We thank all three reviewers for their detailed and constructive comments, and the editors for the opportunity to revise the manuscript.

We revised the manuscript for clarity, assumptions, and mathematical consistency. Our responses are in bold green; verbatim text from the revised manuscript is in *bold italic blue.* The main changes are:

- Removed the reforestation analysis because its results were dominated by assumptions about root cohesion. We retain restoration as a potential application, but it requires a separate analysis.
- Expanded the hillslope–channel connectivity analysis, added new river-connectivity figures, and revised the title accordingly.
- Expanded the treatment of uncertainty and added sensitivity experiments for soil strength, model resolution, and volume estimation.
- Placed the detailed robustness analyses in the Supplementary Information (SI), including comparison to landslide inventory, bootstrap uncertainty in the KS metric, spatially correlated soil-strength perturbations, and a comparison of empirical and soil-depth-based LSO volumes. The main manuscript summarizes the implications of these tests; the full methods and results remain in the SI because they support model interpretation but are secondary to the paper's main contribution of introducing the connectivity framework.
- Audited and corrected the post-processing workflow, reran the complete analysis, and regenerated all figures and tables. Some numerical results therefore differ from the original submission; these updates reflect corrected and internally consistent aggregation of model outputs rather than changes to the conceptual framework.
- Discussed the apparently high modeled failure and unstable-terrain rates, showing that they remain broadly consistent with published susceptibility estimates and critical-rainfall thresholds. As this is primarily a framework paper, we do not attempt to resolve the broader controls on long-term landslide sediment production. Nevertheless, we find the discrepancy between precipitation-driven mobilization capacity and plausible sediment availability scientifically important and therefore added a focused analysis of how soil production and soil-mantle recovery may constrain realized sediment supply. We see this as an important direction for future model development and empirical research.
- Substantially revised the public code and clearly separated the pixel-level slope-stability calculations from the COHESION modules for LSO delineation, volume, runout, connectivity, and exposure. This modular structure clarifies which components belong to the susceptibility input and which constitute the COHESION framework.

Contents

Reviewer 1

In this manuscript, Schmitt and colleagues present COHESION, a connectivity-aware framework that augments landslide susceptibility mapping with empirical runout modelling. The article addresses an important gap by coupling susceptibility with runout to quantify cascading downslope hazard and exposure at landscape scale. In general, the workflow is clearly staged with helpful schematic figures and an explicit data-needs tables. Overall, the proposed approach is novel but yet reminded me of the framework currently used in New Zealand by Earth Science Institute (ex-GNS) for estimating landslide runouts and further risks (see Brideau et al. (2021) and de Vilder & Massey (2024)).

While I found the topic to be of high scientific and practical interest, I was disappointed by the quality of the paper. I have a feeling that the authors submitted the manuscript in a rush, omitting the proofreading step. This resulted in typos, a lack of measurement units in equations (line 176), inconsistency in abbreviations (for example, both FS and FoS were used for Factor of Safety), incorrectly rendered bibliographic links (see lines 84 and 204), and most importantly, the absence of a Conclusion section. Before applying for the next round of the pipeline, all these issues should be fixed.

That being said, I enjoyed reading this manuscript and found it very thought-provoking. I believe that the manuscript could be published in NHESS after major revisions and a second review round. See my comments below.

We thank Reviewer 1 for the encouraging assessment and detailed suggestions. We revised the writing and framing accordingly.

Major issues

1. While the authors explicitly specified that LSO is the term borrowed from image classification (see line 116), I found it difficult to understand how the LSO corresponds to the landslide scar (or landslide initiation area). That is, as far as I understood it, the LSO is generally bigger than the landslide itself, as it tends to merge several scars together due to the coarse spatial resolution (30 m). It would be beneficial to the reader to illustrate the difference between the digitised landslide and the LSO on a figure (as early as possible in the text, but figure 4 suits this purpose nicely).

We agree that an LSO is a raster representation and is therefore constrained by the 30 m grid. A real scar may occupy only part of a cell, whereas the model treats the full cell as part of the LSO. We clarified this in the text:

> ***It should also be noted that the delineation of LSOs needs to follow the somewhat arbitrary boundaries of raster cells. Thus, even if a landslide scar might only cover a fraction of a cell in reality, this cell would count as fully unstable in the model.***

We also added examples to Figure 3b.

2. The study area is located in the Nepal Himalayas, a region notorious for its landsliding processes. Indeed, the authors provide extensive information about the hydrometeorological setting (lines 129–136), but almost no data about the landslide processes. The reported quantitative information about the “larger” landslides (see line 143) is too vague from my perspective. It would be nice to read more about the nature of the landsliding processes, what the main drivers, sizes, and volumes are. Such a section in the first half of the manuscript may benefit the reader in terms of better understanding of the (quite) complex model (see my following comments). Yet, I understand that some of the questions I am asking are answered further in section 4.7, which is acceptable, but too late, in my view.

We agree and expanded the Study Area section with a concise summary of regional landslide types, triggers, sizes, and geomorphic relevance:

> ***Due to an intersection of natural factors, the Nepal Himalayas are a hotspot of landslide occurrence (Marc et al., 2023). Rapid uplift produces steep topography, while intense orographic precipitation is concentrated in a few months of monsoonal rain. Landslides range from small, shallow, frequent soil failures to much larger and less frequent deep-seated failures. Monsoonal precipitation is the dominant trigger, although earthquakes are also important; the 2015 Gorkha earthquake triggered more than 4,000 landslides (Kargel et al., 2016). Together, landslides and rockfalls are major agents of erosion and sediment supply to Himalayan rivers, while dense settlement and road construction create high exposure.***

3. My next question evolves from the previous one. Overall, I have an impression that the authors are proposing a framework which, by their design, is agnostic to the landslide-triggering event and landslide type. With a larger dataset and a slightly different approach, I would not mention that as an issue. However, since some of the modelling steps involve empirical equations by Rickenmann (2005) for runout length and Larsen, Montgomery, and Korup (2010) for landslide volume estimation, these were developed for a specific type of landslide. Rickenmann (2005) was super explicit that his empirical equation fits only debris flows. Larsen, Montgomery, and Korup (2010) developed equations for bedrock and soil landslides, and from the current manuscript, it is unclear which one the authors used and why. Moreover, the recent findings by Jones et al. (2025) indicate that empirical equations from the global and event datasets, the one the authors used, tend to critically underestimate landslide volume. I would like to see a discussion on this matter in the article.

We agree that the empirical relationships were developed for specific landslide types and that their applicability must be stated clearly.

COHESION is applied here to shallow, rainfall-induced soil failures. We therefore use the soil-landslide exponent from Larsen et al. (2010), rather than the bedrock exponent, and corrected the parameter values in Section 3.5:

> ***Larsen et al. (2010) studied a dataset of soil and bedrock landslides, comprised of*** **more than 4000 observations. Based on that dataset, they propose that the selection of the exponent γ should depend on the landsliding mechanism, with values of γ_{rock}=1.35 $\pm$ 0.01 and γ_{soil}=1.145 $\pm$ 0.008. As COHESION is designed to model landslides that are driven by saturation-excess in the soil column, we herein use γ=1.145 and $log\alpha$=0.86 (i.e., α=$10^{0.86}$=7.24) and .**

We also clarify that Rickenmann (2005) was derived for debris flows and is most applicable to rapid, flow-like shallow mass movements:

> ***Runout length is estimated using the empirical scaling relationship proposed by Rickenmann (2005), originally derived for debris-flows, i.e., rapidly mobilizing mass movements. This formulation relates vertical drop to travel distance and implicitly assumes rapid, flow-like mobilization. While computationally efficient for basin scale screening, this approach does not resolve detailed flow dynamics or rheology, and therefore should not be interpreted as a process-resolving runout simulation. Specifically, we follow Rickenmann's (1999, 2005) empirical formula linking distance traveled by a debris flow is controlled both by volume, which determines the available kinetic energy, and by downslope gradient.***

We additionally discuss evidence that global area–volume relationships may underestimate volumes in tectonically active settings (Jones et al., 2025). We treat this as structural uncertainty affecting volume and runout, and emphasize that COHESION is a screening framework rather than a process-resolving model.

4. The Factor of Safety estimation (section 4.1) and hillslope hydrology (section 4.2) require a lot of very detailed data. However, Tables 1 and 2 show readers that the data sources usually have a pretty coarse resolution, up to 1 km. While it is totally acceptable to use global datasets in data-scarce regions for regional studies (I mostly work in ungauged basins myself, no judgement at all), it is a must to discuss the potential implications of input data resolution on landslide susceptibility and connectivity. This is especially vital for the assessment of threshold subsurface flow, water balance, and so on, which I assume are the most influential factors for landslide modelling.

We agree that input resolution matters. The model runs on the 30 m DEM; coarser inputs are resampled to this grid, so the minimum represented cell area is 30 × 30 m.

