

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, as well as the editor for their supportive and constructive comments. We have written a detailed response to all the questions below and have edited the manuscript to incorporate these useful suggestions.

Referee 1: Anonymous

Comment	Response
<p>My main issue is that -- while temporal cross validation is applied -- a corresponding spatial cross-validation is missing.</p>	<p>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.</p> <ol style="list-style-type: none"> 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. <p>We added a few sentences discussing these issues in the limitations.</p>
<p>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 contributed GPS tracks to the database? Is the number of users per year stable? What happened to the data from 21/22/23?</p>	<p>The track data unfortunately doesn't come with user IDs, therefore the number of unique users cannot be estimated.</p> <p>The data was purposefully sampled by Skitourenguru GmbH and was not collected in the years 2021-2023. We understand the confusion and have elaborate more on the data collection process in the revised manuscript.</p>
<p>164 "Therefore, only data from 2021 onwards is included for modelling and prediction," ... so for</p>	<p>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</p>

<p>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.)</p>	<p>between the two prediction models. We have made this explicit in the data section.</p>
<p>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?</p>	<p>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 wanted to keep the calculation of variables equal for both data sources we opted to use the 100m- elevation band.</p> <p>We agree that we could have limited 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.</p> <p>In practice, since warning regions have an area of around 200km², 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 emphasise this point.</p>
<p>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.</p>	<p>Thank you for the comment, we adjusted the sentence to make it more readable.</p>
<p>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 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.</p>	<p>We edited this as suggested to use the present tense.</p>
<p>386 "Figure 9 shows the importance for each variable for the</p>	<p>Thanks for the comment, we adjusted this.</p>

performance of the model" ... Should probably be "importance of each variable".	
387 "from each cross-validation seasons" ... Should probably be "season".	Thanks for the comment, we adjusted 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 Skitouenguru 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.	<p>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.</p> <p>The GPS tracks that we use were compiled by Skitouenguru, but they come from 3 different data sources: camp2camp.org, skitouenguru.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.,:</p> <ul style="list-style-type: none"> • 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). <p>We made this more clear in the revised manuscript.</p>

	<p>Further, we applied for data from Strava, but were not granted access.</p> <p>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 emphasise in the limitations.</p>
<p>Methods</p>	
<p>Line 141 to 147 - Does Skitouren guru 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 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.</p>	<p>Thank you for this important remark.</p> <p>No, Skitouren guru doesn't require a subscription to use. Anyone can access the website and search for tours, or upload GPS tracks. We made this clear in the manuscript.</p> <p>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 added a brief remark on this in the revised manuscript.</p>
<p>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, 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.</p>	<p>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.</p>
<p>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 Skitouren guru after 2021. Wouldn't the same assumption apply to the GPS tracking data?</p>	<p>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 added this information to the revised manuscript.</p>

<p>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.</p>	<p>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, and looking at relative and normalized differences wouldn't change with the corrected values. Additionally, the unique ID for clicks is not necessarily congruent with unique users, due to users browsing in private mode or clearing history/cookies.</p> <p>What we did do is to estimate the distance between the clicked routes per user and day, assuming that the unique session ID approximately corresponds to a user ID. 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.</p> <p>We have added a few sentences discussing this in the data section.</p>
<p>Line 188 to 190 - Wind speed seems like a worthwhile variable to consider because it impacts avalanche hazard conditions, snow quality, and how enjoyable the experience of being in the mountains is for the day.</p>	<p>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 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 added a comment with respect to the potential importance of wind in the limitations.</p>
<p>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.</p>	<p>We changed this sentence and highlighted the importance of pleasure and the desire to ski an untracked slope in the skier's route selection.</p> <p>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.</p>

<p>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.</p>	<p>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 behaviour in terrain, which are outwith the scope of this paper, see Techel, 2014: https://doi.org/10.5194/nhess-15-1985-2015). We added a comment on this in the discussion.</p>
<p>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 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.</p>	<p>This is a good input and we have incorporated it into the outlook. It is however not in the scope of the current study.</p> <p>Regarding the source of the GPS data, see comment before.</p>
<p>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</p>	<p>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</p>

