The missing drifts: Widespread and systematic underestimation of heterogeneity in mountain snow density and SWE across scales

Elijah N. Boardman, Karen L. Boardman, Christopher A. Jones, Sean D. Shipman, John A. Whiting, Joseph W. Boardman, Adrian A. Harpold

Response to Reviewer Comment RC1

We appreciate the reviewer's time commitment to engaging with the publication process.

However, we feel that the reviewer has severely misconstrued the scope, novelty, and relevance of our study, to the detriment of the potential TC readership.

In this response, we clarify numerous aspects of the study that have not been demonstrated in the prior literature, and we demonstrate how the reviewer overlooked these components of the manuscript. We appreciate that the novelty of our study needs to be further detailed throughout the manuscript to better guide the reader's understanding, and we intend to revise the title and manuscript language for greater clarity on the points that are novel.

Although many studies have addressed "the missing heterogeneity problem" (reviewer quote) for snow in general, our rigorous Bayesian statistical framework is the first to directly quantify the (non)representativeness of existing snow <u>density</u> data in a comparable region of the Rocky Mountains. No other studies have evaluated SNOTEL <u>density</u> representativeness despite widespread reliance on SNOTEL density data for real-time SWE estimation throughout the U.S. Rocky Mountain region. The reviewer only discusses (and cites) studies addressing snowpack heterogeneity broadly, not snow <u>density</u> heterogeneity specifically, which is a unique strength of our study given our novel deep drift <u>density</u> profiles, high-elevation <u>density</u> profiles, etc.

In a revision, we plan to further emphasize <u>density</u> heterogeneity in the title and manuscript.

This study directly addresses what I would call the measurement and estimation problem: using 36 snow and repeat airborne lidar across the Wind River Range to show that widely used snow density and SWE datasets systematically underrepresent heterogeneity. A key finding is that complete profiling of a 5.9 m drift in a nivation hollow yields a bulk density of 585 kg/m³ — exceeding prior empirical bulk snow density model predictions by 14–35%. At the watershed scale, three near-real-time gridded SWE datasets underestimate total SWE storage in a heavily drifted 1 km² glacial cirque basin by 65–73%, despite underestimating watershed-average SWE by only 2–17%.

The reviewer correctly identifies two findings of the paper. However, the reviewer fails to describe how we apply our findings to basin-scale sensitivity tests. Moreover, the reviewer fails to note the unique heavily wind-drifted context of our study region, which contributes to the novelty of our investigation into the representativeness of existing monitoring stations and gridded datasets. Reducing our findings to these two facts (one about the nivation hollow and one about the glacial cirque) ignores the multi-scale, multi-dataset, basin-scale components of our study and misconstrues the scope of our results (see subsequent bullet points).

The same research group recently published two other papers dealing with snow heterogeneity in the Wind River Range, Wyoming.

It is natural that a multi-million dollar, multi-year research campaign in a novel geographic setting would produce multiple publications.

The other Wind River Range papers deal with entirely different concepts and use different datasets from different years. One of the papers deals with multi-year glacier change and streamflow patterns (neither of which are analyzed in the present study). The other paper introduces the "snowshed" concept for tracking snow blowing across ridges in an ML framework (which is completely unrelated to the present study). The present study deals with topics that have not been addressed in either of the previous papers, namely, spatial snow density heterogeneity representativeness and comparing the propagation of wind-drifted density and depth heterogeneity into watershed-scale SWE metrics using different methods.

As a sidenote, we note that our previous manuscript submitted to TC was critiqued for being "monograph in scope" (i.e., too many new results). We believe that papers should contain the right amount of data necessary to make the relevant arguments, and that may be more or less data for different papers.

Compared to these papers, this study shows:

1. **Density profiling of extreme drifts.** The study shows results of a 5.9 m drift, where the measured bulk density is 585 kg/m³. It is shown that widely used snow density models fail badly such deep snow.

