Revisions for “Tracking the slopes: A spatio-temporal prediction model for backcountry skiing activity in the Swiss Alps using User Generated Content”

Leonie Schäfer, Frank Techel, Günter Schmudlach, Ross Purves

Thanks to both referees for their supportive and constructive comments. We have written a detailed response to all of the questions below and would be happy to edit the manuscript to incorporate these useful suggestions.

Referee 1: Anonymous

Comment	Response
My main issue is that -- while temporal cross validation is applied -- a corresponding spatial cross-validation is missing.	We agree that a spatial cross-validation would be interesting. We did consider this, and there are several issues which make carrying out spatial cross-validation challenging. 1) The track data (GPS tracks) are a relatively small dataset. Carrying out a spatial cross-validation with these data would be challenging - especially if we went beyond a leave one out approach. 2) Randomly leaving out single warning regions, which would allow us to carry out spatial cross-validation, is unlikely to show much. We already know that click and track behaviour is spatially autocorrelated, so it would be necessary to think through a more complex sampling strategy, where (1) becomes a problem. 3) Our results show that the model output varies spatially (Figure 10b) suggesting that we are capturing spatial variation in the relationship between clicks and model inputs. We propose adding a few sentences discussing these issues in the limitations.
149 "Between 2013 and 2024, over 6’800 GPS tracks were uploaded by backcountry recreationists throughout all seasons except 150 for seasons 21/22 and 22/23." ... I assume this means that the last season included in the track dataset is 23/24. Can you include information on how many individual skiers	The track data unfortunately doesn’t come with user IDs, therefore the number of unique users cannot be estimated. The data was purposefully sampled by Skitourenguru GmbH and was not collected in the years 2021-2023. We understand the

contributed GPS tracks to the database? Is the number of users per year stable? What happened to the data from 21/22/23?	confusion and will elaborate more on the data collection process in the revised manuscript. See also comment of referee 2 regarding line 61 on GPS data representativeness.
164 "Therefore, only data from 2021 onwards is included for modelling and prediction," ... so for prediction, we only have an overlap between click and track data in 20/21 and 23/24? (This seems to be confirmed by Table 2 & 3 but might be worth making explicit in the data section.)	Yes, for prediction we have only 2 seasons that overlap. Given the spatial and temporal granularity of each season, we believe that this is enough for a comparison between the two prediction models. We will make this explicit in the data section.
209 "mean values were calculated based on the grid points that lie in an elevation band within ±100 m of the mean track elevation (track data), respectively the mean route elevation in a given region (click data)" ... Wouldn't it make sense to further limit the weather grid cells using a maximum distance to actual skiing routes?	For the track data, we don't have any skiing routes, just the elevation of the track, due to obfuscation before we receive the data. Because we want to keep the calculation of variables equal for both data sources, we opted to use the 100m- elevation band. We agree that we could limit the elevation belt to regions near official skiing routes (since we could represent these with another dataset) and assume that most people will stick to these routes. However, due to the coarse resolution of the meteorological data (1 km), we believe that the resulting differences will be minor. In practice, since warning regions have an area of around 200km^2, the maximum possible distances from the actual locations (assuming dimensions of 10 x 20km) are likely to be of the order of 10km even if we assume no topography, and much less given that we incorporate mean elevation bands (c.f. Figure 2b) - we have added a 1km grid for our meteorological data to this figure to emphasize this point.
272 "This approach resulted in four (nine) training runs, each cross-validated using four (nine) different seasons for the click (track) data." ... To make this sentence easier to for the reader, I suggest to reword it instead of putting the track model info in brackets.	Thank you for the comment, we will adjust the sentence to make it more readable.
348 "The underlying driver for the systematic overprediction of the track model lay in the modelling process itself, as artificially balanced 350 numbers of presence and absence points were used for training. When	We'll edit this as suggested to use the present tense.

verified with real-life and therefore unbalanced data, the model predicted more presence than was observed." ... Please check if the use of past tense "lay" is appropriate or if present tense "lies" should be used since the model was not adjusted after the issue was discovered and all presented results are from the overpredicting model.	
386 "Figure 9 shows the importance for each variable for the performance of the model" ... Should probably be "importance of each variable".	Thanks for the comment, we will adjust this.
387 "from each cross-validation seasons" ... Should probably be "season".	Thanks for the comment, we will adjust this.

