Responses to Reviewers' Comments for Manuscript egusphere-2025-2143

Machine learning for automated avalanche terrain exposure scale (ATES) classification

Addressed Comments for Publication to

Natural Hazards and Earth System Sciences

by

Kalin Markov, Andreas Huber, Momchil Panayotov, Christoph Hesselbach, Paula Spannring, Jan-Thomas Fischer and Michaela Teich

We would like to thank the reviewers for the insightful comments made on our manuscript. The constructive and stimulating feedback and suggestions will help to improve the quality of our research paper. We have carefully read and discussed the reviewers' comments and will incorporate the feedback in a revised version of the manuscript. This document contains our responses to the comments of both reviewers following the chronological order of received reviews.

Sincerely,

Kalin Markov, Andreas Huber, Momchil Panayotov, Christoph Hesselbach, Paula Spannring, Jan-Thomas Fischer and Michaela Teich

Authors' Response to Reviewer 1

The originally posted comments by John Sykes can be accessed at `https://doi.org/10.5194/egusphere-2025-2143-RC1`. We reprint the comments here along with our response to each of the comments.

> **General Comments.**
> This manuscript presents a machine learning approach to classifying avalanche terrain using the Avalanche Terrain Exposure Scale (ATES). The authors developed a novel and meaningful validation approach and tested performance of several iterations of random forest models in a study area with limited avalanche information available. Overall, the research is well written, the methods are explained well, figures and tables are easy to digest and visually capture the key points of the research, and the results and discussion are sound.
> I recommend this manuscript be published after minor revisions. Specifically, there are a few methodological questions that need to be clarified and I would ask that the authors reconsider how they are wording their conclusion that forest canopy cover is not an important feature for automated ATES classification in light of the limitations of the validation data and quality of the forest data used. Feature importance in a random forest model is highly dependent on the training and testing data, so this conclusion may be specific to the study area of this research. Further, an optional addition that would be useful to situate these results in the broader field would be to compare the accuracy of the RF approach to the previously published 'deterministic' autoATES method.

Response: We thank John Sykes for the detailed, constructive, and encouraging feedback.

Here are our responses to the two main comments:

- **PCC feature importance**

We acknowledge and agree that the low feature importance of PCC observed in the final ATES classification stage may be influenced by several factors. As you noted, this could be related to the imbalance between forested and non-forested pixels in the training dataset, as well as to the potentially lower quality and resolution of the input PCC data. To investigate this observation further and avoid premature conclusions, we conducted several supplementary analyses.

First, we modified our Random Forest training script to preprocess the dataset and ensure a balanced representation of forested and non-forested pixels. Specifically, we determined the total number of forested pixels in our training polygons (considerably fewer than non-forested ones) and randomly sampled the same number of non-forested pixels, achieving a 50/50 ratio. Pixels with a PCC value of 0 were classified as non-forest, while those with a PCC value greater than 0 were considered forested. Using this balanced dataset, we trained model RF1 with its four features: slope, PCC, binary-thresholded PRA, and alpha angle. While the original unbalanced model assigned PCC a feature importance of 0.073788, the balanced dataset only slightly increased this value to 0.088802, leaving PCC as the least important feature.

We also explored alternative methods for generating PCC values using RapidEye imagery and DEM data combined with machine learning, as described in our in-review article for Forestry Ideas (Markov et al., 2025) and similar to approaches presented by Panayotov et al. (2024). When retraining our best-performing model (RF2) with these alternative PCC layers, we again observed consistently low PCC importance in the final ATES classification stage.

Based on these investigations, we have expanded the methods and discussion sections of our manuscript to document the balancing experiment and have rephrased our conclusion to be more cautious, suggesting that the low PCC importance may be a possibility that warrants further research rather than a definitive finding. To conclusively assess whether the PCC data quality is a contributing factor, future work should construct a high-accuracy, ground-truth PCC layer, for example using LiDAR-derived forest metrics, and re-evaluate feature importances.

Finally, we emphasise that the PCC layer plays a critical role in the earlier processing steps of our workflow, particularly in PRA calculations and `com4FlowPy` avalanche simulation runs. Our related studies (Markov and Panayotov, 2024; Markov et al., 2024) and the forthcoming Forestry Ideas paper (Markov et al., 2025) demonstrate that omitting PCC at these stages leads to substantial changes in PRA and runout modelling results. Our current hypothesis is that PCC's strong influence on these intermediate layers makes its direct inclusion in the final ATES classifier less impactful, as its effects are already implicitly captured by derived features. The manuscript has been revised to better explain this reasoning with appropriate references.

- **Comparison with deterministic AutoATES model**

  Thank you for this valuable suggestion. While we initially decided to omit this comparison from the manuscript, we agree that it provides important additional context. The "deterministic" AutoATES approach previously applied to the study region by Panayotov et al. (2024) represents a slight adaptation of the original AutoATES Austria model (Huber et al., 2023), which itself is a variation of AutoATES 2.0 (Toft et al., 2024).

  We conducted a basic, preliminary comparison between this deterministic model and our best-performing model, RF2, focusing on two representative subregions: a sparsely forested area and an alpine region above treeline. We added a figure in the appendix illustrating this visual comparison, as well as a table summarising the total area predicted for each ATES class by both models for the entire study area. Overall, the class distributions are very similar across the study area for both models. RF2 seems to predict more *challenging* terrain in steep, sparsely forested slopes, whereas the deterministic AutoATES model classifies these areas more frequently as *simple* terrain. Differences above treeline are minor. We have added some discussion about this to the manuscript. We believe that the overall similarity with a deterministic approach supports our hypothesis in research question 1 - namely that machine learning techniques can be used as an alternative to deterministic ones for the final ATES classification stage (even outperforming them in certain cases).

Intro

Line 15:

Global fatality numbers are much higher than I've seen in other recent publications (from Acharya et al 2023). Worth double checking, typical records from Europe and North America estimate closer to 140 annual fatalities. This additional research on Himalayan events is very meaningful, but differs from past estimates.

Response:

Thank you for pointing this out. We checked the references and could find the following (approximate) numbers.