In fact, we show multiple profiles from deep drifts, not just the one 5.9 m drift. There is also a 4.0 m drift (the highest-elevation major drift of this depth ever profiled in the Rocky Mountain region, at 3820 m or 12,500 ft. elevation). There is also a partial profile from a second lower-elevation deep nivation hollow (and we demonstrated in this study that partial profiles are likely representative since there is no depth-density trend in the nivation hollows).

Again, the reviewer fails to appreciate the full scope of our study by only mentioning one of our multiple deep drift profiles.

2. **Benchmarking operational gridded SWE products.** This study directly evaluates the performance of existing near-real-time SWE products and shows the heterogeneity problem propagates into operational datasets. It is shown that that watershed-average SWE can look well-estimated even while the spatial distribution — and especially the high-SWE tail — is dramatically wrong.

This is a correct interpretation of our findings. We are curious, though, why the reviewer apparently finds such an investigation "not worthy of publication in TC." The datasets we investigate (SNODAS, CU-SWE, SWANN) are widely disseminated and discussed in the US water supply forecasting community with high-profile funding initiatives and burgeoning operational reliance. These datasets have not previously been evaluated using comparable lidar-derived datasets in a comparable environment to the Wind River Range. Directly evaluating the heterogeneity of these datasets (as done in our study) is quite different from the reviewer's generalized assertion that "most current approaches underestimate the heterogeneity."

Our study has a unique geographic context that is much more heavily wind drifted than most areas with unique multi-kilometer snow transport (doi.org/10.1029/2024GL113599). Moreover, our study has a unique real-time context, as all of our datasets were available in near-real-time (including the lidar-based SWE map and density data). We clearly emphasized this real-time component throughout the existing manuscript, and this real-time context is essential for the relevance of snow data to water supply forecasting, which is the primary funding source for applied snow science in the western U. S.

Thus, our study provides a unique opportunity to directly <u>test</u> the specific operational datasets that are relied on by water managers, instead of simply asserting that "most approaches underestimate heterogeneity," which is otherwise a <u>hypothesis</u>.

The reviewer seems to be confusing generalized hypotheses about all snow data (reviewer quote: "most current approaches underestimate the heterogeneity") with concrete statistical evidence concerning specific datasets in a particular setting and context (presented in our study).

3. **SNOTEL sampling bias.** The study notes that all 88 SNOTEL sites in Wyoming are located in forested areas, structurally excluding them from observing alpine snowpack heterogeneity.

Correct. However, the reviewer fails to appreciate that the representativeness of SNOTEL has <u>never been assessed previously</u> with respect to <u>snow density</u>, not snow depth or SWE.

There is only a single previous study that directly quantified missing SNOTEL heterogeneity at the basin scale (Molotch and Bales 2006, cited by the reviewer below). This single previous study is from 20 years ago in a different geographic region that is not comparable to the alpine Wind River Range (only 11% of the Molotch and Bales watershed was above timberline, compared to 76% in our study). This previous study also does not leverage modern Bayesian statistical approaches, as we have done.

Most importantly, this previous study <u>did not evaluate snow density</u> (in fact, snow density is never even mentioned in that paper). This means that our study is the first to explicitly evaluate the representativeness of SNOTEL snow density at the basin scale using a quantitative framework (Fig. 12).

After reading the large preprint (42 pages without references) I was disappointed as there is unfortunately only little new scientific insight in this study. Based on the above three points, only the first is new, i.e. showing the density distribution in such a deep snow drift and providing an explanation for the vertical and lateral differences within the nivation hollow.

This is factually incorrect and raises the question of whether the reviewer overlooked the entire body of results and figures that aren't contained in the abstract.