Referee 2: John Sykes

Comment	**Response**
Introduction	
The introduction provides a well written broad overview of the existing literature for estimating base usage rates of backcountry skiers. The literature encompasses a variety of techniques and identifies strengths and shortcomings of each approach.	Thank you for this helpful and positive feedback.
Line 52 to 56 - The knowledge gap is clearly identified.	
Line 61 - The research questions are well defined	

One question is why only use the Skitourenguru app as input data? This could introduce significant bias to the data set based on the characteristics of the users of the app. Broadening the data to include multiple apps (e.g. White Risk, Strava) could provide an interesting comparison and help determine if patterns apply generally or are specific to the user group of one specific application.	Thank you for the comment. We agree that the specific user base of a certain application has an impact on the data, and that the ideal solution would be to combine multiple data sources. The GPS tracks that we use were *compiled* by Skitourenguru, but they come from 3 different data sources: camp2camp.org, skitourenguru.ch and gipfelbuch.ch. These are three popular mountaineering platforms in Switzerland, where people upload condition reports (climbing, ski touring), often together with a GPS track. We will add this information to the data section and also add the technical documentation of the GPS data as a reference (ARPD_Manual_3.1.2.pdf). Further, subsets of the data were already used in previous publications, e.g.,: ● Techel et al., 2014: https://doi.org/10.5194/nhess-15-1985-2015 ● Schmudlach et al., 2018: Proceedings, International Snow Science Workshop 2018, Innsbruck, Austria) ● Winkler et al., 2021: https://doi.org/10.1016/j.coldregions.2021.103299 ● Degraeuwe et al., 2024: https://doi.org/10.1016/j.coldregions.2024.104169). We will mention this in the revised manuscript. Further, we applied for data from Strava but were not granted access. We fully agree that we have very limited knowledge on the user base represented in the data and that this is a limitation of our work which we will emphasize in the limitations.
Methods	
Line 141 to 147 - Does Skitourenguru require a paid subscription to use? Is the data from this study collected only from paid subscribers? Does the app only cover the Swiss Alps or does it also cover other areas? This type of information about the app is relevant to the	Thank you for this important remark. No, Skitourenguru doesn't require a subscription to use. Anyone can access the website and search for tours, or upload GPS tracks. We will make this clear in the manuscript.

sample demographics and could give a better sense of how accessible the website is to different users. For example, individuals just getting into backcountry skiing or those visiting from other regions may be less likely to pay for a paid application specific to Switzerland and therefore could be systematically excluded from the sample.	The data analysis was limited to Switzerland for several reasons: Most importantly, GPS data was only available for Switzerland. In addition, there are some inconsistencies on how the avalanche danger levels are used in the Alpine countries as shown by Techel et al. (2018) (https://doi.org/10.5194/nhess-18-2697-2018). We will add a brief remark on this in the revised manuscript.
Line 155 - Do you extract the terrain characteristics of the GPS tracks prior to representing them as a single data point? Additional terrain information such as slope incline, aspect, runout exposure, percent of track in forested areas could be meaningful for more detailed understanding of the terrain characteristics. These could still be summarized to the level of GPS track or the clicked route to preserve anonymity.	Yes, we have detailed terrain information for the GPS tracks, such as slope and aspect. However, because we predict the activity to the relatively broad spatial resolution of the warning regions, we did not focus more closely on terrain characteristics, since the goal was not to predict activity at a single location or on a single slope. slope.
Line 162 - Why do you assume that the increase in popularity of the website in 2021 impacts the click data only and not the GPS tracking data? I assume that this decision was made because you assume that a much larger, and potentially more representative, proportion of backcountry users are engaging with Skitourenguru after 2021. Wouldn't the same assumption apply to the GPS tracking data?	This is a good point. We looked at the temporal distribution of the datasets and found no such trend for the GPS data. This is likely due to the fact that the GPS data was purposefully sampled from multiple data sources, which was not clearly elaborated in the manuscript. We will add this information.
Line 169 to 171 - I would assume that users engaged in trip planning might click on multiple routes to compare options before selecting their destination. Do you have a way to account for the fact that the ratio of clicks to actual ski tours is likely biased heavily towards clicks? Such as tracking the number of clicks per website user and assuming each user is only going to actually complete a single ski tour on the following day.	Thank you for this input. Yes, we could correct for the average number of clicks per unique user. We didn't correct this because we did not try to estimate absolute numbers from clicks. Looking at relative and normalized differences wouldn't change with the corrected values. What we did do is to check the distance between the clicked routes per user and day. This distance was on average smaller than the usual size of a warning region, suggesting that most people focus on one region while planning, which made a correction for location not necessary.
Line 188 to 190 - Wind speed seems like a worthwhile variable to consider because it impacts avalanche hazard conditions, snow	Yes, we agree that wind definitely plays a role both for avalanche hazard and for quality of skiing. However, we did not consider wind due to complexity of wind fields and the

