Responses to the reviewers September 2024

Christoph Schaller[1,2], Luuk Dorren[1], Massimiliano Schwarz[1], Christine Moos[1], Arie C. Seijmonsbergen[2], and Emiel E. van Loon[2]

[1]Bern University of Applied Sciences - HAFL, Länggasse 85, 3052 Zollikofen, Switzerland
[2]University of Amsterdam UVA - IBED, Sciencepark 904, 1098 XH Amsterdam, The Netherlands

Correspondence: Christoph Schaller (christoph.schaller@bfh.ch)

Date: 13 September 2024

Dear Reviewers, dear Editor,

Thank you for your feedback and the decision on our response to the reviewers' comments. In accordance with your decision, we have made major revisions to our manuscript based on our previous response. We have integrated most of the points as explained in the detailed text below. We tested the NFI forest type as an additional covariate but did not include it in the final model. We improved several visualisations taking into account some of the reviewers' suggestions. We also provide additional high-resolution variants for most of the graphics. The first two chapters were combined and partially revised to address the concerns stated by the reviewers. We also revised the methods section aiming to emphasise the explored variables and methods more clearly. The results have been updated based on the modifications stemming from the adaptation of the 2 m limit for landslide depth in the reference data. Finally, we revised and amended the discussion, with a focus on discussing uncertainty. Please refer to our individual replies in the sections below for details.

With best regards,
Christoph Schaller

1 Anonymous Referee RC1

Dear Authors,, I apologize for the late upload of my review report. The good news is that my review will be very positive: I liked the manuscript and I think it addresses an important topic, well inside the aims and scopes of NHESS. It has an ambitious goal (soil thickness prediction at national scale is very challenging and, to the best of my knowledge, never attempted before in these terms), which is pursued with a robust and original approach. The manuscript is well structured, well written and very clear. Overall, the manuscript surely deserves publication and wit will be a valuable contribution to the journal. I just have a few comments. None of them is a big issue, and the manuscript can be accepted after minor revisions.

1.1 GENERAL COMMENTS

1 - The thing I miss the most in your study is a better linkage with the geology.

1a- I would add a test site description section, to make clear that you work at national scale (an amazing feature of your work) and briefly describing the main features of Switzerland (geology, climate).

1b- Geology is important in influencing both landsliding and soil thickness distribution. You use geology/lithology in your study by means of bedrock density, which you consider a proxy for bedrock lithology. In my opinion, this strategy needs a better justification. I guess you did this to have a numerical variable instead of a categorical variable (e.g. lithology classes). However, rock/soil cohesion, internal friction angle or hydraulic conductivity (just to name a few) may sometimes have a better relation with soil properties or slope stability. Some more reasoning on your strategy would be welcome.

1c- Building on previous comment, I recommend adding a table in which you list the main lithologies and the density value you assigned to them (or viceversa, as it suits you better). This may be also linked to a figure with main lithologies (see comment about test site description). Does Swisstopo contain thematic layers about other geotechnical properties? If yes, why didn't you consider them as well? When accounting for morphology, you took into account many parameters, so geology could be also accounted for by different parameters.

1d- In section 6.1 you analysed landslide distribution by slope class. However, the first thing that came up in my mind was to evaluate their distribution across lithologies. Is it possible to quickly perform a similar analysis?

>> 1a) We have added new section "Study area" with a brief description of the geological and climatic conditions in Switzerland.

>> 1b) and 1c) We have added additional details on the tested soil property data to the manuscript and the accompanying Jupyter notebook. We also tried to make the criteria for the variable selection clearer. After careful consideration, we opted not to include the soil property variables in the final model. We also added some additional details on the rock density data to the manuscript. However, we refrained to add an additional visualisation or table on the lithologies and densities, since thy are already sufficiently documented by Zappone and Kissling (2021). Therefore, we opted to add references to Zappone and Kissling (2021) and the visualisation of the data in the map.geo.admin.ch web portal.

>> 1d) we tested a visualisation with the landslide depths grouped by the lithology label from the rock density data (see Fig. 1 below). In our opinion the visualisation does not contribute enough to warrant an inclusion in the main part of the paper. If requested by the editor, we can include it as additional material. For now it is included in the accompanying Jupyter notebook.