<p>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.</p>	<p>skill scores, which indicates that there is little difference between seasons.</p>
<p>Results</p>	
<p>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?</p>	<p>Yes, actually, two regions that did not contain any GPS tracks were removed from the data analysis entirely. We now mentioned 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.</p>
<p>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.</p>	<p>Thanks for the remark, we adjusted this in the revised version.</p>
<p>Line 387 - 'validation season.'</p>	<p>We edited as suggested.</p>
<p>Line 387 - "variable importance was calculated'</p>	<p>We edited as suggested.</p>
<p>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.</p>	<p>Thank you.</p>
<p>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.</p>	<p>Thank you for the comment, this sentence was poorly worded and we adjusted this in the manuscript.</p> <p>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.</p>
<p>Line 555 - Is there any data available from Skitounguru about the general demographics of their user base? You are claiming a few times in the discussion that click</p>	<p>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</p>

data captures a broader set of users but there is no direct evidence about the sample characteristics from this data set.	think it is a fair assumption that clicks portray a wider audience.
Line 565 - 'Lastly, we found that online engagement...'	We edited as suggested.

Editor Comment: Pascal Hägeli

Comment	Response
L61: Research objectives In the description of your research objectives, you state "(a) find out if and how real-world behaviour corresponds to online behaviour". Given that your dataset of GPS tracks is rather small and its representativeness and relationship to the click data is unclear, I would like you to be more precise in the description of your objective. Your dataset does not allow you to draw general conclusions about "real-world behaviour" in backcountry skiing. Instead, you are comparing the clicks of people that use the Skitounguru for trip planning to GPS tracks that have been uploaded to the site. You can hypothesize what these datasets and your analysis might say about general behaviours in the discussion section, but it is beyond your research question. Please be clear about this throughout the manuscript.	Thanks for this important comment, we have reformulated the research objective to make it clear that real-world behaviour is captured by GPS tracks. These GPS tracks do not only originate from Skitounguru, but 3 popular websites in Switzerland. We now have made this clear in the data section. We have also made clear throughout the paper the difference between planning behaviour and skiing activity as captured by clicks and GPS tracks respectively.
L156: Representation of tracks Why did you choose average elevation to represent your GPS tracks? Minimum and maximum elevation of a tracks seems more informative as it would allow you to identify the likely minimum temperature that they would experience and relate the track to the avalanche prone area described in the bulletin. I am sure you had your reasons for the mean value, but please explain.	We now calculate the temperature as suggested based on minimum elevation as this captures the warmest section of a tour and therefore whether the route remains above the snowfall line. As we are using elevation only for the calculation of meteorological variables, and given the coarse resolution of meteorological data, we believe that the mean elevation range is a fair representation for precipitation and visibility.
L165: Use of dataset Studies using different parts of their dataset for different analyses	We aimed to have meaningful variables that can also be used to predict future seasons. Adding the

<p>always raises warning signs. You could have included an additional “season” variable in your click model to account for this.</p>	<p>season variable would make it impossible to predict for a future season.</p>
<p>L175 and Figure 4: Normalizing of counts In my opinion, normalizing unnecessarily obscures the data you are presenting, and it creates an inconsistency between Figures 4 & 8. You can produce the exact same presentation by showing the clicks and counts on two different scales. This allows the reader to see the count numbers. Furthermore, please use the full width of the page for these charts and use gridlines that highlight the weekends. This will allow the reader to see the patterns more clearly. Finally, one should use consistent representations for the same data type. Both of your time series are count data, and they should therefore be presented the same way.</p>	<p>Thank you for this comment. We adjusted the Figure as suggested using a double y-axis representing absolute counts. For figure widths we followed the journal Latex template - this specifies a width of 12cm for two-column figures and 8.3cm for one-column figures.</p>
<p>L220: Characterization of regions Why did you calculate ski route density by dividing the total length of all ski routes in the region by the area of the region. I assume this variable aims to represent the options that backcountry users have. However, your approach makes longer routes count more than shorter routes, which does not seem quite right. Why not using number of routes divided by area? Please also explain the thinking behind your accessibility variable in a little bit more detail.</p>	<p>Thank you for this remark about the ski route density, we adjusted this as suggested. For accessibility, we used road density and census density as an approximation of how easily a region is accessible by how many people. We followed the approach by Stahl Olafsson, 2022, https://doi.org/10.1016/j.landurbplan.2022.104511</p>
<p>L295: Location of detailed explanation of presentation All of the employed methods should be described in the methods section. However, it might be easier for the reader if you moved some of the detailed explanations of the pros and cons of the normalized PD plots to the results section where readers are looking at these charts. You want to present the information where it is most useful for the reader. This</p>	<p>This is a good idea. We have moved the explanations of PD pots to the results section.</p>