The main resolution-related uncertainty is terrain smoothing, which affects both slope and subsurface routing. A 10 m DEM may be preferable locally, but 30 m remains practical

for basin-scale applications using global data. We expanded this limitation in the Discussion.

Detailed analyses of DEM resolution and alternative volume estimates are provided in the SI. We retain the main implications in the Discussion while placing the full diagnostics in the SI to keep the main text focused on the framework.

We added the following discussion and references:

> ***The 30 m DEM used in this study is consistent with many basin-scale slope-stability assessments, but remains coarse relative to individual landslide processes. It sets a minimum mapped source area of 900 m^2 and smooths terrain gradients and subsurface routing. Finer DEMs would better represent these controls, but very high-resolution regional models may be impractical and would still be constrained by coarser subsurface inputs.***

> Validation is limited to comparing slope distributions via the Kolmogorov-Smirnov statistic, which is generally fine. Nevertheless, I have never seen that in susceptibility-related literature. Quantitative skill assessment against the inventory (e.g., AUROC, F1 measure, and so on) is more common. I would recommend the authors either use the generally accepted quantitative metrics (see Petschko et al. (2014)) or apply a bootstrap analysis of the KS statistic to make your quality assurance process more robust and defendable. In that case, it would be beneficial to the paper to mention why you chose not to use them.

We agree that AUROC and F1 are common, but they require event-specific binary predictions. COHESION estimates long-term rainfall-exceedance probabilities, whereas the inventory covers only a few monsoon seasons; classification metrics would therefore conflate long-term susceptibility with event occurrence. We instead use the KS distance descriptively to compare observed and modelled slope distributions and explain this choice in the Methods:

> ***Traditional landslide susceptibility studies often evaluate predictive performance using classification metrics such as AUROC or F1 scores (e.g. Petschko et al., (2014)). However, these metrics require binary predictions of failure occurrence for specific events. In contrast, our model estimates long-term average failure probabilities derived from multi-decadal rainfall statistics, while the landslide inventory represents a limited time window (2010–2015). Direct classification-based validation would therefore conflate event-specific triggering with long-term susceptibility.***

We also added two robustness checks: bootstrap estimates of the KS distance and 100 runs with spatially perturbed soil-strength fields.

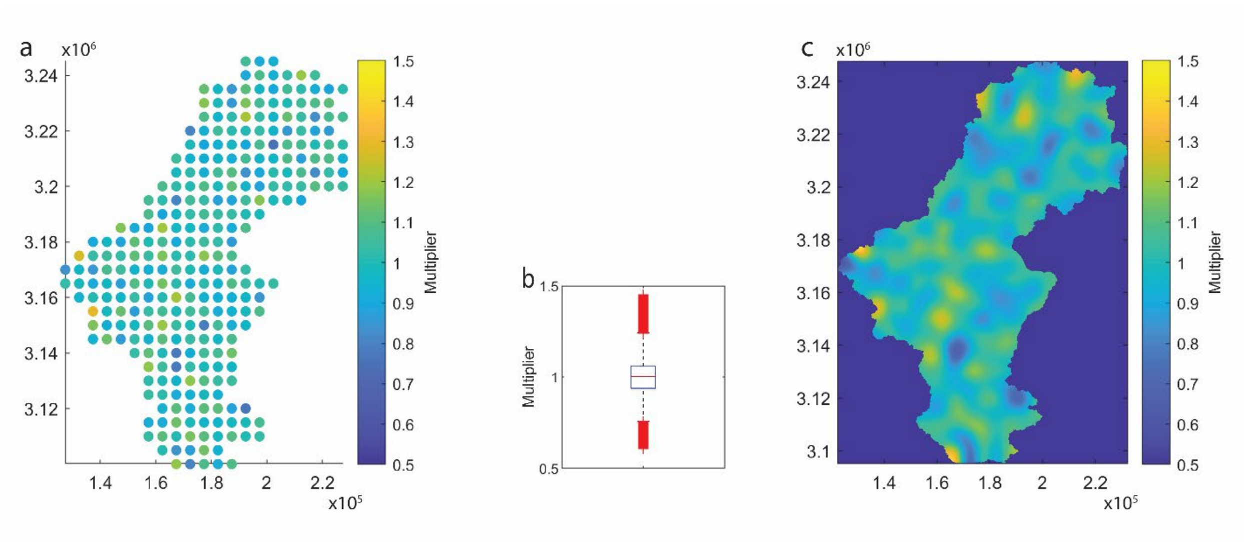
The full methods and results for these robustness checks are provided in the SI because they evaluate model robustness rather than define the COHESION workflow; the main manuscript reports the key outcomes.

The bootstrap analysis is described below:

> ***While the observation database includes landslides from only a few seasons, the model identifies all potentially unstable slopes. To account for that imbalance and estimate confidence intervals for the KS value, we perform 100 bootstraps, using 1000 random samples from modelled (using parameters resulting in the lowest KS value) and observed landslides, and report the mean and standard deviation of the KS value across those 100 comparisons.***

The perturbation analysis is described in the Supplemental Methods and Results:

> ***We devised an additional experiment to test sensitivity to spatial variability in subsurface strength that is not represented in the available data. We assume that unobserved spatially heterogeneous factors can make soils locally stronger or weaker than indicated by the input data.***
>
> ***We generated 100 spatially correlated perturbation fields from control points spaced 5 km apart across the basin. Point values were drawn from normal distributions centered on 1, interpolated to continuous rasters, and smoothed with a Gaussian filter.***
>
> ***For each run, the same perturbation field was applied to cohesion and friction angle, so both varied in the same direction and represented locally stronger or weaker soils. This avoids compensating perturbations and tests sensitivity to aggregate soil strength rather than independently propagating uncertainty in c and ϕ.***



> ***Figure 1: Random perturbation grids are used to study model sensitivity to uncertain spatial distributions of subsoil parameters. a) sampling points spaced***

with 5km distance across the model domain. Each point is assigned a perturbation value drawn from a normal distribution centered around 1, with the resulting distribution of perturbance values shown in b. Point values are interpolated to a spatially continuous disturbance field (c) using a Gaussian filter. Each of 100 generated disturbance grids is then multiplied with cohesion and friction angle values to represent spatial distribution of stronger (perturbance >1)/weaker (perturbance <1) soils

5. Last but not least, I found the discussion to be very limited, and the authors did not provide any comparison with existing models and tools for estimating runout distances, susceptibility, and connectivity. It would be nice to see how their approach fits within the existing frequentist frameworks developed in Europe (Steger et al. 2021; Steger et al. 2022) and New Zealand (Spiekermann et al. 2022) for joint landslide connectivity-susceptibility predictions. On the runout side, tools such as Flow-Py (D'Amboise et al. 2022), GPP (Wichmann 2017), Grfin Tools, and others should be discussed, and how COHESION complements or differs from these.

We agree and substantially expanded the Discussion to position COHESION among existing susceptibility-connectivity and runout models.

We now compare COHESION with the approaches of Steger et al., Spiekermann et al., Brideau et al., and de Vilder and Massey, as well as Flow-Py, GPP, and Grfin.

The revised text frames COHESION as a computationally light basin-scale screening framework that links probabilistic slope stability, object delineation, empirical volume estimation, and DEM-based routing. It complements rather than replaces detailed runout models.

We believe this addresses the requested comparison.

> *Taken together, these limitations underscore that COHESION is intended as a basin-scale screening framework, with its primary contribution being the integration of susceptibility and hillslope connectivity to represent cascading hazard pathways, while remaining computationally tractable for regional analyses. Within the broader landscape of landslide hazard assessment approaches, COHESION occupies a position between statistical susceptibility mapping approaches and more detailed, mechanistic runout simulation tools. Several recent studies have explicitly combined susceptibility mapping with morphometric connectivity metrics to estimate joint release and sediment delivery probabilities (Spiekermann et al., 2022; Steger et al., 2021, 2022). Similar to COHESION, these approaches recognize that landslide hazard cannot be fully characterized without accounting for downslope connectivity. However, release susceptibility in these frameworks is typically inferred statistically from mapped inventories, and connectivity is evaluated using terrain-based indices rather than through explicit estimation of landslide magnitude and*

runout distance. In contrast, COHESION derives release probability from an infinite-slope formulation driven by probabilistic rainfall forcing and aggregates conditionally unstable cells into landslide objects from which volume and runout are estimated explicitly. Approaches, such as those developed by Brideau et al. (2021) and de Vilder and Massey (2024) similarly address the relationship between landslide magnitude and mobility by deriving statistical relationships between volume and travel distance or DH/L ratios, and by providing probability-of-runout-exceedance curves for different landslide types. These methods offer valuable constraints on expected mobility and are well suited for regional hazard assessment for locations with available landslide inventories. However, they do not explicitly couple probabilistic triggering, object-based landslide delineation, and DEM-based routing within a single spatially distributed framework, as implemented in COHESION. In parallel, physically based or semi-empirical tools such as Flow-Py (D'Amboise et al., 2022), the Gravitational Process Path model (Wichmann, 2017), and Grfin (USGS, 2023) simulate landslide routing and, in some cases, entrainment and debris-flow growth using simplified physical formulations. These models provide more detailed representation of flow paths and deposition patterns and are well suited for site to catchment-scale applications with detailed parameterization. COHESION differs in that it prioritizes basin-scale applicability and computational efficiency, representing first-order cascading effects of rainfall-induced slope failures using globally available datasets rather than resolving detailed flow mechanics.