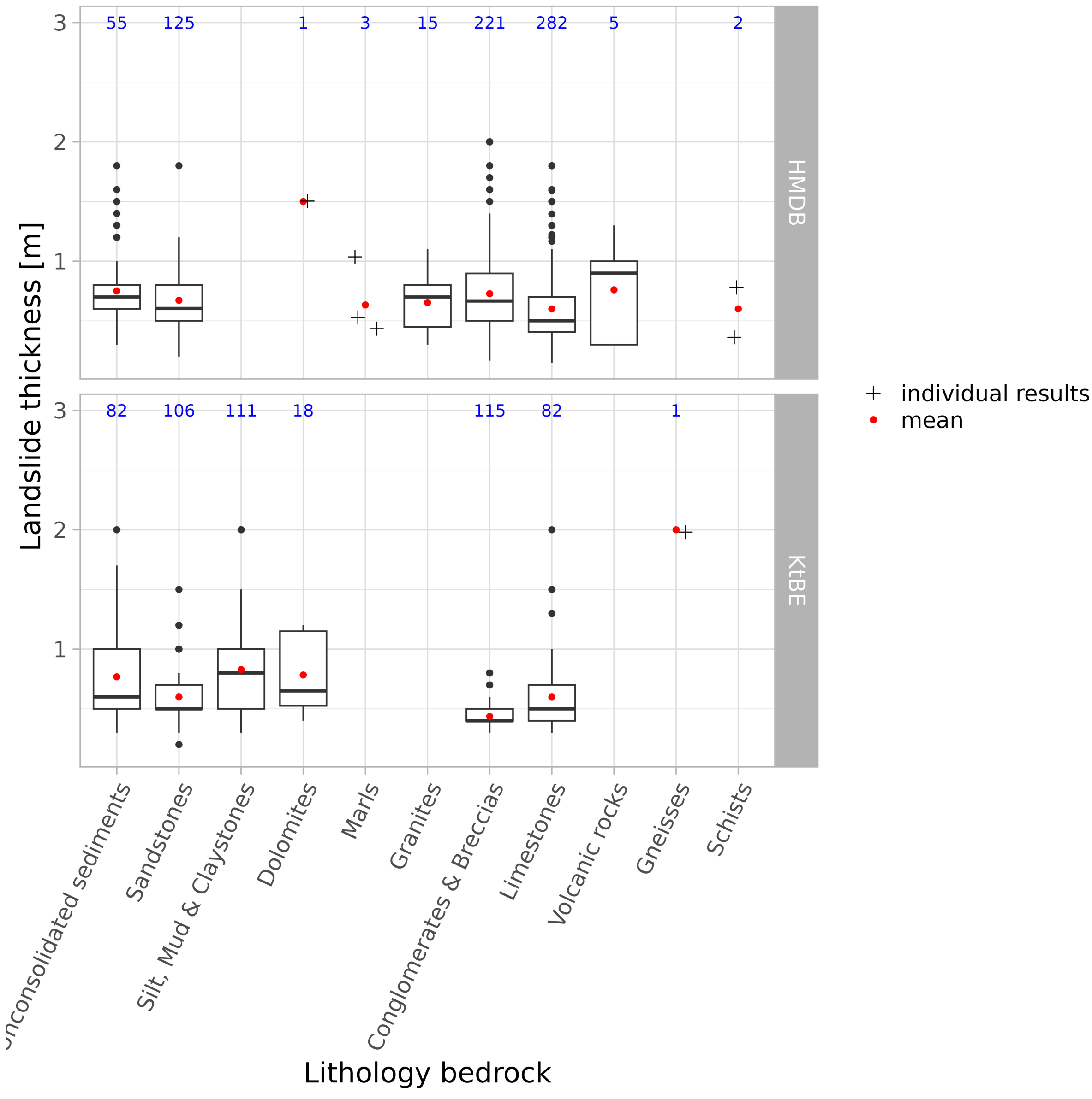

Figure 1. Box plots showing the distribution of the landslide thickness by the lithology class assigned in the bedrock density data based on Zappone and Kissling (2021). The lithology classes are ordered by ascending value of the mean bulk density (from left to right). The red dots represent the mean value while the blue numbers show the number of records per slope class.

1.2 SPECIFIC COMMENTS

L2: I would mention the areal extension of Switzerland. I think the width of your test site is a big constrain to the work; as such, it should be emphasized as an additional point of strength (you may consider to stress it also in the discussion and conclusion).

>> We added information on the total area of Switzerland and the area of the cantons to the study area description and to table 2.

L13: Would you consider adding also https://doi.org/10.1016/j.jeem.2024.102942 ? I think it is pertinent, as it reports on indirect impacts of hydrogeological processes, which are rarely accounted for (most studies focus only on casualties or direct economic damages).

>> Thank you for pointing out this publication. However, after careful consideration we opted not to include it, since its content is not directly linked to our argumentation.

L7 - I suggest to also convert CHF amount to USD. A few European readers may be aware of the value of CHF, but maybe not all the international readers are familiar with this currency.

>> The amounts have been converted to USD while retaining the original CHF amounts in brackets.

L21 - My English is not better than yours, but isn't "carried out within" more appropriate than "carried out at within"?

>> You are absolutely correct. We have removed the superfluous "at" in the manuscript.

L25 - I found odd to read the landslide definition in the middle of the introduction... Isn't it better to move it earlier in the text?

>> When combining and restructuring the former sections 1 and 2 we moved the definition into the new subsection 1.1 "Background on landslide failure thickness" where it is combined with other definitions and background information on the context of the study.

L51 and L55 are mentioning three and two landslide inventories, respectively. At this stage, this is confusing.

>> This mention was removed together with the exclusion of the StorMe inventory from the main analysis.

