The authors would like to sincerely thank Anonymous Referee #2 for their thoughtful comments and careful review of our manuscript.

Per your first major comment, we agree that the LiDAR field-of-view being unable to capture the entirety of the runout zone is an unfortunate byproduct of our study site and scanning location (also see reply to Referee #1, paragraph 2). Though this limitation required an ad hoc adjustment of friction coefficient values to limit runout extent, because it was consistently applied throughout all simulations, we maintain that it supports the central point of the communication: that runout distance and area, as well as all other modeled outputs (max velocity, flow height, etc.), can vary considerably when LiDAR data is incrementally incorporated for initialization. We also agree that it’s likely the extent would have varied even more between simulations had the slides been allowed to reach flatter terrain, which we believe further highlights the need for additional investigations on the merit of LiDAR data’s use in avalanche dynamics modeling. While the magnitude and nature of this variability may differ between avalanche locations and scenarios, we found the sensitivity observed in our study notable, and worth sharing with the broader avalanche science and remote sensing community as a Brief Communication.

With regards to your second major comment, related to sources of variability, these are both insightful critiques, which we address below:

1) You are correct that in RAMMS, friction coefficients vary spatially with topography, and thus altering the sliding layer changes not only the topographic input, but also the dependent friction coefficient distribution, both of which impact simulation results. We believe this to be a relevant consideration, but beyond the scope of our study. In the work presented, we are only focused on investigating the sensitivity of an avalanche dynamics model (RAMMS) to LiDAR-derived inputs for initialization, as opposed to using generic inputs and assumptions. Therefore, we do not seek to quantify, investigate, or pontificate on the strengths and weaknesses of the model itself. For example, when using RAMMS, a different DEM might sometimes have a greater effect on results via friction coefficient distributions (associates, as you say), while other times the varied topography itself may play a larger role. It is not our intent to distinguish between these two, but rather to show that holding all other variables equal, the incorporation of LiDAR-derived inputs can produce drastically different simulation results. The same can be said of vegetation delineation and its influence on simulation results. In RAMMS, areas designated as vegetated have a scalar value added to what the unvegetated friction coefficient would
have been, based on the local topography. Therefore, in the case of varied vegetation delineation (between manually identified, LiDAR inputs, etc.) the mechanism leading to differing simulation outputs is entirely related to altered friction coefficients by RAMMS, but this is an inherent property of the model, and therefore also not within the scope of our study. Based on our results, we suggest that regardless of how a model handles vegetated areas or topographic inputs for flow dynamics simulations, the sensitivity of the model to variations in its initialization must always be a primary consideration; particularly if the model is ever to be used as an operational tool.

2) With regards to numerical artifacts, though an interesting and important consideration, we contend this was not a major discrepancy in the cases presented here. First, at no point were we comparing between simulation sets with automated vs. increased friction values, nor across site locations; we were only comparing simulation cases to themselves when LiDAR data was incrementally incorporated (see Table 1). Second, we have no control over the time-domain of RAMMS. Intuitively, RAMMS will produce larger and faster avalanches with a larger mass and steeper slopes. This was especially the case in the ‘upper’ ridge site location, where max pressure, velocity, and height were all reached at the pinch points of the pairing hourglass features at mid-slope. In this case, the time step at which these maxima were achieved is not a relevant consideration, and we therefore maintain our position that differences in these maximum values are solely due to differences in the simulation inputs. Furthermore, in simulations with automated friction coefficients, stopping criterion (5% momentum threshold) were met when enough mass exited the field-of-view to drop the simulated momentum below the threshold, so any variation in the time step of that occurrence (which was minimal) wouldn’t impact maxima values recorded mid-simulation. Even if the entire field-of-view was observable and the stopping criterion were lowered to a 0% momentum threshold, debris at this point was spreading and decelerating, and we find it very unlikely that new maxima in any output variable would be recorded. Similarly, in simulations where friction coefficients were manually increased, the stopping criteria was met long after movement of the debris flow front had largely plateaued (see simulation .gifs), and thus differing final time steps would not have played a significant role in the final debris extent at simulation completion. For these reasons, we respectfully refute the suggestion that numerical artifacts played a significant role in our results and analysis.

Responses to your minor/technical comments:

- p1 l21: “common DEM” – what does common refer to?