Minor issues

table 1 — as far as I know, SoilGrids has a resolution of 250 m, not 1 km.

We used the previous 1 km SoilGrids product. We updated the model and Table 1 to the current 250 m product.

figure 3 — it is not related to the figure per se, but I would like to read some discussion on this topic somewhere in the text. Why did you use reanalysis data, such as ERA-5L, for example? The northern part of the basin is represented with a very limited number of meteostations located mostly in the valley bottom. The quality of interpolation in the north-eastern part of the basin is questionable due to the clustering of stations.

We retained gauge observations because ERA5-Land would require bias correction and downscaling, introducing other uncertainties. We now note this as a useful future comparison.

figure 9 — the d facet has never been discussed in the text, and it is, in my opinion, the most controversial. Why does the cumulative percentage of roads for the "reforestation" scenario start from 20 something? Why not from zero? In the current form, I can interpret the results as that for the cumulative percentage of roads of

approximately 37%, the failure probability equals roughly 0.6 for both reforestation and deforestation scenarios.

We removed the reforestation scenario and used the space for the analyses requested by the reviewers.

line 293 — specify the kriging approach you used as well as the quality assessment. In other words, how confident are you that your interpolation was good?

We now specify ordinary kriging and the software used. We also acknowledge that formal kriging cross-validation was beyond the scope of this framework paper.

We added the following clarification:

> ***We then used ordinary kriging (Wolfgang Schwanghart, 2010) to interpolate $\mu_i \wedge \sigma_i$ for each cell in the Kaligandaki basin (Figure 3). Note that we did not perform an error analysis for the kriging, but such an analysis could be performed in future iterations. Similarly, it could also be studied if using gridded reanalysis products for precipitation could improve model results.***

References

de Vilder SJ, Massey CI. 2024. Guidelines for natural hazard risk analysis on public conservation lands and waters – Part 4: a commentary on analysing landslide risk to point and linear sites. Lower Hutt (NZ): GNS Science. 81 p. Consultancy Report 2024/38.

Brideau, Marc-André, Saskia de Vilder, Chris Massey, Andrew Mitchell, Scott McDougall, and Jordan Aaron. 2021. "Empirical Relationships to Estimate the Probability of Runout Exceedance for Various Landslide Types." In Understanding and Reducing Landslide Disaster Risk: Volume 2 From Mapping to Hazard and Risk Zonation, edited by Fausto Guzzetti, Snježana Mihalić Arbanas, Paola Reichenbach, Kyoji Sassa, Peter T. Bobrowsky, and Kaoru Takara, 321–27. ICL Contribution to Landslide Disaster Risk Reduction. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-60227-7_36.

D'Amboise, Christopher J. L., Michael Neuhauser, Michaela Teich, Andreas Huber, Andreas Kofler, Frank Perzl, Reinhard Fromm, Karl Kleemayr, and Jan-Thomas Fischer. 2022. "Flow-Py v1.0: a customizable, open-source simulation tool to estimate runout and intensity of gravitational mass flows." Geoscientific Model Development 15 (6): 2423–39. https://doi.org/10.5194/gmd-15-2423-2022.

Jones, Katie E., Jamie D. Howarth, Chris I. Massey, Biljana Luković, Pascal Sirguey, Corinne Singeisen, Caleb Gasston, Regine Morgenstern, and William Ries. 2025. "An Alternative to Landslide Volume-Area Scaling Relationships: An Ensemble Approach Adopting a Difference Model to Estimate the Total Volume of Landsliding Triggered by the 2016 Kaikōura Earthquake, New Zealand." Landslides 22 (7): 2219–36. https://doi.org/10.1007/s10346-025-02479-x.

Larsen, Isaac J., David R. Montgomery, and Oliver Korup. 2010. “Landslide Erosion Controlled by Hillslope Material.” Nature Geoscience 3 (4): 247–51. https://doi.org/10.1038/ngeo776.

Petschko, H., A. Brenning, R. Bell, J. Goetz, and T. Glade. 2014. “Assessing the quality of landslide susceptibility maps – case study Lower Austria.” Natural Hazards and Earth System Sciences 14 (1): 95–118. https://doi.org/10.5194/nhess-14-95-2014.

Rickenmann, Dieter. 2005. “Runout Prediction Methods.” In Debris-Flow Hazards and Related Phenomena, edited by Matthias Jakob and Oldrich Hungr, 305–24. Springer Praxis Books. Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-27129-5_13.

Spiekermann, Raphael I., Hugh G. Smith, Sam McColl, Lucy Burkitt, and Ian C. Fuller. 2022. “Development of a Morphometric Connectivity Model to Mitigate Sediment Derived from Storm-Driven Shallow Landslides.” Ecological Engineering 180 (July): 106676. https://doi.org/10.1016/j.ecoleng.2022.106676.

Steger, Stefan, Volkmar Mair, Christian Kofler, Massimiliano Pittore, Marc Zebisch, and Stefan Schneiderbauer. 2021. “Correlation Does Not Imply Geomorphic Causation in Data-Driven Landslide Susceptibility Modelling – Benefits of Exploring Landslide Data Collection Effects.” Science of The Total Environment 776 (July): 145935. https://doi.org/10.1016/j.scitotenv.2021.145935.

Steger, Stefan, Vittoria Scorpio, Francesco Comiti, and Marco Cavalli. 2022. “Data-Driven Modelling of Joint Debris Flow Release Susceptibility and Connectivity.” Earth Surface Processes and Landforms 47 (11): 2740–64. https://doi.org/10.1002/esp.5421.

Wichmann, Volker. 2017. “The Gravitational Process Path (GPP) model (v1.0) – a GIS-based simulation framework for gravitational processes.” Geoscientific Model Development 10 (9): 3309–27. https://doi.org/10.5194/gmd-10-3309-2017.

Reviewer 2

General comment

The paper proposes a probabilistic framework for landslide risk assessment in large areas, which couples landslide susceptibility with hillslope connectivity, so to better estimate landslide volumes and runouts, which allows a more reliable identification of assets at risk. The topic is of high interest for the readership of NHESS. The manuscript is well structured, and the English language is good, so that the reading is easy.

Although the idea of linking hillslope connectivity with landslide susceptibility is valuable, the manuscript suffers from some methodological flaws that make the present version not suitable for publication. Specifically, the major issues deal with the probabilistic modelling of rainfall;

the formulations used for the assessment of slope stability and the calibration of the parameters therein; the choice of treating only rainfall with a probabilistic approach, while other parameters which affect the results even more than rainfall are also uncertain but have been assumed deterministically known; how runout has been evaluated; the analysis of the assessed risk, which omits details seemingly important without explanation.

Some of the above listed major issues pose doubts about the quantitative results presented and the conclusions drawn. Besides, there are several minor points that should be improved. Both major and minor issues are explained in detail hereinafter.

We thank Reviewer 2 for the detailed comments, particularly on notation, probabilistic treatment, and subsurface uncertainty. We corrected the equations and terminology and added new sensitivity experiments. We also make clearer that COHESION is a regional screening framework with unavoidable data and process limitations.

Detailed comments

Line 68 – “Numerical” should be “mathematical”, as the term numerical refers to techniques to solve the (PD) equations of mathematical models.

We agree and replaced the term with “computational modelling,” which is more accessible in the Introduction.

Line 100-101 – Susceptibility maps refer to the proneness of a slope to failure. Thus, it is not surprising that they don’t consider connectivity and runout. I totally agree with the authors, but the point here should be that mapping susceptibility is not enough to assess landslide risk, rather than stating that susceptibility does not consider exposure.

We agree and revised the paragraph as follows:

> ***Typically, susceptibility maps are derived in a gridded model domain (“rasters”) consisting of individual grid cells. Susceptibility values are then determined for each of the individual cells. By design, the resulting maps represent the spatial distribution of landslide hazards in a spatially continuous grid but without considering the connectivity between adjacent grid cells. This seemingly minor technical distinction has major implications for management and science.***

We also added the following clarification:

> ***Thus, while practical for landscape-scale applications, mapping susceptibility alone is not sufficient to evaluate landslide hazards and risks.***

Line 141 – The second “(B)” should be “(C)”.