Section 2.1 - Just a comment: another point clearly highlighting the importance of soil thickness is the math formulae used in slope stability models. As scientific knowledge advances, the complexity of models has always been increasing, integrating new parameters and new processes in the stability equation. However, soil thickness has always been there: since the first pioneering equations (e.g. Skempton, A. W., & DeLory, F. A. (1957). Stability of natural slopes in London clay. Thomas Telford Publishing, London, UK, 15, 378-381.), soil thickness has always been there, among the uncontested key parameters!

>> Thank you for this comment. We have included the reference, albeit at a slightly different place due to the restructuring of the first two sections.

Tab1 - GIST-RF was recently applied to another case of study. If you want, you can add https://doi.org/10.1016/j.catena.2024.108024 along with Xiao's work.

>> Thank you for pointing out this publication. We have included it in the table as well as in the discussion.

Section 3.1 - I suggest to be clearer on one point: the geometry used to map landslides in the inventory. Are they mapped as polygons or points? In case they are points, it would make a big difference if the mapped point is the triggering point or the impact point, especially in case of shallow landslides with large runout distances.

>> We slightly reformulated the text in order to make clear that the HMDB data includes only the coordinates of the failure point while the KtBE inventory contains polygons of the landslide envelope.

L133 - Geological bedrock instead of geological underground?

>> Thank you for the comment. This have replaced this by "geological substratum".

L148-153 - see comment 1c.

>> The manuscript was amended to include the tested soil property data (please refer to the reply on comment 1c for additional information).

L212 (and elsewhere) - Shouldn't it be CARET, with capital letters? If yes, please adjust all occurrences in the text.

>> The explanation in brackets corresponds to the explanation in the introduction chapter of the caret package. However, the package authors render the name in lower-case in all other instances. Therefore, we have kept the lower-case rendering.

L311 I was surprised in seeing errors higher than the maximum expected value (2.5m) Didn't you applied a upper constrain to the modeled thickness values?

>> The adoption of the 2 m limit and subsequent removal of the deeper cases lead to changes in the trained results. Some predictions still exceed 2 m but the maximum is now at 2.07 m. The results have been updated accordingly.

L330 I would make reference to Fig B1

>> We added a reference to figure B1 in the supplementary materials.

L376-379- Could it be that by adding many 0 values in calibration the overall average modeled thickness is lowered?

>> We ran a test of the same model without the 0 m depth points. For the ML models as well as the SFM model the addition of the points lowers the mean of the predicted values by 2.4 cm to 6.4 cm while the mean values of the Simple-S and Simple-Z models increased by 13.1 cm to 21.6 cm. We have added information on this in the discussion.

2 Anonymous Referee RC2

This relevant and interesting manuscript by Schaller et al. presents a statistically based approach to predict the thickness of shallow landslides in Switzerland.

The authors state four objectives:

(i) To present descriptive statistics of shallow landslide thickness and related explanatory variables based on three Swiss landslide databases.

(ii) Develop and test three different machine learning approaches, ranging from linear regression models to generalised additive models and random forest models, to predict shallow landslide thickness based on different geospatial datasets and their derivatives.

(iii) Evaluate the performance of the models.

(iv) Compare the developed model with three existing models (focusing only on elevation, slope and cumulative slope distribution).

As shallow landslide thickness is a key variable in shallow landslide susceptibility modelling and in a further step of run-out modelling, i.e. shallow landslide hazard indication mapping, the prediction of shallow landslide thickness is of high relevance for further model development, but also for practitioners in natural hazard management. Thus, the authors address a highly relevant topic, especially as they develop models for larger areas and are not limited to smaller catchments. This last aspect in particular highlights the study, as there is no other large-scale approach to predicting shallow landslide thickness over a large area and in high mountains, at least to my knowledge.

Predicting the thickness of shallow landslides is a very difficult task, not least because of the very small-scale heterogeneity of the influencing factors. The authors test many geomorphometric properties derived from a digital elevation model. However, two very important properties are analysed in less detail: The geology/soil and the vegetation/forest.

The latter two in particular can influence the depth of landslides on a very small scale. Although the authors use Zappone and Kisslings rock density dataset, it is questionable to what extent other layers with geological properties could be taken into account to significantly improve the models. In addition, comprehensive information on soil properties is available for Switzerland (forest and agricultural areas; e.g. Baltensweiler et al. 2021 doi:10.1016/j.geodrs.2021.e00437).

>> As mentioned in the reply to 1a to 1d by reviewer 1, information on the tested soil property data was added back to the manuscript. We also tried to extend the information on the variable selection in the methods and discussion in order to make the selections more transparent and to better substantiate them. However, depending on the point of view, the substantiation might still be unsatisfactory.

