Predicting the thickness of shallow landslides in Switzerland using machine learning

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

Abstract. Landslide thickness is a key variable in various types of landslide susceptibility models. In this study, we developed a model providing improved predictions of potential shallow landslide thickness for Switzerland. We tested three machine learning (ML) models based on random forest (RF), generalized additive (GAM), and linear regression (LM) models. Next, we compared the results to three simple models that link soil thickness to slope gradient (Simple-S and SFM) and elevation (Simple-Z). The models were calibrated using data from two field inventories in Switzerland ("HMDB" with 709 records and "KtBE" with 515 records). We explored 39 different covariates including metrics on terrain, geomorphology, vegetation, and lithology at three different cell sizes. To train the ML models, 21 variables were chosen based on the variable importance derived from RF models and expert judgement. Our results show that the ML models consistently outperformed the simple models by reducing the mean absolute error by at least 20%. The RF models produced a mean absolute error of 0.25 m for the HMDB and 0.20 m for the KtBE data. Models based on ML substantially improve the prediction of landslide thickness, offering refined input for enhancing the performance of slope stability simulations.

1 Introduction

Rainfall-induced spontaneous landslides pose serious threats to infrastructure and inhabited areas worldwide (Froude and Petley, 2018; Emberson et al., 2020). In Switzerland, landslides regularly cause extensive infrastructure damage and closures, evacuations and even fatalities. For instance, 74 people died as a result of 40 different landslide events between 1946 and 2015 (Badoux et al., 2016). In August 2005, shallow landslides and the resulting hillslope debris flows caused damage amounting to USD 167 million across Switzerland within 48 hours (Bezzola and Hegg, 2007). Approximately USD 17 million are spent each year on landslide protective measures in Switzerland (Dorren et al., 2009). To mitigate this risk, regional landslide hazard mapping and modelling provide an important basis for indicating potential hazard areas (Dahl et al., 2010; Kaur et al., 2019; Shano et al., 2020; Di Napoli et al., 2021). In Switzerland, national-scale shallow landslide modelling was carried out within the SilvaProtect-CH project (Dorren and Schwarz, 2016). Results from this project have provided an important basis for, amongst others, the delimitation of landslide protective forests across the country and the preliminary risk analyses for national roads (Arnold and Dorren, 2015) and railways.

1.1 Background on landslide failure thickness

In general, landslides can be defined as the movement of a rock or soil mass along a slope (Varnes, 1978; Hungr et al., 2014).
The term "shallow landslides" typically refers to translational sliding movements of soil material (earth and/or debris) from the upper soil layers, characterised by a well-defined sliding surface (Cruden and Varnes, 1996; Hungr et al., 2014). Often, shallow landslides result in hillslope debris flows, which can be very destructive because of their velocity and resulting impact pressure (Zimmermann et al., 2020). Landslides are usually classified as shallow landslides if the thickness of the instable mass does not exceed 2 m, which is also used as a definition in Switzerland (Lateltin et al., 2005). However, in some cases, the failure plane lies within the top three meters (Sidle and Ochiai, 2013; Rickli et al., 2019; Li and Mo, 2019). With a median release area of around 200 m^2 and an average thickness of 0.5 m to 1 m, shallow landslides generally fall into the category of small (10^0–10^3 m^3) landslides as proposed by McColl and Cook (2024). Occasionally, they can fall into the category of very small (10^{-3}–10^0 m^3) or medium (10^3–10^6 m^3) landslides. In this study, we define landslide thickness as the average thickness of the instable mass measured perpendicular to the original slope surface down to the failure plane (T_L in Fig. 1).

The basic disposition for shallow landslides is determined by the slope gradient, the slope morphology, and the geotechnical properties of the soil, which have a strong link to the underlying geology (Hungr et al., 2014; Watakabe and Matsushi, 2019; Chinkulkijniwat et al., 2019). The definition of what constitutes soil varies between disciplines. In the context of landslides, soil is primarily viewed from an engineering perspective, which considers soil to be identical to the regolith cover (i.e. the entire unconsolidated material above the bedrock) (Huggett, 2023). Most landslides occur in wet, partially saturated soils, and are triggered by water input due to e.g. rainfall or seismic activity (Leonarduzzi et al., 2017; Schuster and Wieczorek, 2018). An increasing water content in the soil induces a reduction of soil shear strength leading to the failure of soil material within a shear band. This is usually situated at the interface of below-ground discontinuities such as between regolith and bedrock (Catani et al., 2010; Zhang et al., 2017; Xiao et al., 2023) or between layers with different soil characteristics (Li et al., 2013; Ali et al., 2014; Ran et al., 2018; Chinkulkijniwat et al., 2019). This boundary defines the failure plane of shallow landslides and is as such implemented in physically-based models (Ran et al., 2018). The hydrological properties of the soil and the local hydrological conditions influence the occurrence and depth of the failure plane of shallow landslides. Examples are the rainfall characteristics and the runoff disposition of the upslope area, as well as the groundwater table and the pore-water pressure (Caine, 1980; Iverson, 2000; Guzzetti et al., 2008b; Li et al., 2013; Chinkulkijniwat et al., 2019).

In many studies, soil thickness is used as a proxy for landslide thickness (e.g. Montgomery and Dietrich, 1994; Pack et al., 1998; Iida, 1999; Baum et al., 2002; D'Odorico and Fagherazzi, 2003; Segoni et al., 2012; Ho et al., 2012; Merghadi et al., 2020). When describing or modelling landslides, care should be taken to define the thickness unambiguously. Sometimes the terms soil thickness and depth are used interchangeably. However, in most studies, depth refers to a measurement in the vertical direction, while thickness refers to a measurement perpendicular to the surface of the slopes surrounding the release area (e.g. Meisina and Scarabelli, 2007; Catani et al., 2010; Jia et al., 2012; Ho et al., 2012; Lanni et al., 2012; Patton et al., 2018). Some studies use a reverse definition of the two terms, using the term thickness for vertical measurements and the term depth for perpendicular measurements (e.g. Cruden and Varnes, 1996; Pack et al., 1998). Similarly, the depth of the failure plane may be

defined in the vertical direction (Watakabe and Matsushi, 2019; Meier et al., 2020; Chang et al., 2021) or perpendicular to the slope (Iida, 1999; Schwarz et al., 2010; Li et al., 2013). The perpendicular direction seems to be favoured if the thickness is used for calculating the landslide volume (WSL and BAFU, 2018; Jaboyedoff et al., 2020; Hählen, 2023). In this study, we use the term depth when measuring in the vertical direction (depth at scarp D_F and soil depth D_S in Fig. 1) and thickness when measuring perpendicular to the slope (T_L in Fig. 1).

T$_L$: Landslide thickness
D$_F$: Vertical depth at the scarp
D$_S$: Soil depth
β_0: Slope above the failure scarp
β_F: Slope of the original surface in the failure area
dashed line: failure plane
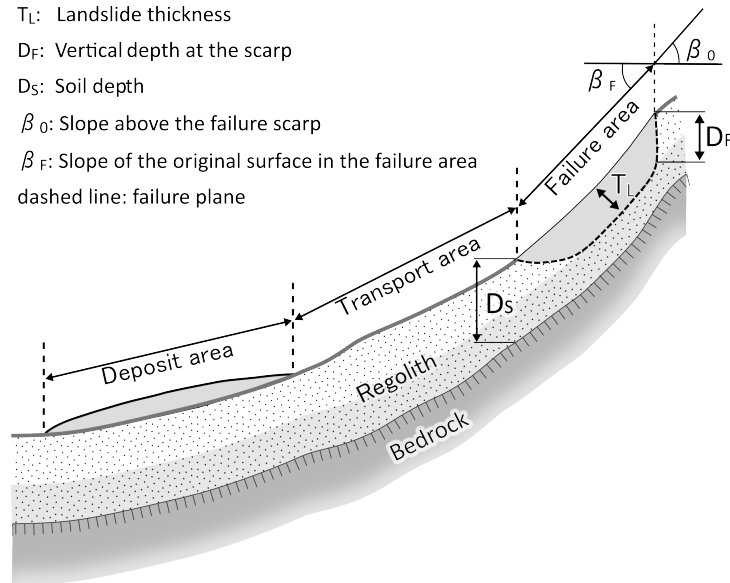
Figure 1. Schematic representation of a shallow landslide with its failure, transport, and deposit area, including important definitions used in this study. Variations in shading reflect differences in soil characteristics.

1.2 Models for estimating soil thickness for landslide modelling

Landslide thickness or soil thickness is a key variable in various types of models for simulating landslide susceptibility reaching back to the first pioneering equations (e.g. Skempton and deLory, 1957). The references used for calibrating these models are usually based on field measurements (e.g. by digging soil pits (Catani et al., 2010) or by drilling (Xiao et al., 2023)) or on data from landslide inventories (van Zadelhoff et al., 2022). However, dense field measurements are only available for small extents. Even then, the resulting soil thickness maps have high uncertainties due to the large heterogeneity of soil variables (Cohen et al., 2009; Jia et al., 2012; Lanni et al., 2012). To deal with the uncertainties in landslide thickness, different modelling approaches have been adopted. These models can be grouped into three categories: conceptual models, physically-based models, and empirical models (data-driven models, e.g. machine learning (ML)) (Murgia et al., 2022). Conceptual models aim to provide a simplified methodology for estimating changes in slope stability (Murgia et al., 2022), e.g. using cellular automaton models (Piegari et al., 2006). The physically-based models can be further divided into deterministic (e.g. Montgomery and Dietrich, 1994; Baum et al., 2002) and probabilistic models (e.g. Pack et al., 1998; Horton et al., 2013; van Zadelhoff et al., 2022).

Empirical models predict landslide occurrence based on factors that can be directly or indirectly linked to slope instability (Reichenbach et al., 2018). Such models are gaining interest and have become more commonplace in predicting inputs for landslide models due to improved data availability/quality and increasing research on ML and other computational techniques (Hengl et al., 2017; Merghadi et al., 2020; Wadoux et al., 2020; Xiao et al., 2023).

Table 1 gives an overview of models for predicting soil thickness found in the literature with a focus on those applied for shallow landslide simulations. The explanatory variables used in these models have informed the choice of covariates for this study. Furthermore, they are good examples providing an overview of the model types used for predicting soil depth. Most of these models are either deterministic (e.g. Montgomery and Dietrich, 1994; Baum et al., 2002) or probabilistic (e.g. van Zadelhoff et al., 2022).

The accuracy of the potential landslide thickness is of paramount importance for the performance of slope stability models (cf. Iida, 1999; Larsen et al., 2010; Milledge et al., 2014; van Zadelhoff et al., 2022). Consequently, improving the estimation of landslide thickness is key for enhancing the performance of slope stability models. With the overall aim of developing a model that provides a more accurate prediction of the potential shallow landslide thickness compared to pre-existing simple models, the four main objectives of this study are:

(1) to present descriptive statistics on data on the thickness of shallow landslides and additional potentially explanatory data from two field inventories in Switzerland,

(2) to develop and test new models for predicting the potential thickness of shallow landslides in Switzerland based on ML (random forest models (RF), generalized additive models (GAM), and linear regression models (LM)) using input variables including terrain metrics and vegetation,

(3) to evaluate the performance of the developed models using data from the two shallow landslide inventories,

(4) to compare the performance of the developed models with three previously published models that predict shallow land-slide thickness based on altitude, slope, and cumulative slope distribution.

Table 1. List with a summary of soil thickness models found in the literature.

Model used	Explanatory variables	Relationship with soil depth	Reference
Constant soil thickness across the study area	Mean soil thickness, possibly combined with standard deviation	Constant value	Montgomery and Dietrich (1994); Baum et al. (2002); Ho et al. (2012)
Elevation	elevation	Linear scaling	Saulnier et al. (1997); Catani et al. (2010); Segoni et al. (2012)
Slope gradient linear	slope gradient	Linear scaling	Saulnier et al. (1997); Catani et al. (2010); Segoni et al. (2012)
		Linear relationship statistically fitted	Lanni et al. (2012)
Slope gradient exponential	slope gradient	Exponential law	Segoni et al. (2012)
Slope gradient cumulative	thickness measurements and slope values from a landslide inventory	Cumulative normal distribution of slope values	van Zadelhoff et al. (2022)
Hillslope curvature	Curvature	Linear relationship; varies across different landscapes as a function of the standard deviation in catchment curvatures	Patton et al. (2018)
Topographic Wetness Index (TWI)	TWI	Linear relationship	Ho et al. (2012)
Linear model with terrain indices	slope gradient, plan curvature, profile curvature, specific catchment area, and relative position on the hillslope	Linear relationship established by multi-linear regression	Jia et al. (2012)
Process-based models	time since last the denuding by landslides	linked to soil development over time by approximating the development using a logarithmic function	Iida (1999)
	bulk density of the rock, elevation, slope	mass balance between soil production from underlying bedrock and the divergence of diffusive soil transport by solving the evolving soil depth using a finite difference model under varying initial conditions	Dietrich et al. (1995)
GIST (Geomorphologically Indexed Soil Thickness)	slope gradient, horizontal and vertical slope curvature, relative position within the hillslope profile	the gradient and curvature are connected to the kinematic stability of the regolith cover while the distance from the hill crest accounts for the position within the soil toposequence	Catani et al. (2010); Segoni et al. (2012); Xiao et al. (2023)
sGIST	slope gradient, horizontal and vertical slope curvature, relative position within the hillslope profile	GIST without geomorphological indexing	Catani et al. (2010); Segoni et al. (2012)
GIST-RF	GIST inputs plus altitude, plan curvature, and Terrain Roughness Index (TRI) as covariates	Random forests regression	Xiao et al. (2023)
GIST-MCS	GIST method additionally incorporating Latin Hypercube Sampling-Based Monte Carlo Simulations to model uncertainties in determining the influence of the slope factor	GIST plus Monte Carlo Simulations	Gupta et al. (2024)
Machine learning ensemble	158 remote sensing-based soil covariates (primarily derived from MODIS land products, SRTM DEM derivatives, climatic images and global landform and lithology maps)	ensemble including random forests and gradient boosting and/or multinomial logistic regression	Hengl et al. (2017)

2 Study area

The study area is distributed across Switzerland (Fig. 2), which has a total area of 41'291 km^2 (FSO, 2021). The Swiss landscape can roughly be divided into the Alps covering around 58% of the country, the Central Plateau covering 31% and the Jura covering 11% of the area (FDFA, 2023). The Jura mountains in the northwest are mainly characterised by limestones

100

and marls (Pfiffner, 2021; Zappone and Kissling, 2021). The Central Plateau extending from Geneva towards the northeast is a molasse sedimentation basin covered by thick quaternary deposits resulting from the erosion of the Alpine chain (Reynard et al., 2021). With the Helvetic, Penninic, and Austroalpine nappes, the Alps include three major nappe systems. Parts of these nappes can be referred to as Pre-Alpes (cf. Fig. 2), which include lower mountain ranges and in the most northern parts even foothills. The alpine nappe piles led to the amalgamation of very different rock types. These include continental and oceanic basement rocks (granites, gneisses, and schists), shallow marine carbonates (limestones and marls), deep marine clastics (sandstones and conglomerates), which locally resulted from turbidites known as Flysch deposits, and radiolarian chert (Pfiffner, 2021). Switzerland is characterised by a temperate semi-continental climate that is strongly influenced by its altitude and its complex topography (Fallot, 2021). The mean annual air temperature varies with the altitude showing 8.5-11.9 °C at 500 m above sea level, 6.2-9.6 °C at 1000 m, and 3.9-7.3 °C at 1500 m over the period from 1981–2010 (Fallot, 2021). The mean annual rainfall increases with altitude with values from 900-1300 mm for the Central Plateau and gradually increases from the southern piedmont of the Jura mountains towards the northern side of the Alps where these amounts exceed 2000 mm and reach up to 3000 mm per year or more on the wettest summits in the Central Alps.