<p>detailed description is difficult to understand for the reader in the methods section since they have not seen the charts yet.</p>	
<p>L341+: Order of presentation in results section It seems odd that you talk about the model performance before you have even described the models themselves. I strongly recommend that you describe the models and the influence of the variables first. Once the reader has a sense of this, it is much more meaningful to discuss performance.</p>	<p>We have re-arranged the order of the results as suggested.</p>
<p>L341: Details on model performance Your section on the model performance is very detailed, and one can get easily get lost in it. While interesting, is all this information really necessary for the addressing the research questions you specified in the introduction? Instead of documenting everything possible, being more selective can help direct readers' attention to the most important parts of the analysis.</p>	<p>We shortened the section as suggested and have concentrated on key results..</p>
<p>L378 and Fig. 8: Examination of prediction error It is interesting to look at the time series of the prediction errors. Similar to my earlier comment about Figure 4, I would use the full width of the page to display the information, and I would use the same presentation for the same type of data (counts represented as lines). In addition, however, why are you presenting the errors as weekly and not daily summaries like the original count data? Since weekday/weekend is an important predictor, this presentation obscures any related patterns. Given the structure in the time series of the prediction error for the click model, it seems like there is something at play (most likely weather related) that is not captured by your predictor variables. It would be useful to</p>	<p>From the response before:</p> <p>For figure widths we followed the journal template - these specify a width of 12cm for two-column figures and 8.3cm for one-column figures.</p> <p>We have changed the figure as suggested to represent daily counts and marked the weekends in the plot. We've kept the area/ line plots since otherwise the difference in the tracks plot between predicted and observed will be very difficult to read.</p> <p>We also have commented on the pattern of the click deviation in the results.</p>

<p>comment on this in either the results or discussion section.</p>	
<p>L412 and Fig. 10: Impact of different variables I very much enjoyed the discussion of the impact of the different variables. I think that your description of the spatial pattern in the response to increasing danger ratings is interesting, but your current presentation of the effect in Figure 10b is making it difficult to properly see it. In essence, you are saying that there is an interaction between the increase in danger rating and some of the variables that describe the forecast regions (accessibility, wildlife areas, and ski route density) which results in a response that is spatially variable. However, since the change in danger ratings in your scenario is not homogenous across the Alps, it is difficult to separate the two patterns. I think two maps that show the change in response to a consistent change from a) moderate to considerable, and b) considerable to high would be more insightful for showing what is going on. In addition, I think it would be good to flip the color scale to have positive values at the top and negative values at the bottom.</p>	<p>We added this Figure as suggested, showing a consistent change from 1 to 2 and from 2 to 3. Further, we switched the colour legend so that positive values are shown at the top.</p>
<p>L419+: Discussion I think your discussion could be much more powerful and insightful. Right now, a considerable amount of text repeats background information explaining the nature of the data (e.g., L438+), and the thinking behind some of the design choices of your study. In addition, you repeat information from the results section and in some circumstances, you introduce additional results (e.g., L446+: discussion of spatial patterns) without further discussing its higher-level meaning. Furthermore, many aspects of the discussion section touch on different limitations of the analysis without</p>	<p>We have restructured the discussion completely, based on your suggestions and we think that the implications both scientifically and practically are now more clearly formulated.</p>

<p>ever giving the reader a complete picture.</p> <p>There are lots of interesting insights coming out of your study, and I think that tightening and slightly restructuring the discussion section would help the main results and implications of your study come out more strongly. The current structure of the section (data, methodology, and variables) makes it hard to get the full picture. I suggest the following subsections for the discussion:</p> <ul style="list-style-type: none"> • A synthesis of the results that very explicitly answers the research questions outlined at the beginning of the paper. • A dedicated limitation section that summarizes the limitations from all sources (data, variables, and methodology) to produce a comprehensive picture of that these models can and cannot do. • A more targeted implications section that focuses on what the results mean for avalanche safety initiatives and research. What important insights and new opportunities does your study create (e.g., highlight areas that warrant extra avalanche forecast resources because people tend to flock there under certain conditions) and what are the steps required to take this research to the next level (e.g., integrating information about the terrain characteristics of the route and tracks to better understand how people change their choices based on the avalanche conditions). 	
<p>The conclusion section is very short and does not seem to add much value right now. It would be useful if you could provide a bit of an outlook that looks at next steps for this line of research.</p>	<p>We have re-written the conclusion as suggested and incorporated a more comprehensive outlook.</p>
<p>Editorial Comments</p>	
<p>Comment</p>	<p>Response</p>
<p>L65: Calling Section 2 “Background” instead of “State of the Art” would make the structure</p>	<p>We changed this as suggested.</p>