In addition to the deepest complete alpine snow density profile ever reported in the literature, the following aspects of our analysis are also novel:

- First quantitative analysis of SNOTEL snow density representativeness at the basin scale, an issue of broad relevance since these measurements are widely relied upon but rarely validated across large alpine regions (Sect. 4.3 and Fig. 12)
- Independent evaluation of widely cited SNOTEL-based density models with respect to observed variability (e.g., first evaluation of new Hill et al. 2019 model against deep alpine snow drift data), which can inform the adoption or refinement of these models
- Exposition of a new representativeness-weighted statistical approach for spatial density extrapolation from point measurements (Sect. 2.2.1, Eq. 2, and Fig. 12)
- First quantitative analysis of the impact of wind-packing snow density on basin-scale SWE heterogeneity (Fig. 10 and Table 2)
- First parametric regression model for empirically estimating the effect of forests on snow density from gridded canopy cover datasets without having to run computationally complex process-based snowpack models (Sect. 4.2 and Fig. 7)
- Novel Bayesian statistical quantification (with credible intervals) for the relationships between snow density and predictor variables (Fig. 7); note that previous studies have assessed things like simple snow depth-density correlations, but ours is the first study to quantify these effects from a Bayesian statistical modeling perspective using the representativeness of each measurement across the entire lidar-surveyed domain
- Presentation of numerous snow density profiles from unique environments, including high elevations (up to 4040 m or 13,255 ft. elevation), avalanche debris, and steep slopes (up to 43°). The novelty of our dataset is not just the deep drifts—few (if any) studies have previously sampled or analyzed complete snow density profiles from comparable extreme alpine environments in the Rocky Mountain region (note that snow properties are inextricably linked to the regional geographic and climatological context)
- First intercomparison of complete snow density profiles from the top, middle, and bottom edge of a nivation hollow, showing how the density is spatially variable depending on the position within the hillslope concavity (Fig. 4)

- First computational fluid dynamics (CFD) simulation of wind speed in a nivation hollow comparing the wind speed over the (1) bare ground surface and (2) lidar-based snow surface, providing a process-based interpretation of our density results
- First demonstration that CFD-simulated wind speed can predict nivation hollow snow depth based on the match between spatial patterning of simulated CFD wind speed and lidar-based snow depth (Fig. 5)

Unfortunately, it is only one profile from one winter.

This is factually incorrect.

Our study includes 36 profiles, including 3 profiles from deep drifts (3.5, 4.0, 5.9 m) in addition to comparison with prior years' nivation hollow density profiles (see comparison and discussion to 2024 WRR empirical density model), which explicitly shows how similar patterns repeat across years. The reviewer completely ignores our analysis of other profiles, such as the top and bottom edges of the nivation hollow or the high elevation profiles.

Additionally, while the present study was in review, we have collected 14 additional snow density profiles from May 2026. Two of these new profiles are repeats of the nivation hollow measurements from last year, further supporting our results. We can add these additional new profiles to supplemental information in a revision.

However, already the finding that empirical models of snow density underestimate the measured bulk density is not surprising, because, as the authors themselves write: "Many of the models tested are being used beyond their intended scope."

Nevertheless, these models are widely used to inform predictions of density variability. Our comparison between previous models and our dataset provides insight into the level of departure from baseline expectations, not a fundamental critique of the other models.

Moreover, some of the models (like Hill et al. 2019) are explicitly claimed to "allow snow depth measurements to be converted to SWE" (abstract of Hill et al. 2019), including in the context of lidar-based snow depth surveys (page 1768 of Hill et al. 2019). Thus, while we acknowledge that the prior models aren't expected to be perfect, it is important to demonstrate the potential limitations of these prior approaches.

Similarly, regarding the second of the above points. It is not surprising that most of current available products fail to reflect the spatial heterogeneity, because, as the authors themselves write "Gridded SWE products that do not explicitly consider snow transport processes are unlikely to capture the spatial variability". However, they were trained to reflect watershed-average, where they perform mostly quite well.

It is debatable whether a dataset with 17% watershed-scale bulk SWE error is an acceptable level that the reviewer describes as "perform mostly quite well" (Table 2, U. Colorado SWE).

Moreover, these gridded SWE datasets are widely used to reflect spatial heterogeneity in other studies, even if this is not a proper use of the datasets. For example, the recent highly publicized study by Raleigh et al. (2025) on snow hotspots (doi.org/10.1038/s43247-025-02660-z) used 2 of the 3 datasets in our study (SNODAS and U. Arizona SWE) as supplemental datasets to explicitly investigate snow spatial heterogeneity. Thus, our demonstration that these datasets completely fail to capture heterogeneity in a heavily wind-drifted alpine environment is timely and important.