3 Materials

3.1 Landslide inventories

This study used reference datasets compiled from two different Swiss landslide inventories. The first dataset (hereafter "HMDB") is based on a comprehensive database of shallow landslides and hillslope debris flows (WSL, 2024) that occurred between 1997 and 2021 created by Rickli et al. (2016, 2019). Most of the HMDB records were collected after heavy rainfall events within defined perimeters. The data in this inventory are based on field surveys performed with identical protocols that include relevant variables such as the dimensions of the landslides, site characteristics, and runout characteristics (Rickli and Graf, 2009; WSL and BAFU, 2018). If values could not be measured in the field (e.g. because of terrain changes since the event), the database may contain estimated values that are marked accordingly (Rickli et al., 2016). The location of the landslides is recorded by geographical x and y coordinates of the failure point in the Swiss coordinate system LV95 measured at the upper edge of the scarp. Of 760 entries in the inventory, 75 comprised measured and 199 comprised estimated thickness values in the field. These were recorded perpendicular to the original slope surface (T_L in Fig. 1). For 435 records we were able to estimate the thickness by dividing the recorded landslide volume by the recorded failure area. This approach was chosen based on an exchange with the author of the database (Rickli, 2023) since the mean landslide thickness in the HMDB is mainly used for the estimation of the failure volume (Rickli et al., 2016). In the end, we had 711 records with thickness values. Since we removed two records with an estimated value larger than 2 m from the dataset, the resulting dataset comprised 709 records.

The second dataset (hereafter "KtBE") is based on an inventory created by the office for natural hazards of the canton of Bern (Hählen, 2023) and comprises 519 landslides, 6 of which are also recorded in the HMDB inventory. The landslides in the KtBE inventory were recorded between 2005 and 2021. For the inventory, the failure zones and runout envelopes of shallow landslides were digitized as polygons from orthoimages. The mean thickness of the landslides is based on expert estimates

based on the orthoimages. The author estimated that there is a possible error of between 25% to 50% for the thickness (Hählen, 2023). Since only events were recorded for which the entire process area could be reliably reconstructed from the orthoimages, many events in forests or intensively cultivated areas were not included in the inventory. This limits the possibility of making comparisons between landslides within and outside of forests. Four records were filtered out from the dataset because of a missing thickness value or a thickness value larger than 2 m, leaving 515 records in the final dataset.

The distribution of the landslide locations across Switzerland (Fig. 2) and the number of landslides per canton (Table 2) are uneven. The fact that the datasets only contain records for 9 of the 26 cantons in Switzerland is not only caused by uneven occurrence of landslides but also by differences in the availability of recorded data. While the recorded landslides are distributed across large parts of Switzerland, there is a clear concentration on the northern parts of the Alps and Pre-Alps. Locally clustered occurrences were mostly caused by specific extreme precipitation events in combination with unfavourable geological substratum such as flysch or molasse (Reynard et al., 2021; Steger et al., 2022).

Table 2. Number of recorded landslides per canton in the HMDB and KtBE shallow landslide datasets

Canton	Abbreviation	Area in km^2	HMDB	KtBE
Appenzell Ausserrhoden	AR	244	103	0
Appenzell Innerrhoden	AI	172	2	0
Bern	BE	5'959	170	483
Fribourg	FR	1'671	0	1
Grisons	GR	7'105	106	0
Lucerne	LU	1'494	86	33
Obwalden	OW	491	240	0
Vaud	VD	3'212	1	0
Zürich	ZH	1'729	1	0
Total		22'077	709	517
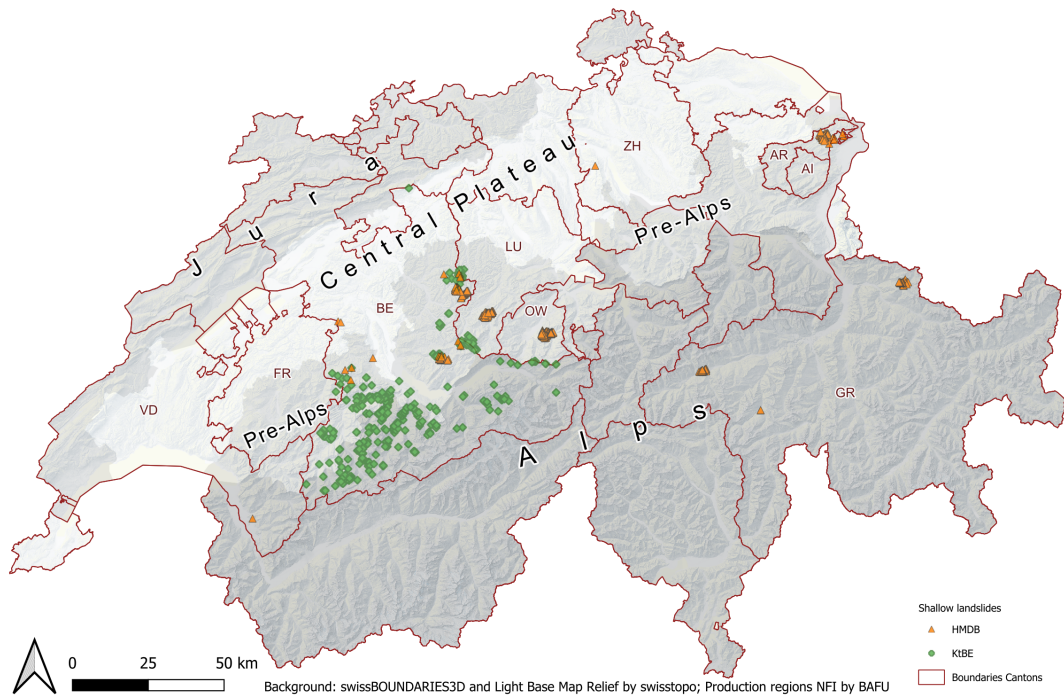
Figure 2. Map of the recorded shallow landslide locations across Switzerland for the HMDB (orange triangles) and KtBE (green diamonds) datasets. The Jura, the Central Plateau, the Pre-Alps, and the Alps are indicated with different shades of grey. Background: swissBOUND-ARIES3D (Swisstopo, 2024b), production regions NFI (BAFU, 2020), and light base map relief (Swisstopo, 2024c)

3.2 Model input data

The following datasets were used as input for the models in this study:

- LiDAR (Light Detection And Ranging)-based swissALTI3D digital elevation model (DEM) with 0.5 m cell size (Swisstopo, 2023a).
- EU-DEM v1.1 DEM with 25 m cell size (EEA, 2016).
- LiDAR-based vegetation height model (VHM) with 1 m cell size (cf. Schaller et al., 2023).
- National Forest Inventory (NFI) forest type raster 2018 with 10 m cell size indicating the proportion of coniferous trees (Waser and Ginzler, 2018).
- Modelled data on extreme point precipitation (Frei and Fukutome, 2022) with 1 km cell size for different return periods (2, 10 30, 50, 100, 200, and 300 years).
- Rock densities across Switzerland in the form of a vector-based dataset (Swisstopo, 2020).
- Soil property maps with 30 m cell size providing predictions for clay, sand, and silt contents at 0–30 cm, 30–60 cm, and 60–120 cm Stumpf et al. (2024).
- Ground cover layer of swissTLM3D (Swisstopo, 2023b).

– Topographic catchment areas of Swiss water bodies (BAFU, 2019).

Section 4 explains why these datasets as well as the derived covariates were included. Note that we used a preliminary version of the soil property maps, kindly provided by the Swiss Competence Center for Soils (CCSol, 2024).

4 Methods

In the first part, we analysed the distributions of the thickness and slope gradients of the landslides in the HMDB and KtBE datasets. We used the R environment for statistical computing (R Core Team, 2022) to calculate the descriptive statistics. These included the minimum, maximum, mean, standard deviation, median, and mean absolute deviation (MAD) for the landslide thickness and slope gradient recorded in the inventories as well as the elevation and slope gradient from the covariates.

The subsequent modelling part is divided into three stages (Fig. 3). In stage 1, intermediate DEM rasters for the subsequent stage were prepared. During stage 2, the data used for the final modelling and analysis stage were prepared. The covariates, such as terrain variables or geology, were derived based on the DEM as well as other input data. Then, the covariates were added to the reference data by sampling the generated rasters at the landslide failure points. The analysis and modelling based on the previously prepared dataset were performed in stage 3 including the training of three types of machine learning (ML) models and the application of three simple model types.
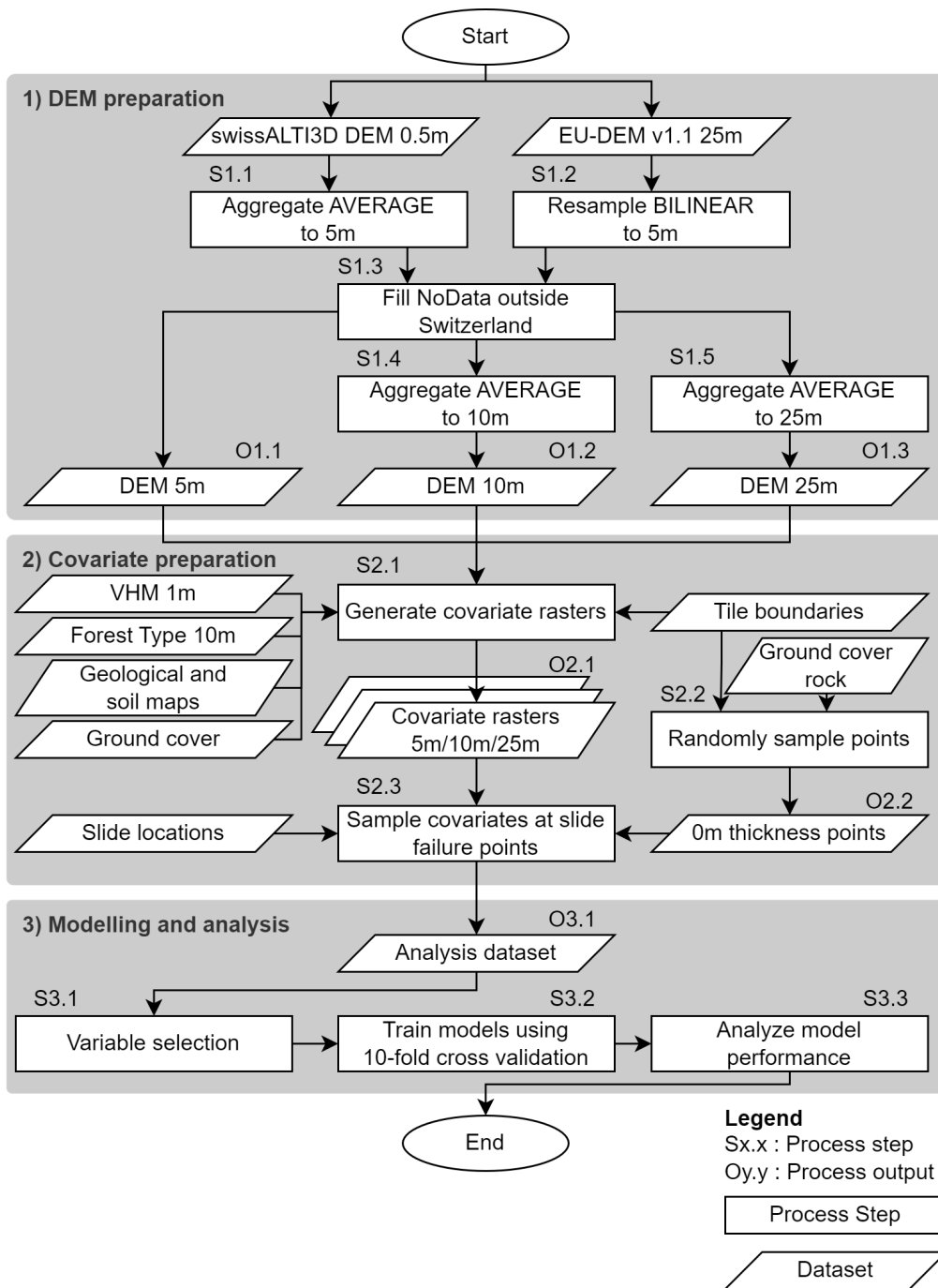
Figure 3. Flowchart of the methodology applied for this study. The three separate grey boxes delineate the 1) DEM preparation, 2) covariate preparation, and 3) modelling and analysis stages.

All three stages were implemented in R. The calculation of some variables was implemented in R using the *sf* (Pebesma and Bivand, 2023) and *terra* (Hijmans, 2023) packages. However, the processing of most variables was implemented using SAGA GIS (System for Automated Geoscientific Analyses, Conrad et al. (2015)) via the *RSAGA* package (Brenning, 2008) or using GDAL (Geospatial Data Abstraction Library) tools (GDAL/OGR contributors, 2021) called from R. The code for the entire process can be found in the repository accompanying this study (Schaller, 2024).

4.1 DEM preparation

The covariates for the ML models were calculated as rasters with a cell size of 5 m, 10 m, and 25 m (outputs O1.1, O1.2, and O1.3 in Fig. 3). The 0.5 m cell size swissALTI3D DEM was used as input for the terrain-based covariates. The original DEM was first aggregated to 5 m cell size using gdalwarp with the average function (step S1.1 in Fig. 3). The areas outside the Swiss borders are not covered by swissALT3D. Covariates calculated over a window would lead to incorrect values near the border. Therefore, the areas outside the border were filled using the EU-DEM. For this purpose, the original extent of the 5 m DEM was first buffered by 4 km to accommodate the largest radius used in the covariate calculations (step S1.2 in Fig. 3). The areas with NoData within that buffered extent were then filled using values from the EU-DEM resampled to 5 m using gdalwarp with a bilinear function (step S1.3 in Fig. 3). This filled raster was the primary input for the terrain analysis. The additional DEM rasters with 10 m and 25 m cell sizes were derived by aggregating the filled 5 m DEM using gdalwarp with the average function (steps S1.4 and S1.5 in Fig. 3). Although the resampling of the EU-DEM introduces a certain error in the elevation values outside the border, we regard this error as an acceptable trade-off for the good data availability and the simplified data preparation process.