Fixed

Line 180 – The equation adopted for calculating FS is an approximation, as (to my understanding) the soil column of thickness z_i is schematized as being totally saturated in the lower part of thickness m_i*z_i and totally dry above. I suggest specifying the hypotheses behind the chosen formulation, especially because the proposed methodology could be applied also with FS calculated with different equations (still with the infinite slope assumption, but with a formulation closer to reality that considers the presence of unsaturated soil). This would help highlight the general validity of the proposed approach.

> **We now state the hydrostatic assumptions behind this standard infinite-slope formulation and note that COHESION could also use alternative pixel-level FOS formulations.**

Line 184 – It would be preferable to define z_i as soil thickness, and to specify that FS is evaluated assuming that the slip surface is at the base of the soil column of thickness z_i.

> **We now define soil thickness as the assumed depth of the failure plane.**

Lines 185-186 – I understand that you are considering that the soil has always the same unit weight (i.e., 15.7 N/m^3, corresponding to about 40% soil porosity if the density of solid particles is assumed around 2700 kg/m^3), even in a catchment as large as 6000 km^2. This assumption looks doubtful, and if the absence of the subscript is not a typo, it should be motivated somehow based on soil characteristics.

> **We agree that a spatially uniform soil unit weight is a simplification. Distributed density data are unavailable, and the selected value is close to values reported for unconsolidated material in Nepal (Kumar and Sengupta, 2024). We now state this regional-scale assumption explicitly.**

Line 187 – In the equation, the friction angle is without the subscript i, as it was constant. I understand that this is a typo. Maybe also γ_s should be γ_{si}?

> **The notation reflects the implementation: friction angle varies with slope, whereas soil unit weight is constant across the model domain.**

Line 193 – The assumption of linking friction angle to slope inclination relies on considerations made in the reference given at line 197. However, in that paper slope stability modelling was not the focus, but just a means to understand what mechanism might be behind observed landslides. Using a relationship like equation 2 for slope stability assessment (even at regional scale) deserves more discussion. From a physical point of view, it makes sense to expect that stronger soil rests on steeper slopes (although cohesion may change this simplistic picture). In any case, this does not imply that soil resting on gentle slopes must be weak, as equation 2 implies (there is no physical reason for this). In any case, even on steep slopes, assuming that friction angle must always be (much) smaller than inclination angle gives the cohesion (both soil effective cohesion and additional cohesion from root reinforcement) the entire task of ensuring slope equilibrium.

We agree that the original relationship forced unrealistically low friction angles on gentle slopes. We changed it to impose a 20° minimum while retaining slope dependence:

This ensures that the friction angle remains at least 20°, including on gentle slopes.

I played a bit with equation 1. For instance, evaluating FS at the depth z=2 m and with the values of soil strength parameters obtained through calibration (i.e., b=0.65 and c=5 kPa) indicates that, without roots, the equilibrium of all slopes with α>28°, even if completely dry (m_i=0), is impossible. If we consider a 50% saturated slope (m_i=0.5) all slopes without roots with α>17° would fail. Root reinforcement, introduced as additional 2 kPa cohesion, totally changes the picture: all slopes up to α=18° are unconditionally stable, even if the soil is totally saturated, and with m_i=0.5 failure occurs for any α>26°. In short, slopes are too weak.

Considering a 30° inclined slope with z=2m and m_i=0.5 (FS=1.01), reducing cohesion from the value with roots, 7 kPa, to 6 kPa makes FS reduce of about 8.2% (FS=0.93). To get the same reduction of FS, m_i should grow up to 0.71 (more than additional 170 mm infiltrated rainwater). Thus, slope stability seems less sensitive to the effects of rain than to (arbitrarily) chosen soil strength parameters.

I understand that the study refers to an area where information about soil properties is uneasy to obtain, and that the proposed methodological framework would remain valid even with a weak modelling of soil stability. But the results are so much affected by the way soil strength is modelled (think about reforestation/deforestation effects, in a model where the additional cohesion due to root reinforcement is by far the major control of slope stability), that I wonder how many of the obtained quantitative results would survive if just a little change (improvement) of slope stability modelling was made.

Line 207 – The subscript i is missing at the left-hand side. The DEM resolution is constant throughout the entire map of the basin, so B should not have the subscript i.

We replaced b_i with b and corrected inconsistent use of lower- and upper-case symbols.

Line 211 – Is K constant everywhere? Or should it be K_i?

Yes. We now state explicitly that K is held spatially constant in this implementation.

Line 218 – It would be clearer if the equality $Q_i=q_i*A_{D,i}$ was given in equation 3 (as it is in equation 12 at line 255). To my understanding, B (the size of a pixel) should be constant so without the subscript j.

We expanded the equation for clarity:

Line 220-221 – The meaning of kf_i and how it is estimated should be described.

We derived k_f values from previous applications in South Asia (Hamel et al., 2020; Mandle et al., 2017), and added the parameter to Table 2.

Line 225 – There is a typo. I guess it should read "$Q_{R,j}$ and $Q_{D,j}$".

Corrected the indices and notation throughout.

Line 227 – The word "runoff" is probably missing at the end of the line.
Fixed.

Lines 254-257 – The width of a pixel was indicated with capital B (here it is b), and it should not have the subscript i, as the DEM resolution is not changing from pixel to pixel.
Fixed, and standardized b throughout.

Line 257 – The equation is incorrect (Q^*_i should be only at the right-hand side).
Fixed and reformatted for legibility:

$$Q_i^* = m_i^* b T_i$$

Lines 271-284 – In equation 15, $p^*_i(e)$ represents the precipitation capable of triggering a landslide at cell i depth during an event e, and rainfall events may have any duration. However, the EV type I (Gumbel) distribution, and the parameters therein, should be calibrated based on a set of measurements of extreme rainfall all sharing the same duration (e.g., Koutsoyiannis et al., 1998). For instance, μ_i and σ_i would be different if the probability distribution refers to hourly rainfall depths or daily rainfall depths (as well as for any other different rainfall duration that might be considered). What kind of rainfall data (hourly? daily?) has been used to estimate the parameters of equations 16 and 17 at the positions of the rain gauges represented in Fig. 3, then used to interpolate them throughout the whole basin area with kriging? This should be specified and, even more importantly, it should be explained how the same probability distribution (obtained from a set of rainfall data with a given duration) can be used to describe the probability of rainfall depths falling during events with different durations. To avoid mistakes in the calculated probabilities, maps like those shown in Figure 3 should be produced for any possible rainfall duration, and the appropriate one should be chosen according to event duration, so to get the correct parameters (and to calculate the correct probability). Alternatively, a fixed rainfall duration might be assumed, but in such a case it would not be correct to refer to rain events (e.g., if the fixed duration is one day, as it may seem from the text at lines 491-495, p* should be defined as the critical daily precipitation rather than event precipitation).

We agree that the extreme-value distribution must refer to a fixed rainfall duration. The Gumbel parameters are estimated from annual maxima of daily rainfall totals, and p_i* is now defined as the critical daily precipitation depth. The exceedance probability therefore refers to annual maximum daily precipitation exceeding this threshold. Other durations would require separate duration-specific distributions and parameter maps.
We also removed the misleading term "event" and the unnecessary subscript e, and revised the notation throughout this section.

Lines 290-293 – Again, case-specific information is mixed with the general methodology.

We retained limited case-specific information where it is needed to define the rainfall formulation, but removed unnecessary implementation detail and clarified the distinction between general method and case-study choices.

Lines 294-300 – To my understanding, all the spatially variable parameters except precipitation have been considered deterministic. This should be better underlined, as the sentence at lines 299-300 may be misleading,
We agree. In this implementation, only extreme rainfall is treated probabilistically; soil depth, cohesion, friction angle, and other inputs are spatially variable but deterministic. The resulting failure probabilities are probabilistic because they derive from rainfall-threshold exceedance. We clarified this distinction and discuss broader uncertainty propagation as future work:

In the present implementation, stochasticity is introduced only through the probabilistic representation of extreme rainfall, while most other parameters are treated as spatially-variable, but deterministic.
and the sentence at the beginning of section 4.3 does not help (lines 241-242: "Herein, we consider for the spatiotemporal variability in soil moisture through a statistical approach"). Indeed, the variability of soil properties suffers the same uncertainty as rainfall, meaning that soil properties are available at a very small number of (sparse) locations, compared to the extension of the basin where the analysis is run.
We agree and now distinguish spatial heterogeneity from probabilistic uncertainty explicitly.

