Overall Comment:

The manuscript is well written and presents many novel methods and findings in a fascinating application. I recommend the work be published with some minor revisions. The majority of my comments are presented with the intention of providing opportunity to improve the interpretation and impact of this important paper.

The study uses recent advances in avalanche terrain modelling, an excellent application of the RAMMS dynamic model to model terrain traps and exposure to potential avalanche hazard, and three modelling approaches used to describe the run list decision-making process employed by the CMH Galena guiding team. The accuracy of the models presented is high and indicates that these models show promise in this application. Note, Bayesian Networks and Machine Learning models is not my area of expertise, and thus, I leave the critical review of the model development and technical details to others.

The manuscript may be considered long at 60 pages. The overall quality and ease of interpretation by readers may benefit if it could be condensed during the review process if possible.

Specific Comments:

In my understanding, avalanche likelihood and size forecasts are often produced based on the character of the avalanche expected, sometimes described as the "Avalanche Problem". Avalanche Problems usually include a broad description of where they are expected to exist within the terrain according to elevation band (i.e. Alpine, Treeline, Below Treeline) and aspect (e.g. N- NE – E, SE, …). Given that runs have varying levels of exposure to these elevation bands and aspects, it is rationale to assume that high likelihood and size ratings (i.e. elevated avalanche hazard) may exist on some runs, but not on other runs for a given day depending on the avalanche problems. Thus, the type and location of avalanche problem is likely an important factor in run list decisions. Could the authors provide an explanation of why avalanche problems were not included in the analysis? Do the authors think that a future study would benefit from including these data?

I understand that mechanized operations often use a snow safety team that gathers snowpack data, investigates conditions on runs (e.g. snow depth, crevasses, snow quality), and conducts explosive avalanche control. Operations often send snow safety teams to gather data from runs that are close to opening (e.g. coded red, black, or yellow) and these data are often critical to run list decisions. Does the variable "last_skied" represent these snow safety investigations? Or would the investigations be included in this variable along with regular skiing of the runs with guests? If snow safety investigations were not included, could the authors provide rationale why this potentially important run list decision factor was not included? Perhaps this type of snow safety investigation is not part of the regular CMH Galena practice?

In my understanding, explosive control work is often used to reduce avalanche hazard, decrease uncertainty in avalanche hazard forecasting, and to protect key features on runs prior to skiing with guests; and hence, this work often has an impact on the run list status. Does the CMH Galena operation use explosive control? If so, is there a reason that explosive control data were not

included in the analysis and modelling? Could the authors postulate the effect on the model results if a suitable dataset representing avalanche control were included?

It appears that the variable “Runlist_prior” has a very high feature importance in both the random forest and XGB models. To reduce this effect and focus the model and analysis on the factors that may lead to change decisions in the run list, did the authors consider removing runs that are rarely coded other than green (i.e. open for guiding)? That is, while I am not intimately familiar with the runs at CMH Galena, I do understand that operations often have “regular routes” or runs that involve predominantly Simple avalanche terrain (see Avalanche Terrain Exposure Scale, Statham and Campbell, 2024) and these runs are only closed for guiding in rare extreme avalanche conditions. This means that output node for these runs is likely not sensitive to the input variables for avalanche hazard conditions. Could the model and analysis provide more insight into the relevant decision factors if the output node focused on runs that often change their status?

As far as I understand, mechanized guides often use a conditional opening coding (typically coded as yellow) where a specific condition must be met prior to opening the run. If the condition is not met, the run remains closed and is not opened for skiing. Could the authors provide rationale why this standard run coding level was not included in the analysis, and postulate on what the effect on the results would have been if it was included?

Suggest increasing the size and / or resolution of the Figures to ensure they are discernible in the final publication. Currently, many of the words and symbols in the Figures are difficult to read. Specific figures where this comment is applicable are:

- Figure 1 (inset map), Figure 2 (legends, run names), Figure 4 (variable names), Figure 6 (variable names), Figure 7 (variable names), Figure 8 (run names)

An important addition to the introduction worth including is that the terrain identified in run lists already presents a significant filter on potential terrain. That is, there is much terrain that is either not skiable, too severe, or inaccessible for some reason that prevents it from even being considered on the run list. For example, a study describing the terrain indicated on the run list from well-established operations would provide value.

Lines 36-37: Suggest re-wording this sentence. It is not clear what “avalanche terrain hazard” refers to. Avalanche terrain is often described by its overall severity. This word could be an option to replace “hazard” here and elsewhere in the manuscript.

Line 64: Is the morning run-coding meeting only 15 minutes at CMH Galena?

Line 65: Suggest adding the word “may” after “… the reasons for not discussing a run…”. There are other reasons why a run may remain uncoded.

Lines 126-127: Providing a list of the state-of-the-art methods or a few examples would be helpful here.

Lines 130 – 133: Please clarify that only the most conservative line on each run was used to extract terrain characteristics used to describe the avalanche terrain on each run.

Line 155: Suggest enlarging Figure 2. The legends are difficult to discern.