4.2 Covariate preparation

4.2.1 Calculating covariate rasters

As can be seen in the overview of the processed variables and their corresponding cell sizes (Table 3), some covariates were calculated for all three cell sizes while others were only calculated for specific cell sizes. Vector-based input data were rasterised, including the generation of separate 0/1 encoded rasters for categorical variables. The terrain-based covariates calculated from the DEM included metrics commonly used in terrain analysis, hydrology, and geomorphology that were also used in other studies aiming to predict failure or soil depth. To represent the influence of the forest, the VHM was used to calculate statistics on vegetation height while the NFI forest type raster was used as an indicator for species composition, which can influence the stabilising influence of forests. The rock densities across Switzerland (Swisstopo, 2020) served to represent the underlying lithology. The density of rocks varies based on their chemical composition and the structure of their minerals (from crystalline to amorphous) (Zappone and Kissling, 2021). Each entry in the dataset is assigned to one of the 21 defined lithological groups and shows the expected range in which the mean bulk density of the local lithology varies (Swisstopo, 2020). For an overview of the lithology groups and densities please refer to the visualisations in Zappone and Kissling (2021), the documentation included in Swisstopo (2020) or its visualisation in the Swisstopo web map portal (Swisstopo, 2024a). Similarly, the soil

property maps were used to test their influence on landslide depth. The modelled data on extreme point precipitation (Frei and Fukutome, 2022) was used as a proxy variable assuming that locations with potentially extreme rainfall amounts would experience increased erosion and landsliding activity, which would lead to reduced soil cover. The swissTLM3D ground cover layer (Swisstopo, 2023b) was tested as a categorical covariate and was used to identify areas with rock cover.

To optimize the calculation of covariates for the entire area of Switzerland, the processing of the covariates was parallelized (step S2.1 in Fig. 3). For this purpose, Switzerland was divided into tiles based on aggregated catchments. The tiles based on catchments ensure that the values calculated for hydrological covariates are correct within the catchment area. We used the pre-defined aggregation level of the topographic watershed dataset (BAFU, 2019), which defines watersheds with a size of approx. 150 km^2. These correspond to a tile size with acceptable processing times while keeping the number of tiles at a manageable level. For the actual processing, the extents of the catchments were buffered by 500 m (5 m and 10m rasters) or 4 km (25 m rasters) to ensure data availability for covariates calculated with window sizes that go beyond the catchment border. Only the covariate values within the catchment borders were used for further processing since the values outside the border may be incorrect. Note that variables connected to hydrology were not calculated based on the original DEM raster, but a sink-filled version of the DEM generated using the SAGA GIS Tool "Fill Sinks" based on the method by Planchon and Darboux (2002).

Table 3. Variables explored as covariates for the machine learning models along with their input data and the tools used for their processing.

Covariate name	5 m	10 m	25 m	Used tool	Input data
	Cell size				
Elevation above sea level	x[1]	x	x		
Slope in degrees	x	x[1]	x		
Aspect in degrees	x	x	x		
General curvature	x	x[1]	x	SAGA "Slope, Aspect, Curvature"	
Profile curvature	x	x	x[1]		
Plan curvature	x	x	x		
Aspect northness	x	x[1]	x		
Aspect eastness	x	x	x		
Elevation percentile with 50 m and 200 m radius	x				
Topographic Position Indexes (TPI) with radius 15 m	x			R with *terra*	
TPI with radius 50 m, 200 m		x			
TPI with radius 500 m, 1 km, 2 km, and 4 km			x[1]		
Toposcale	x	x	x		
Catchment area	x	x	x[1]		
Catchment slope	x	x	x[1]		
Modified catchment area	x	x	x	SAGA "SAGA Wetness Index"	
Topographic wetness index	x	x	x		swissALTI3D (Swisstopo, 2023a), EU-DEM (EEA, 2016)
Topographic wetness index SAGA	x[1]	x	x	SAGA "Topographic Wetness Index (TWI)"	
Flow accumulation (top-down) contributing area	x	x	x	SAGA "Flow Accumulation (Top-Down)"	
Negative openness with 50 m and 200 m radius	x[1]			SAGA "Topographic Openness"	
Multiresolution Index of Valley Bottom Flatness (MRVBF)	x	x[1]	x	SAGA "Multiresolution Index of Valley Bottom Flatness (MRVBF)"	
Multiresolution Index of the Ridge Top Flatness (MRRTF)	x	x	x[1]		
Convergence index	x	x	x	SAGA "Convergence Index"	
Upslope curvature	x	x	x	SAGA "Upslope and Downslope Curvature"	
Downslope curvature	x	x	x		
Morphometric protection Index	x	x	x	SAGA "Morphometric Protection Index"	
Vector Ruggedness Measure (VRM)	x	x	x[1]	SAGA "Vector Ruggedness Measure (VRM)"	
Terrain Ruggedness Index (TRI)	x[1]	x[1]	x	SAGA "Terrain Ruggedness Index (TRI)"	
Slope height	x	x	x[1]		
Valley depth	x	x	x[1]		
Normalized height	x	x	x	SAGA "Relative Heights and Slope Positions"	
Standardized height	x	x	x[1]		
Mid-slope position	x	x	x		
Geomorphon with 50 m and 200 m radius	x			SAGA "Geomorphons"	
Groundcover	x			gdalrasterize	swissTLM3D (Swisstopo, 2023b)
Extreme point precipitation values for 60 min duration with 10-year return period			x	gdalwarp	Frei and Fukutome (2022)
Maximum of vegetation height within cell based on 1 m cell size VHM	x[1]	x	x		Schaller et al. (2023)
Percent coniferous trees		x			Waser and Ginzler (2018)
Mean bulk density of the underlying bedrock[2]	x[1]			gdalrasterize	Zappone and Kissling (2021)
Clay, sand, and silt content at three depth levels (0–30 cm; 30–60 cm; 60–120 cm)[3]	x			gdalwarp	Stumpf et al. (2024)

[1] Variables and cell size used in the final ML models. [2] Sampled directly from the original vector data. [3] Sampled directly on the original 30 m cell size raster.

4.2.2 Sampling covariate values

After calculating the covariate rasters, we prepared the dataset for the subsequent modelling and analysis stage. This was achieved by merging the event data from the inventories with the covariate values by sampling the cell values at the failure

point of the landslides (step S2.3 in Fig. 3). In the case of the HMDB dataset, the reported coordinates of the failure point were directly used. For the KtBE dataset, only the event envelopes were available as input. Therefore, the failure point was approximated based on the envelope polygon. First, the elevation for the vertices was determined by sampling the 5 m cell size DEM. Then the x and y coordinates of the approximate failure point were calculated as the average for the corresponding coordinates of the vertices with an elevation equal to or higher than the 95th percentile of the elevation of all vertices of the envelope polygon.

The inventories only contained entries for locations with actual landslides and therefore a sufficient soil thickness. To better reflect the fact that landslides cannot occur in areas with bare rock, additional points with a landslide thickness of 0 m in rocky areas were added to the analysis dataset (step S2.2 in Fig. 3). The points were randomly generated within the areas marked as rock or loose rock in the ground cover layer of the swissTLM3D landscape model (Swisstopo, 2023b) using the spatSample() function of the *terra* package in R (Hijmans, 2023). The points were generated separately for each dataset and only within the rock signatures in the catchments containing slide events. After generation, the points were manually checked for plausibility using the SWISSIMAGE orthoimages (Swisstopo, 2024d). The number of generated points actually included in the datasets is proportional to the percentage of the catchments covered by the rock signature, namely 34 points for the HMDB dataset (4.1% rock cover) and 53 points (9.7% rock cover) for the KtBE dataset.

4.3 Modelling and analysis

4.3.1 Models and covariate selection

We tested three different types of ML models for predicting the potential failure thickness of shallow landslides. The R *caret* package (Classification And REgression Training) (Kuhn, 2008) was used to fit these models. Due to the promising performance of random forests (RF) in other studies, our development efforts were mainly focused on RF models (Breiman, 2001), implemented using the *ranger* package (Wright and Ziegler, 2017). In addition, we tested linear regression (LM) using the built-in lm function of R as well as generalized additive models (GAM) via the *mgcv* package (Wood, 2011), which allow to model non-linear relationships. We chose those model types to evaluate if they could achieve similar performance to RF models while maintaining explainability (James et al., 2021).

All three model types were trained using the same input data and validation procedures. The covariates included in the models were chosen based on a combination of exploratory analysis, inputs from the literature (especially the works listed in Table 1), and expert knowledge of the authors (step S3.1 in Fig. 3). The exploratory analysis included test-fitting RF models with both the HMDB and KtBE datasets to determine variable importances and attempts at automatic variable selection using recursive feature elimination. The inputs from the literature and the expert knowledge of the authors influenced the overall selection of the tested covariates. From the pool of candidates, we aimed to select covariates with high importance, whose combination resulted in good model performance, while keeping the number of covariates and therefore the model complexity low. At the same time, we tried to balance the inputs from the test fittings with the expert-based input, leading to the following choice of covariates:

- **aspect_nness_10**: northness of the aspect calculated as cosine(aspect) at 10 m cell size
- **curvature_10** and **curvature_profile_25**: general curvature at 10 m cell size and profile curvature at 25 m cell size as calculated by the SAGA module "Slope, Aspect, Curvature"
- **h_5**: altitude above sea level sampled at 5 m cell size
- **mrrtf_25**: multiresolution index of the ridge top flatness at 25 m cell size
- **mrvbf_10**: multiresolution index of valley bottom flatness at 10 m cell size
- **openness_neg_200_5**: negative openness calculated with a radius of 200 m at 5 m cell size
- **rhob_m**: mean bulk density of the local lithology
- **slope_10**: slope gradient in degrees at 10 m cell size
- **slope_height_25**, **standardized_height_25**, and **valley_depth_25**: slope height, standardized height, and valley depth at 25 m cell size calculated by the SAGA module "Relative Heights and Slope Positions" (Böhner and Selige, 2006)
- **tpi_500m_25** and **tpi_4km_25**: topographic position indexes 500 m and 4 km window size at 25 m cell size
- **tri_r5_5** and **tri_r5_10**: terrain ruggedness index with a radius of 5 cells for 5 m and 10 m cell sizes
- **twi_5**, **catchment_area_25**, and **catchment_slope_25**: topographic wetness index, catchment area and catchment slope at 5 m and 25 m cell size as calculated by the SAGA module "SAGA Wetness Index" (Böhner et al., 2002; Böhner and Selige, 2006)
- **vhm_max_5**: maximum vegetation height at 5 m cell size
- **vrm_r5_25**: vector ruggedness measure with a radius of 5 cells at 25 m cell size

Additionally, three simple models for predicting soil depth were applied as a comparison to the three ML models. The first two models that we adapted were proposed by Saulnier et al. (1997) and predict soil thickness, which is often used as a proxy for landslide thickness. They use a predicted soil depth based on either elevation (hereafter Simple-Z) or slope gradient (hereafter Simple-S) by applying a linear interpolation based on the minimum and maximum values in a set of reference data. The third method used for comparison, proposed by van Zadelhoff et al. (2022) (hereafter SFM), also uses the slope gradient to predict soil depth. However, to account for the shallow soils on steep slopes the method derives the soil thickness from a log-normal distribution and multiplies it by a correction factor, which is a function of the slope gradient. For the Simple-Z model, the elevation sampled at 5 m cell size was used to train the model. For the Simple-S and the SFM model, the slope sampled at 10 m cell size was used since it showed the best fit with the slopes at the failure point recorded in the inventories.

4.3.2 Training and validation

Separate models were trained for the HMDB and KtBE datasets, resulting in 12 models overall (step S3.2 in Fig. 3). Both datasets were split into training data (80%) and validation data (20%). All models were trained using 10-fold cross-validation (James et al., 2021). The built-in *caret* functions were used for the RF, GAM, and LM models. Custom cross-validation methods were written for the Simple-Z, Simple-S and SFM models. The structure of the RF, GAM and LM models was automatically determined by *caret* during the training process.

The hyper-parameters of an RF model can significantly influence model performance (Huang and Boutros, 2016; Probst et al., 2019). Therefore, the training of the RF model was combined with a hyper-parameter tuning leveraging a modified version of the tune-grid functionality included in the *caret* package. The function performs a grid search over a defined set of parameters and values (cf. Table 4). Similarly, the standard hyper-parameters for GAM and LM available in *caret* were tuned. The final RF model showed clear differences in the chosen hyper-parameters. The HMDB-trained model used the splitrule "extratrees", a minimum node size of 2, 2 variables to split at, and 100 trees with a maximum depth of 50. The KtBE-trained model used the splitrule "maxstat", a minimum node size of 10, 8 variables to split at, and 50 trees with a maximum depth of 150. For the GAM models of both datasets, the "REML" method was used for smoothing parameter estimation. The KtBE-trained model was fitted with possible penalisation of terms (i.e. each term can potentially be removed from the model during fitting by adding an additional penalty) while the HMDB-trained model was fitted without. For the LM model, the hyper-parameter tuning resulted in a fit without intercept for the HMDB-trained model and a fit with intercept for the KtBE-trained model.

Table 4. Parameters and values used for tuning the hyper-parameters for the random forests (RF), generalized additive model (GAM) and linear regression (LM) models.

Model	Parameter	Default value	Values
RF	Number of variables considered to split at each node	Square root of the number of variables	2, 6, 8, 10, 22
	Splitting rule	'variance'	'variance', 'extratrees', 'maxstat'
	Minimal node size to split at	5 (for regressions)	2, 6, 8, 10, 12
	Number of trees	500	50, 100, 200
	Maximal tree depth	unlimited	50, 100, 150, 200, 250
GAM	Smoothing parameter estimation method	Generalized Cross-Validation (GCV) with Mallows' Cp for scale	'GCV.Cp', 'REML'
	Possible removal of terms through penalisation	FALSE	TRUE, FALSE
LM	Whether to fit with or without intercept	Fit with intercept	TRUE, FALSE

Performance measures were calculated based on the application of the trained models to the respective validation data in order to assess model performance (step S3.3 in Fig. 3). In addition, the models trained with the HMDB dataset were cross-applied to the KtBE validation data and vice versa to evaluate the transferability across datasets.

We used the MAE of predicted landslide thickness versus landslide thickness from the inventory as the primary performance assessment measure, which was calculated as:

$$MAE = \frac{1}{n_{slides}} \Sigma_{i=1}^{n_{slides}} \left| y_{Actual_i} - y_{Predicted_i} \right| \tag{Eq. 1}$$

Where:

MAE: Mean absolute error of the landslide failure thickness

y_{Actual_i}: Landslide failure thickness in m for landslide i according to the inventory

$y_{Predicted_i}$: Landslide failure thickness in m for landslide i predicted by the model

n_{slides}: Total number of landslides in the reference data

Additionally, we calculated the coefficient of determination R^2 to judge how good the fit between the predicted and actual data is:

$$R^2 = cor(y_{Actual}, y_{Predicted})^2 \tag{Eq. 2}$$

Where:

R^2: Coefficient of determination

y_{Actual}: Landslide failure thicknesses in m according to the inventory

$y_{Predicted}$: Landslide failure thickness in m predicted by the model

5   Results

5.1   Statistical properties of landslide inventories

The distributions of the landslide thickness in the HMDB and KtBE datasets are very similar with mean and median values that lie close to each other (see Table 5). The histogram of the landslide thickness (Fig. 4a) shows that 90% of the landslides had a thickness smaller than or equal to 1 m in the HMDB and the KtBE dataset and both show peaks at 0.5 m and 1 m. This is especially the case for the KtBE dataset, which also shows additional small peaks at 1.5 m and 2 m landslide thickness. In both the HMDB and KtBE datasets, landslides mainly occurred at slope angles above 15°, with a tendency for the number of occurrences to increase with the slope angle and a sharp decrease above 40° (Fig. 4b). Half of the landslides occurred above 36° in the HMDB and above 31° in the KtBE dataset. Almost no landslides occurred on slopes with more than 55° or below 15° slope angle. Visual verification of the landslides outside these boundaries using orthoimages and maps confirmed that the failures are plausible. The landslides above 55° slope angle (HMDB: n=4) were all located in forested areas and most landslides below 15° (HMDB: n=16, KtBE: n=5) were located at a transition between flat and steep terrain such as a terrace or a road. The comparison of the recorded slope values with those from the covariates sampled at 5 m cell size showed slightly lower mean and median values for the HMDB dataset and very similar values for the KtBE dataset. The elevation range where shallow

landslides occurred, as well as its distribution, is similar in both datasets (Table 5). In the distribution of the landslide thickness over the slope classes (Fig. 5), the KtBE dataset shows a slight tendency towards decreasing landslide thicknesses and lower variance with increasing slope steepness. The HMDB dataset shows a similar, albeit even weaker, tendency towards such a decrease and no discernible pattern in the variance.