- **US:** around 25 fatalities/yr ☞ `https://files.nwac.us/wp-content/uploads/2021/02/09091435/US-by-State-Avalanche-Fatalities-over-last-30-years.pdf`
- **CAN:** around $10-15$/yr ☞ `https://doi.org/10.1503/cmaj.081327`
- **JAP:** around 10/yr ☞ `https://arc.lib.montana.edu/snow-science/objects/ISSW2023_P2.15.pdf`
- **Russia:** ≈ 15/yr ☞ Seliverstov et al. (2008)
- **European Warning Services:** ≈ 120/yr ☞ `https://www.avalanches.org/fatalities/fatalities-statistics/`
- **High Mountain Asia:** ≈ 62/yr over last 50 yrs but 175/yr in the period 2010-2019 ☞ Acharya et al. (2023)
- **global estimates:** 250/yr ☞ Schweizer et al. (2015) and Acharya et al. (2023)

These numbers suggest that the combined annual number of avalanche fatalities in Europe (EAWS member countries) and North America (United States and Canada) is likely in the range of 150–160 fatalities per year, which aligns with the number you mentioned. When adding approximately 15 fatalities per year in Russia (including the Northern Caucasus), plus probably several cases from the Southern Caucasus (Georgia, Azerbaijan), and additional incidents in high-mountain Asia (an average of 62 per year over the past 50 years, excluding high-mountaineering accidents, according to Acharya et al. (2023)), as well as around 10 fatalities annually in Japan and likely a smaller but non-negligible number from the Andes region in South America, the global total plausibly exceeds 200 fatalities per year. A figure approaching 250 fatalities annually therefore appears reasonable, particularly when considering potential under-reporting in remote mountain regions outside Europe and North America. Also Schweizer et al. (2015) come up with a similar global estimate. However, it is also apparent that numbers of fatalities are subject to considerable intra-annual variations.

We have revised the introduction to emphasise that the number of global fatalities represent estimates rather than exact numbers and have added additional references to support this statement.

Lines 63 - 70:

If autoATES is working well for open terrain this points to a specific discrepancy in how forest data is captured. Therefore, another alternative to making the application of autoATES more consistent for novel regions would be improving consistency of input data sets (forest cover, DEM).

Response:

We agree with this point and have added more clarification in the manuscript stating that for forested regions, besides modifications to the classification tree, new methods to improve the quality of the input data, especially the PCC, can be tried out and evaluated.

Lines 72-78:

You describe the autoATESv2.0 approach to classifying terrain as expert based, which is accurate, but it is also a physically based model that uses the output of PRA and runout models to explicitly categorize terrain. Shifting to a machine learning approach is more of a statistical approach to classifying terrain using ATES. Can you add some discussion about the trade off of physical versus statistical models. I understand that machine learning is much easier to implement if you have high quality training data, but there are limitations in terms of generalizability and reliance on a limited set of training data. These trade-offs are critical to highlight to paint the full picture of switching from a physically based to statistically based modelling approach.

Response:

We generally agree on this point. However, we somewhat disagree on the classification of AutoATES as a physically-based model chain, since several steps of the model chain (PRA, runout modelling, and also the final ATES classifier) are also empirically motivated and involve heuristics and thresholds based on statistical analysis of observed avalanches. Nevertheless, we agree that ML methods strongly depend on the availability and quality of training data and that limitations exist with regard to model generalisability. We have added a paragraph at the end of the introduction to discuss the trade-offs involved in switching from a (more) physically-motivated to a more statistical approach.

Methods

Line 120-125:

Good description of the overall study area topography. You could move the discussion about the distribution of different ATES classes to this section if you want. I found myself looking for that information while reading this section, but I see that you have it nicely summarized in Table 1 and paragraph 1 of section 2.3.

Response:

We believe that the distribution of ATES classes is not an inherent characteristic of the study area but rather a mapping and modelling result. Therefore, we think that it is best if it remains in the section where we describe the training data. However, we added an additional sentence describing the general characteristics of avalanche terrain that can be found in the study region, without going into detail on distribution of ATES classes.

Figure 1:

Figure 1 is an excellent overview of your study area for those unfamiliar with the local geography. I assume the light green/teal shading is a rough approximation of treeline elevation, which may be worth including in the legend.

Response:

Thanks for pointing this out. Yes, the light green/teal shading in the overview map in Figure 1 presents a rough approximation of the areas covered by forests. The upper boundary of these areas corresponds to approximate tree-line level. The lighter green/teal shading represents areas mainly covered by dwarf mountain pines. We updated the figure legend in accordance with your suggestion to also provide information on these areas.

Lines 135 - 140:

Can you add information about how these DTMs were created. For example, using LiDAR, photogrammetry, or radar? Was the DEM data produced from satellite, UAV, or drone based remote sensing?

Response:

The digital terrain models (DTMs) used in this study were derived from historical military topographic maps of Bulgaria and are based on geodetic surveys and manual cartographic techniques. The DTM from Geopolymorphic Cloud, which covered a small portion of the study area, was created from digitising 1:5000 topographic maps, while the DTM from Pirin National Park was created from slightly lower resolution maps. We added more information about this in Sect. 2.2 in the manuscript, as well as an acknowledgement that these DTMs may have more limited accuracy in certain places, particularly in densely forested or rugged areas where direct observations were difficult and smaller terrain features may have been generalised or omitted.

Lines 122 - 148:

In my experience the quality of the input forest data has a very large impact on the quality of the output of the autoATES model. Did you consider creating your own forest data using free satellite imagery such as Sentinel 2? Considering the known limitations of the Copernicus forest data you mentioned, creating your own forest data could significantly improve autoATES performance.

Response:

We did investigate alternative methods for creation of PCC data - i.e. using Random Forests and RapidEye imagery and also working with UAV data (Panayotov et al., 2024). However for this study we wanted to focus on the ATES classification and keep everything else as simple as possible, eliminating new methodology and experimentation for those parts. Therefore, we chose to use the simple and ready to use, openly available Copernicius tree cover density dataset. In our in-review publication for the Forestry Ideas journal (Markov et al., 2025), we take a more in-depth look at the affects of creating and using different forest PCC layers for ATES classification. We have added a brief mention of this to the discussion part of our paper.

Line 149:

This sentence probably does not need its own paragraph.

Response:

We agree and we merged it with the previous paragraph.

Line 161:

Relying on one local expert to create the training data introduces a high degree of subjectivity to the machine learning approach. Prior research has shown that there can be major differences in how avalanche experts categorize terrain and apply the ATES scale. By relying on one local expert and using a machine learning approach you are putting a heavy emphasis on the skill of the local expert in driving the accuracy of your automated model. I recommend adding a statement along these lines to recognize the potential bias in your training data.

Response:

We fully agree. Unfortunately, due to limited resources in our region, our training data was only drawn by a single local expert. There is a brief acknowledgement of this as a potential limitation to our study in the discussion. We have also added a comment on the potential bias in this section that describes the training data as well.

Line 170:

Why did you decide not to include ATES class 0 terrain (non-avalanche?)

Response:

We chose to omit class 0 - *non-avalanche terrain* - because of the specific setting of our study area and also because of general concern regarding the automated delineation of a class representing terrain that is completely safe under all circumstances. In our study area, there exist only small areas that could be classified as *non-avalanche* terrain, most of which would have to be accessed by traversing higher graded terrain first. Also, as stated in Statham and Campbell (2025), the delineation of *non-avalanche* terrain requires a high level of confidence in the assessment, which is difficult to achieve. Therefore, most current AutoATES classifiers chose to leave out this terrain class, and we have done the same in our study.

Figure 2

This is a very interesting and novel approach to precisely define many small polygons and not create a continuous validation data set. The limitation of mapping continuous areas at high resolution is a significant challenge for developing validation data for autoATES.