And I believe that also soil thickness z_i at each node of the grid is a mere estimate. The uncertainty affecting soil and slope properties can be treated with a probabilistic approach in a similar way to what you do with rainfall, although with a different probability distribution (e.g., Roman Quintero et al., 2025). Given the poor results obtained in terms of identification of unstable zones (lines 608-610), likely due to the way soil strength parameters have been assigned (see some of my comments, but also your discussion at lines 614-618), probably treating also the uncertainty of soil parameters with a probabilistic approach (e.g., Almeida et al., 2017; Roman Quintero et al., 2025) would be worth.
We agree that a full probabilistic treatment of soil parameters would be valuable but is beyond this framework paper. As a first test, we added a perturbation experiment that creates spatially coherent stronger and weaker soil fields. Cohesion and friction angle

vary in the same direction to avoid compensating changes; the experiment is therefore a sensitivity test of aggregate soil strength, not full independent uncertainty propagation.

The detailed perturbation methods and results, together with the bootstrap and resolution/volume diagnostics, are provided in the SI. We summarize their implications in the main manuscript because these analyses qualify uncertainty in the case-study implementation rather than form core steps of the framework.

Perturbation analysis

We devised an additional experiment to test sensitivity to spatial variability in subsurface strength that is not represented in the available data. We assume that unobserved spatially heterogeneous factors can make soils locally stronger or weaker than indicated by the input data.

We generated 100 spatially correlated perturbation fields from control points spaced 5 km apart across the basin. Point values were drawn from normal distributions centered on 1, interpolated to continuous rasters, and smoothed with a Gaussian filter.

For each run, the same perturbation field was applied to cohesion and friction angle, so both varied in the same direction and represented locally stronger or weaker soils. This avoids compensating perturbations and tests sensitivity to aggregate soil strength rather than independently propagating uncertainty in c and ϕ.

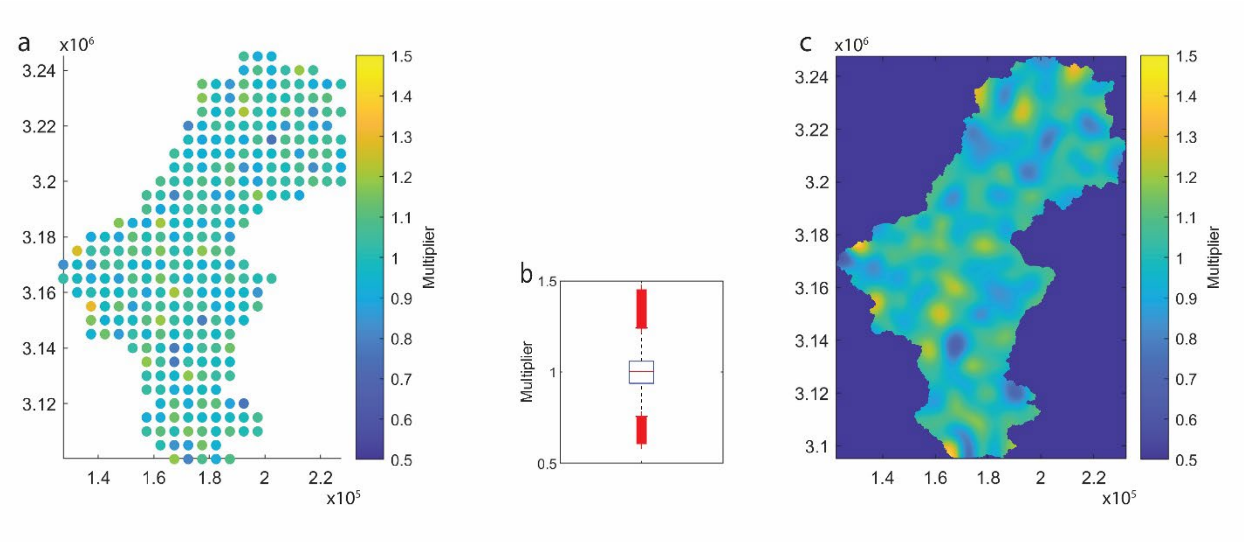


Figure 2: Random perturbation grids are used to study model sensitivity to uncertain spatial distributions of subsoil parameters. a) sampling points spaced with 5km distance across the model domain. Each point is assigned a perturbation value drawn from a normal distribution centered around 1, with the resulting distribution of perturbance values shown in b. Point values

are interpolated to a spatially continuous disturbance field (c) using a Gaussian filter. Each of 100 generated disturbance grids is then multiplied with cohesion and friction angle values to represent spatial distribution of stronger (perturbance >1)/weaker (perturbance <1) soils

Lines 304-305 – The last words in the figure caption seem to be a typo and should be removed (i.e. "Connected slope failure assessments").

We moved this text to a separate subsection, which also addresses the next comment.

Line 306 – From this line onward, the part of the methodology dealing with connectivity and landslide objects is described. This is an essential part of the proposed methodology and has nothing to do with how the probability of failure is calculated; thus, it would be better to start here a separate subsection 4.4.

Line 329 – More details about how the function works and how it has been used (it has a parameter to be set) should be given.

We added a short description of the connectivity criterion:

> ***To delineate connected zones of instability, we first derive the D8 flow directions using functions from TopoToolbox. Topotoolbox is also used to group conditionally unstable cells that are downslope into a landslide object.***

COHESION uses D8 flow directions in TopoToolbox to group downslope-connected conditionally unstable cells. We clarified that this grouping has no additional user-set parameter.

Line 342 – There is a typo: a "T" is missing at the beginning of the line (it should be "This").
Fixed.

Line 367 – The statement is physically not correct: landslide volume determines the mass, but to calculate the kinetic energy also the velocity is needed, which in turn depends directly on the elevation drop from the unstable zone to the toe where the kinetic energy is evaluated (e.g., the available potential energy) and only indirectly on the slope gradient (owing to energy loss along a longer pathway). Indeed, empirical equation 24 considers only the elevation drop and neglects the gradient.

We rewrote the paragraph to correct the physical interpretation:

> ***As a final step, we relate landslide runout length to the volume of mobilized material. Runout length is estimated using the empirical scaling relationship proposed by Rickenmann (2005), originally derived for debris-flows, i.e., rapidly mobilizing mass movements. This formulation relates vertical drop to travel distance and implicitly assumes rapid, flow-like mobilization. While computationally efficient for basin scale screening, this approach does not resolve detailed flow dynamics or rheology, and therefore should not be interpreted as a process-resolving runout simulation. Specifically, we follow***

Rickenmann's (1999, 2005) empirical formula linking distance travelled, landslide volume, and elevation drop. . Consequently, the more material is mobilized and the greater the elevation drop, the longer travels landslide material . This relationship has been established through empirical studies (Rickenmann, 1999, 2005). Following the equations presented in Rickenmann (1999, 2005) we calculate runout downslope of an LSO as

$$L_{LSO}=1.9V_{LSO}^{0.16}H_{LSO}^{0.83}$$

Lines 373-376 – It is unclear how the runout distance is considered: does it start from the toe of the slope, or does it also include the path travelled along the slope?

We now state explicitly that runout begins at the lowest cell of each LSO:

L_{LSO} ***is the runout length and*** H_{LSO} ***is the vertical drop between the starting point of runout (the lowest point of the landslide scar) and the downslope end point along the runout path (both in meters)***

In this respect, the statement "These two quantities are not independent, as longer runout distances naturally involve greater vertical drops" sounds at the same time obvious (the dependence is in equation 24) and misleading (do the authors here mean anything different from the link empirically expressed by equation 24?).

We rewrote these paragraphs to remove the ambiguous statement:

While computationally efficient for basin scale screening, this approach does not resolve detailed flow dynamics or rheology, and therefore should not be interpreted as a process-resolving runout simulation. Specifically, we follow Rickenmann's (1999, 2005) empirical formula linking distance travelled, landslide volume, and elevation drop. This relationship has been established through empirical studies (Rickenmann, 1999, 2005). Following the equations presented in Rickenmann (1999, 2005) we calculate runout downslope of an LSO as

$$L_{LSO}=1.9V_{LSO}^{0.16}H_{LSO}^{0.83}$$

1

L_{LSO} ***is the horizontal distance from the toe of the landslide (in the model defined as the lowest cell in an LSO) and*** H_{LSO} ***is the vertical drop between that cell and the elevation at the end of the runout path. In a gridded model, it is practical to evaluate the condition in Equations 24 sequentially for all cells along each potential downslope path. Each step will increase*** L_{LSO} ***either by*** b ***or by*** $\sqrt{(2b)}$***, depending on whether flow proceeds orthogonally or diagonally between grid cells, and where*** b ***is the grid-cell size. Simultaneously, the elevation drop*** H_{LSO}***between the initiation cell and the current downslope cell is updated. For each downslope step, the empirical***

relationship is evaluated using the current value of H_{LSO}. Runout propagation terminates when the cumulative travel distance exceeds the distance predicted by Equation 24, or when the flow path intersects a river channel. It should be noted that Rickenmann's empirical relation is built on specific inventories of debris flows, but builds on a well developed body of literature linking volumes of slope failures (e.g., from landslides, debris flows and rockfalls) to runout distance (Brideau et al., 2021) and was shown to be relatively consistent with measured landslide and debris flow runout in Nepal (Dahlquist and West, 2019).