Table 5. Summary statistics including the total number of records (n), n within forests, the minimum, median, median absolute deviation (MAD), mean, standard deviation (SD), and maximum for important characteristics of shallow landslides derived from the landslide inventories as well as from covariates derived from swissALTI3D (Swisstopo, 2023a). Note that lower sample sizes result from missing values in the inventory.

Source	Variable	Dataset	n	(n in forest)	Min	Median	MAD	Mean	SD	Max
Inventory	**Landslide thickness [m]**	HMDB	709	(319)	0.1	0.7	0.3	0.6	0.3	2.0
		KtBE	515	(20)	0.2	0.6	0.1	0.5	0.3	2.0
	Slope [°]	HMDB	648	(295)	1.0	35.1	7.4	36.0	9.0	80.0
		KtBE	515	(20)	8.3	31.0	5.3	31.1	5.8	52.5
Covariates	**Elevation [m a.s.l.]**	HMDB	709	(319)	523	1214	346	1188	290	2257
		KtBE	515	(20)	624	1334	406	1283	470	2288
	Slope 5 m [°]	HMDB	709	(319)	6.0	33.3	6.9	34.3	7.8	58.9
		KtBE	515	(20)	8.2	31.3	6.1	31.8	6.6	48.3
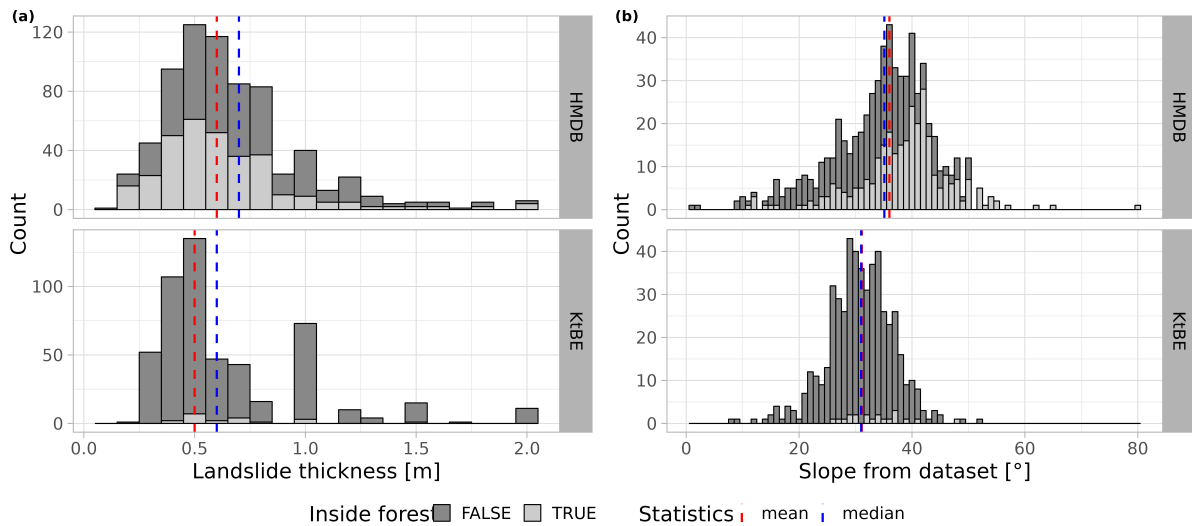
Figure 4. Histograms showing the distribution of (a) the landslide thickness and (b) the mean slope values in the release areas recorded in the HMDB (upper row) and KtBE (lower row) datasets.
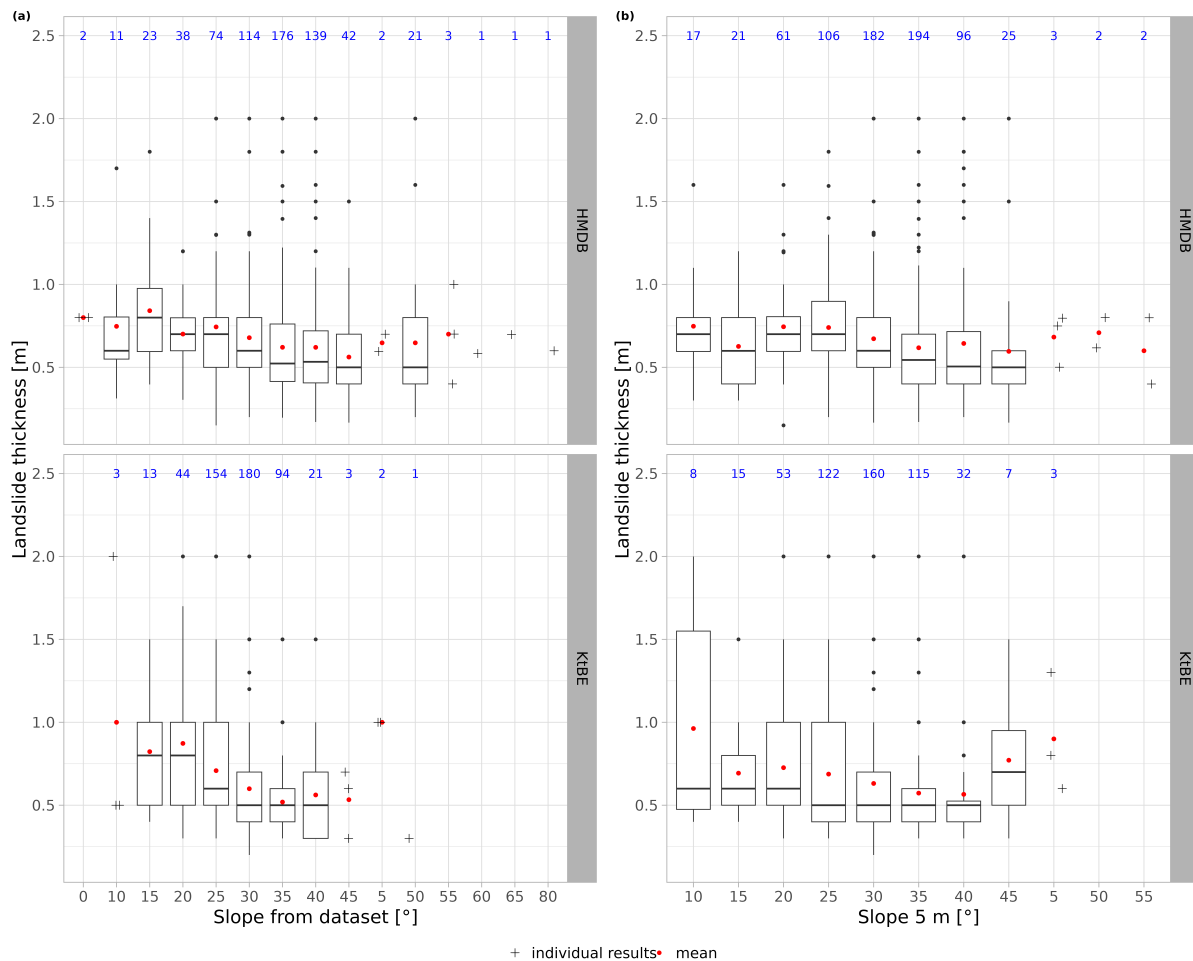
Figure 5. Box plots showing the distribution of the landslide thickness over slope classes based on (a) the reported slope for the different datasets and (b) the slope values sampled at 10 m cell size. The red dots represent the mean value while the blue numbers show the number of records per slope class. Individual, cross-shaped points are shown for classes with less than 5 data points.

5.2 Modelling results

Overall, the RF models performed best with an MAE of 0.25 m for the HMDB and 0.20 m for the KtBE dataset (see Table 6). While the MAE values of the GAM and LM models are comparable for the respective datasets, their R^2 values tend to be lower than the RF models. Furthermore, both the GAM and LM models show some outliers with negative prediction values of up to -0.73 m. The Simple-Z and Simple-S models clearly showed worse performance with MAE values exceeding 1 m. The SFM model achieved MAE values of 0.37 m for HMDB and 0.38 m for KtBE. When comparing the performance between the datasets, the models for the KtBE dataset performed slightly better than the HMDB-trained models. The results for the

cross-application are variable. In most cases, MAE values were comparable to those of the model trained with the same dataset

345 with differences between 0 cm to 7 cm for the ML models and between -7 cm and 9 cm for the simple models.

Table 6. Performance of the trained models. Results are shown for the application to the validation data of the respective dataset as well as for the cross-application to the other dataset. n_{test} denotes the number of records in the validation data. $Prediction_{min}/Prediction_{max}$ correspond to the minimum/maximum predicted values while $Error_{min}/Error_{max}$ correspond to the minimum/maximum error. AE_{p90} denotes the 90th percentile of the absolute error.

Model type	Training dataset	Validation dataset	n_{test}	MAE	R^2	$Prediction_{min}$	$Prediction_{max}$	$Error_{min}$	$Error_{max}$	AE_{p90}
Random Forest	HMDB	HMDB	138	0.25	0.08	0.01	0.86	-0.58	1.43	0.54
Random Forest	BE	BE	114	0.20	0.32	0.02	1.36	-0.86	1.34	0.39
Random Forest	BE	HMDB	743	0.25	0.07	0.01	1.16	-0.70	1.65	0.53
Random Forest	HMDB	BE	576	0.28	0.18	0.00	0.91	-0.82	1.52	0.51
GAM	HMDB	HMDB	138	0.25	0.09	0.08	0.79	-0.48	1.49	0.50
GAM	BE	BE	114	0.21	0.30	-0.48	1.23	-0.71	1.39	0.45
GAM	BE	HMDB	743	0.25	0.08	-0.29	1.17	-0.67	1.68	0.56
GAM	HMDB	BE	576	0.26	0.23	-0.29	0.80	-0.64	1.48	0.42
LM	HMDB	HMDB	138	0.25	0.09	0.16	0.89	-0.41	1.51	0.49
LM	BE	BE	114	0.23	0.23	-0.73	1.26	-0.67	1.47	0.45
LM	BE	HMDB	743	0.25	0.07	-0.53	1.25	-0.95	1.68	0.51
LM	HMDB	BE	576	0.26	0.20	-0.45	0.93	-0.68	1.70	0.46
Simple-S	HMDB	HMDB	138	0.90	0.00	1.26	1.93	-1.93	0.63	1.25
Simple-S	BE	BE	114	1.09	0.06	1.11	1.96	-1.96	0.45	1.37
Simple-S	BE	HMDB	743	0.99	0.05	1.04	2.07	-2.07	0.58	1.33
Simple-S	HMDB	BE	576	1.02	0.13	0.96	1.93	-1.93	0.50	1.32
Simple-Z	HMDB	HMDB	138	0.87	0.09	0.50	2.01	-1.46	0.42	1.31
Simple-Z	BE	BE	114	0.87	0.20	-0.13	1.97	-1.45	0.82	1.29
Simple-Z	BE	HMDB	743	0.92	0.08	-0.11	2.07	-1.77	0.71	1.31
Simple-Z	HMDB	BE	576	0.88	0.16	-0.02	1.98	-1.83	0.87	1.31
SFM	HMDB	HMDB	138	0.37	0.00	0.01	1.43	-0.92	1.79	0.73
SFM	BE	BE	114	0.38	0.00	0.00	1.54	-1.24	1.70	0.81
SFM	BE	HMDB	743	0.37	0.01	0.00	1.62	-1.10	1.99	0.74
SFM	HMDB	BE	576	0.35	0.02	0.00	1.71	-1.31	1.89	0.75

The data points of the measured vs. the ML-predicted landslide thickness are located closer to the identity line when compared to the three simple models (cf. Fig. 6). However, all three models showed a tendency to slightly overestimate lower landslide thicknesses and distinctly underestimate higher thicknesses. The Simple-S model tended to overestimate the landslide thickness clearly across the entire range. The Simple-Z model showed a similar tendency but with a higher variance. The

350 SFM model also exhibited a high variance but with predictions closer to the identity line. In particular, the added points with 0 m landslide thickness showed high variances in all three of the simple models.