Response:

Thank you for the comment - we agree that this is an interesting and novel approach. We believe that in future work, we could also try out the continuous approach for delineating training data. It would be very interesting to compare results from models trained on the small polygons with those trained on a continuous ATES map. We have briefly mentioned this as potential future work in the discussion.

Line 178

Why did you choose a 50/50 split for your training and testing data? To my knowledge a 80/20 or 70/30 split is more typical of the machine learning field.

Response:

We acknowledge that 50/50 train-test splits are less common than 70/30 or 80/20; however, they are an established and valid approach in machine learning research, as documented by Joseph (2022) and Afendras and Markatou (2019). While some models (e.g., neural networks) may require a larger training proportion, we believe that Random Forests can be effectively trained even with half of our dataset, leaving ample data for model evaluation. Importantly, allocating more pixels to the test set allowed for more robust validation and a better estimate of generalisation performance on new, unseen terrain—a key focus of our study, as we aim for these models to be transferable to other regions. Additionally, using a slightly smaller training set reduced computational requirements and helped limit potential overfitting.

Line 207

Assigning a value of 0 to slope angles below 28 degrees puts a lot of faith on the accuracy of your DTM. This could lead to missing PRA on small slopes where adjacent lower angle terrain can smooth the slope angle due to the neighborhood function used to calculate slope angle. Further, avalanches on slope angles below 28 degrees are possible with persistent weak layers, especially surface hoar which is notorious for causing avalanches on slope angles of 25 degrees or less. I would consider decreasing this cutoff value or removing the cutoff value entirely and fine tuning the slope angle cauchy function so that the 'fuzzy and' operator can handle these low angle slopes with consideration of forest cover and wind shelter.

Response:

Our motivation for using the cutoff value at 28° for the slope membership function was largely based on ensuring consistency/comparability with previous AutoATES studies in the study area (Panayotov et al., 2024) and other studies conducted by co-authors of the manuscript in Austria (Huber et al., 2023; Hesselbach, 2023). The mentioned studies also used this cutoff value in their applications of AutoATES and resulting binary PRAs were found to be in good agreement with observations (e.g. Hesselbach, 2023) and local expert assessment. Limiting the extent of PRAs to areas above 28° or even 30° is also in accordance with PRA models suggested by Bühler et al. (2013), Bühler et al. (2018), and Veitinger et al. (2016). While we only expect avalanches in our study area to release on slopes below 28° on rare occasions, we agree with your assessment of the potential problems associated with using the cutoff value in combination with DTMs of limited quality and resolution - especially potential misses of steep but short slopes.

In response to your comment we investigated the influence of using the slope cutoff at 28° vs. not using the slope cutoff (as proposed by Toft et al., 2024; Sykes et al., 2024) for our study region. With respect to the binary PRAs (with the threshold set at 0.3), we did not see any changes to modelled binary PRAs between the two methods (Fig. 1). We did see differences in the continuous PRA membership values between the two variants for our study area, with around 10 % of the pixels with a difference greater than 0 having a difference larger than 0.04 (see Fig. 2). We also retrained our model RF2 with the PRA_{cont} layer produced without the slope cutoff, but did not see any significant changes to (i) reported feature importance (ii) validation metrics on the test set or (iii) the final predictions for the whole study area and associated predictive confidences of the model. This is expected, since the absence of differences in the binary PRAs (PRA_{bin}), which are used as input to the runout and intensity modeling step, means that the differences do not affect other features and only introduce minor differences to the PRA_{cont} feature.

In light of these results, we see limited value in re-running the model and validation chain with a reduced or removed cutoff for the PRA-membership function for slope - also provided that this is not the main focus of the manuscript. Instead, we adapted the methods section (l. 206 ff.) to better reflect, that the choice of using the slope cutoff was motivated by maintaining consistency with previous applications of AutoATES in the study area. In addition we included a short paragraph in the discussion section (4.3. Limitations and potential for improvement), which (i) specifically summarizes potential drawbacks of using the slope cutoff (especially in combination with limited quality DTMs) in our study and (ii) more broadly discusses the transferability of models and parameterisations between study areas and across input data with varying resolution/quality.

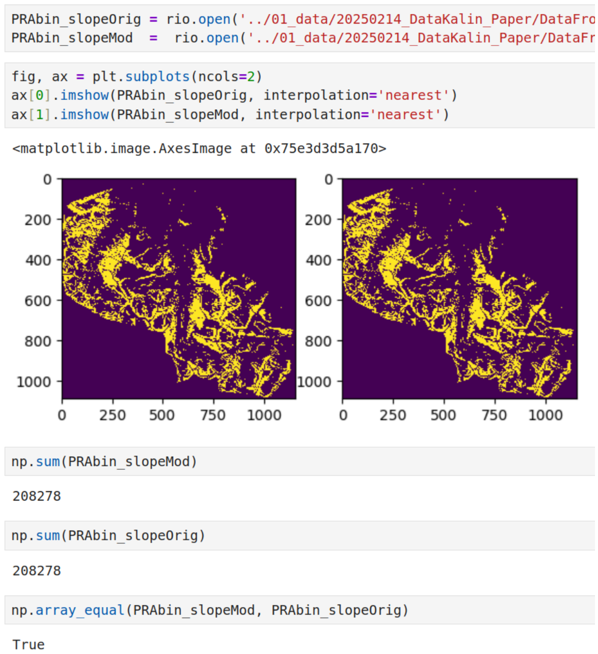
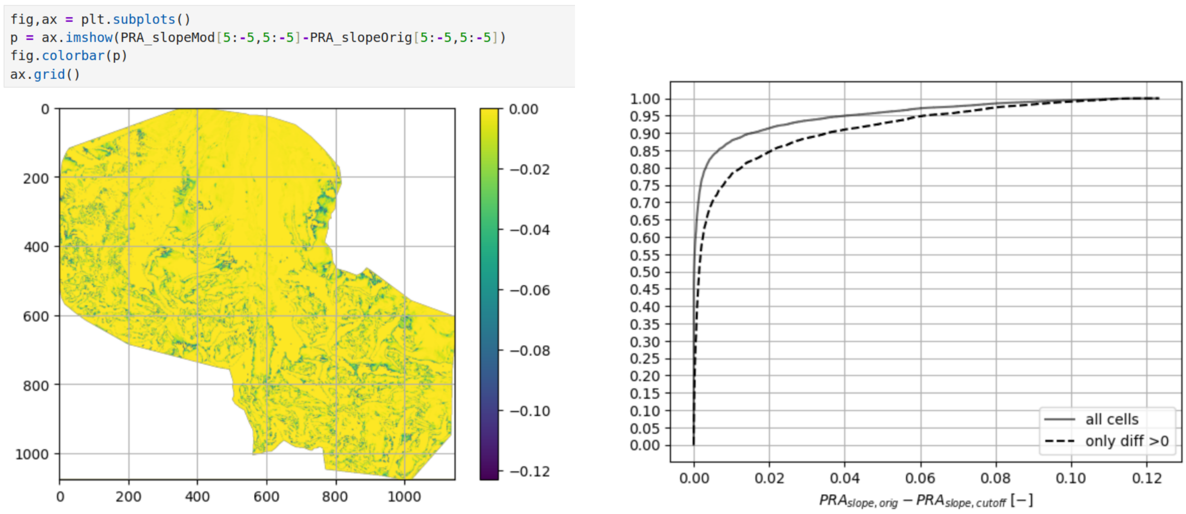


Figure 1: Comparison of binary PRAs calculated using the slope cutoff at 28° (right) vs. not using the cutoff (left). Both methods produce identical binary PRAs.