Dahlquist, M. P. and West, A. J.: Initiation and Runout of Post-Seismic Debris Flows: Insights From the 2015 Gorkha Earthquake, Geophysical Research Letters, 46, 9658–9668, https://doi.org/10.1029/2019GL083548, 2019.

Lines 377-380. The description of the procedure is not clear, and I have doubts about its correctness (at least, from what I could understand). If the remaining runout distance is reevaluated with equation 24 at each downslope step with a new elevation drop, this implies that the already attained velocity of the mass sliding from upslope is neglected, as equation 24 refers to the position of the landslide scarp, where the mass is initially still. The sketch in the corner of Fig. 5 does not help: for the γ_4 path (which consists of three pixels), a drop δ_h is represented (it should be δH_h, if I correctly understand), which should be the "the elevation drop from LSO_k to i" (as explained at lines 382-383). To improve the clarity, besides correcting all the inaccuracies in the symbols and in the figure, I suggest indicating the position of node i (it should be where the vertical segment marking δ_h touches the ground line). However, the corresponding horizontal segment should be δL_h (explanation at line 383: "δL_h is the horizontal distance from LSO_k to I") and not L_4, as indicated in the figure.

We revised the figure and caption to show the node positions, distances, and stopping criterion.

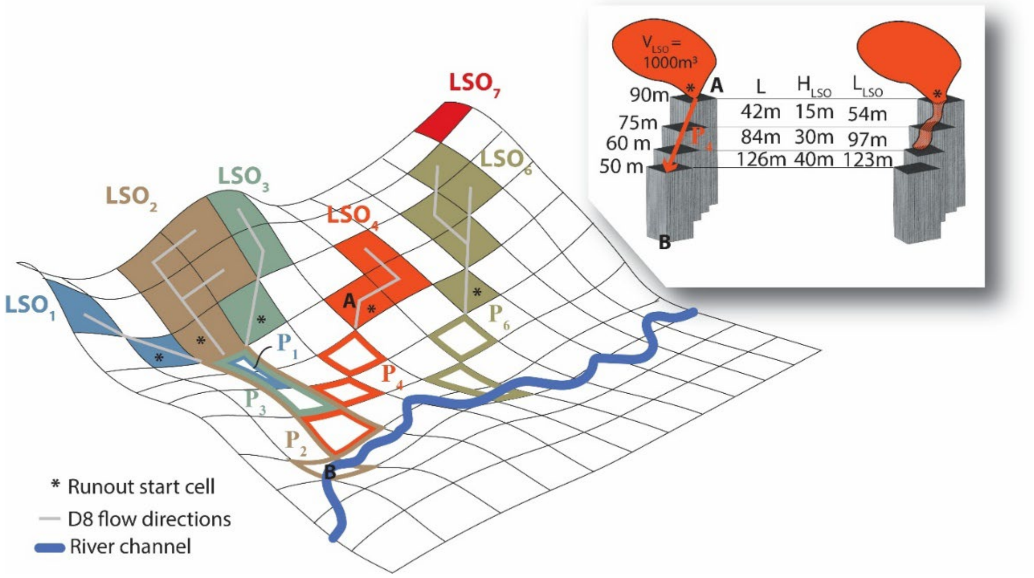


Figure 3: Schematic of runout modelling in a gridded domain. Filled cells indicate different LSOs and colour-matched hollow cells indicate the associated runout paths (P). Start cells for runout calculations are marked with a *. The inset details the runout calculation for LSO_4 following Rickenmann (2005) and assuming a landslide volume V_{LSO} of 1000m3, 30 m cell resolution, and with absolute elevation values of each cell denoted to the right. The runout would stop in the 3rd cell, as the calculated travel distance L_{LSO} becomes smaller than the actual distance to the next cell.

Lines 377 and 379 – Equation 30 does not exist. Please correct.

Corrected.

Line 380 – Please, clarify what you mean with “channel network”. The nearest river branch?

Yes—the nearest downstream river branch. We clarified this.

It seems that Figure 5 has typos, as the runout path of LSO_4 is indicated as γ_3 and the exponent of V is missing in the equation.

Please see the corrected figure.

Lines 410-412 – The KS test is usually carried out comparing the KS statistics with a threshold value, which depends on the number of elements, m and n, belonging to the two compared sets (

significance. I understand that here you are not really trying to check if predicted and observed failures share the same distribution of slope inclinations, but you have m=1170 observed landslides and n, the number of modelled LSO, which is not given in the paper, but should be very high looking at the smoothness of the cumulative blue curves of Figure 6c and 6d (and to the too large extension of unstable areas discussed at lines 609-610). Assuming for instance n=10000, it results $KS_{lim} = 0.014$, meaning that observed and modelled failed slopes are very differently distributed, even when the minimum value KS =0.1 is obtained through calibration. I think that this bad performance, which does not hamper the validity of the proposed approach (a better slope stability modelling could be implemented within the same framework), relies on the friction angles used for assessing slope stability, i.e. the inadequacy of equation 2, the relationship linking friction angle to slope inclination (as indeed you argue in the discussion). Correcting that relationship would be easy, so to get better results.

We agree that formal KS significance is not useful here because the very large model sample would make even small differences significant. We therefore use KS only as a descriptive distance between the observed and modelled ECDFs and as a relative metric across parameterizations, not as a hypothesis test.

We also avoid labelling any KS value as intrinsically good or bad because no accepted benchmark exists for this application and the inventory is incomplete. The selected value is simply the lowest among physically plausible parameterizations. We clarified this and added bootstrap estimates of KS variability.

Line 413-414 – Soil specific weight γ_s has been used in the FS equation. Although soil density and specific weight are obviously related, it would be better to run the sensitivity analysis with reference to the symbols and quantities that appear in the equations used to calculate FS.

We agree that a formal global sensitivity analysis would be valuable, but it is beyond this frame-work paper. We added a spatial soil-strength perturbation experiment as a first robustness test.

Table 2 – The last row of the table reports the friction angle as a calibration parameter, while it is calculated through the empirical relationship given in equation 2, and indeed the parameter b of that equation is indicated as calibration parameter. I suggest removing the last row of the table, as this might appear inconsistent.

Corrected. We renamed the coefficient η because b is already used for DEM cell size.

Figure 6 – The black arrows plotted above a dark blue are hardly visible. I suggest changing the color of the arrows to a lighter one.

We retained the arrow because it crosses both dark and light backgrounds; no single colour remains visible throughout.

Lines 510-513 – The volume of a landslide may not only reduce while it moves downslope, but it may also increase, due to entrainment of more soil and debris along the path. I suggest rephrasing this sentence to consider also this instance.

We agree and revised the sentence to:

> ***It should also be noted that the current model does not account for changes in sediment load along the runout path as sediment might be deposited or entrained. As a function of such deposition or entrainment, an asset located more downslope on runout pathway might be damaged more or less than a more upslope asset.***

Lines 526-529 – Please, double check the references to the curves and the panels of figure 9 for evaluating the percentage of exposure of roads and buildings to landslide and runout, as it seems that there are some discrepancies.

Please note that the relevant sections were removed or significantly reworked

Lines 531 and 532 – So far, the word "building" was used, while now it is "structures". To avoid misunderstanding, I suggest keeping the same word.

Changed to "building" throughout.

Lines 530-535 – The description is unclear. From the graphs of figure 9a and 9b a clear different behavior of building and roads as regards LSO and runout exposure appears, which is not clearly discussed in the text (i.e., buildings are more exposed to LSO with small probability and uniformly exposed to runout with any probability, while the opposite holds for roads). Also, the English language should be improved here.

We removed the deforestation/reforestation analysis. Its results were dominated by assumed root cohesion, while baseline cohesion already reflects a vegetated landscape. Addressing this properly would require a dedicated uncertainty analysis, so we refocused the manuscript on the core framework.

Line 549 – I guess there is a typo in the text in the parentheses, as "on" should probably become "only". This is confirmed by figures 9c and 9d, where the effects of deforestation/reforestation have been evaluated only for LSO exposure. However, this choice should be better motivated (see one of my following comments): what does it mean "for clarity"? Are the effects on runout exposure less clear? And why?

Lines 550-554 – I think that the curves plotted in fig. 9c and 9d deserve a deeper comment. I see that deforestation not only affects very slightly the total % of exposed buildings (as noted by the authors), but it even reduces the buildings exposed to failure probabilities smaller than

0.5. From the cumulative curves, it is difficult to understand what is behind this result: a histogram could allow the reader to better understand if, and to what extent, the buildings exposed to small probability of failure (<0.5) have moved to higher probabilities. This effect looks clearer for the exposure of roads, where clearly deforestation moves assets from low-probability exposure to p>0.9, as indicated by the increase of the slope of the cumulative % in the interval (0.9, 1.0).