The distribution of the residuals per slope class (Fig. 7) was similar for most slope classes where training and validation data stem from the same dataset. Some classes like e.g. the 40° class of the HMDB dataset show higher variance. The simple

models generally also had higher variances than the ML models. Looking at the cross-application of the models (Fig. A1 in

355 Appendix A1), there were evidently more outliers and higher variances across all slope classes. In particular, the 15° class of

the HMDB-trained models applied to the KtBE data consistently showed high variances.
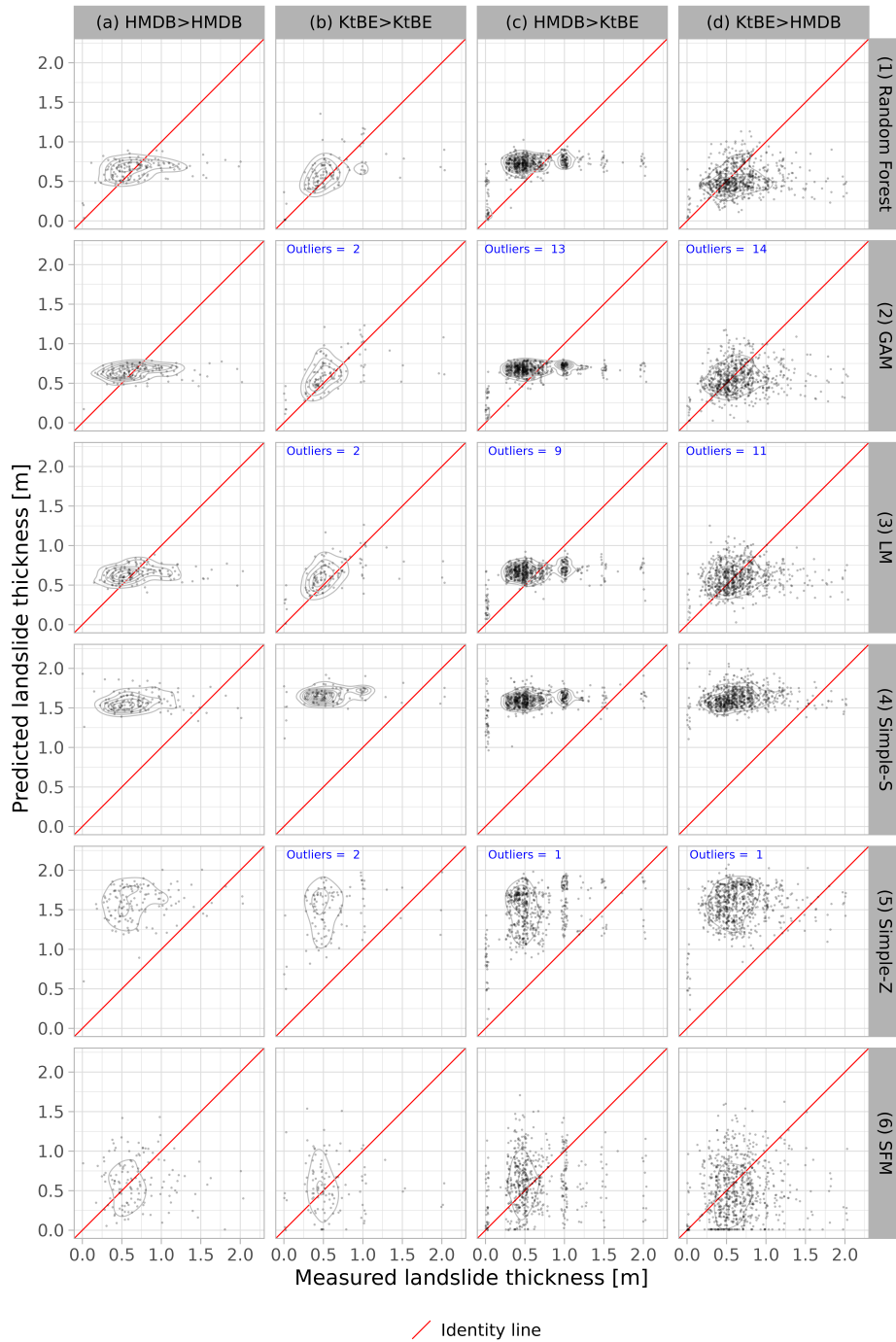
Figure 6. Scatter plots showing the measured vs. the predicted landslide thickness differentiated by model type (rows) and dataset (columns): (a) model trained and tested with HMDB dataset, (b) model trained and tested with KtBE dataset, (c) model trained with HMDB and tested with KtBE dataset, and (d) model trained with KtBE and tested with HMDB dataset. The diagrams have 2D kernel density contours in the background as well as the identity line in red. Where present, the blue text denotes the number of outliers outside the display range of the plots.
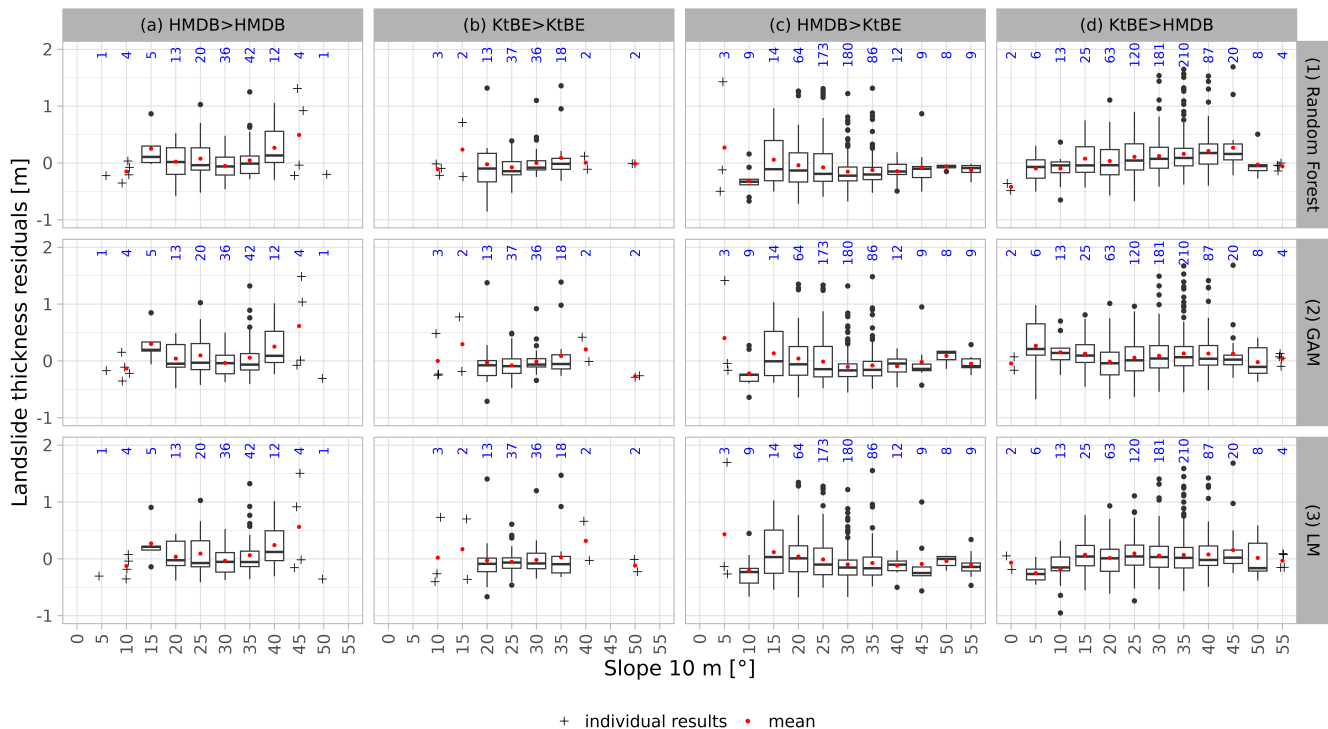
Figure 7. Box plots showing slope classes sampled at 5 m cell size vs. the residuals of the predicted landslide thickness for the best three models differentiated by model type (rows) and dataset (columns, see caption Fig. 6 for details). The red dots represent the mean value while the blue numbers show the number of entries per slope class. Individual, cross-shaped points are shown for classes with less than 5 data points. For readability, those that occurred above the 55° slope gradient class are not displayed here.

The analysis of the variable importance of the ML models showed clear differences between the models. However, the variables elevation, terrain roughness index at 10 m cell size, mean density of the local lithology, negative openness, and multiresolution valley bottom flatness are found more often among the top-ranked ones (compare Fig.B1 in supplementary materials). The hyper-parameter tuning resulted in a reduction of the MAE of up to 2 cm for certain models. The chosen parameters were the same across most of the datasets.

6 Discussion

6.1 Landslide inventories

The slope values recorded in the HMDB and KtBE landslide inventories (cf. Fig. 4) largely match the ranges found in the literature, with reported values from 5° to 35° (Guzzetti et al., 2008a), from 20° to 35° (Meier et al., 2020), from 22° to 40° (Dahl et al., 2010), and from 19° to 50° with predominant values from 25° to 45° (Rickli and Graf, 2009).

Larsen et al. (2010) compared data on landslide thickness from several landslide inventories located across the globe, some of which showed peaks around 0.5 m, 1 m, and 2 m similar to the ones observed in the HMDB and KtBE inventories. Particularly in the KtBE dataset, the observed peaks are a possible bias of the expert estimations. For the KtBE dataset, the errors in the landslide thickness were estimated to be 25% to 50% (Hählen, 2023). In general, the data quality of the inventory data is crucial for its suitability for ML. A contrasting example is the Swiss national register of natural hazard events ("StorMe") which is maintained by the Federal Office for Environment (FOEN) and the cantons (Burren and Eyer, 2000). The distribution of the landslide depth of a selection of 2988 shallow landslides recorded in StorMe (Fig. 8) shows similar peaks but in a much more pronounced manner. We also attribute these peaks and the gaps in between to rough estimates by many different experts performing the recording of event characteristics in the field. We tested the StorMe data for possible use in this study. However, the anomalies in the landslide thickness distribution together with other data quality issues like missing or unrealistic values led to the exclusion of the dataset. This illustrates that initiatives like the HMDB and the acquisition of more high-quality data play a key role in obtaining the input data necessary for future ML-based modelling.
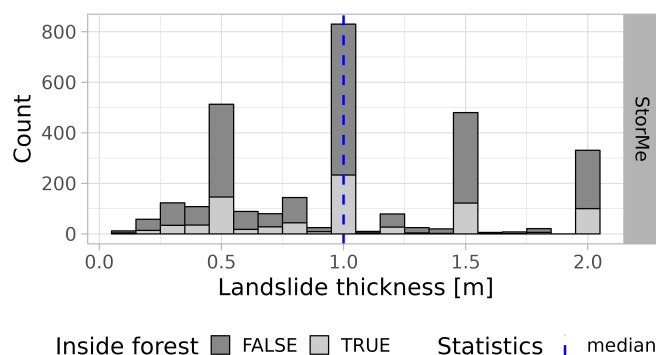
Figure 8. Histogram showing the distribution of the landslide thickness recorded for 2988 shallow landslides in the Swiss national register of natural hazard events (StorMe).

6.2 Model performance

Among the three ML models tested, our study focused on RF because of its potential. The LM and GAM models served as reference models with different properties. RF models can flexibly capture non-linear relationships and interactions between variables and covariates (Breiman, 2001; James et al., 2021). However, their complex model structure and ensemble nature hinder model explainability and result interpretation (Breiman, 2001; James et al., 2021). LM have the advantage of a transparent model structure that is simple to implement but can not capture non-linear relationships (James et al., 2021). GAM are an extension of LM that allows to model non-linear relationships by fitting smoothing functions to the explanatory variables (Hastie and Tibshirani, 1990; James et al., 2021). However, they are more susceptible to overfitting (Wood, 2011). In addition, interactions must be explicitly specified for LM and GAM, since they are not captured automatically (James et al., 2021).

The results from the model tests reflect some of the different model properties and, overall, confirmed the potential of the ML models for predicting the potential failure thickness of shallow landslides.

390 The ML models clearly outperformed the two simple models based on elevation and slope gradient proposed by Saulnier et al. (1997). However, it has to be noted that those models were originally intended to be used in smaller areas like single catchments with more uniform terrain, whereas our application covers a much larger area in Switzerland with more diverse terrain conditions. Although the SFM model proposed by van Zadelhoff et al. (2022) calibrated with the values from the inventories produced better results, the RF model still yielded MAE values that were 20% to 44% lower. While the ML models

395 exhibited low errors, they all tended to underestimate higher landslide thickness values. We attribute this to the low number of records with a landslide thickness of more than one meter in the datasets used for model training. The cross-application of the ML models showed that they are mostly transferable between the HMDB and KtBE datasets. One limiting factor might be the spatial and temporal hierarchy in the HMDB data resulting from the fact that most of the data were recorded in perimeters after defined heavy rainfall events. However, tests showed no significant signs of spatial autocorrelation in the model

400 predictions. Another limiting factor may be the missing events in forests or in intensively cultivated areas not recorded in the KtBE inventory. Nevertheless, there appears to be enough similarity in the distributions of the landslide thickness over the slope classes and the covariates in both datasets to conclude that the models are transferable. This is supported by an additional test with a cross-application of the HMDB-trained model to the StorMe dataset, which showed a clear decline in performance. We attribute this mainly to the distinctly different distribution of the landslide thickness in the StorMe inventory with the dominant

405 peaks at 0.5 m, 1 m, and 2 m. This indicates a limited generalizability of the trained models, although this may also be due to differences in data quality.

  Although comparability is limited due to differences in the study area extents and the use of soil depth landslide thickness instead of soil depth as the target variable, our results show similar tendencies as comparable studies. Catani et al. (2010) applied two simple models based on elevation and slope proposed by Saulnier et al. (1997) along with the geomorphology-

410 based GIST model in the Terzona catchment (24 km^2) in Italy. The Simple-Z and Simple-S models performed worst with MAE values of 0.94 m and 0.54 m, while the GIST model had an MAE of 0.11 m. Subsequently, Segoni et al. (2012) applied the same models in the Armea catchment (37 km^2) in Italy. The Simple-Z and Simple-S models again performed worse with MAE values of 0.78 m and 1.03 m compared to an MAE of 0.23 m for the GIST model. An additional model linking the soil depth to the slope using an exponential function showed an MAE of 0.45 m but tended to give unrealistically low prediction

415 values. Xiao et al. (2023) applied the GIST model and the random forests-based GIST-RF model to generate soil depth maps in a section along the Yangtze River in Wanzhou County (27 km^2), where soil depth ranges from 0 m to 40 m. The MAE values of 10.6 m for the GIST showed that the original model cannot deal with the complex geological settings and high variability in soil depth at the study site. However, the GIST-RF model showed an MAE of 3.52 m, demonstrating the potential for improvement through ML techniques. (Gupta et al., 2024) applied GIST, GIST-MCS and GIST-RF models in a study assessing

420 soil thickness along three important roads (673 km^2) in the Joshimath region (Indian Himalaya). The GIST, GIST-MCS, and GIST-RF models showed MAE of 3.94 m, 2.86 m, and 1.64 m, thus further confirming the potential of GIST-RF.

The results of the ML models are promising. The RF model showed the overall best performance with the best fit and without any undesired outliers. The performance of the GAM models and LM models was comparable, while they had an overall worse fit and showed undesired outliers (especially negative ones). The partial effect plots for the GAM models showed that the degree of smoothing is low or close to zero for most terms. This is, at least in part, a result of the automatic estimation of the smoothing functions performed by *caret*, which does not allow any manual intervention. We expect that the GAM models could be improved by manually building a model with individualised terms and smoothing functions. Overall, additional tests and optimisation may be advisable. In particular, we would expect the inclusion of additional records with a landslide thickness between 1 m and 2 m to improve the model performance. The hyper-parameter tuning only yielded improvements that did not exceed a reduction of the MAE by 1 cm. However, the difference of up to 2 cm between some of the hyper-parameter combinations for the RF models shows that hyper-parameter tuning is generally worthwhile. The results of the hyper-parameter tuning for the RF models also showed tendencies that are mostly in line with the findings reported by Probst et al. (2019). Additional model variants with and without additional points with 0 m landslide thickness in rock signature have been explored. However, while the addition of the points did increase the overall MAE of the ML models by about 1 cm to 5 cm, most of the R^2 values clearly improved. At the same time, the points with 0 m landslide thickness also influenced the mean of the predicted values. For the ML models as well as the SFM model there was a decrease of 2.4 cm to 6.4 cm of the overall mean of the predicted values. For the Simple-S and Simple-Z models the mean of the predicted values increased by 13.1 cm to 21.6 cm. Due to this influence and given the large variance of the predictions for the 0 m thickness points, it is still not clear to us whether the addition of the points is recommendable. The scheme used for random sampling in space could potentially introduce a bias and class overlap (i.e. overlap between conditions associated with occurrences of events and pseudo-absences), which may be mitigated by adopting a uniform sampling approach (Da Re et al., 2023).