(a) (b)

Figure 2: Differences between PRA_{cont} without ($PRA_{slope,orig}$) and with ($PRA_{slope,mod}$) slope cutoff at 28°. Map of differences for the study area (a) and cumulative distribution of differences (b).

Line 225:

Why are you targeting/limiting your runout simulations to size 3 avalanches? The ATES v2 classification scale specifies return frequencies for avalanches greater than size 3. Based on your description of the terrain in the study area, with some slopes having 1000 m of vertical relief, there is a strong possibility of avalanches larger than size 3. How do you factor these very large to historic avalanche events into your ATES classification?

Response:

Thanks for pointing out this discrepancy. Your observation is correct - the terrain in our study area is certainly capable of producing avalanches larger than typical size three and has done so in the past (Panayotov and N. Tsvetanov, 2024). The reported targeting/limitation of the `com4FlowPy` parameterisation to large avalanches is a remnant of previous studies (Huber et al., 2023; Hesselbach, 2023; Spannring, 2024) which have produced AutoATES maps that were also compared to Swiss CAT maps (cf. Harvey et al., 2018; Harvey et al., 2024), which specifically target size 3 avalanches and below.

The used α angle in our study corresponds to the one reported by Huber et al. (2023). In that study the authors showed that although they were targeting size ≤ 3 avalanches, the best-fit parameterisation (based on a limited set of observed avalanches) tended to produce avalanches $>$ size 3 in $1/3 - 2/3$ of predictive simulations (depending on the used criterion for size classification - e.g. runout length, affected area, ...). Underestimation of predicted avalanche sizes was reported only in a minority of cases ($< 5\%$). This is very likely an effect of the previously reported correlation of α with characteristics of single avalanche paths - specifically the average track inclination β (Lied and Bakkehøi, 1980; Bakkehøi et al., 1983; Larsen, 2021; Toft et al., 2023) - that is not captured by a global α angle.

We used the parameters reported by Huber et al. (2023) and also previously used for producing AutoATES maps in the study area by Panayotov et al. (2024) as a starting point for our model parameterisation. In addition we utilised (i) empirical relationships proposed by Bakkehøi et al. (1983) (α angle in dependence to average track inclination β and total vertical drop H) and (McClung and Gauer, 2018) (z^{δ}_{lim} or equivalent v^{lim} in relation to path length S_0 and vertical drop H_0), (ii) comparative simulations carried out with the physically-based avalanche model `avaframe::com1dfa` (Tonnel et al., 2023), and records of past avalanches in the study area (e.g. Panayotov and N. Tsvetanov, 2024; M. Tsvetanov N. P., 2024) to refine our parameterisation. The utilised parameterisation for `com4FlowPy` also successfully reproduces runouts and intensities of avalanches larger than size 3 (see figures A.1. (c) to (f) in the manuscript) and modelled avalanche outlines are in reasonable qualitative agreement with outlines of historic avalanches (e.g. Panayotov and N. Tsvetanov, 2024) and local expert assessment.

We will update our description on the parameterisation of the runout model around line 223 of the manuscript to better reflect the used parameterisation process. Specifically, we will remove the misleading *"with a focus on capturing typical runouts of large avalanches"* and provide additional clarification on the parameterisation process and additional references.

It is also important to note that the training data was drawn by the expert to take into consideration potential avalanches larger than size 3 and label the potentially affected runout regions appropriately (usually as *challenging* terrain).

Table 4:

The RF model with the most input features only has 7 features. Why did you limit your feature selection to a relatively sparse set? Do you think it would be worthwhile to expand the set of features to include additional output from the PRA, additional forest data, or additional runout simulation information? One of the main benefits of machine learning methods is that they can handle very high dimensional data, which would support testing an RF model with more features.

Response:

We decided to limit the scope of the new features we added and keep it simple - one model with the input features used in AutoATES Austria, one with an extended set, and one with limited features based on the feature importances from the second model. In the extended model, we decided to include all major types of input features, without repetitions or derived features (except slope, PCC and PRA, where slope and PCC are also used to calculate PRA, but we left them, as slope is logically one of the most important features, while PCC is used in the deterministic AutoATES classifiers).

Our reasoning in having a maximum number of 7 features was based on:

i) keeping the models simple and in principle "backward-compatible"/"comparable" with classical decision-tree based classifiers in a sense that each feature would likely also be helpful for human interpreters, while also avoiding multicollinearity between features as much as possible.

Increased multicollinearity, for example, would have likely made interpretation of relative feature importances more complex.

In general, we think that exploring a larger feature set by including more input data used in the PRA creation and more output data from the runout modelling as well as additional forest data could potentially cause improvement in model performance and is worth trying out in future work. However, this would likely require the development of strategies to deal with potential overfitting and reduce ”backward compatibility” with traditional approaches. We have added more discussion on this topic in the manuscript, where we describe the features used for the models. We also added a section in the discussion stating that extending the input features used could be an area of future work.

Table 4:

The figures and tables in this section do a very good job of illustrating how the autoATES output looks on the terrain and providing a statistical summary of each model.

Response:

We agree, thank you for the comment!

Figure 8

It is interesting that slope is consistently the second highest in feature importance while PRA is near the middle or end of the feature importance list. The PRA output is largely driven by slope angle distribution, which makes me wonder why slope is so dominant here. Could there be an impact of your local expert using slope angle maps to create the training/testing data and therefore your RF models contain some bias towards weighting slope angle more heavily?

Response:

The training data was originally drawn by the expert based on his "expert opinion" of the terrain and visually inspecting satellite imagery, after which he was given a slope map to check that his polygons are precisely drawn and refine them if needed. Therefore, there is a chance that this has caused slight bias towards the slope feature having more importance in the Random Forest model training. However, we believe that slope should definitely be one of the most important features that determine avalanche hazard potential anyway, so this high importance is quite logical as well. We have added more details in the training data section on how the training data was drawn and refined by the expert and we have expanded the discussion on the feature importance of slope to acknowledge that there could be some bias towards its high feature importance due to its partial usage in the training data refinement.

Line 495

Do you think the limitation you mentioned in your forest data in the intro could be contributing to the lower performance for challenging terrain?

