Reply to Reviewer 1

We thank both reviewers for their constructive comments. Below we address each question/comment (blue text). In the revised manuscript we will add more details about our reasoning and the chosen methodology in particular the statistical approach.

This paper examines landslide occurrences along a busy highway in India. The authors generate a landslide inventory and analyze corresponding landslide density with respect to some external factors including precipitation, slope and lithology. Overall, it is a publishable work but there are so many points which are not clear to me at all. I have listed all my comments below line-by-line. My main problem with the paper is that the authors do not clearly introduce their methodology and they do not explain reasoning behind choices they made. Therefore, I had a hard time understanding why/how they did some analyses. Following moderate to major revisions, I believe the paper could be publishable.

Lines 29-34. Please cite the relevant literature.

We added a reference to this a recent article:

Boora, S., & Karakunnel, M. T. (2024). The SDG conundrum in India: navigating economic development and environmental preservation. *International Journal of Environmental Studies*, *81*(2), 961–976. https://doi.org/10.1080/00207233.2024.2323321lWe added a reference to a newspaper article.

https://timesofindia.indiatimes.com/india/Govt-plans-Bharat-Mala-a-5000km-road-network/articleshow/47102122.cms#

Lines 35-37. Please cite the relevant literature.

We added a reference to the article:

Sharma, A. K., Parkash, S., & Roy, T. S. (2014). Response to Uttarakhand disaster 2013. *International Journal of Scientific and Engineering Research*, *5*, 1251-1256.

Line 37. Considering the flow of your text, the last line of this paragraph is a bit off.

We'll change it to "Ensuring accessibility and connectivity during such events is thus of live-saving importance, yet requiring considerable maintenance efforts (Uniyal, 2021)."

Line 51. Please cite those "few" studies, cite at least some of them.

We added two references:

Huat, B. B., & Jamaludin, S. (2005). Evaluation of slope assessment system in predicting landslides along roads underlain by granitic formation. *American Journal of Environmental Sciences*, *1*(2), 90-96.

Ching, J., Liao, H. J., & Lee, J. Y. (2011). Predicting rainfall-induced landslide potential along a mountain road in Taiwan. *Geotechnique*, *61*(2), 153-166.

Lines 61-62. "Limiting the inventory to road blocking landslides was required to cope with the overwhelming number of landslides and to ensure that we account for the landslides that detached most recently." Could you please elaborate this further? How do you define those "recently occurred" landslides (e.g., "the ones that blocked roads in the last five years")? And why do you want to put such a restriction to your inventory?

We define "recently occurred" as having occurred during the preceding monsoon season and the unusually heavy rainfall in the two weeks before our survey. In the revised manuscript we specify the periods of high rainfall as the monsoon season May - Aug and the heavy rainfall during 5-12 October (see attached pdf DISTRICT_RAINFALL…., page 6 Uttarakhand).

We initially tried to attribute the landslides to the latter period by matching our data with high-resolution images available in Google Earth but their temporal resolution for many stretches of the road was insufficient. Moreover, to a large part, historic imagery has subsequently been made unavailable over India.

Line 64. "previous studies" Please cite at least some of them.

We added the citations:

Das, I., Stein, A., Kerle, N., and Dadhwal, V. K.: Landslide susceptibility mapping along road corridors in the Indian Himalayas using Bayesian logistic regression models, Geomorphology, 179, 116–125, https://doi.org/10.1016/j.geomorph.2012.08.004, 2012.

Devkota, K. C., Regmi, A. D., Pourghasemi, H. R., Yoshida, K., Pradhan, B., Ryu, I. C., Dhital, M. R., and Althuwaynee, O. F.: Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya, Nat. Hazards, 65, 135–165, https://doi.org/10.1007/s11069-012-0347-6, 2013.

Line 65. "in two spatial dimensions" No need to indicate this, please remove.

We would rather stick with our formulation to stress the distinctiveness of our approach.

Line 65. "conceptualizes landslides as a network-attached spatial point pattern" Could you please elaborate this further, what do you mean by this? It seems like this is an important element in your methodology and thus, it deserves a bit more attention. What are the main findings of Baddeley and others and how do you think they contribute to the existing literature?