6.3 Covariate selection

The selected covariates in the ML models mostly describe the terrain and its geomorphology. This shows parallels to models aiming to predict soil depth in general based on geomorphology (Catani et al., 2010; Xiao et al., 2023). The elevation, slope, TWI, curvature, and aspect included in our covariates were also identified as important causal factors of shallow landslides in grassland regions of Switzerland by Zweifel et al. (2021). In addition, the inclusion of covariates on the vegetation and underground are worth mentioning. The influence of the vegetation (especially the forest) was realised by including the maximum of the VHM. This variable consistently showed higher importance for the models trained with the HMDB data compared to the KtBE-trained models. This may be due to the low number of landslides within the forest present in the KtBE data. While the high importance of the maximum of the VHM hints at a significant influence of the forest, the NFI forest type raster with the proportion of coniferous trees was not included in the final model. The reason was the very low or even negative importance values suggesting a lack of relationship with the target variable, which could be a result of the distribution of the landslides inside the forest. The geological substratum was included by the mean local density of the underlying bedrock serving as a proxy of the lithology. While the importance of lithology varied significantly between the different model types and dataset combinations, it was high enough to warrant inclusion in the model. The covariates on soil properties were ultimately not

included in the final model. On the one hand, the importance of the clay, sand, and silt content at the different depths varied considerably from negative importance to moderately important depending on the dataset. We tested several covariate selections including different combinations of reasonably important soil property variables and the mean bedrock density. While some of these combinations showed an MAE close to or even slightly better than the final selection, they also showed lower R2 values. In addition, adding up to 9 different variables would have increased the model complexity considerably. Together with the uncertainties connected with the soil property data, this led to the decision not to include any soil property covariates in the final model.

A variant of the ML models including the rainfall amount for a duration of 60 minutes for an extreme point precipitation event with a 10-year return period as an additional variable was also explored. The results, however, were almost identical to the model without this additional variable. We speculated *a priori* that this variable might be related to soil thickness due to an erosion effect, but this is apparently not the case or not detectable. Aiming for a model with less complexity, we finally opted to exclude this variable.

6.4 Uncertainty

Although the performance of the model is promising, it is subject to uncertainties, which are a challenge in natural hazard modelling and prediction. Uncertainty quantification is critical for improving the reliability and interpretability of predictive models, particularly in applications that inform risk management and mitigation. While various methods exist for quantifying uncertainty, they differ significantly across disciplines and model types, with no universally applicable solutions for machine learning models (Beven, 2018; Jalaian et al., 2019; Simmonds et al., 2022). Unlike classical statistical models such as linear regression, which offer established confidence interval formulas, machine learning approaches often lack analogous tools, creating additional challenges for systematic uncertainty evaluation (Jalaian et al., 2019). In this study, we did not apply uncertainty quantification due to the heterogeneous nature of the tested models and the inherent difficulties in quantifying uncertainties in both reference and covariate data. These data constitute a significant source of uncertainty, as inaccuracies in inputs propagate through the models, ultimately influencing predictions (Simmonds et al., 2022). Addressing these data uncertainties is crucial for enhancing the robustness of ML-based models and should be prioritized in future research similar to the work of Meinshausen (2006) or (Wager et al., 2014) providing confidence intervals for RF.

The main uncertainties in the covariate data lie in the soil, geological substratum, and forest type datasets. For the soil and the geological substratum, which (for the most part) cannot be measured directly, making data subject to uncertainties based on model assumptions and mapping precision. Similar to the model in our study, the values in the tested soil property maps were predicted using a Quantile Regression Forest (Meinshausen, 2006) based on sparse field measurements and a large number of covariates derived from remote sensing (Stumpf et al., 2024). The resulting predictions are already subject to a considerable degree of uncertainty, especially due to the 30 m cell size and the high local heterogeneity in soil properties. In the geological data, uncertainties stemming from different surveyors and methodologies cannot be ruled out. This is exemplified by the discontinuities in the data stemming from different map sheets in the original data on bedrock and unconsolidated deposits from the GeoCover dataset (Swisstopo, 2023c) used in preliminary tests. The bulk density dataset of local lithology explicitly

attempts to explicitly quantify variability and uncertainty by including percentile values of the density distributions (Swisstopo, 2020; Zappone and Kissling, 2021). While this approach enhances transparency, the uncertainty still reflects natural variability within lithologies, errors in physical property measurements, and sparse sampling coverage (Zappone and Kissling, 2021). Forest type data, modelled using RF and neural networks using remote sensing data (Waser et al., 2017), are similarly subject to uncertainty. Significant deviations from the national forest inventory, particularly in the underestimation of broadleaved trees, highlight the limitations of current methods(Waser et al., 2017). The authors primarily attributed these deviations to errors in the input image data rather than the classification approach itself (Waser et al., 2017).

Uncertainties in the field inventory data used to train the models also influenced the accuracy of model predictions. Positional errors in the failure points, e.g. due to imprecise GPS measurements, can result in inaccurate sampling of covariates, reducing model reliability. This issue is partially mitigated by the averaging effects of the larger cell sizes and windowed calculations used for the covariate rasters. Additionally, errors in estimating landslide thickness present a challenge, as these inaccuracies can render datasets unsuitable for ML applications. For example, the limitations of the StorMe dataset due to unreliable thickness estimates highlight the necessity for improved field data protocols tailored to ML requirements.

The uncertainty of the model results is also largely influenced by the choice of model type, training data and covariates used during the training process (Jalaian et al., 2019). Different combinations of training records and covariates can lead to significant variability in the structure and outputs of the used RF, GAM and LM models. For instance, cross-validation introduces a random factor that can result in variability, particularly when models are trained on small datasets. This variability highlights the need for systematic sensitivity analyses to identify robust covariate combinations that optimize model performance. Moreover, the selection process revealed numerous covariate combinations with similar performance metrics, suggesting a potential for overfitting or underutilization of critical variables (cf. Merghadi et al., 2020). Given that optimal covariate combinations may vary depending on cell sizes, input datasets, and model types, future studies should explore adaptive methods for covariate selection that are tailored to specific data and model contexts. This would align model configurations more closely with the specific characteristics of the study area.

7 Conclusions

In this study, we presented an ML-based approach to predict the potential thickness of shallow landslides. The new machine learning models consistently performed at least 20% better when comparing the MAE to simple models based on slope gradient and elevation. We conclude that the selected set of covariates, including metrics on terrain, geomorphology, vegetation, and lithology, is a suitable basis for predicting shallow landslide thickness using ML. Considering the overall performance and the lack of outliers in the predictions, we consider the RF model to be the most accurate approach to generate improved inputs for slope stability models. For future work, we plan to adapt this study's RF model, which is built for predictions on single sample points, for the generation of rasters covering large extents. Additionally, the model can be further developed, especially by improving the input dataset with additional field data, by testing variants with additional sample points and sampling schemes for locations with 0 m landslide thickness, and by further refining the selection of covariates.

Code availability. The code for this study is available on Zenodo (Schaller, 2024)
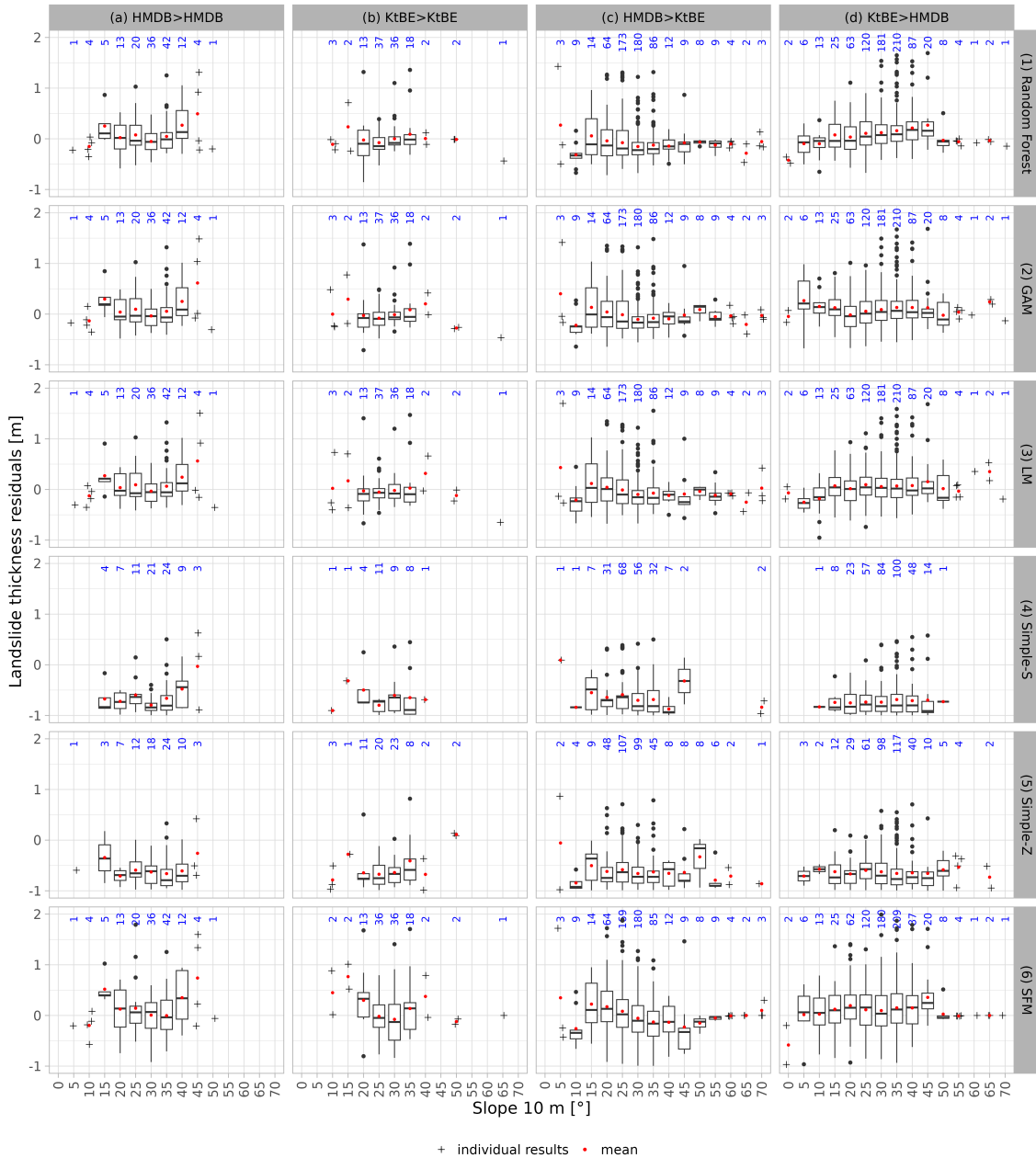
Figure A1. Box plots showing slope classes sampled at 5 m cell size vs. the residuals of the predicted landslide thickness for all models differentiated by model type (rows) and dataset (columns): (a) model trained and tested with HMDB dataset, (b) model trained and tested with KtBE dataset, (c) model trained with HMDB and tested with KtBE dataset, and (d) model trained with KtBE and tested with HMDB dataset. The red dots represent the mean value while the blue numbers show the number of entries per slope class. Individual points are shown for classes with less than 5 measurements.

A1   Variable importance and model tuning

Variable importance values were extracted for the ML-based models (Fig. B1). The values of the RF and LM were extracted from the overall result of the 10-fold cross-validation result. Since no meaningful overall values could be extracted from the GAM fit, the importance values from the best model were extracted. The results show that there are differences in the importance of the individual variables depending on the dataset and the model. Nevertheless, several variables are more often among the top-ranked variables including the elevation, the terrain roughness index at 10 m cell size, the mean density of the local lithology, the negative openness, and the multiresolution valley bottom flatness.

530
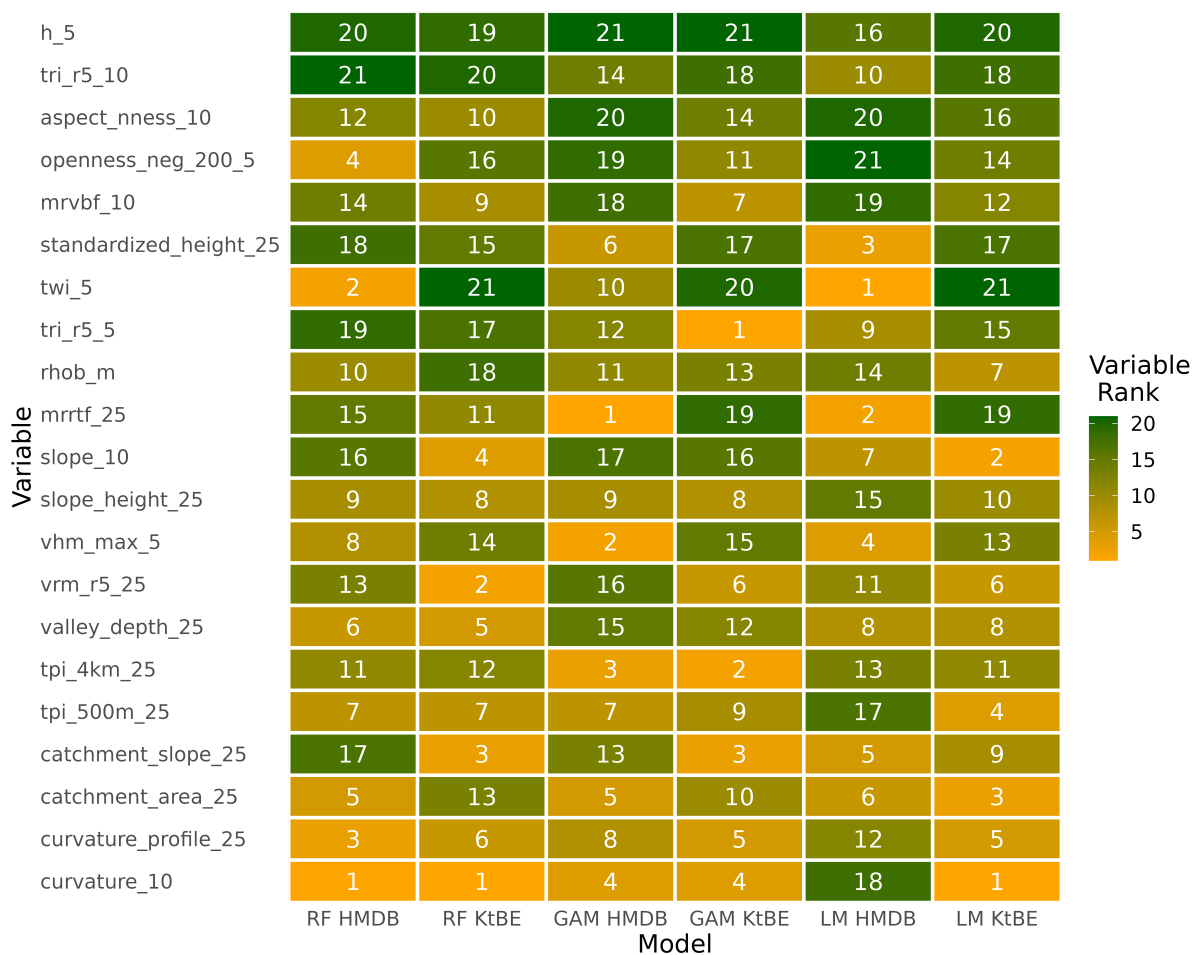
Figure B1. Heatmap showing the overall variable importance extracted from 10-fold cross-validation for the RF and LM models and the importance of the best fit for the GAM model. The number in the cells and their colour correspond to the importance rank of the variable within the model, with 21 (green) being the most important and 1 (yellow) being the least important variable. The variables are sorted by the sum of the ranks of each row.