Line 156: Figure caption. Suggest revising the figure caption to the following or similar: "Figure 2: Comparison of PRA polygons (upper images) and runout impact pressure (lower images) for frequent (left upper and lower) and large runout (right upper and lower) simulations. The frequent PRA and impact pressure simulations represent smaller storm snow avalanches, whereas the large PRA and impact simulations represent deeper more connected persistent weak layer avalanches." Note, the images could be labelled "a, b, c, d" and referenced accordingly.

Line 190: Figure 3. Could the authors please explain or postulate why or how there are conservative clusters of lines mixed with non-conservative lines? For example, the run "Gorilla" appears to show the black (conservative lines) mixed in similar terrain with the green coloured (non-conservative lines). The figure could be improved by explaining what the differentiation between colours represents and communicating this in the legend (i.e. why does the legend only show orange and black lines when the figure shows many colours?). Lastly, the outlines for the ski run polygons could be a different colour than black because black is used for the conservative lines.

Lines 194 – 204:

- In the reviewer's understanding, the output of PRA is probabilistic. What does the extraction of PRA values along a GPS track result in? For example, is the mean PRA value the mean of all probabilistic values that the GPS track intersects which could mean that the track averages values of 0 (non-PRA raster cells) to 1 (complete PRA raster cells)?
- Slope incline is the core factor determining PRA which means that PRA and slope incline are strongly correlated. Could the model and analysis be simplified by using only PRA rather than PRA and Slope Incline as variables? Forest cover is also a strong input for PRA determination, so a similar question as to where the value is in using both forest cover and PRA in the analysis? If the model and analysis benefits from including these correlated variables, could you please explain or postulate what this value is? Further, avalanche hazard ratings will be strongly correlated with recent avalanche activity, similar to above could the authors explain the influence of including these types of strongly correlated variables in the models (note, this is more for the general knowledge of the reviewer and does not necessarily need to be included in the manuscript)?
- Similarly, avalanche runout depth, runout velocity, and runout impact pressure are strongly correlated. Could the analysis and model be simplified by using only avalanche impact pressure as an indicator of exposure to potential avalanches? Could you clarify if avalanche runout depth is used as an indicator of terrain traps or something else?

Line 197: Are the relevant GPS tracks the "conservative cluster" tracks? If so, suggest adding this slight clarification.

Lines 209 – 210: In my understanding, the destruction of weak layers by skier traffic is highly dependent on conditions. How is this variable constant?

Lines 211 – 212: In this reviewer's understanding, ski quality of a run is highly condition dependent. How is the ski quality for a run a constant variable?

Lines 402 – 403: Suggest defining the node acronyms in the legend or caption. It becomes tedious to go back and forth from the text to the figure to determine what the nodes refer to.

Line 780: Figure 8 caption. Suggest expanding the caption to explain what the yellow outlined runs refers to. For example, is it the case that the yellow outlined red runs were green the previous day and now are changing to closed?

Line 781: Limitations. In my understanding, run list coding practices vary by operation often based on the nature of the terrain, common avalanche hazard conditions, typical guests (e.g. level of ability, preferences), and experience of the guiding team. Given that the expert guide author involved in this study has an intimate understanding of the CMH Galena run list coding practices, the discussion and limitations section would benefit from some thoughts of how well these models are actually capturing the Galena run coding decisions. Are the key factors influencing run list decisions identified and do the feature importance match the expert guide author's intuition? Are there other key factors not included (e.g. explosives, snow safety investigation, avalanche problems) that would add to the decision models? Given the expert guide likely has an awareness of the run list coding practices at other operations, the discussion and limitations section would benefit from thoughts on how the results may relate to other operations.

Appendix A: In the reviewer's understanding, likelihood of avalanches is typically communicated and assessed on five level ordinal scale (i.e. Unlikely, Possible, Likely, Very Likely, Almost Certain). Could the authors please explain why the "Persistent avalanche likelihood" variable only includes the ratings None, Unlikely, Possible, Likely, and Very Likely; whereas, the "non-persistent avalanche likelihood includes the ratings Possible, Likely, Very Likely and Almost Certain?

Technical Corrections:

Lines 46 – 48: Suggest a slight modification:

"While these tools can be effective for general recreationists, their simplicity - particularly their focus on the public avalanche danger rating - limits their value for more complex decision-making contexts such as professional guiding or advanced amateur recreation.

Line 127: Add "," after the word "tenure".

Line 133: Add "," after the word "tools". Or use an active voice in the sentence. Note, this sentence could be combined with a slight revision of the preceding lines to clarify the terrain description process.

Line 143: Change “at al.” to “et al.”.

Line 108: Replace the “,” after the reference with “and”.

Line 196: Please check that the word “area” is appropriate after “PRA”. If using the acronym in full this would equate to “… potential release area area…”.

Line 566: Change “in” to “is”.

Appendix A: The labels on the histogram are not readable. Please ensure the final publication uses suitable font size for readability.

Lines 218 – 221: Suggest revising these sentences to avoid the double “In addition”, however this is only a stylistic writing suggestion.

Line 577: Suggest adding the numbered heading “3.3.3”.

Line 607: Suggest adding the numbered heading "3.3.4”.