Landslide susceptibility analysis commonly relies on discretizing the study region into pixels where each pixel indicates either the presence or absence of landslides. Subsequently, techniques such as logistic regression are used to predict the probability of a landslide as a function of a number of predictor variables. A pixel-based logistic regression analysis is approximately equivalent to a Poisson point process model if the landslides are originally stored as point features (Baddeley et al. 2010). We take this approach to modeling landslide susceptibility using Point point process models a step further by conceptualizing landslides as point features that are located on or along-side the road network (Okabe and Sugihara 2012). This means that we do not analyze road-attached landslides as events that occur on a continuous and unbounded plane, but rather as events that are along-side of roads. The term "along-side" indicates a somewhat broad spatial relation, but "implies that the physical unit of

the event [...] has an access point on a network" (Okabe and Sugihara 2012, page 7). In our case, this means that a landslide intersects with the road and blocks it. These intersections are, at the scale of our analysis, represented by point features, i.e. points on a network. Our approach relies on the analysis of the spatial arrangement of the points (the point pattern) along the line, and we hope to reveal important features (e.g. trends in point density) with respect to geo-environmental variables.

Baddeley. M. Berman. N.I. Fisher. A. Hardegen. R.K. Milne. D. Schuhmacher. R. Shah. R. Turner. "Spatial logistic regression and change-of-support in Poisson point processes." Electron. J. Statist. 4 1151 - 1201, 2010. https://doi.org/10.1214/10-EJS581

Okabe, A., & Sugihara, K. (2012). Spatial analysis along networks: statistical and computational methods. John Wiley & Sons.

Line 66. "simplified approach" If you provide more explanation we can also have an idea if this is a simplified approach or not.

The comment above also addresses this point. The term "simplified" is not an appropriate term that we will avoid in a revised version of the manuscript because every model entails simplification and abstraction.

Lines 71-72. "We present our results and discuss uncertainties and potential shortcomings of our approach. We conclude with recommendations for refinement of the approach and further research avenues" No need for these lines, you need to have a discussion, conclusions and interpretations anyway.

We will delete these sentences in a revised version of the manuscript.

Line 72. "Study site" You can convert this heading to something like "study area and data" because in the method section you mention your data sources although, for instance, in figure 1 you already present your geology map. Thus, you can move those sections here.

Thanks for pointing out this issue. We will separate the description of data sources more stringently in a revised version of the manuscript.

Line 75. "1000–1200"mm

We will correct the error in the revised version of the manuscript.

Line 115. Method section: Please first provide a summary of your methodology. There are several elements there. If you systematically present your methodology as step1, step2 and so forth, your readers could follow you more easily. For instance, you start with field work and jump into how you mapped landslides. However, it would be better if you list your steps and then explain what you have done (and what datasets you have used) in each of those steps. This is to say that, please first give us a structure and then we can better understand what you did and why you did. In this regard, you can first describe different landslide categories (Figure 1) and indicate those corresponding periods. We need to know the set of criteria you used to label those landslides.

We thank the reviewer for suggesting a better way of presenting and structuring the methods. In a revised version of the manuscript, we will rewrite these sections so that readers are better able to comprehend the approach and its detailed implementation.

Lines 115-130. It seems like you mapped landslides as points but polygons, right? Please indicate this. Could you also mention where you put those points, to the crown of each landslide? Could you please make a figure showing a close-up view showing examples for each landslide category you mentioned? Please show those examples by providing multi-temporal images, so we can see corresponding conditions in different periods. You are saying you collected some GPS locations during your field work. Then I think you should also show us how you assigned those coordinates to corresponding hillslopes.

The reviewer is right in assuming that the landslides were mapped as points using a handheld GPS rather than as polygons. These points were at locations along the road, where landslides intersected and blocked the road. These points were not assigned to the corresponding hillslopes, rather hillslope conditions upslope of the road were assigned to the road network. For example, we calculated slope with a 210 m bufferzone at the side of the road which is located higher than the road, and mapped the average slope within this zone to the road. As such, this value becomes an attribute of the road which serves as a predictor variable during modeling. In a revised version we will clarify this point better.

As mentioned in the manuscript, unfortunately Google Earth has removed all images from the public archive that we used to categorize the landslides. Notwithstanding, we will try to retrieve high-res imagery to showcase the different landslide conditions in different periods (see next answer).

