Yes, for an average snow density of 275 kg/m³, a 10 % error would be 27.5 kg/m³ which agrees exactly with the RMSE of the snow pit validation.

Line 405: Consider revising this sentence "which suggests that depth is primary to SWE in this environment"

Revised. "Using an average measured value performs slightly better."