of your manuscript more in line with standard scientific papers.	
L70 (and others): Mccammon should be spelled McCammon	We changed this as suggested.
L73: Surveys and questionnaires are generally analyzed with quantitative methods.	We changed this as suggested.
L86: Replace “emergency avalanche beacons” with “avalanche transceivers”	We changed this as suggested.
L133: Why are you providing the mean elevation of the Swiss Alps? Is this a meaningful/necessary statistic for your study? The fact that 50% of the area is above 1500m seems more informative. Furthermore, precision of these numbers seems unnecessarily high. It would be sufficient to say that the Swiss Alps cover roughly 26,000 km ² .	We changed this as suggested.
L153: “Obfuscated” is not a commonly used term and therefore makes the explanation more difficult to understand. You can easily provide the same explanation without it. E.g., “To protect the privacy of the contributing users, we converted/transformed/simplified the GPS track to ...”	We would like to retain this term, as it is a known concept used in research about mobility data and privacy. We however introduced and explained the term the first time that we mention it to make it more clear for the reader.
L175 (and others): Please avoid using dashes in scientific writing. I know that ChatGPT likes using dashes, but they are really not commonly used in scientific writing.	We have retained dashes where we found they improved the readability of our text.
L186: I would not use the term “human factor” here. This term is generally used to refer to specific challenges in the human decision-making process, and not any non-weather, snowpack, or terrain factor that affects people’s choices. In general, I think it is an overused term, and we should use it in a more precise way.	We changed this to ‘patterns of human behaviour’.
L273 (and others): As pointed out by Reviewer 1, please do not use brackets like this. It creates hard to read text. Saving a single sentence is not worth making readers read the same sentence several times.	Thanks for this remark, we have re-written the sentences for better readability.
L314 (Fig. 3 and Table 2): I would integrate the correlations presented	Good catch! The correlation values in the plot were accidentally calculated using Pearson’s correlation

<p>in Table 2 directly into Fig. 3 and deleted Table 2 to save space. Am I correct in assuming that the correlations presented in Table 2 should be consistent with the correlation values shown for a 1-day time lag in Fig. 3. This is currently not the case. Please explain.</p>	<p>instead of Spearman. We updated the plot. We have now added the correlation values to Fig. 3 and discarded Table 2.</p>
<p>L320 (and others): Since you are not calculating spatial autocorrelations, I would describe the observed patterns in a more qualitative way. Like the tracks are spatially clustered or concentrated. It would also be useful to describe these patterns in more detail (e.g., along main transportation corridors, key tourism destinations). The same applies for the description of the spatial pattern of the click data.</p>	<p>We changed this to 'spatially clustered'. However, linking the pattern to main transportation routes is difficult given the coarse resolution of warning regions.</p>
<p>L337: The description that "tracks were more abundant than clicks in some areas" is incorrect because the click dataset is so much larger than the track dataset that I assume it was larger in all regions. Please describe differently.</p>	<p>Thanks for this comment. We have now adjusted the sentence and made clear that 'more abundant' refers to a relative comparison. It means the region's track count was in a higher quantile than the region's click count.</p>
<p>Fig. 5: Including the two variables into a single color scale makes this chart difficult to read. It would be much easier to have two separate charts, one for click and one for tracks. The color saturation currently also does not properly line up with the magnitudes of differences between the click and track data (i.e., more saturated does not mean bigger difference). You also do not explain what you mean with "correspondence".</p>	<p>We have modified the figure and added more information to make it easier to read.</p>
<p>Fig. 7: The labels for the four count categories are not quite right. The way it is presented right now, the top category (e.g., < 294) includes all data points. Is it necessary to group the values into four categories or could the size of the circles just represent the actual values? Would it make sense to turn the color scale upside down to have lower values at the bottom and higher values at the top?</p>	<p>Good point. We have adjusted the legend as suggested so that it shows exclusive categories. We also adjusted the colour legend to show positive values at the top.</p>

<p>L384: “weekly periods of over- and under predictions ...” is not quite right because the periods were actually 3-4 weeks long.</p>	<p>We have changed this to “periods of under/over predictions “.</p>
<p>L387: Replace “classification model” with “track model” to make it easier for the reader to relate it to the other information about this model.</p>	<p>We have adjusted this as suggested.</p>
<p>Appendix B: I do not think this appendix is necessary. All these equations are well established in the scientific community.</p>	<p>We agree, we have removed the Appendix B and added references for the equations in the text.</p>