What is the importance of labelling landslides as (1) new landslide, (2) road-blocking landslide visible before the Sep–Oct 2022 rainfall anomaly, and (3) reactivated landslide. In your model, there is no difference between them, right? Here my main concern is that if you used the accumulated precipitation for the corresponding period. For instance, you have some landslides that occurred before Sep-Oct 2022 and you do not know when they really occurred. And yet, you used accumulated precipitation between Jan-Oct, 2022. Also, you have some new landslides, which have obviously occurred in a different time window compared to the former category but you still use the same period to calculate the cumulative precipitation. The same is also valid for the reactivated landslides. I think you should clearly indicate time spans where you think landslides either occurred or reactivated and then you should calculate the cumulative precipitation for the given temporal windows for different landslides separately .

Btw, I think I mentioned this above already, but it is worth mentioning again that you should better describe your landslide categories and their corresponding time windows. You can simply add a table so we can clearly see what you refer to in each category .

We have now repeated the classification of landslides based on the currently available Google Earth (GE) images. Because a lot of historic images are still missing, we used only landslides that most probably occurred between Jan22 and Oct22, encompassing the entire monsoon season of that year and the heavy rainfall period at the beginning of Oct22. This is the same period, for which we accumulated the rainfall/precipitation of the different rainfall products shown in Fig.3. Landslides that certainly occurred before 2022 and those that were hidden in shadows in the GE images were omitted from the analysis. This reduces the number of landslides in the analysis to 250. Among these are reactivated ones that we highlight in Fig.1.

Line 137. Your readers do not have to know what PPS is, please give the long version of it.

PPS is the name of a numeric class in TopoToolbox to study point processes on river networks which we adopted here to work with road networks. PPS stands for "Point patterns on stream networks" and a reference has been provided. Still, we agree that most readers will not be familiar with it and thus we will provide more detail in a revised version of the manuscript.

Line 137. How did you calculate your landslide density btw? This is your target variable and it needs further explanation? First of all, why did you target landslide density and why not landslide occurrences? In this case, you have a density value depending on your moving kernel. It could be small but I assume you will have a value for a large portion of the study area. This means that even if there is no landslide you will have a value. Am I correct? I am asking these questions because I could not find answers in the manuscript. Could you show us the distribution of density values via a histogram. What is the spatial resolution that you create for your landslide density map? In the result section, you should show us that landslide density map. I guess you calculated landslide density around the road with 210m-wide buffer zone (?) Is this the boundary of your study area anyway?

Apparently, it has not been clear from a description that we model the occurrence of landslides along the road, casting the problem into one spatial dimension, i.e. distance along the road. LS density in Fig. 5 is shown as a 1D kernel density, but it could also be shown with a histogram. The pattern will be the same and depend on bin width, rather than kernel bandwidth. If there is no landslide, and a sufficiently small bandwidth, the density will be zero. In fact, a 15 km bandwidth may be too high and we will consider to reduce the bandwidth (or use a histogram instead).

Line 145. Could you please explain why you chose a log-linear model

The reviewer is referred to Baddeley et al. (2010). Our approach is a spatial logistic regression, where the relation between presence probabilities p and explanatory variables X are controlled by the form of the logistic link function (logit $p = \ln(p/(1-p))$. As pixel size tends to zero, we have $p \to 0$ and $\ln(p/(1-p)) \to \ln(p)$. The limiting Poisson intensity is thus a loglinear function of the covariates (adopted from Baddeley et al. 2010, page 1166).

Line 165. You do not need to hypothesize this, so you better cite the relevant literature.

From a strictly statistical standpoint it remains a hypothesis that rainfall influences landslide occurrence. Yet we reformulate the sentence and include references to work that have previously shown this to be true.

Line 166. "accumulated rainfall" Be careful that some of those satellite products do not provide rainfall data but precipitation. And these are different things .

We change "rainfall" to "rainfall/precipitation".

Lines 180-184. Please show this layer in one of your figures.

It is shown in Fig. 1 with a white line with a dashed signature.

Line 197. Please cite the relevant literature and explain how it works.

We will add additional information and references in a revised version of the manuscript.

Line 203. I need further explanation for this section. If you use Poisson regression, then I do not know how you convert it into AUC (You have a continuous array but AUC mainly works for binary conditions). I remember I saw it also before in the literature but you do not either cite the literature nor provide a detailed explanation. Therefore, it is not clear for the moment.

The reviewer confuses Poisson regression and Poisson point process models. Both are actually related if there are possibly more than one landslide at a specific location. However, here we are predicting landslides using the latter, which is most similar to a spatial logistic regression. In this case, we are dealing with a binary variable and thus the AUC is warranted.

Also, I would expect to see a correlation between predicted landslide density and calculated value. There you can assess your prediction performance via some other statistical measures.