Other studies also use the same gridded datasets that we analyzed to explicitly investigate snowpack heterogeneity, for example SNODAS is used to optimize monitoring station locations (https://doi.org/10.1016/j.jhydrol.2018.04.037), SNODAS is used for calibration of distributed hydrological models (doi.org/10.5194/hess-28-1127-2024), and U. Arizona SWE is used to investigate spatial interactions between snowpack and forest fires (doi.org/10.1088/1748-9326/ac6886).

Even though the reviewer does not find this result "surprising" enough to be "worthy of publication in TC," our investigation of missing heterogeneity is nonetheless fundamental for the interpretation of other studies' results using (or perhaps misusing) these gridded datasets.

The third of the above points is connected to the second point. The finding that SNOTEL sites are not representative on the watershed scale is not only valid for Wyoming. It is valid also for other regions and other in-situ networks and has been demonstrated already many times (e.g. doi.org/10.1002/hyp.6128, doi.org/10.1002/hyp.9355)

It is factually incorrect that this "has been demonstrated already many times."

The non-representativeness of SNOTEL snow <u>density</u>, to the best of our knowledge, has <u>never</u> been investigated previously at the basin scale, and definitely has not been demonstrated previously using as extensive of a deep drift dataset in a comparable environment.

The reviewer provides 2 citations to support their assertion of "many times."

(Molotch and Bales 2006); this is cited prominently throughout our manuscript. As discussed above, the Molotch and Bales paper <u>does not consider snow density</u>, and their study site was very different from ours (southern Rockies, minimal area above treeline, no major glacier cirques, no kilometer-scale snow transport, etc.). Only a small fraction (11%) of the Molotch and Bales study watershed is alpine, compared to the majority of our study region.

(Meromy et al. 2012); this study is irrelevant because it deals with sub-grid variability at fine scales, not systematic variability across mountain ranges. This previous study shows that SNOTELs might not be ideally positioned within individual grid cells. Our study shows that SNOTELs systematically fail to capture the dominant hydroclimate of the northern Rocky Mountain region, which is an entirely different finding with important implications. It is concerning that the reviewer thinks this citation is relevant, since it shows that the reviewer does not apparently appreciate the different spatial scales and types of representativeness in our study.

Given this fact (just one single new scientific finding), I believe that the study, in its current form, is not worthy of publication in TC. If at all, I strongly suggest changing the focus and the title so that the weight is on the density profile in the nivation hollow. I do not like the current title for two reasons: First, not only deep snow drifts are missing in current approaches, but probably also the opposite like ridges with bare ground. Second, it is not all new that most current approaches underestimate the heterogeneity of snow in mountainous terrain (e.g. doi.org/10.5194/hess-19-1339-2015)

It is disingenuous to characterize our study as "one single new scientific finding" given the broad scope of novelty outlined in the bullet points above. We also disagree that "one single new scientific finding" would be necessarily unpublishable given the relevance and context of the finding (implications for basin-scale SWE heterogeneity in the Rocky Mountain region).

Finally, the author totally misses mentioning the current literature, which tries to tackle the missing heterogeneity problem. Here is just a small set of examples…

Our manuscript has 125 references (TC lists "maximum of 80 references"), so it is disingenuous and factually incorrect to say that we "totally miss mentioning the current literature." We are well aware of the modeling literature presented by the reviewer, and have cited much of it in other publications. However, it is not particularly relevant to the present study.

doi.org/10.5194/tc-15-743-2021 (minimal relevance—model development paper)

doi.org/10.5194/tc-18-3533-2024 (minimal relevance—model development paper)

doi.org/10.3389/feart.2024.1393260 (minimal relevance—model development paper)

doi.org/10.3389/feart.2023.1308269 (minimal relevance—model development paper)

doi.org/10.1002/2016WR019872 (minimal relevance—small sub-grid spatial scale)

doi.org/10.5194/tc-18-4315-2024 (minimal relevance—model development paper)

In determining the references to include, we must weigh the relevance of the reference to the arguments we are presenting, the context of our study, and the total number of references (already we are more than 50% beyond TC guidelines for maximum number of references).

