Manuscript Number: Preprint egusphere-2024-3545
Title: Leveraging Citizen Science, LiDAR, and Machine Learning for Snow Depth Estimation in Complex Terrain Environments
Subject: Response to Manuscript Review Comments

Thank for your thoughtful and constructive feedback on our manuscript. We appreciate the time and effort invested in reviewing our work and have implemented several revisions to enhance the clarity, transparency, and accuracy. We hope the revisions adequately address the comments and remain open to further suggestions or questions.

Response to Reviewer #1:

The study tried to characterize and model the snow distribution using the Gaussian Process Regression (GPR) method and the Gaussian-based machine learning model (GMM). GMM and GPR method application for the Lidar observed snow distribution seems novel although they were tested using only one snapshot snow distribution. The land surface characterization in Figure 2 is nice, and the finding of elevation variability requirement for this method is interesting. However, since the transferability of this method may still be arguable, I recommend "major revision" for this review cycle for clarifications and further possible improvements. I have a few major points listed below:

Comment 1:

The presentation of the data used in this study should be improved. The observed snow distribution by the airborne Lidar may be visualized and presented somewhere in the manuscript, perhaps instead of Figure 3. It will be informative for readers to see the variability and the extent of the dataset.

Response to comment 1:

Thank you for your comment. Figure 3 (shown below) has been replaced with a map of the snow-on Lidar, as well as the distribution. Figures 7 and 8 (below) have also been updated to show the lidar snow depth and depth estimate maps along with the error maps. Note the scale of some figures have been shrunk to fit in this document.

New Figure 3:

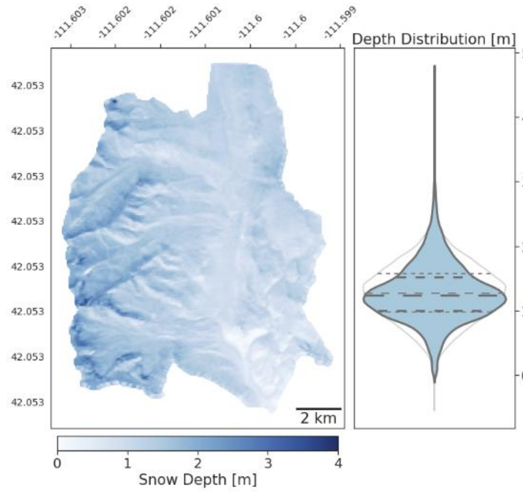


Figure 3. Map and corresponding density distribution of the LiDAR-measured snow depth in Franklin Basin. The violin distribution of the snow depth is overlayed above a Gaussian distribution of 20000 randomly generated samples to express the near-normality of the snow depth distribution. Quartiles of the snow depth are shown as bold hashed lines, and the light grey hashed lines represent the Gaussian distribution quartiles.

New Figure 7:

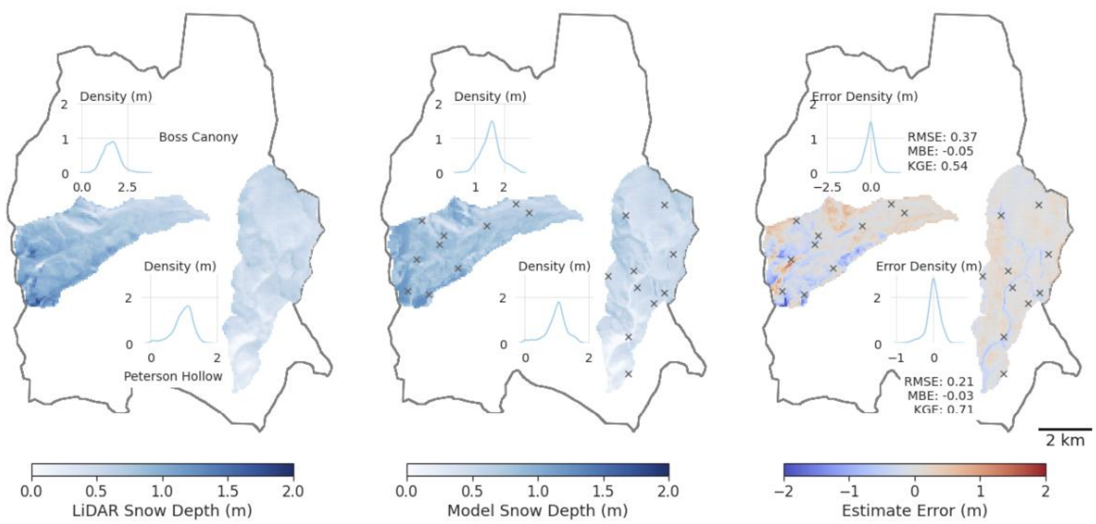


Figure 7. Subbasin Lidar snow depth, estimated snow depth, and estimation error for Boss Canyon and Peterson Hollow for 10 model-identified optimal sampling locations (x) and associated accuracy metrics.

New Figure 8:

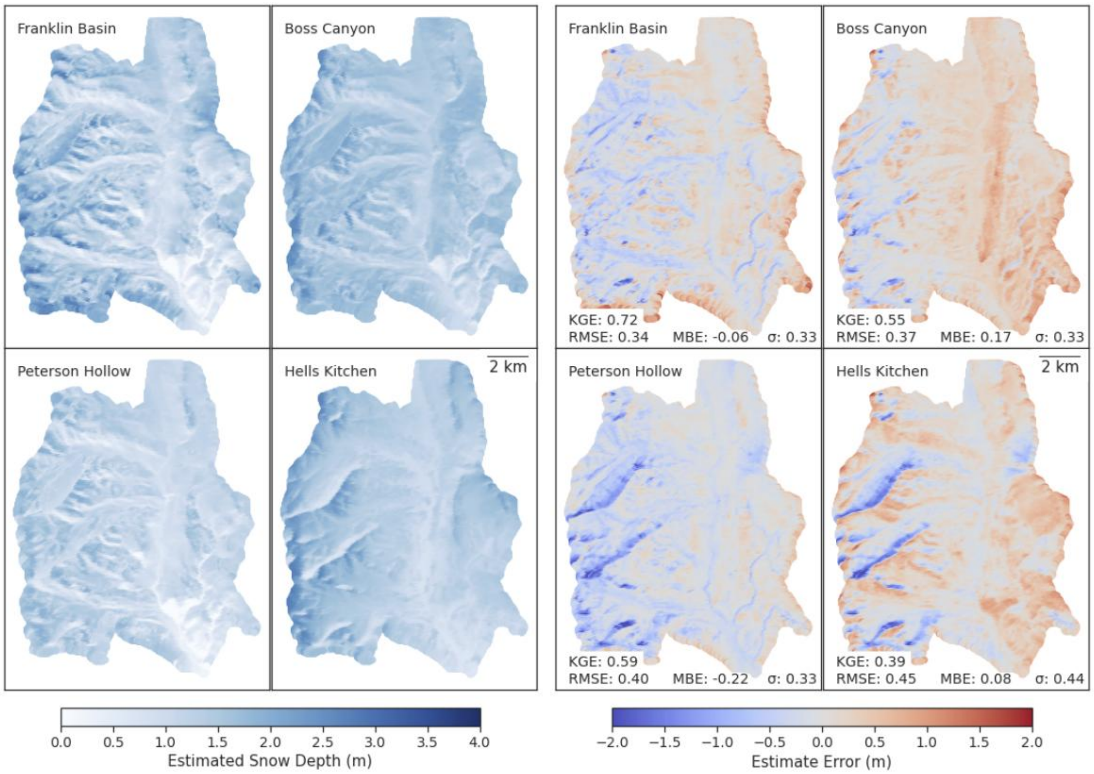


Figure 8. Basin-scale snow depth estimates (left), and snow depth estimate errors (right) with associated accuracy metrics derived from the sampling domains of: Franklin Basin (top left), Boss Canyon (top right), Peterson Hollow (bottom left), and Hell's Kitchen Canyon (bottom right). The estimates were derived with 10 GMM-identified, optimal sampling locations outside of avalanche-prone terrain to train the GPR model.