This correlation is actually shown in Fig. 7c. Correlations between individual predictors and landslide density are shown in Fig. 6.

Line 227. "AIC" Akaike Information Criterion. Please first explain this in your method section and cite the literature as this is not your own methodology.

We have explained AIC in line 210-211 and provided reference to the original publication (Akaike 1974). Nowhere do we claim that this is our own methodology.

Lines 231-234. This needs to be numerically shown, otherwise you can mention it as a source of uncertainty in your discussion section.

We will show this in a revised version of the manuscript.

Lines 238-241. This is not presented in any figure nor via a table. Where are the outputs of the Bayesian feature rank algorithm? I would like to see how you come up with this conclusion .

Thanks for pointing this out. You are right that, except in the text, there is no information about the Bayesian model outputs (except the posterior densities shown in Fig. 4). We'll provide these as a table in a revised version of the manuscript.

Lines 260-266. This needs to be moved to the discussion section.

We would argue against it because these are still results that we are reporting.

Line 287. "this means that spatial variables characterizing the source area (e.g., hillslope gradient) are projected onto the road" What does this mean?

Regarding the hillslope gradient for example, this means that each road pixel is assigned with the hillslope gradient of the adjoining pixels. To be specific we used the mean hillslope gradient of the 7 nearest uphill pixels.

Lines 288-290. I do not think this is relevant, so please remove this. Even if you worked with thousands of landslides that would not be computationally challenging I would say, no?

Line 298. "studies come to different conclusions" But in the following lines you do not mention those different conclusions.

We added a sentence that specifies these different conclusions.

"In a validation study of different gridded rainfall products in the Eastern Himalayas Kumar et al. (2021) report that CHIRPS-2.0 overestimated the monsoon but underestimated annual precipitation."

Lines 324-343. "However, we cannot exclude that small-scale topographic changes due to construction or land use changes" You do not know, you are just speculating because you were not able to capture that signal in your model. Therefore, please rewrite this line and make it smoother instead of saying "we cannot exclude".

We will replace it with: "However, small-scale topographic changes due to construction or land use changes (e.g. abandonment and degradation of agricultural fields and terrace systems) may as well exacerbate road-side slope failures (Jacquet et al., 2015; Mauri et al., 2022)."

Line 357. Please remove footnotes and give citations to online sources.

Changed accordingly

Figure 1. "The highest density of landslides occurs between Rishikesh and Srinagar within lithozone 2 and between Pipalkoti and Joshimath in lithozone 1." Please keep this line for your main text and remove it from here.

done

"For description of the lithozones see Table 2."In lines 190-191, you mention five lithologic units: (1), phyllite and shale (2), quartzite (3), quartzite and igneous rocks (4) and crystalline high grade metamorphic rocks (5). Please indicate the same units in Figure 1.

We add the descriptions in the figure caption instead of pointing to Table 2.

"Note that lithozones 0 and 6 are not crossed by the road and are therefore omitted from the description." Refer to them with their names, please do not use ids.

See answer above. We would rather refrain from adding the unit names to the legend because of the space limitation.

"We subdivided the landslides into new ones, reactivated ones and those that were blocking the road before September 2022." Figure captions are just to describe what you present in that figure, it is not to present what you have done. Please say it in the main text not here.

We think that this piece of information is important to fully understand the pie chart given in the figure and would thus choose to keep it.

"Stars indicate locations of the 1999 Mw 6.6 Chamoli earthquake (Kayal et al., 2003; USGS, 2022) and the 2021 Chamoli rock and ice avalanche" Which stars are showing which events?

This should be clear from the dates that are placed besides the stars. We decided to increase the font size within the figure to improve readability.

Figure 2. This figure is not enough to present how you mapped landslides. I made some comments above.

We hope our answers to the comments above have clarified our approach. We think these images are fully suitable to portray the typical landslides we encountered.

Figure 3. Rainfall or precipitation?

We added the respective information in the text and the caption to this figure.

Figure 4. "Lithozone 1 is missing since the parameter is encapsulated in the intercept." I did not understand what you meant here. Please elaborate this in the main text.

Fitting a generalized linear model (or linear model) including a categorical variable with k levels will return k-1 parameters for this variable, instead of k. This happens to avoid multicollinearity and is achieved by setting one of the level-parameters to zero (a reference category). This finally means that the reference category is represented by the model intercept.