For vegetation, a vegetation height model by Schaller et al. is used - which obviously also influences the model and the prediction. However, the VHM does not provide any insight into the forest structure, which has a decisive influence on soil stability (the authors cite e.g. Rickli et al. 2019). A spruce-dominated mountain forest will have a different influence on the prediction of the thickness of shallow landslides than a mixed forest in the lowlands - even though both sites may have the same geomorphometric characteristics. As a first step, one could, for example, use the mixed forest layer of the Swiss National Forest Inventory, which is available for the whole of Switzerland. I think, this should definitely be included in the discussion.

>> We tested the NFI forest type (Waser and Ginzler, 2018) as an additional covariate. However, due to the low or negative importances, we chose not to include them in the final model. We also added some points on this fact to the discussion.

As the prediction of shallow landslide thickness is intended to be used to generate raster data, i.e. maps, it would be of general interest to know what these rasters will look like. Therefore, I strongly suggest including one or two case study sites and showing how the approach will work for a spatial prediction (including a critical discussion of uncertainties).

>> As outlined in our initial response, we have decided not to include a case study, since we consider the generation and use of the rasters as part of a subsequent study and do not see enough additional value in such a case study to justify its inclusion in the manuscript.

The manuscript is clearly written. However, with the four objectives in mind, I feel that there is room for shortening and more precise wording in a number of places.

Some of the figures are difficult to read and extended captions could help the reader to follow more quickly (figures and tables are not always self-explanatory).

In the methods section, some important aspects are lost or neglected, which makes it difficult to fully evaluate the results. I think, the discussion chapter lacks a more critical discussion of the methods and data used, especially with regard to uncertainties and sensitivities of the models.

Based on the relevance of the topic (which is clearly within the aims and scope of NHESS) and first promising results, I recommend accepting this manuscript after major revisions.

2.1 General Comments

2.1.1 Introduction and Theoretical Background

I suggest combining Chapter 1 and Chapter 2. Otherwise the four objectives seem a bit lost between the two chapters. The introduction sounds more like an extended motivation, but lacks a description of the existing research and the subsequent research gap. Combining the two chapters could easily solve this. However, I suggest shortening the first part "motivation" and adding a short overview of the ML models chosen and why you chose them for your presented study. In chapter 2.2 you describe the existing soil thickness estimation models for landslide modelling and provide Table 1 for a more detailed overview. I think this is a very good idea. However, is it possible to add one or two sentences about them (common basis/differences; your list is "non-exhaustive" -> on what basis did you select the models/studies presented?).

>> We have combined and restructured chapters 1 and 2. The initial paragraph was kept as a general motivation it is followed by two subsections on "Background on landslide failure thickness" and "Models for estimating soil thickness for landslide modelling" which combine elements from the former chapters 1 and 2. The text was amended to make clear

that the content of table 1 is a list with a focus on models used in the context of shallow landslide simulations whose explanatory variables have informed the choice of covariates for our study.

2.1.2 Materials

Most of the records in the HMDB were collected after heavy rainfall events within defined perimeters. I suggest mentioning this in the text as it may influence the choice of statistical models.

>> We have added this in the material description of the HMDB data.

The thickness of landslides in the KtBE dataset was estimated by experts and orthoimagery. Hählen (2023) estimated an error of up to 50%, which is very high. - How well can you fit an ML model with such a high error in the data? Is this a suitable dataset for training the ML models? I suggest including this in the discussion chapter.

>> We slightly amended the material description to mention the possible error of 25% to 50% and added a mention of the propagation of this uncertainty into the model result in the discussion.

You mention that the StorMe dataset was excluded from the model development because of doubts about the data quality (L122). However, you use it for descriptive statistics. How robust are the descriptive statistics? I am not familiar with the StorMe dataset in detail. However, I understand that landslides from the HMBD are included in the StorMe dataset. If so, have you removed the 709 HMDB records from the StorMe dataset?

>> We removed StorMe from the descriptive statistics and only include this inventory as an additional example in the discussion.

Model input data:
What about other geological data? What about soil maps? (Please, see my comments above).

>> We added the tested soil property data Stumpf et al. (2023, 2024) to the list.

In the methods chapter you mention the Topographic Landscape Model TLM (ground cover rock / e.g. Fig. 3). If so, I suggest you include it in your list of input data used.

>> The swissTLM data was added to the list of input data.

2.1.3 Methods

I have some unanswered questions in the methods chapter. This is also where my main concerns about the study lie. These concerns mainly relate to:

(i) Sampling of covariates at the failure points of the slides (S2.3). Did you use a buffer around the failure points? For example, the HMBD was recorded with less accurate GPS systems (many records from the 90s/00s). Uncertainties may range from 5 to 20 metres. How did you extract the points - with a grid size of 5m for example? Wouldn't it be necessary to use a buffer and use the values of several grid cells (as has been done in other studies)? As far as I know, the reliability of the older StorMe database records is even more important (even if they were only used for descriptive statistics).

>> We added a mention of this uncertainties to the discussion.