In this case we are referring to an off-snow DEM of the ground surface. We will revise this line in the edited manuscript.

- p1 l25: does the “snow-snow” interface automatically correspond to the sliding layer – what about different avalanche types?

The snow-snow interface does intuitively correspond to the sliding layer within a start zone. In the transition and runout zones, the top snow surface represents the initial condition of the sliding layer, though as we discuss in our responses to Referees #1 and #3, this surface changes during the avalanche event due to erosion and entrainment. The point is, either is a better representation of the sliding surface initial conditions when compared to a ground DEM. Please see paragraph 1 from our response to Referee #1.

- p2 l30: I am not sure you can refer to “more realistic” in that sense – it could be “more accurate” if a case study with documented observations would be available.

You are correct. We meant to convey that the LiDAR inputs better represent sliding topography, relevant vegetation, and spatially variable snow depth relative to traditional
RAMMS initialization. The operational capacity of these inputs and their potential to improve RAMMS simulation results has yet to be verified (see paragraph 2 of our response to Referee #1). We will correct this line in our revised manuscript.

- p2 l38: what kind of “error”

Using a point measurement of snow depth to an interface of concern rather than accounting for spatial variability in snow depth atop the interface will inevitably result in a less precise estimate of release volume/mass. We will incorporate this context into that section.

- p3 l70: I very much appreciated the idea of data availability but was unable to open the corresponding link: https://doi.org/10.5061/dryad.z8w9ghx9z

The DOI is to be made available upon publication. In the interim, you may access the data assets at this link: https://datadryad.org/stash/share/W4ZoRKMTNE6QIXBjdDFdB_NQZpyxm230EbxR2fycNOU.

- p3 Fig 1: could you indicate where the U and L test cases are located?

Yes, we will add outlines of the U and L start zone boundaries to Figure 1 in the revised manuscript.

- p4 l 82 & 97: could you specify what “traditional” means in this context and which “automatically generated values” you refer to? Please comment on how vegetation is treated in terms of friction coefficients (scenario 4)?

In line 82 the sentence continues to elaborate on what we refer to as traditional: “...“traditional” inputs were used, assigning a ground DEM as the sliding layer, a uniform snow depth across the entire start-zone, and a vegetated area delineated manually from photographs.” There is an automated process in RAMMS that computes the spatial distribution of friction coefficients based on topography. In areas delineated as vegetated, a scalar value is added to account for roughness by increasing friction. We will add this description to the revised manuscript.

- p4 l 82-95: The scenarios need to be more clearly defined. To me it remains unclear what impact on the simulation input the delineation of vegetation has (different friction parameters in this areas? Which ones?), see also major comments below.

Altering the sliding surface DEM changes both the sliding topography and the spatial distribution of topography-dependent friction coefficients. Accounting for spatial variability in snow depth atop an interface alters the release volume/mass. to vegetated areas, and thus variation in vegetation delineation alters the spatial distribution of friction coefficients. We will better describe how the incorporation of LiDAR data at each step is dealt with within RAMMS in the revised manuscript.

- p5 l 106: How do you define and measure runout (please check more recent literature on model evaluation)? Which simulation results do you use? What is your reference and how is it measured?

Runout was defined as the linear distance (when viewed aerially) from the stauchwall of the release area to the furthest extent of avalanche debris at the final time step in the simulation. We will state this definition in the revised manuscript.

- p6 l119: Is this due to the redistribution of the starting mass or due to changes in topography (or changes in friction coefficients (which ones??))? What are the differences
between your test cases?

Presumably a combination of changes in the topography as well as changes in the spatial distribution of friction coefficients dependent on the changed topography. Parsing the individual contributions of these is beyond the scope of our preliminary study. Also, see the answer above regarding lines 82-95.

- pt Fig 3: What does "final" deposition height refer to – the flow depth at the last (which?) time step – to my knowledge the underlying flow model does not consider deposition?

Correct; we were referring to the flow depth at the final time step when the momentum threshold stopping criterion were met. We will state this in the revised manuscript.

- p7 l133: I think that it is only possible to show that change in input data implies a "dramatic change" in the simulation results. A dramatic "improvement" would only be possible if you compare to observation data.

We agree, and will revise the wording of that sentence in the updated manuscript.