Lines 555-564 – Here the reason behind the different effects of deforestation and reforestation: reforestation is made where some landslide risk was present, so it affects failure probability exactly where reducing failure probability is needed. Differently, deforestation occurs where forests, and hence neither buildings nor roads were present, and there is a slight indirect effect only because pixels belonging to the same LSO affect the mean value of the failure probability of that LSO. This makes sense, but it underlines that probably the effects of deforestation on risk should be evaluated on the exposure to the runout rather than on LSO. Why was this choice made?

Figure 9c – Why the green curve ends at p=0.9? It is a cumulative curve, so it should cover the entire probability interval (up to p=1). Even more strangely, the orange dashed curve goes beyond P=1, which is meaningless. Please, correct.

We removed the scenario analysis to make space for the additional model tests and avoid further lengthening the paper.

Line 578 – A typo: “occurs on” instead of “occurs one”.

Corrected.

Lines 614-618 – This paragraph would better fit in the introduction than in the is a repetition of what already discussed in the introduction, and it has

The reviewer comment appears truncated, so we could not identify a requested change.

ines 630-634 – This statement (“well-fitting model”) is not supported by the results, and the sensitivity to soil strength parameters confirms that their estimation is the major weak point of this study, that would deserve a deeper discussion.

We reframed this as: “Expressing friction angle as a function of slope angle is needed to improve the fit between modelled and observed landslide locations.”

References

Almeida, S., Holcombe, E. A., Pianosi, F., and Wagener, T.: Dealing with deep uncertainties in landslide modelling for disaster risk reduction under climate change, Nat. Haz. Earth Syst. Sci., 700, 225–241, https://doi.org/10.5194/nhess-17-225-2017, 2017.

Koutsoyiannis, D., Kozonis, D., and Manetas, A.: A mathematical framework for studying rainfall intensity-duration-frequency relationships, J. Hydrol., 206(1-2), 118-135, https://doi.org/10.1016/S0022-1694(98)00097-3, 1998.

Roman Quintero, D. C., Marino, P., Abdullah, A., Santonastaso, G. F., and Greco, R.: Large-scale assessment of rainfall-induced landslide hazard based on hydrometeorological information: application to Partenio Massif (Italy), Nat. Hazards Earth Syst. Sci., 25, 2679–2698, https://doi.org/10.5194/nhess-25-2679-2025, 2025.

Reviewer 3

This study proposes a GIS-based analytical framework for landslide hazard and risk assessment that incorporates the concept of hillslope connectivity. By analyzing the proximity of unstable slopes and their connectivity to landslide runout pathways, the authors map the probability that mobilized sediment reaches buildings and roads. The work has clear potential to provide valuable insights that could inform landslide mitigation measures and future land-use and river-basin management.

We thank Reviewer 3 for the detailed comments and have addressed them in the revised manuscript.

However, I have the following concerns regarding the methodology:

Major concern — Potential inconsistency between landslide-volume estimation and slope-stability analysis

In this study, the factor of safety is computed using the soil depth assigned to each grid cell when evaluating landslide probability. This setup implicitly fixes the rupture (slip) surface at a specific depth—namely, the base of the soil layer. If so, the landslide volume associated with each landslide object (LOS) should be calculated in a manner consistent with that assumed failure depth (e.g., by integrating soil thickness over the grid cells contained within each LOS).

Instead, the authors estimate landslide volume for each LOS using the global area–volume scaling relationship (Larsen et al., 2010), independent of the soil-thickness distribution used in the infinite-slope stability analysis. This approach could imply a failure depth that differs from (i.e., is shallower or deeper than) the rupture-surface depth assumed in the stability analysis.

If the authors wish to retain Eq. (21) for landslide-volume estimation, they should explicitly discuss and justify its consistency with the assumed rupture-surface depth (i.e., the soil–bedrock interface) used in the slope-stability analysis.

In addition to the major concern outlined above, I identified several minor points that require revision to improve clarity and consistency. Please see the specific comments below.

We agree that the failure depth used in the FOS calculation and the empirical volume estimate are not mechanically identical. Integrating soil depth over every LSO cell would be fully consistent with the assumed failure surface.

However, we interpret each LSO as an envelope of potentially failure-prone terrain, not as a single failure mobilizing the full soil column in every cell. Direct integration could therefore overestimate mobilized volume.

We retain the Larsen et al. (2010) relationship because it provides an observation-based first-order volume estimate for regional screening. We now state this structural inconsistency and its implications explicitly.

We also compare the Larsen-based volumes with estimates obtained by integrating gridded soil depth across LSOs (Supplemental Results 2; Supplemental Figure 6). We retain the detailed comparison in the SI because it is a diagnostic of structural uncertainty, while the main manuscript now states the inconsistency, rationale, and implications.

We added the following clarification:

A conceptual inconsistency exists between the failure depth assumed in the infinite-slope stability analysis and the subsequent landslide-volume estimation. While slope stability is evaluated using the modeled soil thickness at each grid cell, landslide volumes are estimated using the empirical area–volume scaling relationship of Larsen et al. (2010). We adopt this approach because landslide objects (LSOs) are interpreted as outer envelopes of potentially failure-prone terrain rather than uniformly mobilized failures extending to the full modeled soil depth across all included cells.

We also clarify why we use the soil-landslide form of Larsen et al. (2010):

Larsen et al. (2010) studied a dataset of soil and bedrock landslides, comprised of more than *more than 4000 observations. Based on that dataset, they propose that the selection of the exponent* γ *should depend on the landsliding mechanism, with values of* $\gamma_{rock} = 1.35 \pm 0.01$ *and* $\gamma_{soil} = 1.145 \pm 0.008$*. As COHESION is designed to model landslides that are driven by saturation-excess in the soil column, we herein use* γ=1.145 *and* $log\alpha$=0.86 *(i.e.,* α=$10^{0.86}$=7.24*)*.

Specific Comments

Page 1, line 74–75

"(see, e.g., reviews in (Intrieri et al., 2019; Jiang et al., 2022))." → "(see, e.g., reviews in Intrieri et al. (2019) and Jiang et al. (2022))."

Corrected throughout.

Page 2, line 84

"Montgomery David R. and Dietrich William E., 2010;" → "Montgomery and Dietrich, 2010;"

I noticed similar issues elsewhere; please check the manuscript throughout for consistent citation formatting.

We corrected the affected citation metadata and checked the manuscript for similar cases.

Page 3, line 117

“(Ghorbanzadeh et al., 2022)).” → remove the extra parenthesis.

Also, please clarify the intended distinction between volume and mass in this context, as they are not interchangeable.

Corrected to distinguish the two quantities: for each LSO, we calculate both volume and mass, as well as failure probability and runout.

Page 4, line 137–142 (Figure 1)

“and roads (B)” → “and roads (C)”.

Fixed.

Page 5, line 161–162

“Marc et al. (2019a), Figure 2).” → “Marc et al. (2019a) (Figure 2).”

Fixed.

Page 6, line 176–177

The variable defined as τm, i is written as τmi in the subsequent equation. Please standardize notation throughout.

Also, this equation (the first equation appearing in the manuscript) should be numbered.

Both corrected.

Page 6, line 184

The sentences appear broken; “While we do not account for deep percolation,. We use Kent’s (1973) curve number model to estimate surface”

Please correct as appropriate.

Page 7, line 226–227

“fricvtion” → “friction” appears elsewhere. Please standardize.

Fixed.

Page 8, line 237

"(Rawls et al., 1992).." → "(e.g., Rawls et al., 1992)."

Fixed.

Page 8, line 256–259

Eq. (13) appears incorrect: as written, mi*×bi×Ti would always equal 1. Please re-check the equation.

Also, "(Vogl et al., 2019a, b; World Bank, 2019)" should be cited immediately after the phrase "a threshold subsurface flow," before the equation.

Both corrected.

Page 10, line 300

"roduce" → likely a typo (perhaps "reduce" or "reduces"). Please confirm.

Fixed.

Page 10, line 301–305 (Figure 3)

Figure 3 would be easier to interpret if it also showed the spatial distribution of annual maximum precipitation.

We considered this, but retained the figure because adding another raster would make it difficult to read.

Page 10, line 314

Remove the unnecessary comma in "unstable, cells".

Done.

Page 10, line 316–320

Please confirm whether FSi is maximized at mi=0 and minimized at mi=1. As written, the explanation may be reversed.

Also, please number these two equations.

We removed the redundant equations and expanded the explanation of the three stability classes instead.

Page 11, line 331

Consider italicizing k (landslide object index) for consistency with standard mathematical notation.

Done.

Page 13, line 361–363

Eq. (23) has overlapping superscripts/subscripts and is difficult to read. Please rewrite in a clearer format (e.g., by grouping terms in parentheses). The same applies to Eq. (24).