(ii) You used regression models. How did you build / structure these models? As already mentioned, the HMBD, for example, was recorded in perimeters after defined heavy rainfall events. This results in a spatial and temporal hierarchy in the data. Can simple linear models be used or should more complex models be used (e.g., linear mixed effect models)? (iii) How was the quality of linear and generalised additive models checked? Did you perform diagnostic plots of the residuals? Was transformation of the data necessary to meet the assumptions of the models? Based on your list of input variables, there may be interactions between some variables. Have you included an interaction term in your analysis? Perhaps you can include a table/chart of the models you fitted?

>> In the methods section we stressed the point that the models were automatically built by caret during the training process. We decided not to include detailed tables or plot on the diagnostics of the GAM model in the manuscript, since the models are not the focus of the study. However, we added these plots back to accompanying Jupyter notebook and added a few words on the results together with the possible shortcomings of the GAM models in the discussion.

(iv) You mainly use MAE and R$\hat{2}$ for model validation. What about a confusion matrix, ROC curves or AUC values to evaluate your RF model in more detail (particularly in case of an imbalance)?

>> Our study uses machine learning models for regression i.e. we predict numerical values. Confusion matrices, ROC curves and AUC values are only applicable for classification tasks.

(v) Can you provide more information and results on the importance of your covariate selection and fitted models?

>> We already included a heat map with the most important results in the appendices of the manuscript. Some additional details on variable selection were added to the methods and discussion sections.

2.1.4 Discussion

The discussion of landslide inventories is very brief (Chapter 6.1). It would be desirable to address the uncertainties and problems of the databases, such as the accuracy of the measured coordinates or the derivation of depth from orthophotos, in the context of the descriptive statistics and model performance presented.

>> We opted to only slightly expand the section with the discussion on the inventories, especially by the inclusion of StorMe as an negative example. In exchange, we added a separate subsection discussing different factors of uncertainty.

Chapter 6.3 discusses the selection of covariates. I suggest adding a few lines, i.e. a critical discussion, on the effect of geology / soil characteristics and vegetation / forest effects.

>> We slightly expanded the discussion on the inclusion/exclusion of the VHM/forest type as well as the rock density/soil properties.

I suggest adding a critical discussion of the ML models used (see also my comments above on the methods chapter). Have you considered other models (e.g. logistic regression)? Could the robustness be increased by a bootstrap approach?

>> We added some points on the downside of the automatically created GAM models and short passage on the uncertainty connected to the choice of models and covariates.

I also think that an extended critical discussion of the uncertainties is missing.

>> As mentioned above a new subsection on the uncertainties in the inventory data, the covariates, and model choices was added.

2.2 Minor Comments

L31: Change meters to m to be consistent.

>> This was modified.

L35: Can you add a reference for your chosen definition of landslide thickness?

>> We kept this as is since this definition was formulated by the authors as the definition to be used in our study.

L46/47: I suggest rewording the sentence and removing the "according to our recent paper".

>> We completely removed this sentence.

L55: At this stage it is not clear to the reader why you have chosen only two landslide inventories and not all three?

>> The mention of StorMe was be removed at this point.

L60: What do you mean by "landscapes"? - Is that topography? Geomorphometry? L60: Soil types in terms of pedological soil types (e.g. Cambisols) or in terms of a more geotechnical description (e.g. USCS).

>> This is a statement of Hungr et al. (2014). However, the passage was modified during the revision such that the term "landscapes" isn't directly mentioned anymore.

>> We revised and amended the text to make clear, that we refer to soil from an engineering perspective and mean geotechnical soil properties.

L81: I suggest avoiding 'in' or 'see' before a figure reference. This occurs frequently throughout the manuscript, and I suggest it be changed.

>> After consulting several style guides, we removed the "in" and "see" in some places while retaining it in others.

Fig1: You mention the failure area, transit zone and deposition zone in the figure caption. I suggest you highlight these in the figure. In the figure you highlight the regolith. In the text you write "soil / soil type". What are you referring to? Is regolith = soil? Do you define soil as the entire weathering mantle from the surface to the weathering front / bedrock? Probably you can change the word "regolith" to be consistent with your text.

>> We amended the figure to explicitly show the failure/transit/deposit areas. We also amended the text to make clear, that soil refers to the entire regolith in the context of this study.

Tab1: I suggest adding a semicolon after each reference in column 4 and changing the parenthesis in the fifth line. Can you give a brief explanation of the parameters in the table heading? In the last row you mention 158 remotely sensed covariates. Can you somehow classify and mention them?

>> We modified the references accordingly.

>> We modified the header from "parameters" to "explanatory variables", which should make the nature if the content clear.

>> We added a short specification on the general nature of the covariates based on the abstract by Hengl et al. (2017): "158 remote sensing-based soil covariates (primarily derived from MODIS land products, SRTM DEM derivatives, climatic images and global landform and lithology maps)"

L96: You have filtered your data and kept only entries with a thickness of up to 2.5m. In your introduction you refer to the "Swiss definition" of shallow landslides, where the thickness of the instable mass does not exceed 2m. You are working with Swiss landslide inventories. Why did you finally choose a threshold of 2.5m? Can you give some information on how many landslides (n) occur between 2 and 2.5m? L110: Did you remove the six landslides (from tab 2 it seems that you left them in the dataset)? Why - especially since you then cross-validate.