Comment 2:

Also, it is unclear when the LiDAR data collected. The snow distributions are highly dependent on season and year. From the snow distributions (Figure 9), I speculate that it must be late spring. Moreover, observation dates of the in-situ snow depth survey must be presented as well. Were they exactly same day? How good were they? Were they (field data vs. Lidar) consistent each other? It is unclear how the authors use actual field measured data. I would suggest adding a data list table.

Response to comment 2:

We apologize for the confusion of the data metadata. The snow-on LiDAR and snow surveys were all performed on the same day in late spring (Mar 28, 2021). Text has been added to Sections 2.2 and 2.4 to address this, as well as a table provided of the data used in the study.

The field measured snow depths were used as the training data for the Hell's Kitchen Canyon subbasin model. This was the only basin where we performed sampling, and it was not covered by the lidar flight pattern. The samplers followed a rigorous protocol to reduce errors in measurements, however because the lidar data does not cover the sampling locations, we are unable to directly compare field samples to lidar. Text has been added to section 2.4 to clarify

how the field samples were used, and to the Discussion section addresses the potential for error and uncertainty from the lidar data.

Comment 3:

The assumption for the methodology must be further clarified. Based on my understanding, Gaussianity in local snow distribution is required while it may not be true. I recall a recent publication in the same journal (TC) discussing non-Gaussianity of snow distribution (https://tc.copernicus.org/articles/18/5139/2024/). Assumption of local Gaussianity may be addressed in the limitation statement in the discussion as a reminder.

Response to comment 3:

Thank you for the comment and the article recommendation. The GPR does rely on the target variable (snow depth) to be of a Gaussian distribution, and can lose robustness particularly for large outliers or heavy skewness. We justify the use of GPR in our study area based on the near normality of the LiDAR snow depth. We have added a figure of the LiDAR snow depth distribution (Figure 3) to Section 2.2 and provided statistical justification for the assumption. Text has been added to the Discussion section to expand on the Gaussian assumption, and to explain this methodology may suffer when snow depth does not follow this assumption.

Comment 4:

I understand that there was no improvement by increasing sample number from 10 to100. It would be more useful if the author could quantify the ideal snow data point density (for instance, # of data point per unit area, perhaps). I understand it may be beyond scope of this study while lacking statement on potential transferability made this work just a case study based on single instantaneous snow distribution, which is rather weak.

Response to comment 4:

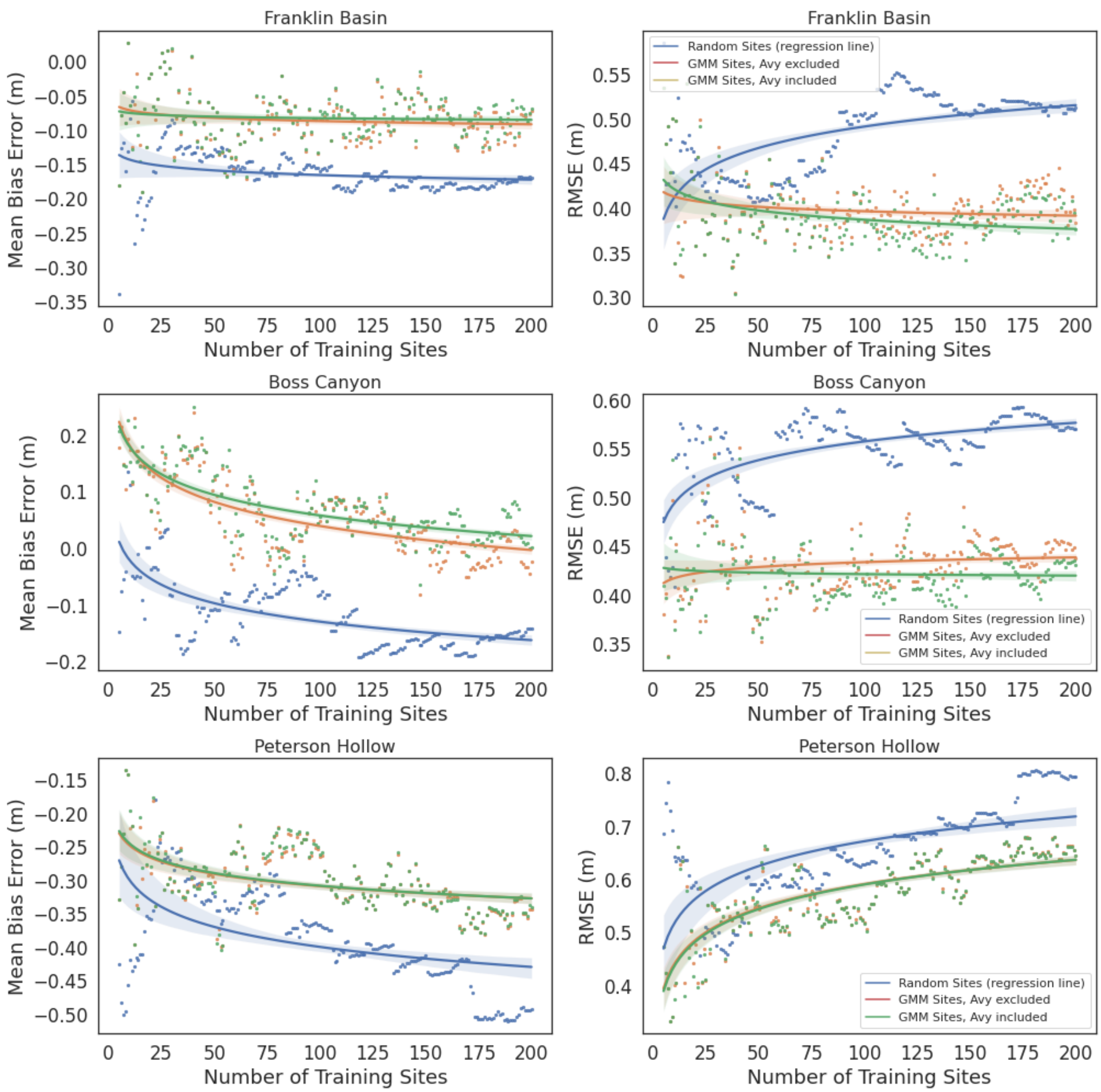
Thank you for the comment. We agree the optimal number of sites is an interesting question, however we see it as different from the focus of this study. We aim to show that a small number (reasonable to be sampled in one day) of samples can effectively model a broader region and a large sampling campaign or dataset is not required. This is why we highlight only the 10 vs 100 site results (Table 2).

We have compiled results of the site sensitivity analysis in the supplemental materials (Figure S1).

We hypothesize the optimal number of sites would be highly dependent on basin size, region, terrain etc. and we did not have the snow LiDAR availability to test a large number of basins. We saw that there is large variation in the individual point-to-point increase of a small training dataset in the 4 basins analyzed, but the trend across all basins is similar; that a large number of points does not greatly improve performance. This can be seen in Figure S1.

We have included more text on this in the discussion, however we chose to provide the analysis results as supplemental materials as not to shift the focus of the study.

Supplemental Figure S1:



Comment 5:

It is good to define the variables in the equations (2 through 4) as physical quantities (e.g. x = snow depth). Capital sigma (=covariance?) may be avoided because you use “summation” as same symbol.

Response to comment 5:

Per the suggestion, we have added the physical property to the variable description, as well as using lower case sigma in equations 2-4 instead of capital sigma to avoid confusion.