quality, and how enjoyable the experience of being in the mountains is for the day.	heterogeneity of wind speeds across larger regions such as warning regions. We therefore believe that including a spatially aggregated wind value per tour would not add meaningful additional information, since winds speeds can differ greatly, even in a single warning region. We will add a comment with respect to the potential importance of wind in the limitations.
Line 195 to 197 - Characterizing the desire for skiing untracked snow simply as a potential heuristic trap seems like an oversimplification. When backcountry skiers decide to undertake the risks of traveling in avalanche terrain there has to be a reward side of the equation that justifies the personal risk. While seeking untracked snow can lead skiers to make ill informed decisions, it is also a fundamental driver of what makes the activity worth pursuing. I think it would be worthwhile to consider the reward side of the decision-making process in selecting variables for your models to help balance out the focus on risk oriented factors. This is illustrated in the results by the RF importance of sunshine on the number of users.	We will change this sentence and highlight the importance of pleasure and the desire to ski an untracked slope in the skier's route selection. So far, we already have two variables that may function as proxies for potential for the pleasure: New snow and the sunshine duration. Although we could integrate the previous day's activity into the click model (giving a proxy for untracked snow) this would not be possible for the track data, and we aimed to use the same input variables for both models to aid interpretation.
Line 218 to 220 - Are there additional avalanche hazard characteristics from the public forecast that could be used to give a more complete picture of the avalanche conditions. I am not very familiar with the Swiss avalanche bulletin, but examples from the North American avalanche products would include avalanche problem type, potential avalanche size, and avalanche likelihood. While the danger rating provides a useful summary, these additional avalanche characteristics provide much more nuance to the current conditions which can significantly impact backcountry skiers terrain selection and risk assessment process.	There are indeed additional variables that could be incorporated in the models, which as you say influence terrain selection. However, these choices are likely less important in the selection of a warning region - our primary units. It's worth noting that these are much smaller than in North America, and there is ample evidence that in Switzerland danger level alone strongly influences the choice of region (not the individual routes chosen or behavior in terrain, which are outwith the scope of this paper, see Techel, 2014: https://doi.org/10.5194/nhess-15-1985-2015). We'll add a comment on this in the conclusions.
Line 259 to 261 - Are there local experts you could consult to verify whether absence of evidence actually implies evidence of absence? For example consulting with local mountain guides to estimate whether the	This is a good input, and we incorporate it into the outlook. It is however not in the scope of the current study.

absence of track and click data correlates with their experience travelling in specific regions. This seems like a strong assumption based on the fact that you are using data from only 1 app, especially for forecast regions with only a few tracks/clicks throughout the period of record. I understand that this approach of inferred absence is necessary to make the models work for the present study, but acquiring all the sample data from a single source is a significant limitation. Maybe you could make a recommendation for how this assumption could be tested in future research.	Regarding the source of the GPS data, see comment before.
Line 269 to 273 - How did you select the season that was held out as testing data? This approach to splitting testing and training data makes sense given the nature of the dataset. However, the performance evaluation of the model could be highly dependent on the characteristics of the weather and snowpack from the testing data. If the snowpack depth, 24 hour new snow, etc. were outside the values in the training data it may skew the performance metrics.	We ran different models with alternating test seasons. We calculated the skill scores of the prediction for each season separately, which are provided in Table 3 of the manuscript. Standard deviation lies below 5% for all skill scores, which indicates that there is little difference between seasons.
Results	
Line 321 and 326 - Does the skewed density of clicks and tracks to a small subset of the forecast regions justify limiting the analysis to these most populated regions? Have you considered filtering out regions that do not have a minimal threshold of track or click data to reliably estimate usage patterns?	Yes. Two regions that did not contain any GPS tracks were removed from the data analysis entirely. We will mention this more clearly in the data section. Further, we experimented with excluding regions with only little data, but this did not improve the performance of the model. Low-data regions usually come with small errors, as can be seen in Figure 7, so they do not negatively impact performance of the models.
Figure 8 - It is pretty hard to see the observed values in panel a. Perhaps you could add a black outline to the observed area or somehow increase the contrast compared to the darker green line of the predicted values.	Thanks for the remark, we will adjust this in the revised version.
Line 387 - 'validation season.'	We will edit as suggested.
Line 387 - ''variable importance was calculated'	We will edit as suggested.

Figure 10 - This is a very useful visualization of the underlying distributions and activity for the two models. The example in panel b and c clearly illustrate the spatial correlation associated with specific conditions.	Thank you.
Line 432 - I'm not sure what you mean by GPS tracks providing limited spatial detail. In terms of providing detail on where individuals are travelling GPS tracks are probably the best type of data available.	Thank you for the comment, this sentence was poorly worded, and we will adjust this in the manuscript. What we meant here is that because the GPS dataset is so small, there are only a few tracks over the whole area when disaggregated to the daily temporal resolution, providing only information for a few regions.
Line 555 - Is there any data available from Skitourenguru about the general demographics of their user base? You are claiming a few times in the discussion that click data captures a broader set of users but there is no direct evidence about the sample characteristics from this data set.	There is no data on general demographics of the click data. However, as people visiting the site automatically contribute to the data set, the data set is two or three orders of magnitudes larger than the GPS data, we think it is a fair assumption that clicks portray a wider audience.
Line 565 - 'Lastly, we found that online engagement…'	We will edit as suggested.