>> We adopted the 2 m limit for the revised study. This change was adapted in the text and the analysis. We tried to make clear that the landslides mentioned on former line 110 were removed based on the filtering due to the depth limit. Adopting the 2 m limit resulted in the removal of 2 additional landslides for the KtBE dataset, which lead to changes in the models and results. These changes are reflected in the revised manuscript.

L97: Please give the URL for the WSL landslide database.

>> We added an additional reference for the homepage of the database.

L110: (Hählen 2023).

>> Thank you for pointing this out. We corrected the citation style.

L126: I suggest starting a new paragraph and shortening the paragraph as you provide Tab2 and the maps (Fig. 2).

>> This paragraph was revised when the chapters 1 and 2 were combined. It was restructured when it was integrated with parts of the original introduction, in order to remove some redundancies. In addition, the motivation behind the selection in the table was added.

L133: Please replace geological underground (e.g. with bedrock or geological condition).

>> Thank you for the comment. This was replaced by "geological substratum".

Tab2: Please include the full name of any abbreviations used in the captions (HMBD, KtBE, StorMe).

>> To avoid an overly long caption, we opted to retain the abbreviations.

Figure 2: Change "slides" to "shallow landslides". Remove "showing the locations" in the figure caption. Please add the sources/background maps in the reference list (swissBOUNDARIES, and relief map).

>> We have revised the figure taking into account these suggestions and added corresponding references.

L141: Why do you need the areas outside the Swiss borders if you are using inventory data from Switzerland? This is unclear.

>> We amended the text to make clear, that the fill operation is necessary to prevent erroneous values with some covariates that are calculated with a window or radius.

L146/147: This sentence is unclear, especially the last part ("assuming that this leads to less soil cover"). Could you please rephrase the sentence?

>> We rephrased this sentence to "This input was used as a proxy variable under the assumption that locations with potentially extreme rainfall amounts would experience increased erosion and landsliding activity, which would lead to reduced soil cover".

L153: Why do you use catchment areas, since you use them as the basis for tiling / parallel processing? Why not use the Swiss reference grid (e.g. SwissSURFACE divisions)? Do you influence the values of some geomorphometric derivatives by tiling? For example, the TWI, if you do not process the whole grid at once? Later you mention that you create a buffer around the catchments for data processing. Do the intersecting areas have the same values (e.g. for TWI)?

>> We amended the method description to make our choices clearer.

L158: The covariates such as terrain variables or geology were derived from the DEM. What do you mean by that? What geology do you derive from the DEM? I thought you used the rock density layer?

>> We have reformulated the text to say: "The covariates, such terrain variables or geology, were derived based on the DEM as well as other input data."

L159: By reference data you mean the inventory data?

>> Yes. This was made explicit.

L160: You mention Stage 4. However, in L156 and Fig3 you highlight three stages.

>> This has been changed to "(stage 3)".

Fig3: (i) Please give the full spelling of the abbreviations S and O used. (ii) In Stage 2 you refer to geological maps. This is the density map? What about the geomorphometric derivatives? (iii) I assume there is a typing error with O3.1 and O3.2 in Stage 2? (iv) S1.4: You average the average? Does this make sense? Why don't you use the original data to do the averaging? Also, if I understand correctly, you resample the EUDEM to 5m that you aggregate/average again to 25m (S1.5)?

>> (i) We added a small legend with the spellings to the figure.

>> (iii) Thank you for pointing out this mistake. It is indeed a typo. We will correct this.

L162: Please use the correct reference for software R.

>> We modified the entry in the bibliography such that the reference is now rendered as (R Core Team, 2022) by the Copernicus LaTex Template.

L169ff: See my comment on Fig 3(iv).

>> We revised the text trying to make the motivation behind the process clearer.

L180: Please delete the first sentence of chapter 4.2.1 as it is repetitive.

>> We deleted this sentence.

Tab3: (i) You use a number of explored covariates. However, you have somehow forgotten to provide the references. Could you please add a column with the references and the input datasets used? For example, there are several ways to calculate flow accumulation and/or TWI (based on multiple flow directions / single flow directions / weighted / ....). (ii) You highlight the variables and cell size used in the final ML models. Following this description, you use the TPI for 15m, 50m, 200m, ..., 4km with a cell size of 25m. Is this correct? However, I am confused as you only mention tpi_500m_25 and tpi_4km_25 in L224. (iii) I suggest adding the other covariates examined as VHM and rhob_m to the table as well.

>> (i) We have extended the table with information on the input data and the exact tool used to calculate the respective covariate.

>> (ii) The different radii of the TPI are ow listed on different lines to make the link to the cell size clear.

>> (iii) We tried to make this clear in the revised version.

L200-222: The approximation of the failure point is not clear. Could you please rephrase the sentence?

>> We rephrased this description.

L205: You have randomly generated points. How did you do this? What procedure did you use?