Response:

We believe the main reason *challenging* terrain is more difficult to predict is that the class encompasses a wider range of possible terrain - from runouts to short steep sections within otherwise not-so-steep slopes, steep forest, etc. This "middle ground" and variability makes it more difficult to predict. However, there is a possibility that particularly in forested areas, especially small forest gaps, the lower quality and resolution of the forest data may causes ATES classification errors there (so small gaps not being detected and being classified as *simple* instead of *challenging*). However, these narrow forest opening are usually very steep and the classifier has learned to classify such very steep areas in forests as *challenging* instead of *simple* terrain (due to the way the training data is drawn), so in most cases it should still catch these areas and classify them correctly.

However, in order to obtain a real answer to this question, we would need to run these models with different forest data that is known to be of higher resolution and quality and then comment on the differences based on the statistics. We have added a paragraph about this in the discussion section of the manuscript.

Line 500 to 503:

Yes! So maybe it would be worthwhile to include even more features to try and bump that accuracy even higher?

Response:

We agree with you that it would be worthwhile to try and add even more features. Please take a look at our reply to your previous comment on this subject. We have added discussion on this in the manuscript and believe that this is a great idea for future work.

Line 515 to 516

Agreed, this is a critical consideration for evaluating model performance for autoATES.

Response:

Yes, absolutely. It is important that hazard mapping models tend to err on the side of caution, leaning towards overestimation rather than underestimation of potential hazards.

lines 525 to 540:

There are several other potential reasons that could cause PCC to have a lower feature importance. First is the distribution of total forested versus non-forested validation pixels. Everywhere that there is no forest cover in your training data the PCC feature would not be useful for classification. Therefore, the relatively low ranking of PCC in feature importance is likely due to the fact that much of the terrain you trained and tested on are not forested. The second cause is that you highlight significant limitations in the forest data in the intro section. The quality of the input forest data will be directly related to how useful it is for ATES classification. I agree that the thresholding approach for forest cover in the current autoATES classification approach is cumbersome and could be improved, but that does not mean that forest cover is not a critical feature to include in the ATES classification model. Finally, just because a feature has lower importance doesn't mean it is not contributing to a better overall classification. As you stated on line 500, machine learning models excel at incorporating many features into the classification. So even if forest cover is only useful for 5-10% of the pixels in your data set, that does not mean that excluding it is the correct approach. The fact that RF2 is the most accurate and includes the most features (including PCC) is a good justification for keeping it. Adding additional features beyond what is included in RF2 might not produce new features with very high ranking feature importance values, but it could incrementally improve accuracy for specific types of terrain where the current model is lacking. Overall, I would consider these factors carefully before making a general statement that forest cover is not an important feature for automated ATES classification.

Response:

Thank you for the valuable comment. We understand your reasoning and agree with your points. Please see our response in the top section of the review file, where we address this topic in more detail. We have updated the manuscript to take a more cautious stance in our conclusion and revised the wording accordingly, as we agree that even features with relatively low importance values can still provide useful information in machine learning models.

Section 4.2.2

This adaptation of using isolated polygons to validate autoATES is a novel and meaningful addition to the field. However, this is also a very different approach from traditional ATES mapping using linear or zonal features which may have some challenges in regards to defining boundaries between classes and incorporating the ATES elements of exposure and route options into the terrain ratings.

Response:

Yes, definitely. We believe that the key advantage is that this approach allows the expert to shift the focus from "determining ATES class boundaries" and instead draw areas where he/she is very sure about the classification, leaving the boundary determination work to the machine. We think that it would be very useful to also perform training of Random Forest classifiers using continuously drawn training data and then compare the results. We have added this a potential future work in the discussion of our paper.

Line 590 to 600

This is a huge advantage of the RF approach to ATES classification. Taking advantage of efficiency of the automated approach while being able to manually validate specific pieces of terrain that are identified as low confidence. Great example of merging automated and manual mapping to create the best possible output for a lowest possible cost.

Response:

Yes, definitely, we agree. This could be very helpful for manually tuning automatically produced maps.

Section 4.3

One additional limitation could be that the RF model will probably be limited to working with input data that is very similar to what it is trained on. Using a lower resolution DEM or a forest cover dataset that captures a different forest characteristic (e.g. basal area, stem density) would likely not work with the RF model developed in this research. Therefore, the RF autoATES model presented here is likely limited to application in regions with similar topography, forest characteristics, and input data availability.

Response:

Yes, definitely, we agree. The forest dataset needs to be expressed as PCC (ranging from 0 to 100%), as this is the format on which the model was trained; using other forest metrics such as basal area or stem density would not be compatible with the current models. The trained RF models presented in this study are likely to perform best in regions with similar topographic and data characteristics. However, it would be valuable to test their applicability in different mountainous regions worldwide to identify where they perform well and where they fail, thereby enabling future improvements. Retraining would be necessary if alternative forest data or input datasets of substantially different quality or resolution were to be used. While this represents a potential limitation of the current approach, it also demonstrates that with proper retraining, the general methodology developed here should be transferable and applicable globally. We have added a section in the discussion acknowledging this limitation.

Line 647 to 651

See comments from discussion section about limitation of feature importance rankings. I think the lack of importance is more reflective of the training/testing data used in this study and quality of the input forest data and not a sign that forest cover is not an important parameter for autoATES mapping.

Response:

Thank you for the valuable comment. We understand your reasoning and agree with your points. Please see our response in the top section of the review file, where we address this topic in more detail. We have updated the manuscript to take a more cautious stance in our conclusion and revised the wording accordingly, as we agree that even features with relatively low importance values can still provide useful information in machine learning models.

Authors' Response to Reviewer 2

The originally posted comments by Cameron Campbell can be accessed at https://doi.org/10.5194/egusphere-2025-2143-RC2. We reprint the comments here along with our responses

General Comments. This preprint manuscript summarizes the development and validation of an automated Avalanche Terrain Exposure Scale (ATES) classification algorithm using random forest machine learning models. The study area encompassed popular backcountry ski-touring destinations in the Pirin Mountains of Bulgaria, with limited information available to the public regarding current snowpack stability and avalanche danger or the spatial distribution of avalanche-prone terrain. The random forest machine learning approach was investigated as a potential data-driven method to improve classification performance over previous automated ATES (AutoATES) mapping for the area that relied on expert-driven classification trees.

Three different iterations of the machine learning model were developed using a different selection of input features informed by a training dataset consisting of isolated manually classified ATES polygons. A selection of established statistical methods was then used to assess the agreement between the resulting AutoATES classifications and an independent test dataset. The results were used to evaluate the utility of random forest machine learning for AutoATES classification and optimize model input features for the study area.