Figure 5. First present your predictive variables and then target variables and then finally show you prediction. What are those blue ticks in panel b? How is it possible that accumulated precipitation shows variation in very small distances? The minimum spatial resolution should be 5 km but somehow I see that it changes even within 1 km. Btw, I do not remember if you mention any downscaling but please clarify that point as well if you did not do that.

We think it is fully appropriate to first show the observation (panels a & b), then the predictor variables and eventually the prediction. The blue ticks in panel b) indicate the landslides.

We resampled the rainfall grids to the spatial resolution of the DEM (30 m) as stated in Lines 176 – 178. The wiggles occur because the road "meanders" through the precipitation field.

Table 1. How did you aggregate these units? I see quite similar lithologic units in different categories.

Because the initial units have many individual lithologies incorporated, an overlap between the aggregated lithozones was inevitable. We tried to separate them as best as possible according to a dominating rock type, i.e. carbonate rocks (1), phyllite and shale (2), quartzite (3), quartzite and igneous rocks (4) and crystalline high grade metamorphic rocks (5). We admit that this involves some degree of uncertainty. We presume however, that reassigning a small number of arguable IDs would not change the overall outcome of our analysis.

We thank the reviewer for the constructive criticism. Our responses are highlighted with blue text below.

The research by Mey and co-authors proposes a nice analysis on the occurrence and modelling of landslides along a highway in India. Such research focus is clearly needed in the context of the Anthropocene and the growing concern of the influence of human activities on Earth surface processes. This research is therefore timely and well within the scope of NHESS.

Having had a look at the extensive review already carried out by a first referee; I can only agree with him/her that the manuscript must be improved quite substantially in several aspects. In addition to the issues evidenced by the first reviewer, especially with respect to the methodological aspects, I would like to stress extra points that I hope would help in improving this work.

We hope that our responses to the first referee's comments address most of the reviewer concerns.

- The introduction is somehow a bit strange as it starts with a focus on the study area, without really providing the broader picture of human-induced problem on landslides, and especially that of road. Besides a need of a broader literature (as stressed by the other referee), on would benefit from a better organization of the introduction structure I think.

We see the point that the reviewer made and would start the introduction along the following lines:

"The global increase and expansion of the human population will inevitably promote the transformation of the natural environment. The construction of roads as a key infrastructural element that accompanies this development, impacts the landscape. Particularly in mountainous regions this may entail for example enhanced hillslope erosion and slope instability (Muenchow et al., 2012), affecting the fast growing montane communities especially in the Andes and the Himalayas. A worldwide surge in the proposals for new roads thus warrants the investigation of the link between road network expansion and landslide exposure (Laurance et al., 2014)."

- Although the case study on that highway in India is a nice one, we miss somehow the reason why this case study could be of special interest for a broader audience. In other words, why, for example, would someone working in the Andes be interested in such a study. Here some work could be needed to improve the introduction and discussion.

We have addressed the link to other similarly affected regions in our addition to the introduction. Besides, our methodological approach based on point pattern analysis on linear networks is innovative and can be applied to other regions. We thus believe that our study has a broader significance beyond our study site.

- As the referee says, knowing about the age of the landslides is quite important. However, the age of the road sections is also something that is important. Recent road sections could be much more damaging/impactful on hillslope instability due to, for example, de-buttressing effect that would be more intense in the recent years that follow a new road cut that along cuts of older roads. However, older roads mean also older outcrops, and therefore potentially increased weathering conditions. Maintenance activities and infrastructures to stabilize/drain the road environments can also vary along the road. Overall, I think that information (discussion) on these issues is needed.

We thank the reviewer for pointing this out. We add to our discussion the following:

"The propensity of landsliding along individual road segments may also be affected by their age. Fresh road cuts might initially suffer more from debuttressing than older ones, whereas the long established road cuts will have experienced longer periods of weathering, potentially affecting their stability. Moreover, remedial measures such as slope enforcement or building of retention walls or artificial drainage can be expected to have a stabilizing effect, although we encountered many of such structures that have failed eventually. We appreciate that such small scale disturbances influence the landslide occurrence, their inclusion in the model, however, would require much more data, which are currently unavailable to us."

- 30 m DEM is used to calculate surface (slope?) gradient. At such a resolution we clearly miss the subtle topographic characteristics that we find along roads. In addition, road cuts would definitely be missed. This issue needs to be clarified in the analyses and/or discussion.

We have recognized this fact in the discussion lines 375-384. We hope that with the answer to the previous comment we have satisfactorily addressed the reviewers concern.