>> We added a mention or the sampling function of the terra package in R to the text.

L208/209: Are the numbers of the points generated correct? (HMBD 29? / KtBE 50?).

>> Yes, the numbers are correct. The number depends on the ratio of rock signature within the catchments where the inventory points are located. Since the locations of the HMDB are more closely clustered and differently located, fewer catchments with less rock signature are affected, thus resulting in less points.

L218: You have chosen a combination of exploratory analysis based on literature and landslide experts. Could you please refer to the literature you used? Who are your experts and how many experts did you contact? On what criteria did they suggest the covariates?

>> We revised the text to make clear, that the choice of variables was informed by the works in Table 1 and by the expert knowledge of the authors.

>> The changed the text to make clear, that the expert knowledge was provided by the authors.

L219: "both datasets", i.e. HMDB and KtBE?

>> Yes. This was made explicit in the text.

L263: Why did you fit the models without intercept?

>> We revised the text in order to make clear, that this is a consequence or the hyper-parameter tuning.

L285-306: This paragraph seems rather long. I would suggest shortening it and referring more to the table and figures.

>> The paragraph was slightly shortened.

Tab5: (i) Why do you highlight the sampling depth in the forest? What are the consequences of that? Are there differences in the descriptive key figures? (ii) You give the standard deviation for the arithmetic mean. However, you also show the median. Please add the corresponding MAD to be consistent. (iii) In general, it might be useful to replace the table with a figure. My suggestion is a combination of violin plots and box plots, as they contain the robust characteristic values on the one hand, and show the data distribution on the other. As mentioned in the introduction, geology and vegetation/forest are two key factors that could significantly influence the prediction of shallow landslide thickness. So why not split the data set further (see comment (i)).

>> (ii) The MAD was added to the table.

>> (iii) We tested a visualisation of the statistics as violin plots as shown below. While violin plots allow for a somewhat easier comparison of distributions, we feel that in our case the contrast between the violin plots and the existing table and histograms is not enough to justify changing the graphics. This is particularly because not all details from the table can be meaningfully integrated into the violin plots without overloading them and because the density functions of the violin plots hide some details on the distribution of the landslide thickness present in the histograms. Therefore, we opted to retain the current table and visualisations while only including the violin plots in the accompanying Jupyter notebook.