Terrain classification for the study uses ATES v.2 with four avalanche terrain classes ranging from Class 1 – Simple to Class 4 – Extreme; however, the optional Class 0 – Non-avalanche Terrain class is not used. The manuscript would benefit from a discussion on the importance of identifying non-avalanche terrain for the study area and potential end-users, and the decision to exclude it from the study. The manuscript could also benefit from more analysis and discussion focused on the areas of disagreement between the AutoATES classification and the test dataset, especially areas where the disagreement is more than one ATES class, areas near the boundaries between ATES class zones, or critical terrain features for backcountry recreational route-planning and safe navigation (e.g., ridge crests, valley bottoms, high mountain passes, and terrain traps).

General Comments. Overall, the manuscript is clear, concise, and well-structured, and addresses the research questions well. The scientific and technical approaches and the applied methods are valid, and the results are discussed in an appropriate and balanced way. The work represents a substantial contribution to the understanding and communication of avalanche terrain severity in Bulgaria, and the development and validation of AutoATES algorithms worldwide.

Response:

We would like to thank Cameron Campbell for the detailed and constructive feedback. Here are our responses to the two main comments:

- **Exclusion of the *non-avalanche terrain* class**

  We chose to omit class 0 - *non-avalanche terrain* - because of the specific setting of our study area and also because of general concern regarding the automated delineation of a class that represents terrain entirely safe under all circumstances. In our study area, there exist only small areas that could be classified as *non-avalanche terrain*, most of which would have to be accessed by traversing higher graded terrain first. Also, as stated in Statham and Campbell (2025), the delineation of *non-avalanche terrain* requires a high level of confidence in the assessment, which is difficult to achieve with automated methods (reported overall accuracies in our study and previous studies are in the range of 80 %). Therefore, most current AutoATES classifiers choose to leave out this type of terrain as a separate class, and we have done the same in our study. We have added few sentences with information on our decision to the training data section. It is important to note that the *simple* ATES terrain class in the training data also includes some *non-avalanche terrain*, and in this way the classifiers learn to treat *non-avalanche terrain* as *simple*, without having to classify it as its own class.

- **Analysis of model errors**

  We fully agree that it is important to take a more detailed look into model classification errors to better understand the behaviour and limitations of the machine learning approach used in this study. In methods such as Random Forest classifiers, as applied here, tracing the decision logic behind individual pixel-level predictions can be challenging. These models rely on a large number of decision trees, each comprising hundreds of rules and thresholds, making it difficult for a human analyst to fully reconstruct the exact reasoning that led to a specific classification outcome. Nevertheless, we have made a concerted effort to investigate misclassifications in greater detail, with particular attention to cases where the predicted ATES class deviated by more than one level from the test set label (hereafter referred to as "severe misclassifications").

As stated in the manuscript, the majority of misclassifications are off by only a single ATES class. Errors exceeding one class level are rare, accounting for less than 1% of test pixels for each class. The most common confusion occurs between the *challenging* and *complex* classes—a finding consistent with previously published AutoATES classifiers, as these two intermediate classes often represent the most difficult terrain categories to distinguish. Table 5 illustrates that, for all models tested, misclassifications within the *challenging* terrain class predominantly represent overclassifications (i.e., predicted as *complex* rather than *simple* terrain). Such overclassification is preferable in hazard detection tasks, as it errs on the side of caution. Similarly, the majority of errors for *complex* terrain are confusions with the *challenging* class. The *simple* and *extreme* terrain classes, representing the two ends of the ATES spectrum, are classified with higher accuracy, as expected, and almost all misclassifications in these categories are off by only one class level.

In addition, we conducted an in-depth analysis of the rare cases where model predictions deviated by more than one ATES class. We identified two main underlying causes:

1) Training and test data generalization. Despite refinement using slope maps, we noticed that some polygons in the original ATES training data contained minor inaccuracies in their boundaries, particularly in shaded areas with abrupt transitions between steep gullies and adjacent mellow or forested terrain. In

these cases, human-drawn boundaries could be offset by a few meters, leading to apparent misclassifications by the model that, in reality, reflect small cartographic inaccuracies in the reference dataset rather than true model errors.

2) Model limitations at sharp ridges and cliff edges. In other cases, the expert-drawn map accurately represented terrain conditions, while the model misclassified pixels located along narrow ridge tops adjacent to vertical cliffs. We traced these errors to limitations in the underlying DTM: certain pixels in these boundary zones were assigned artificially flat slopes extending into cliff areas, resulting in lower ATES predictions than expected. These discrepancies are likely due to both DTM resolution constraints and interpolation artifacts in highly rugged terrain, where the narrowness of ridges and abrupt cliff drops challenge pixel-based terrain analysis.

Importantly, these severe misclassifications represent an extremely small fraction of the total test set and correspond to isolated outlier pixels rather than systematic errors.

We have expanded the discussion section of the manuscript to include this additional analysis, clarifying both the rarity of these cases and their underlying causes.

Line 170 - 173:

Small, precisely delineated polygons of manually assessed terrain were used for the training and test datasets in lieu of conventional continuous ATES zone mapping in order to reduce generalization errors. From Figure 2, it appears as though the locations of these polygons are somewhat random, the dimensions of these polygons range from less than 100 m to almost 1000 m, and some seem more precisely delineated than others (i.e., rounded or squared-off boundaries versus complex precise shapes). However, there are no details on the approach used in identifying or delineating these polygons. It is also unclear whether the boundaries of these polygons represent the exact transition between ATES classes, or if the transition is considered to be somewhere outside of these polygons. The manuscript could benefit from more details with this regard.

Response:

Thanks for pointing this out. We agree that the description of the utilised training data set in the original manuscript lacks detail with respect to the practical approach used to delineate the training polygons. As also mentioned in previous comments and responses to comments from Reviewer 1, the construction of training data is a crucial step in the presented study and warrants a more in-depth description.

The training data was originally drawn by the expert based on his "expert opinion" of the terrain and visually inspecting satellite imagery, after which he was given a slope map to check that his polygons are precisely drawn and refine them if needed. Some of the polygons, particularly those representing *extreme* terrain, were iteratively refined using slope maps to ensure they included only very steep terrain and cliffs and not the skiable gullies in between them. While the borders of most polygons do not reflect actual ATES class boundaries but rather places where the expert was very sure of the classification (which represents the advantage of this type of training data rather than a continuous map), a few areas, such as transitions from wide, flat ridgetops to large, steep slopes, were deliberately drawn as adjacent simple and complex terrain, which helped the classifiers better learn these transitional patterns. Also, it is important to note that the different sizes of training polygons correlate to the extent of the associated terrain features. While polygons delineated in gullies or cliff bands (*extreme*, *complex*) can have smaller dimensions, polygons drawn on larger terrain features are larger.

We extended the description of the training data in the manuscript, also providing a more detailed description of the actual training data delineation process.

Line 260:

The manuscript could benefit from discussion regarding the decision to omit the ATES v.2 *non-avalanche terrain* classification from the analysis. This ATES class can default to simple terrain and is generally considered optional as it often requires high confidence in the assessment (and associated level of effort). However, it can provide valuable information to end-users with little or no tolerance for avalanche risk.