We appreciate that the reviewer thinks process-based model development is important, and we agree. However, this is fundamentally separate from our study, which is observation-based with a focus on snow datasets that are available in real-time for the U.S. Rocky Mountain region. Even if these models could capture snow depth heterogeneity perfectly (which they cannot), few (if any) of these models represent the types of snow <u>density</u> variability analyzed in our study, and the outputs from these models are not currently available in real-time in the U.S. for water supply forecasting applications due to computational complexity and missing fine-scale forcing data.

Despite advances in modeling, we think it is still important to observe the snowpack directly.

Most of the studies cited by the reviewer have little or no validation of spatial heterogeneity in snow density, and even the modeled snow depths rarely agree with spatial patterns shown by lidar (e.g., Fig. 5 of Queno et al. 2024 shows remarkably different patterns of measured and modeled snow drift locations). Our study includes unparalleled snow density observations that might be helpful to these model developers, but the implementation or assessment of process-based models is not the primary goal of our data-driven study.

Nevertheless, we are willing to add some of these model development citations (above) in a discussion paragraph about the potential utility of our findings for others' model development and validation.

Here, some input to individual lines in the manuscript:

L21: Accumulation areas of typical mountain glaciers are often at locations which profit from snow drift. Mass balance measurements before the melt season frequently show densities around 600 kg/m3 (e.g. doi.org/10.5194/tc-13-3413-2019)

Yes, we mentioned that glacier firn commonly reaches these densities—and the provided citation shows these densities occurring in the spring, not the winter. Still, this is a useful citation that we will add.

L40: Pfohl et al. in review: Could not be checked.

We will remove this citation since it is still in review.

L291: There are more newer, more enhanced models available, which were developed based on the same data set (doi.org/10.1016/j.coldregions.2025.104435, doi.org/10.5194/hess-25-1165-2021).

While we appreciate there are very many empirical snow density models, we selected a few of the most widely applied ones as illustrative examples. We are neither attempting nor claiming to benchmark the entire suite (or even the "best") snow density models.

L376-389: Belongs to introduction

The description of the dataset resulting from our efforts is inherently a "result" that goes in the "Results" section.

L455: Fig.4: snow stratigraphy and snowpack temperatures would help a lot to explain density trends.

As discussed in the text, the nivation hollow was isothermal (0 °C) at the time of our measurements, so the temperature profile would be a straight line. We illustrated the stratigraphy in the photograph with arrows to prominent ice layers, but due to time limitations in the field we did not investigate grain size or shape of each layer because these are not the goals of our study.

L501: “we there”?

Should be “there was.”

L503: “lower density”. I strongly recommend to consistently use “bulk density” and reserve “mean density” for the average of bulk densities from different profiles.

Agreed, we can make sure the wording is more consistent (“bulk density” throughout).

L543: Fig: 6: snow stratigraphy would help a lot to explain density trends.

See above.

L545: “the other tree snow pits”. I strongly recommend introducing a table providing an indicator, snow depth, elevation, aspect, density, density-trend for each snow pit and provide the indicator in the figures.

With so many snow pits, we think this would only add confusion and unnecessary numeric clutter. The salient comparisons between individual pairs or sets of snow pits are made in the text, and the full dataset is downloadable.

L557: Snowpack temperatures: I wonder if it was always measured in the shade when shoveling took several hours?

Shoveling the shallow snow at high elevations (where the snow was below freezing) only took a few minutes.

L562: “The coldest temperatures in the Downs Mountain summit drift are near the ground surface”. That might have been due to permafrost?

Perhaps, but this is beyond the scope of our study, and we don’t have any data to support such a hypothesis, so it would be pure speculation.