Author contributions. **Christoph Schaller:** Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualisation. **Massimiliano Schwarz:** Conceptualisation, Methodology, Software, Writing - Original Draft, Writing - Review & Editing, Supervision. **Christine Moos:** Methodology, Writing - Review & Editing. **Arie C. Seijmonsbergen:** Writing - Review & Editing, Supervision. **Luuk Dorren:** Conceptualisation, Methodology, Software, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. **Emiel van Loon:** Methodology, Writing - Review & Editing, Supervision.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. **Appendix C: Acknowledgements**

This work was partially funded by the Bern University of Applied Sciences and the Principality of Liechtenstein. We thank both anonymous reviewers for the numerous suggestions for improvements of the paper.

References

Ali, A., Huang, J., Lyamin, A. V., Sloan, S. W., Griffiths, D. V., Cassidy, M. J., and Li, J. H.: Simplified quantitative risk assessment of rainfall-induced landslides modelled by infinite slopes, Eng. Geol., 179, 102–116, https://doi.org/10.1016/j.enggeo.2014.06.024, 2014.

Arnold, P. and Dorren, L.: The Importance of Rockfall and Landslide Risks on Swiss National Roads, in: Engineering Geology for Society and Territory - Volume 6, edited by Lollino, G., Giordan, D., Thuro, K., Carranza-Torres, C., Wu, F., Marinos, P., and Delgado, C., pp. 671–675, Springer International Publishing, Cham, ISBN 978-3-319-09060-3, https://doi.org/10.1007/978-3-319-09060-3_120, 2015.

Badoux, A., Andres, N., Techel, F., and Hegg, C.: Natural hazard fatalities in Switzerland from 1946 to 2015, Nat. Hazard Earth Sys., 16, 2747–2768, https://doi.org/10.5194/nhess-16-2747-2016, 2016.

BAFU: Topographische Einzugsgebiete Schweizer Gewässer Schweiz, Ausgabe 2019, https://data.geo.admin.ch/ch.bafu.wasser-teileinzugsgebiete_2/, 2019.

BAFU: Produktionsregionen LFI, https://data.geo.admin.ch/ch.bafu.landesforstinventar-produktionsregionen/, 2020.

Baum, R. L., Savage, W. Z., and Godt, J. W.: TRIGRS - a Fortran program for transient rainfall infiltration and grid-based regional slope-stability analysis, Open-File Report, https://doi.org/10.3133/ofr02424, 2002.

Beven, K.: Environmental Modelling: An Uncertain Future?, CRC Press, London, ISBN 978-1-315-27350-1, https://doi.org/10.1201/9781482288575, 2018.

Bezzola, G. R. and Hegg, C.: Ereignisanalyse Hochwasser 2005, Teil 1–Prozesse, Schäden und erste Einordnung, in: Umwelt-Wissen, vol. 707, p. 215, Bundesamt für Umwelt BAFU; Eidgenössische Forschungsanstalt WSL, Bern; Birmensdorf, https://www.bafu.admin.ch/dam/bafu/de/dokumente/naturgefahren/uw-umwelt-wissen/ereignisanalyse_hochwasser2005teil1prozesseschaedenundersteeinor.pdf.download.pdf, 2007.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Brenning, A.: Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models, in: SAGA – Seconds Out (= Hamburger Beitraege zur Physischen Geographie und Landschaftsoekologie, vol. 19), pp. 23–32, J. Boehner, T. Blaschke, L. Montanarella, https://fiona.uni-hamburg.de/e2bfe5e6/boehner-et-al--saga-seconds-out.pdf, 2008.

Burren, S. and Eyer, W.: StorMe–Ein informatikgestützter Ereigniskataster der Schweiz, Internationales Symposon, Interpraevent, pp. 25–35, 2000.

Böhner, J. and Selige, T.: Spatial prediction of soil attributes using terrain analysis and climate regionalisation, SAGA-Analyses and modelling applications, 115, 13–27, http://downloads.sourceforge.net/saga-gis/gga115_02.pdf, 2006.

Böhner, J., Koethe, R., Conrad, O., Gross, J., Ringeler, A., and Selige, T.: Soil regionalisation by means of terrain analysis and process parameterisation, Soil Classification 2001, pp. 213–222, https://esdac.jrc.ec.europa.eu/ESDB_Archive/eusoils_docs/esb_rr/n07_ESBResRep07/601Bohner.pdf, 2002.

Caine, N.: The Rainfall Intensity - Duration Control of Shallow Landslides and Debris Flows, Geogr. Ann. A, 62, 23–27, https://doi.org/10.1080/04353676.1980.11879996, 1980.

Catani, F., Segoni, S., and Falorni, G.: An empirical geomorphology-based approach to the spatial prediction of soil thickness at catchment scale, Water Resour. Res., 46, https://doi.org/10.1029/2008WR007450, 2010.

CCSol: Swiss Competence Center for Soils home page, https://ccsols.ch/de/home/, 2024.

Chang, W.-J., Chou, S.-H., Huang, H.-P., and Chao, C.-Y.: Development and verification of coupled hydro-mechanical analysis for rainfall-induced shallow landslides, Eng. Geol., 293, 106 337, https://doi.org/10.1016/j.enggeo.2021.106337, 2021.

580 Chinkulkijniwat, A., Tirametatiparat, T., Supotayan, C., Yubonchit, S., Horpibulsuk, S., Salee, R., and Voottipruex, P.: Stability characteristics of shallow landslide triggered by rainfall, J. Mt. Sci., 16, 2171–2183, https://doi.org/10.1007/s11629-019-5523-7, 2019.

Cohen, D., Lehmann, P., and Or, D.: Fiber bundle model for multiscale modeling of hydromechanical triggering of shallow landslides, Water Resour. Res., 45, https://doi.org/10.1029/2009WR007889, 2009.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Böhner, J.: System for Automated 585 Geoscientific Analyses (SAGA) v. 2.1.4, Geosci. Model Dev., 8, 1991–2007, https://doi.org/10.5194/gmd-8-1991-2015, 2015.

Cruden, D. and Varnes, D. J.: Landslide Types and Processes, In: Turner, A.K., Schuster, R.L. (Eds.), Landslides: investigation and mitigation. National Academy Press,Washington, D.C, 247, 36–75, 1996.

Da Re, D., Tordoni, E., Lenoir, J., Lembrechts, J. J., Vanwambeke, S. O., Rocchini, D., and Bazzichetto, M.: USE it: Uniformly sampling pseudo-absences within the environmental space for applications in habitat suitability models, Methods Ecol. Evol., 14, 2873–2887, 590 https://doi.org/10.1111/2041-210X.14209, 2023.

Dahl, M.-P. J., Mortensen, L. E., Veihe, A., and Jensen, N. H.: A simple qualitative approach for mapping regional landslide susceptibility in the Faroe Islands, Nat. Hazard Earth Sys., 10, 159–170, https://doi.org/10.5194/nhess-10-159-2010, 2010.

Di Napoli, M., Di Martire, D., Bausilio, G., Calcaterra, D., Confuorto, P., Firpo, M., Pepe, G., and Cevasco, A.: Rainfall-Induced Shallow Landslide Detachment, Transit and Runout Susceptibility Mapping by Integrating Machine Learning Techniques and GIS-Based 595 Approaches, Water, 13, 488, https://doi.org/10.3390/w13040488, 2021.

Dietrich, W. E., Reiss, R., Hsu, M.-L., and Montgomery, D. R.: A process-based model for colluvial soil depth and shallow landsliding using digital elevation data, Hydrol. Process., 9, 383–400, https://doi.org/10.1002/hyp.3360090311, 1995.

D'Odorico, P. and Fagherazzi, S.: A probabilistic model of rainfall-triggered shallow landslides in hollows: A long-term analysis, Water Resour. Res., 39, https://doi.org/10.1029/2002WR001595, 2003.

600 Dorren, L. and Schwarz, M.: Quantifying the Stabilizing Effect of Forests on Shallow Landslide-Prone Slopes, in: Ecosystem-Based Disaster Risk Reduction and Adaptation in Practice, edited by Renaud, F. G., Sudmeier-Rieux, K., Estrella, M., and Nehren, U., Advances in Natural and Technological Hazards Research, pp. 255–270, Springer International Publishing, Cham, ISBN 978-3-319-43633-3, https://doi.org/10.1007/978-3-319-43633-3_11, 2016.

Dorren, L., Sandri, A., Raetzo, H., and Arnold, P.: Landslide risk mapping for the entire Swiss national road network, Landslide Processes: 605 from geomorphologic mapping to dynamic modeling, Strasbourg, France, pp. 6–7, 2009.

EEA: European Digital Elevation Model (EU-DEM), http://www.eea.europa.eu/data-and-maps/data/eu-dem, 2016.

Emberson, R., Kirschbaum, D., and Stanley, T.: New global characterisation of landslide exposure, Nat. Hazard Earth Sys., 20, 3413–3424, https://doi.org/10.5194/nhess-20-3413-2020, 2020.

Fallot, J.-M.: Climate Setting in Switzerland, in: Landscapes and Landforms of Switzerland, edited by Reynard, E., pp. 31–45, Springer 610 International Publishing, Cham, ISBN 978-3-030-43203-4, https://doi.org/10.1007/978-3-030-43203-4_3, 2021.

FDFA: Geography – Facts and Figures, https://www.eda.admin.ch/aboutswitzerland/en/home/umwelt/geografie/geografie---fakten-und-zahlen.html, 2023.

Frei, C. and Fukutome, S.: Extreme Point Precipitation, in: Data and Analysis Platform, Hydrological Atlas of Switzerland, https://hydromaps.ch/#en/8/46.830/8.190/bl_hds--precip_24h_2a$4/NULL, 2022.

615 Froude, M. J. and Petley, D. N.: Global fatal landslide occurrence from 2004 to 2016, Nat. Hazard Earth Sys., 18, 2161–2181, https://doi.org/10.5194/nhess-18-2161-2018, 2018.

FSO: Land use in Switzerland - Results of the Swiss land use statistics 2018 | Publication, Swiss Statistics, Federal Statistical Office, ISBN 978-3-303-02130-9, https://www.bfs.admin.ch/asset/en/19365054, 2021.

GDAL/OGR contributors: GDAL/OGR Geospatial Data Abstraction software Library, https://gdal.org, 2021.

620 Gupta, K., Satyam, N., and Segoni, S.: A comparative study of empirical and machine learning approaches for soil thickness mapping in the Joshimath region (India), CATENA, 241, 108 024, https://doi.org/10.1016/j.catena.2024.108024, 2024.

Guzzetti, F., Ardizzone, F., Cardinali, M., Galli, M., Reichenbach, P., and Rossi, M.: Distribution of landslides in the Upper Tiber River basin, central Italy, Geomorphology, 96, 105–122, https://doi.org/10.1016/j.geomorph.2007.07.015, 2008a.

Guzzetti, F., Peruccacci, S., Rossi, M., and Stark, C. P.: The rainfall intensity–duration control of shallow landslides and debris flows: an
625 update, Landslides, 5, 3–17, https://doi.org/10.1007/s10346-007-0112-1, 2008b.

Hastie, T. J. and Tibshirani, R. J.: Generalized Additive Models, vol. 43, CRC Press, 1990.

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,
630 https://doi.org/10.1371/journal.pone.0169748, 2017.

Hijmans, R. J.: terra: Spatial Data Analysis, https://rspatial.org/, 2023.

Ho, J.-Y., Lee, K. T., Chang, T.-C., Wang, Z.-Y., and Liao, Y.-H.: Influences of spatial distribution of soil thickness on shallow landslide prediction, Eng. Geol., 124, 38–46, https://doi.org/10.1016/j.enggeo.2011.09.013, 2012.

Horton, P., Jaboyedoff, M., Rudaz, B., and Zimmermann, M.: Flow-R, a model for susceptibility mapping of debris flows and other gravita-
635 tional hazards at a regional scale, Nat. Hazard Earth Sys., 13, 869–885, https://doi.org/10.5194/nhess-13-869-2013, 2013.

Huang, B. F. F. and Boutros, P. C.: The parameter sensitivity of random forests, BMC Bioinformatics, 17, 331, https://doi.org/10.1186/s12859-016-1228-x, 2016.

Huggett, R.: Regolith or soil? An ongoing debate, Geoderma, 432, 116 387, https://doi.org/10.1016/j.geoderma.2023.116387, 2023.

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

Hählen, N.: Kennzahlen zu spontanen Rutschungen im Kanton Bern mit Schwerpunkt auf Alpen und Voralpen, https://www.researchgate.net/publication/368510037, 2023.

Iida, T.: A stochastic hydro-geomorphological model for shallow landsliding due to rainstorm, CATENA, 34, 293–313, https://doi.org/10.1016/S0341-8162(98)00093-9, 1999.

645 Iverson, R. M.: Landslide triggering by rain infiltration, Water Resour. Res., 36, 1897–1910, https://doi.org/10.1029/2000WR900090, 2000.

Jaboyedoff, M., Carrea, D., Derron, M.-H., Oppikofer, T., Penna, I. M., and Rudaz, B.: A review of methods used to estimate initial landslide failure surface depths and volumes, Eng. Geol., 267, 105 478, https://doi.org/10.1016/j.enggeo.2020.105478, 2020.

Jalaian, B., Lee, M., and Russell, S.: Uncertain Context: Uncertainty Quantification in Machine Learning, AI Mag., 40, 40–49, https://doi.org/10.1609/aimag.v40i4.4812, number: 4, 2019.

650 James, G., Witten, D., Hastie, T., and Tibshirani, R.: An Introduction to Statistical Learning, Springer US, New York, NY, ISBN 978-1-07-161417-4, https://doi.org/10.1007/978-1-0716-1418-1, 2021.

Jia, N., Mitani, Y., Xie, M., and Djamaluddin, I.: Shallow landslide hazard assessment using a three-dimensional deterministic model in a mountainous area, Comput. Geotech., 45, 1–10, https://doi.org/10.1016/j.compgeo.2012.04.007, 2012.

Kaur, H., Gupta, S., Parkash, S., Thapa, R., Gupta, A., and Khanal, G. C.: Evaluation of landslide susceptibility in a hill city of Sikkim Himalaya with the perspective of hybrid modelling techniques, Ann. GIS, 25, 113–132, https://doi.org/10.1080/19475683.2019.1575906, 2019.

Kuhn, M.: Building Predictive Models in R Using the caret Package, J. Stat. Softw., 28, 1–26, https://doi.org/10.18637/jss.v028.i05, 2008.

Lanni, C., Borga, M., Rigon, R., and Tarolli, P.: Modelling shallow landslide susceptibility by means of a subsurface flow path connectivity index and estimates of soil depth spatial distribution, Hydrol. Earth Ssyst. Sc., 16, 3959–3971, https://doi.org/10.5194/hess-16-3959-2012, 2012.