Response:

Yes, we agree that the manuscript can benefit from a discussion on our handling of the *non-avalanche terrain* class. As pointed out by you and also discussed in Statham and Campbell (2025) the delineation/mapping of *non-avalanche terrain* requires a high level of confidence in the assessment and is considered optional with the possibility to implicitly include it in the *simple terrain* class.

We discussed on using *non-avalanche terrain* as a separate class, but refrained from it for several reasons:

- We agree that the definition of a separate *non-avalanche terrain* class might provide valuable information to end-users with little to no risk tolerance; however, we believe that the process of assigning a *non-avalanche terrain* class should ultimately be based on manual assessment, rather than purely automated models. The reason is that the confidence in the assessment has to be high, and also take into account additional factors that are not included in the AutoATES model chain (e.g. avalanche control work, nivo-meteorological variables), which can be incorporated in the decision.

- While we can certainly identify spots of *non-avalanche terrain* also in our study area, we believe that this information is of limited value to end-users, specifically if they have to cross higher rated terrain to get there, i.e., *non-avalanche terrain* is only meaningful if it can be accessed through *non-avalanche terrain*.
- As long as areas for mapping focus on alpine environments (like our study area) and do not include ”definitive” *non-avalanche terrain* (e.g. bottoms of large valleys clearly away from steep slopes) we think it’s ok to include small patches of *non-avalanche terrain* in the *simple* class.

We added a short discussion on our use/omission of the *non-avalanche terrain* class in section 2.3 where we also describe the training data set and its creation. Specifically, we state that *non-avalanche terrain* is included in our definition of *simple* terrain when constructing the training set. AutoATES, AutoATES v.2.0 and applications (Sykes et al., 2024a) also do not include *non-avalanche terrain* as its own class.

line 400:

The manuscript could benefit from further analysis and discussion regarding the classification errors that involved misclassification by more than a single class level. Is this largely attributed to errors in input data or are there specific terrain features that the model grossly misclassifies?

Response:

Thank you for the valuable comment. Please see our response to your general comment above for an in-depth discussion on the topic.

line 427 - 429:

The manuscript could benefit from further analysis and discussion in regard to the finding that predictive confidence tends to decrease near the boundaries between classes. Can this be related to the approach used in delineating the boundaries of the manually derived polygons used for the training and test datasets? I.e., see comment above, it is unclear whether the boundaries of these polygons represent the exact transition between ATES classes, or if the transition is considered to be somewhere outside of these polygons.

Response:

As stated in the comments above, we have added a detailed explanation of the training data creation and its characteristics. Most of the polygon boundaries do not represent actual ATES class boundaries but rather places where the expert was very sure of the classification. However, we do not think that the lower confidence near the ATES class boundaries is a result of this but rather because the feature values in those regions are more in the middle ground and less distinctively associated with a single ATES class in most cases, resulting in the models being less confident, as expected. In order to see if the smaller confidence in these regions could be tied to the training data format, we would need to draw a new type of training data that represents a continuous map and then rerun the models and see if the confidence results change. This is an excellent idea for future work. We have added information about this in the manuscript and suggested that it could be tried out in the future.

line 470 - 472:

It could also be noted that the analysis performed by Sykes et al. (2024) included validation against manually derived benchmark maps with areas classified as non-avalanche terrain that were subsequently reclassified as simple for the analysis. This increased the agreement rate between the AutoATES algorithm and the test dataset for the simple terrain class.

Response:

Thank you for the input. We included this information and updated the paragraph. In our study, some of the *simple* terrain drawn in the training set was *non-avalanche terrain* - in this way, we tried to get the models to learn to classify *non-avalanche terrain* as *simple*.

line 487 - 489:

The manuscript could benefit from further discussion regarding the influence of potentially including non-avalanche terrain classified as simple terrain in the test and training datasets on the model accuracy results. I.e., it is expected that the model would be able to easily classify non-avalanche terrain as simple terrain.

Response:

We acknowledge this point and have clarified in the revised manuscript that certain pockets of *non-avalanche terrain* were intentionally included and labelled as *simple* terrain in the training set. We expect this direction of classification to be reliable, as *non-avalanche terrain* should consistently be identified as *simple* terrain given its characteristic features, which are generally even more distinct from *challenging* terrain than other *simple* terrain areas. This is reflected in the high skill scores achieved for the *simple* class in our results.

Conversely, distinguishing between *non-avalanche terrain* and *simple avalanche terrain* may pose a greater challenge for the model, as these two classes are more closely related and share similar feature distributions. We have expanded the discussion in the manuscript to explicitly address the model's ability to correctly classify *non-avalanche terrain* as *simple* terrain and to highlight potential challenges in differentiating these two categories.

line 595 - 598:

The manuscript could benefit from further discussion regarding the importance of manual quality control and fine-tuning of automatically produced ATES maps prior to final map publication. It is the disagreement found between the AutoATES output and the test dataset that necessitates this.

Response:

We thank the reviewer for this valuable comment. We fully agree that manual quality control and fine-tuning of automatically produced ATES maps is essential before disseminating these products to end-users.

Our machine learning approach using Random Forests achieves an overall classification accuracy of approximately 80%, with maximum accuracy for individual classes reaching around 90%. These values are comparable to those reported in previous studies. However, they also clearly highlight the need for post-processing to ensure high-quality final products. In particular, it is crucial to verify the results and apply manual corrections, especially in areas where more detailed local knowledge and expert information are available. Furthermore, the confidence outputs of the Random Forest model provide valuable guidance for this process: areas with low predictive confidence can be specifically targeted for fine-tuning and manual review, thereby improving both the accuracy and reliability of the final ATES maps. We have expanded the discussion in the revised manuscript to emphasise this point.

Technical Corrections

There is inconsistent use if the acronyms "ML" and "RF" throughout the report. Suggest defining acronyms on first use then consistent use throughout.

There is inconsistent and inappropriate use of capitalization throughout the References section (e.g., lines 691, 699-700, 712-713, 717, 730-731, 756-758, 783-784, 804) and some authors names are missing (e.g., lines 732, 740, 829).

Response:

We thank the reviewer for pointing this out and for the thorough reading of the manuscript. We have reviewed the use of the acronyms *ML* and *RF* throughout the text to ensure consistency and have revised the references section accordingly.

References

Acharya, Anushilan, Jakob F. Steiner, Khwaja Momin Walizada, Salar Ali, Zakir Hussain Zakir, Arnaud Caiserman, and Teiji Watanabe (2023). "Review article: Snow and ice avalanches in high mountain Asia – scientific, local and indigenous knowledge". In: *Natural Hazards and Earth System Science* 23.7, pp. 2569–2592. DOI: 10.5194/nhess-23-2569-2023.

Afendras, Georgios and Marianthi Markatou (2019). "Optimality of training/test size and resampling effectiveness in cross-validation". In: *Journal of Statistical Planning and Inference* 199. DOI: 10.1016/j.jspi.2018.07.005.