L573: “the temperature profile is reversed compared to what we would expect for depth hoar formation”. Not necessarily, as the depth hoar formation typically happens during a period of several days where the temperature measurement is only valid at the time of the measurement.

Yes, and in fact we already stated this caveat in the text (lines 573-574).

L600: That also depends on the aspect of the forest (doi.org/10.5194/hess-30-1691-2026)

Yes, and this is captured by the aspect variables in our snow density model.

L764: “avalanche debris is the densest snow within the high-elevation Downs Mountain vicinity”. Just one sample?

No, as shown in Fig. 6, there are multiple samples from different avalanche blocks.

L768: “Since snow near the surface of the drift exceeds the vertically integrated mean density”. This can easily be caused by wind- or melt-crust.

Yes, and that supports our contention “overburden pressure is a negligible driver of densification” (in the nivation hollow).

L791: “Tabler (2003) emphasizes the pressure of overlying snow”. Tabler developed is method for snow drifts along road fences and was never thinking about special cases like nivation hollows.

Yes, but is there any reason to expect that artificial snow drifts follow different physical laws from natural snow drifts? This comparison to Tabler’s assumptions was actually added because a previous reviewer suggested that the processes responsible for deep drift density were already well-established by Tabler, while the opposite seems to be true, since Tabler assumes overburden pressure is the driver (which would result in a depth-density trend, contrary to observations).

L799: “For the first time, our study quantifies the watershed-scale impact of ignoring wind-packing effects on snow density.” Snow density heterogeneity?

Agreed, this is a helpful minor language clarification.

It is noteworthy that the reviewer does not dispute our claim quoted above, yet still disputes that the paper provides anything novel beyond the single drift profile (see bullet points above).

L814: The best model shows only an underestimation of 3%! It is in sect. 3.5 and not 3.4!

Not sure what is intended here—line 814 is talking about self-limiting drift-trapping and doesn’t seem relevant to the reviewer’s comment.

L882: Current progress in forest-snow is missing: doi.org/10.5194/hess-24-2545-2020, doi.org/10.5194/hess-27-2099-2023 or doi.org/10.1029/2020WR029064

We disagree that these citation are relevant.

(Helbig et al. 2020) addresses snow interception and does not even mention the effect of forests on the seasonal evolution of snowpack density.

(Mazzotti et al. 2023) simulates snow cover evolution, not snow density. In fact, this paper describes the impact of forests on surface layer snow density as “an unexplored research topic to date.” So, this would strongly support the novelty and timeliness of our own study.

(Mazzotti et al. 2021) is a model development study that does not mention snow density.

Again, the reviewer seems to have completely missed the novelty of our research with respect to spatial and vertical heterogeneity in <u>snow density</u>.

L890: “overrepresentation of in-situ stations in forested regions”. This might be true for the SNOTEL network, but it is not true for other mountainous snow networks. Also, the opposite has been observed (doi.org/10.1002/hyp.10295).

This is an interesting citation, but it doesn’t seem particularly relevant to our study. We never claimed that all in-situ networks always underestimate snow depth, and this is much beyond the scope of our study. Rather, we claimed that the specific SNOTEL network underestimates heterogeneity in snow density in our specific Rocky Mountain study region. Nevertheless, we can include this citation as a contrasting example of different variability elsewhere.

L915: “colocation issues between the snow pillow (which measures SWE) and the snow depth sensor (which measures snow depth) could contribute to inconsistent density calculations”. Please provide any literature for this claim.

(Goodison et al. 1981) as cited in manuscript.

L938: “the two deepest snow pits (4.0 and 5.9 m) have the third- and fourth-highest representativeness scores.” Not shown - please provide table/figure.

This is precisely what is shown by Fig. 12, which is why this sentence immediately follows the sentence introducing Fig. 12.

L1022: “Airborne lidar surveys provide the key to accurate quantification of snow depth heterogeneity”. Right, but please also mention that they are expensive and therefore rarely available in other mountain regions.

Our study is not an economic one, so this is not germane to the research questions at hand.