Larsen, I. J., Montgomery, D. R., and Korup, O.: Landslide erosion controlled by hillslope material, Nat. Geosci., 3, 247–251, https://doi.org/10.1038/ngeo776, 2010.

Lateltin, O., Haemmig, C., Raetzo, H., and Bonnard, C.: Landslide risk management in Switzerland, Landslides, 2, 313–320, https://doi.org/10.1007/s10346-005-0018-8, 2005.

Leonarduzzi, E., Molnar, P., and McArdell, B. W.: Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data, Water Resour. Res., 53, 6612–6625, https://doi.org/10.1002/2017WR021044, 2017.

Li, W. C., Lee, L. M., Cai, H., Li, H. J., Dai, F. C., and Wang, M. L.: Combined roles of saturated permeability and rainfall characteristics on surficial failure of homogeneous soil slope, Eng. Geol., 153, 105–113, https://doi.org/10.1016/j.enggeo.2012.11.017, 2013.

Li, Y. and Mo, P.: A unified landslide classification system for loess slopes: A critical review, Geomorphology, 340, 67–83, https://doi.org/10.1016/j.geomorph.2019.04.020, 2019.

McColl, S. T. and Cook, S. J.: A universal size classification system for landslides, Landslides, 21, 111–120, https://doi.org/10.1007/s10346-023-02131-6, 2024.

Meier, C., Jaboyedoff, M., Derron, M.-H., and Gerber, C.: A method to assess the probability of thickness and volume estimates of small and shallow initial landslide ruptures based on surface area, Landslides, 17, 975–982, https://doi.org/10.1007/s10346-020-01347-0, 2020.

Meinshausen, N.: Quantile Regression Forests, J. Mach. Learn. Res., 7, 983–999, http://jmlr.org/papers/v7/meinshausen06a.html, 2006.

Meisina, C. and Scarabelli, S.: A comparative analysis of terrain stability models for predicting shallow landslides in colluvial soils, Geomorphology, 87, 207–223, https://doi.org/10.1016/j.geomorph.2006.03.039, 2007.

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., and Abderrahmane, B.: Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance, Earth-Sci. Rev., 207, 103 225, https://doi.org/10.1016/j.earscirev.2020.103225, 2020.

Milledge, D. G., Bellugi, D., McKean, J. A., Densmore, A. L., and Dietrich, W. E.: A multidimensional stability model for predicting shallow landslide size and shape across landscapes, J. Geophys. Res.-Eearth, 119, 2481–2504, https://doi.org/10.1002/2014JF003135, 2014.

Montgomery, D. R. and Dietrich, W. E.: A physically based model for the topographic control on shallow landsliding, Water Resour. Res., 30, 1153–1171, https://doi.org/10.1029/93WR02979, 1994.

Murgia, I., Giadrossich, F., Mao, Z., Cohen, D., Capra, G. F., and Schwarz, M.: Modeling shallow landslides and root reinforcement: A review, Ecol. Eng., 181, 106 671, https://doi.org/10.1016/j.ecoleng.2022.106671, 2022.

Pack, R., Tarboton, D., and Goodwin, C.: The SINMAP approach to terrain stability mapping. Procedeengs of the 8th congress of the international association of engineering geology, Vancouver, British Columbia, Canada, pp. 21–25, 1998.

Patton, N. R., Lohse, K. A., Godsey, S. E., Crosby, B. T., and Seyfried, M. S.: Predicting soil thickness on soil mantled hillslopes, Nat. Commun., 9, 3329, https://doi.org/10.1038/s41467-018-05743-y, 2018.

Pebesma, E. and Bivand, R.: Spatial Data Science: With Applications in R, Chapman and Hall/CRC, https://doi.org/10.1201/9780429459016, 2023.

Pfiffner, O. A.: The Geology of Switzerland, in: Landscapes and Landforms of Switzerland, edited by Reynard, E., pp. 7–30, Springer International Publishing, Cham, ISBN 978-3-030-43203-4, https://doi.org/10.1007/978-3-030-43203-4_2, 2021.

695 Piegari, E., Cataudella, V., Di Maio, R., Milano, L., and Nicodemi, M.: A cellular automaton for the factor of safety field in landslides modeling, Geophys. Res. Lett., 33, https://doi.org/10.1029/2005GL024759, 2006.

Planchon, O. and Darboux, F.: A fast, simple and versatile algorithm to fill the depressions of digital elevation models, CATENA, 46, 159–176, https://doi.org/10.1016/S0341-8162(01)00164-3, 2002.

Probst, P., Wright, M. N., and Boulesteix, A.-L.: Hyperparameters and tuning strategies for random forest, WIREs Data Min. Knowl., 9,
700 e1301, https://doi.org/10.1002/widm.1301, 2019.

R Core Team: R: A Language and Environment for Statistical Computing, https://www.R-project.org/, 2022.

Ran, Q., Hong, Y., Li, W., and Gao, J.: A modelling study of rainfall-induced shallow landslide mechanisms under different rainfall characteristics, J. Hydrol., 563, 790–801, https://doi.org/10.1016/j.jhydrol.2018.06.040, 2018.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F.: A review of statistically-based landslide susceptibility models,
705 Earth-Sci. Rev., 180, 60–91, https://doi.org/10.1016/j.earscirev.2018.03.001, 2018.

Reynard, E., Häuselmann, P., Jeannin, P.-Y., and Scapozza, C.: Geomorphological Landscapes in Switzerland, in: Landscapes and Landforms of Switzerland, edited by Reynard, E., pp. 71–80, Springer International Publishing, Cham, ISBN 978-3-030-43203-4, https://doi.org/10.1007/978-3-030-43203-4_5, 2021.

Rickli, C.: Personal communication, 2023.

710 Rickli, C. and Graf, F.: Effects of forests on shallow landslides - case studies in Switzerland, Forest Snow and Landscape Research, pp. 33–44, https://www.dora.lib4ri.ch/wsl/islandora/object/wsl%3A15351/, 2009.

Rickli, C., McArdell, B., Badoux, A., and Loup, B.: Database shallow landslides and hillslope debris flows, 13th congress INTER-PRAEVENT 2016. 30 May to 2 June 2016. Lucerne, Switzerland. Extended abstracts "Living with natural risks", pp. 242–243, https://www.dora.lib4ri.ch/wsl/islandora/object/wsl%3A20790/, 2016.

715 Rickli, C., Graf, F., Bebi, P., Bast, A., Loup, B., and McArdell, B.: Schützt der Wald vor Rutschungen? Hinweise aus der WSL-Rutschungsdatenbank, Schweizerische Zeitschrift für Forstwesen, 170, 310–317, https://doi.org/10.3188/szf.2019.0310, 2019.

Saulnier, G.-M., Beven, K., and Obled, C.: Including spatially variable effective soil depths in TOPMODEL, J. Hydrol., 202, 158–172, https://doi.org/10.1016/S0022-1694(97)00059-0, 1997.

Schaller, C.: HAFL-WWI/Landslide_Thickness_Prediction: Release for Predicting shallow landslide thickness using ML v0.1.4,
720 https://doi.org/10.5281/zenodo.14228477, 2024.

Schaller, C., Ginzler, C., van Loon, E., Moos, C., Seijmonsbergen, A. C., and Dorren, L.: Improving country-wide individual tree detection using local maxima methods based on statistically modeled forest structure information, Int. J. Appl. Eearth Obs., 123, 103 480, https://doi.org/10.1016/j.jag.2023.103480, 2023.

Schuster, R. and Wieczorek, G.: Landslide triggers and types, in: Landslides - Proceedings of the First European Confer-
725 ence on Landslides, Prague, Czech Republic, 24-26 June 2002, pp. 59–78, Routledge, London, ISBN 978-0-203-74919-7, https://doi.org/10.1201/9780203749197-4, 2018.

Schwarz, M., Preti, F., Giadrossich, F., Lehmann, P., and Or, D.: Quantifying the role of vegetation in slope stability: A case study in Tuscany (Italy), Ecol. Eng., 36, 285–291, https://doi.org/10.1016/j.ecoleng.2009.06.014, 2010.

Segoni, S., Rossi, G., and Catani, F.: Improving basin scale shallow landslide modelling using reliable soil thickness maps, Nat. Hazards, 61, 85–101, https://doi.org/10.1007/s11069-011-9770-3, 2012.

Shano, L., Raghuvanshi, T. K., and Meten, M.: Landslide susceptibility evaluation and hazard zonation techniques – a review, Geoenvironmental Disasters, 7, 18, https://doi.org/10.1186/s40677-020-00152-0, 2020.

Sidle, R. and Ochiai, H.: Landslides: Processes, Prediction, and Land Use, ISBN 978-0-87590-322-4, https://doi.org/10.1029/WM018, 2013.

Simmonds, E. G., Adjei, K. P., Andersen, C. W., Hetle Aspheim, J. C., Battistin, C., Bulso, N., Christensen, H. M., Cretois, B., Cubero, R., Davidovich, I. A., Dickel, L., Dunn, B., Dunn-Sigouin, E., Dyrstad, K., Einum, S., Giglio, D., Gjerløw, H., Godefroidt, A., González-Gil, R., Gonzalo Cogno, S., Große, F., Halloran, P., Jensen, M. F., Kennedy, J. J., Langsæther, P. E., Laverick, J. H., Lederberger, D., Li, C., Mandeville, E. G., Mandeville, C., Moe, E., Navarro Schröder, T., Nunan, D., Sicacha-Parada, J., Simpson, M. R., Skarstein, E. S., Spensberger, C., Stevens, R., Subramanian, A. C., Svendsen, L., Theisen, O. M., Watret, C., and O'Hara, R. B.: Insights into the quantification and reporting of model-related uncertainty across different disciplines, iScience, 25, 105 512, https://doi.org/10.1016/j.isci.2022.105512, 2022.

Skempton, A. and deLory, F.: Stability of natural slopes in London Clay, in: Proc. 4th Internal Conference on Soil Mechanics and Foundation Engng., London, 1957, vol. 15, pp. 378–381, Thomas Telford Publishing, London, UK, 1957.

Steger, S., Schmaltz, E., Seijmonsbergen, A. C., and Glade, T.: The Walgau: A Landscape Shaped by Landslides, in: Landscapes and Landforms of Austria, edited by Embleton-Hamann, C., pp. 237–251, Springer International Publishing, Cham, ISBN 978-3-030-92815-5, https://doi.org/10.1007/978-3-030-92815-5_15, 2022.

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, 2024.

Swisstopo: Geologie Gesteinsdichte, https://data.geo.admin.ch/ch.swisstopo.geologie-gesteinsdichte/, 2020.

Swisstopo: swissALTI3D, https://www.swisstopo.admin.ch/de/geodata/height/alti3d.html, 2023a.

Swisstopo: swissTLM3D, https://www.swisstopo.admin.ch/en/geodata/landscape/tlm3d.html, 2023b.

Swisstopo: Maps of Switzerland - Rock density, https://map.geo.admin.ch/#/map?lang=en¢er=2599631.84,1188720.65&z=1.549&bgLayer=ch.swisstopo.pixelkarte-grau&topic=ech&swisssearch=Rock+density&layers=ch.swisstopo.geologie-gesteinsdichte, 2024a.

Swisstopo: swissBOUNDARIES3D, https://www.swisstopo.admin.ch/en/landscape-model-swissboundaries3d, 2024b.

Swisstopo: WMTS-FSDI service, layer "Light base map relief", https://wmts.geo.admin.ch/EPSG/3857/1.0.0/WMTSCapabilities.xml?lang=de, 2024c.

Swisstopo: WMTS-FSDI service, layer "SWISSIMAGE Background", https://wmts.geo.admin.ch/EPSG/3857/1.0.0/WMTSCapabilities.xml?lang=de, 2024d.

Swisstopo, B. f. L.: GeoCover v2, https://www.swisstopo.admin.ch/en/geodata/geology/maps/geocover.html, 2023c.

van Zadelhoff, F. B., Albaba, A., Cohen, D., Phillips, C., Schaefli, B., Dorren, L., and Schwarz, M.: Introducing SlideforMAP: a probabilistic finite slope approach for modelling shallow-landslide probability in forested situations, Nat. Hazard Earth Sys., 22, 2611–2635, https://doi.org/10.5194/nhess-22-2611-2022, 2022.

Varnes, D. J.: Slope movement types and processes, Special report, 176, 11–33, 1978.

Wadoux, A. M. J. C., Minasny, B., and McBratney, A. B.: Machine learning for digital soil mapping: Applications, challenges and suggested solutions, Earth-Sci. Rev., 210, 103 359, https://doi.org/10.1016/j.earscirev.2020.103359, 2020.

Wager, S., Hastie, T., and Efron, B.: Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife, J. Mach. Learn. Res., 15, 1625–1651, http://jmlr.org/papers/v15/wager14a.html, 2014.

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

Waser, L., Ginzler, C., and Rehush, N.: Wall-to-Wall Tree Type Mapping from Countrywide Airborne Remote Sensing Surveys, Remote Sensing, 9, 766, https://doi.org/10.3390/rs9080766, 2017.

770 Watakabe, T. and Matsushi, Y.: Lithological controls on hydrological processes that trigger shallow landslides: Observations from granite and hornfels hillslopes in Hiroshima, Japan, CATENA, 180, 55–68, https://doi.org/10.1016/j.catena.2019.04.010, 2019.

Wood, S. N.: Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, J. Roy. Stat. Soc. B, 73, 3–36, https://doi.org/10.1111/j.1467-9868.2010.00749.x, 2011.

Wright, M. N. and Ziegler, A.: ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R, J. Stat. Softw.,
775 77, 1–17, https://doi.org/10.18637/jss.v077.i01, 2017.

WSL: Datenbank flachgründige Rutschungen und Hangmuren, https://www.wsl.ch/de/services-produkte/ datenbank-flachgruendige-rutschungen-und-hangmuren/, 2024.

WSL and BAFU: Dokumentation von Hangmuren und spontanen Rutschungen, https://hangmuren.wsl.ch/static/pdf/Anleitung_HMDB_V9. pdf, 2018.

780 Xiao, T., Segoni, S., Liang, X., Yin, K., and Casagli, N.: Generating soil thickness maps by means of geomorphological-empirical approach and random forest algorithm in Wanzhou County, Three Gorges Reservoir, Geosci. Front., 14, 101 514, https://doi.org/10.1016/j.gsf.2022.101514, 2023.

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.

785 Zhang, S., Xu, Q., and Zhang, Q.: Failure characteristics of gently inclined shallow landslides in Nanjiang, southwest of China, Eng. Geol., 217, 1–11, https://doi.org/10.1016/j.enggeo.2016.11.025, 2017.

Zimmermann, F., McArdell, B. W., Rickli, C., and Scheidl, C.: 2D Runout Modelling of Hillslope Debris Flows, Based on Well-Documented Events in Switzerland, Geosci., 10, 70, https://doi.org/10.3390/geosciences10020070, 2020.

Zweifel, L., Samarin, M., Meusburger, K., and Alewell, C.: Investigating causal factors of shallow landslides in grassland regions of Switzer-
790 land, Nat. Hazard Earth Sys., 21, 3421–3437, https://doi.org/10.5194/nhess-21-3421-2021, 2021.