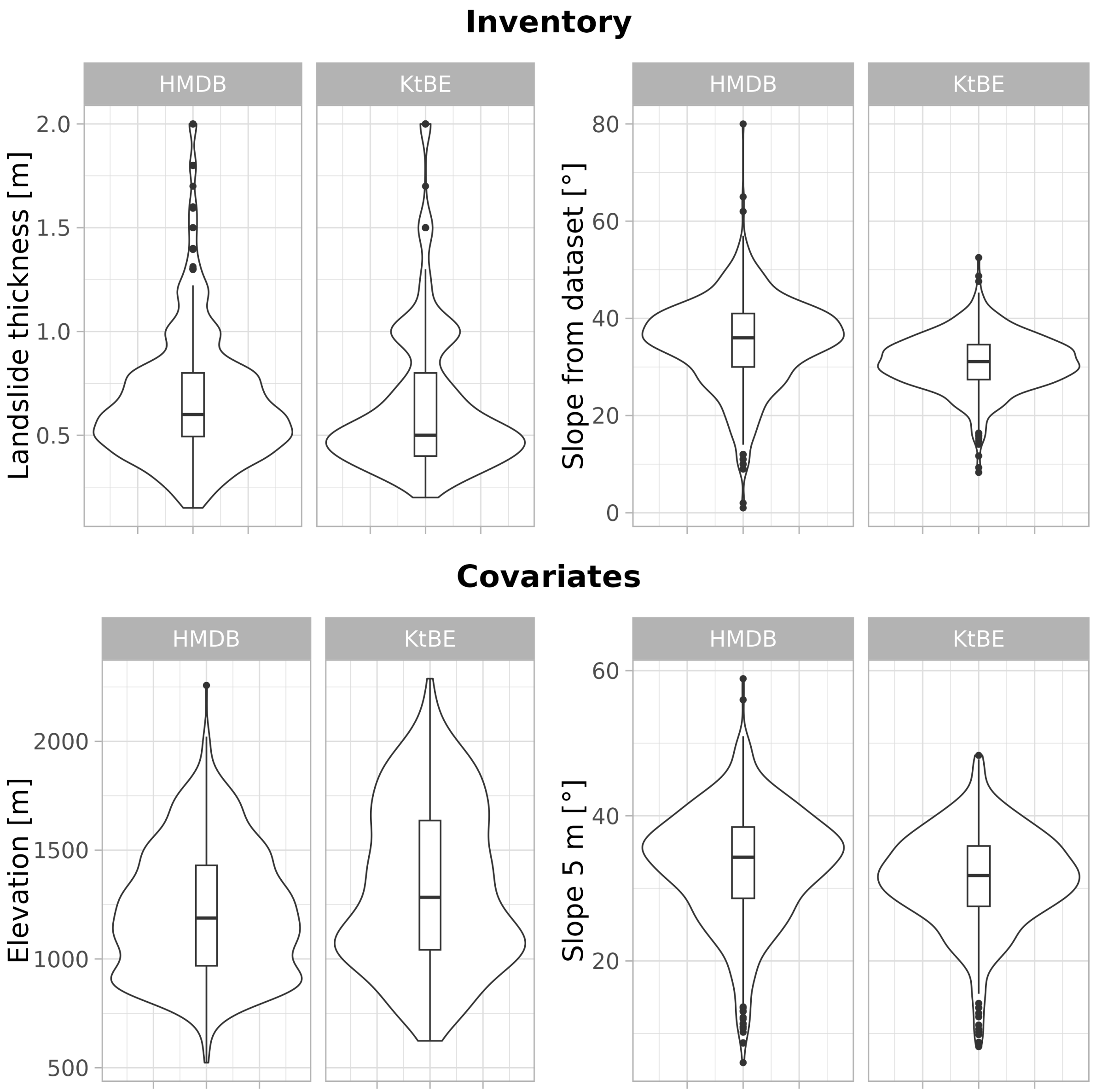

Figure 2. Box plots showing the distribution of the landslide thickness by the lithology class assigned in the bedrock density data based on Zappone and Kissling (2021). The red dots represent the mean value while the blue numbers show the number of records per slope class.

Fig4: Is the mean missing from the bottom left figure (or is it the same as the median)?

>> Due to the changes in the datasets following the adoption of the 2 m limit, the mean and median are now slightly different but still hard to see. The Context in table 5 and the high-resolution variant of the figure shouts, however, be enough to notice the difference.

L308: I cannot find the value 0.24m in Table 6. Is 0.24m correct (or should it be 0.25)? Which is the correct value?

>> Thank you very much for this hint. The discrepancy between the table and the text was corrected.

Fig5 (and all other box plots in the manuscript): (i) Individual boxes contain only one to 10 data points. At least five values are needed just to define the box. Perhaps you could point this out or simply replace the box with a median and MAD? (ii) In some cases, your data have a wide range (outliers and long whiskers). Could you try to transform your data (e.g. symlog) so that the reader gets more insight into the differences/major part of the data (boxes). This is especially the case for Fig7).

>> We have revised to box plots to show the individual measurements as points, along with the mean value if there are less than 5 samples.

>> We tested a variant of the box plots with a symlog transformed y-axis. While the boxes are somewhat emphasised, we ultimately did not see enough additional value in the modified version and opted to keep the original version. The modified versions are, however, available in the accompanying Jupyter notebook.

L318: Compared to?

>> We clarified this by adding "when compared to the other three models" to the sentence.

L328: Please show/refer to a figure.

>> We added a reference to figure B1 in the supplementary materials.

Fig6: (i) It is very difficult to read this figure as the plotted points are very small. The same applies to the 2D kernel density contours. (ii) Figure caption, last line: How many outliers are outside the display range (n=?)?

>> We tried to optimise the raster file for the figure and added an additional high-resolution variant in the PDF format. The figures were also modified to show the number of outliers as text, where applicable.

>> We will either specify the number of outliers in the caption or as an overlay in the plots themselves.

L382: What variables are you referring to (overlap with Zweifel et al. 2021)?

>> We added an explicit mention of the overlapping variables to the text.

References

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLOS ONE, 12, 0169 748, https://doi.org/10.1371/journal.pone.0169748, 2017.

Hungr, O., Leroueil, S., and Picarelli, L.: The Varnes classification of landslide types, an update, Landslides, 11, 167–194, https://doi.org/10.1007/s10346-013-0436-y, 2014.

Stumpf, F., Behrens, T., Schmidt, K., and Keller, A.: Hinweiskarten für Bodeneigenschaften - Landesweit modellierte Karten für Bodeneigenschaften für drei Tiefenstufen, Tech. Rep. 6, Kompetenzzentrum Boden (KOBO), ccsols.ch, Zollikofen, https://ccsols.ch/wp-content/uploads/2023/11/Hinweiskarten_Bodeneigenschaften_CH_final.pdf, 2023.

Stumpf, F., Behrens, T., Schmidt, K., and Keller, A.: Exploiting Soil and Remote Sensing Data Archives for 3D Mapping of Multiple Soil Properties at the Swiss National Scale, Remote Sensing, 16, 2712, https://doi.org/10.3390/rs16152712, number: 15 Publisher: Multidisciplinary Digital Publishing Institute, 2024.

Waser, L. and Ginzler, C.: Forest Type NFI, https://doi.org/https://doi.org/10.16904/1000001.7, 2018.

Zappone, A. and Kissling, E.: SAPHYR: Swiss Atlas of Physical Properties of Rocks: the continental crust in a database, Swiss J. Geosci., 114, 13, https://doi.org/10.1186/s00015-021-00389-3, 2021.