We rewrote both equations and checked the rendered PDF for legibility.

Page 13, line 377

Eq. (30) is not in the text. Please check numbering and citations.

Checked and corrected.

Page 13, line 382

The variable defined as δHh is written as δh in Figure 5. Please ensure consistency.

The runout equations and figure were revised together; see the new Figure 4.

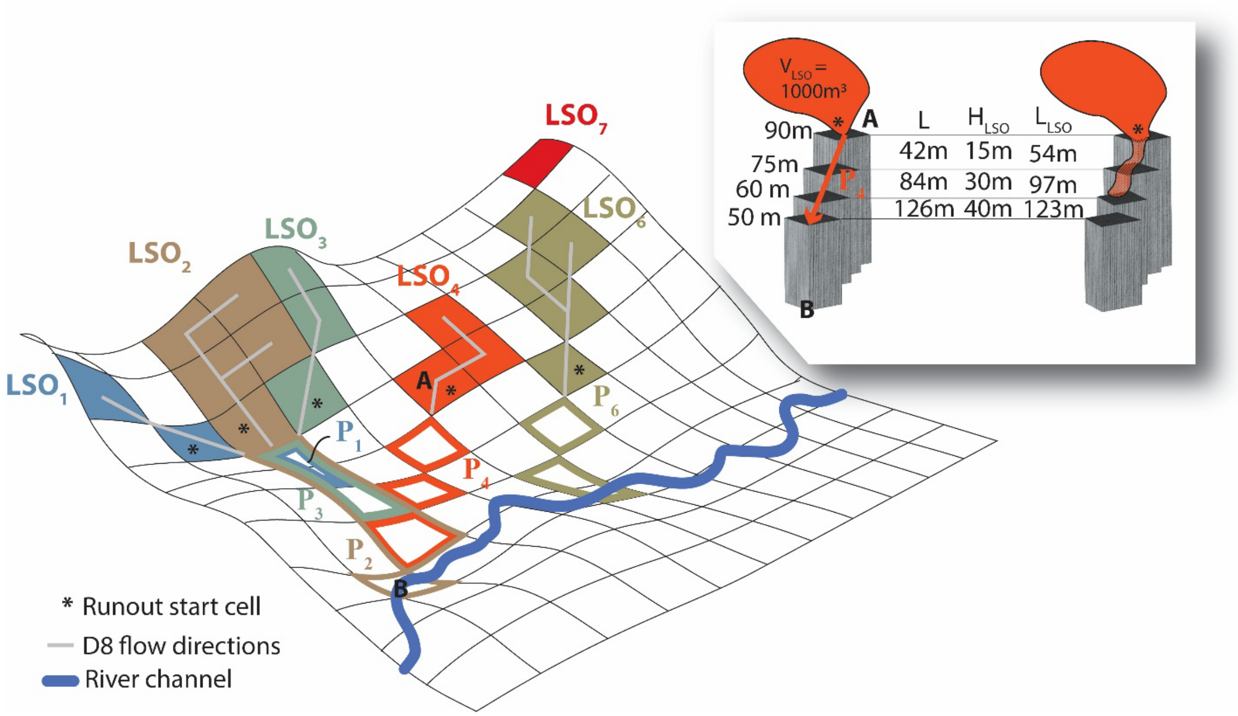


Figure 4: Schematic of runout modelling in a gridded domain. Filled cells indicate different LSOs and colour-matched hollow cells indicate the associated runout paths (P). Start cells for runout calculations are marked with a *. The inset details the runout calculation for LSO_4 following Rickenmann (2005) and assuming a landslide volume V_{LSO} of 1000m^3, 30 m cell resolution, and with absolute elevation values of each cell denoted to the right. The runout would stop in the 3rd cell, as the calculated travel distance L_{LSO} becomes smaller than the actual distance to the next cell.

Page 14, line 385–389 (Figure 5)

The equation shown in the upper-right subfigure differs from Eq. (24) (the V multiplier appears to be missing). Please check which version is correct.

The figure and equation were revised for consistency.

Fixed.

Also, please clarify how the river-line position and extent were determined.

We now state that the river network is DEM-derived using a 5 km² contributing-area threshold:

> ***River networks could be extracted from a DEM or drawn from a global dataset, or propagation of runout could be selected to continue even if cells are classified as river. Herein we used a river network derived from a DEM in a previous study*** (Vogl et al., 2019)

Page 14, line 400

"Marc et al. (2019)" → please specify 2018, 2019a, or 2019b to match the reference list.

Fixed throughout.

Page 17, line 435–439 (Table 2)

"Value source" does not need parentheses.

Replace parenthetical citation style with consistent author–year format where appropriate, e.g.: - (Vanacker et al., 2003) → Vanacker et al. (2003) - (Dhakal and Sidle, 2003; McGuire et al., 2016; Sidle, 1991) → Dhakal and Sidle (2003); McGuire et al. (2016); Sidle (1991)

Also, "Estimated average from Vanacker et al., 2003" → "Estimated average from Vanacker et al. (2003)."

Fixed.

Please add the land-cover coefficient kf to the parameter list.

Added.

Page 18, line 454

Is it necessary to restate an equation already defined earlier as Eq. (2) (line 193–194)?

Consider removing redundancy.

We retained the equation because this is where the three cell-stability classes are interpreted; restating it aids readability.

Page 24, line 566–574 (Figure 9)

Since LOS failure and sediment arrival are more likely at higher probabilities, the plots may be clearer if x-axis is reversed (probability 1.0 → 0.1) and the cumulative curves are shown in descending order.

Also, why does the cumulative curve for the Reforestation scenario in Fig. 9c stop at 0.9?

The reforestation figures were removed with that analysis.

Page 26, line 602–604

This sentence is unclear and contains typos (e.g., “areas area”). Please rewrite for grammatical correctness and clarity.

Revised as follows:

> ***Including runout increases the mapped landslide-hazard footprint by 53.3% relative to LSO source areas alone, showing that downslope runout substantially expands the area identified as potentially affected.***

Page 27, line 639

Delete the double period.

Done.

Page 27, line 663–664

Remove the duplicated wording “such as CASCADE.”

Done.

Rebuttal references

Brideau, M.-A., de Vilder, S., Massey, C., Mitchell, A., McDougall, S., and Aaron, J.: Empirical Relationships to Estimate the Probability of Runout Exceedance for Various Landslide Types, in: Understanding and Reducing Landslide Disaster Risk: Volume 2 From Mapping to Hazard and Risk Zonation, edited by: Guzzetti, F., Mihalić Arbanas, S., Reichenbach, P., Sassa, K., Bobrowsky, P. T., and Takara, K., Springer International Publishing, Cham, 321–327, https://doi.org/10.1007/978-3-030-60227-7_36, 2021.

Petschko, H., Brenning, A., Bell, R., Goetz, J., and Glade, T.: Assessing the quality of landslide susceptibility maps – case study Lower Austria, Natural Hazards and Earth System Sciences, 14, 95–118, https://doi.org/10.5194/nhess-14-95-2014, 2014.

Rickenmann, D.: Empirical Relationships for Debris Flows, Natural Hazards, 19, 47–77, https://doi.org/10.1023/A:1008064220727, 1999.

Rickenmann, D.: Runout prediction methods, in: Debris-flow hazards and related phenomena, Springer, 305–324, 2005.

Spiekermann, R. I., Smith, H. G., McColl, S., Burkitt, L., and Fuller, I. C.: Development of a morphometric connectivity model to mitigate sediment derived from storm-driven shallow landslides, Ecological Engineering, 180, 106676, https://doi.org/10.1016/j.ecoleng.2022.106676, 2022.

Steger, S., Mair, V., Kofler, C., Pittore, M., Zebisch, M., and Schneiderbauer, S.: Correlation does not imply geomorphic causation in data-driven landslide susceptibility modelling – Benefits of exploring landslide data collection effects, Science of The Total Environment, 776, 145935, https://doi.org/10.1016/j.scitotenv.2021.145935, 2021.

Steger, S., Scorpio, V., Comiti, F., and Cavalli, M.: Data-driven modelling of joint debris flow release susceptibility and connectivity, Earth Surface Processes and Landforms, 47, 2740–2764, https://doi.org/10.1002/esp.5421, 2022.

de Vilder, S. J. and Massey, C. I.: Guidelines for natural hazard risk analysis on public conservation lands and waters – Part 4: a commentary on analysing landslide risk to point and linear sites., GNS Science, Lower Hutt (NZ), 2024.

Vogl, A. L., Schmitt, R. J. P., Simpson, R. D., Bryant, B. P., Wolny, S., and Narain, U.: Valuing Green Infrastructure: Volume I: Case Study of Kali Gandaki Watershed, Nepal, The World Bank, Washington DC, 2019.

Wolfgang Schwanghart: kriging, 2010.