Bakkehøi, S., U. Domaas, and K. Lied (1983). "Calculation of snow avalanche runout distance". In: *Annals of Glaciology* 4, pp. 24–29.

Bühler, Y., S. Kumar, J. Veitinger, M. Christen, A. Stoffel, and Snehmani (2013). "Automated identification of potential snow avalanche release areas based on digital elevation models". In: *Natural Hazards and Earth System Sciences* 13.

Bühler, Y., D. von Rickenbach, A. Stoffel, S. Margreth, L. Stoffel, and M. Christen (2018). "Automated snow avalanche release area delineation – validation of existing algorithms and proposition of a new object-based approach for large-scale hazard indication mapping". In: *Natural Hazards and Earth System Sciences* 18, pp. 3235–3251.

Harvey, S., M. Christen, Y. Bühler, C. Hänni, N. Boos, and B. Bernegger (2024). "Refined Swiss Avalanche terrain mapping CATV2 / ATHV2". In: *Proceedings, International Snow Science Workshop, Tromsø, Norway, 2024*. Tromsø, Norway, pp. 1637–1644.

Harvey, S., G. Schmudlach, Y. Bühler, L. Dürr, A. Stoffel, and M. Christen (2018). "Avalanche Terrain Maps for Backcountry Skiing in Switzerland". In: *Proceedings, International Snow Science Workshop, Innsbruck, Austria, 2018*. Innsbruck, Austria, pp. 1625–1631.

Hesselbach, C. (2023). "Adaptation and Application of an Automated Avalanche Terrain Classification in Austria". MA thesis. University of Natural Resources and Life Sciences, Vienna.

Huber, A., C. Hesselbach, F. Oesterle, M. Neuhauser, M. Adams, M. Plörer, L. Stephan, H.B. Toft, J. Sykes, C. Mitterer, and J.-T. Fischer (2023). "AutoATES Austria - Testing and application of an automated model-chain for avalanche terrain classification in the Austrian Alps". In: *Proceedings, International Snow Science Workshop, Bend, OR, USA, 2023*. Innsbruck, Austria, pp. 1272–1278.

Joseph, V. Roshan (2022). "Optimal ratio for data splitting". In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 15.4, pp. 531–538. DOI: 10.1002/sam.11583.

Larsen, H. T. (2021). "Using random forest to predict empircal avalanche runout". Masters' Thesis. University of Oslo, Department of Geosciences.

Lied, K. and S. Bakkehøi (1980). "Empirical calculations of snow-avalanche run-out distance based on topographic parameters". In: *Journal of Glaciology* 26.94, pp. 165–177.

Markov, K. and M. Panayotov (2024). "Avalanches in the Bunderitsa valley in Pirin Mountains". In: Intel Entrans, Sofia. Chap. Defining of avalanche terrain with ATES modelling for the region of Bansko ski resort in Pirin.

Markov, K., M. Panayotov, E. Tcherkezova, and A. Huber (2025). "Forest percent canopy cover influence on automated avalanche terrain exposure scale (ATES) classification in Pirin Mountain, Bulgaria". In: *Forestry Ideas*.

Markov, K., M. Panayotov, E. Tcherkezova, and M. Teich (2024). "Identifying and mapping avalanche terrain using the ATES model for the region around Bansko ski resort, Pirin". In: *Forest Science (Nauka za Gorata)* 60, pp. 99–122.

McClung, D.M. and P. Gauer (2018). "Maximum frontal speeds, alpha angles and deposit volumes of flowing snow avalanches". In: *Cold Regions Science and Technology* 153, pp. 78–85.

Panayotov, M., K. Markov, N. Tsvetanov, A. Huber, C. Hesselbach, and M. Teich (2024). "Avalanche hazard mapping using the Avalanche Terrain Exposure Scale (ATES) in the high mountain ranges of Bulgaria." In: *Proceedings, International Snow Science Workshop, Tromsø, Norway*, pp. 1660–1667.

Panayotov, M. and N. Tsvetanov (2024). "Dating of avalanches in Pirin mountains in Bulgaria by tree-ring analysis of Pinus peuce and Pinus heldriechii trees". In: *Dendrochronologia* 85, p. 126206.

Schweizer, J., P. Bartelt, and van Herwijnen, A. (2015). "Snow Avalanches. Snow and Ice-Related Hazards, Risks and Disasters". In: ed. by J. F. Shroder, W. Haeberli, and C. Whiteman. Elsevier. Chap. Snow Avalanches, pp. 395–436. DOI: doi:10.1016/b978-0-12-394849-6.00012-3.

Seliverstov, Y., T. Glazovskaya, A. Shnyprakov, Y. Vilchek, K. Sergeeva, and A. Martynov (2008). "Assessment and mapping of snow avalanche risk in Russia". In: *Annals of Glaciology* 49, pp. 205–2009.

Spannring, P. (2024). "Comparison of two avalanche terrain classification approaches: Avalanche Terrain Exposure Scale - Classified Avalanche Terrain". MA thesis. University of Innsbruck, Faculty of Geo- and Atmospheric Sciences.

Statham, G. and C. Campbell (2025). “The Avalanche Terrain Exposure Scale (ATES) v.2”. In: *Natural Hazards and Earth System Sciences* 25, pp. 1113–1137.

Sykes, J., H. Toft, P. Haegeli, and G. Statham (2024a). “Automated Avalanche Terrain Exposure Scale (ATES) mapping – local validation and optimization in western Canada”. In: *Natural Hazards and Earth System Sciences* 24, pp. 947–971.

Sykes, J., H. Toft, P. Haegeli, and G. Statham (2024b). “Automated Avalanche Terrain Exposure Scale (ATES) mapping – local validation and optimization in western Canada”. In: *Natural Hazards and Earth System Sciences* 24, pp. 947–971.

Toft, H. B., K. Müller, J. Hendrikx, C. Jaedicke, and Y. Bühler (2023). “Can big data and random forests improve avalanche runout estimation compared to simple linear regression?” In: *Cold Regions Science and Technology* 211, p. 103844.

Toft, H. B., J. Sykes, A. Schauer, J. Hendrikx, and A. Hetland (2024). “AutoATES v2.0: Automated Avalanche Terrain Exposure Scale mapping”. In: *Natural Hazards and Earth System Sciences* 24, pp. 1779–1793. DOI: 10.5194/nhess-24-1779-2024.

Tonnel, M., A. Wirbel, F. Oesterle, and J.-T. Fischer (2023). “AvaFrame com1DFA (v1.3): a thickness-integrated computational avalanche module – theory, numerics, and testing”. In: *Geoscientific Model Development* 16, pp. 7013–7035.

Tsvetanov N.and Panayotov, M. (2024). “Avalanches in the Bunderitsa valley in Pirin Mountains”. In: Intel Entrans, Sofia. Chap. Dendrochronologicalreconstruction of large magnitude avalanches in Bunderitsa valley.

Veitinger, J., B. Sovilla, and R. S. Purves (2016). *Model code of release area algorithm.*