Received: 10 March 2023 | Revised: 22 December 2023 | Accepted: 3 January 2024

DOI: 10.1002/esp.5770

RESEARCH ARTICLE

ESPL WILEY

Spatial transferability of the physically based model TRIGRS using parameter ensembles

Lotte de Vugt[1] | Thomas Zieher[2] | Barbara Schneider-Muntau[3] | Mateo Moreno[4,5] | Stefan Steger[4] | Martin Rutzinger[1]

[1]Institute of Geography, University of Innsbruck, Innsbruck, Austria

[2]Austrian Research Centre for Forests, Innsbruck, Austria

[3]Unit of Geotechnical Engineering, University of Innsbruck, Innsbruck, Austria

[4]Institute for Earth Observation, Eurac Research, Bolzano, Italy

[5]Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, the Netherlands

Correspondence
Lotte de Vugt, Institute of Geography, University of Innsbruck, Innrain 52f, 6020 Innsbruck, Austria.
Email: Lotte.De-Vugt@uibk.ac.at

Present address
Stefan Steger, GeoSphere Austria, Hohe Warte 38, Vienna, Austria.

Funding information
Autonomous Province of Bozen/Bolzano (Südtirol/Alto Adige), Grant/Award Number: D52F19000140003

Abstract

The development of better, more reliable and more efficient susceptibility assessments for shallow landslides is becoming increasingly important. Physically based models are well-suited for this, due to their high predictive capability. However, their demands for large, high-resolution and detailed input datasets make them very time-consuming and costly methods. This study investigates if a spatially transferable model calibration can be created with the use of parameter ensembles and with this alleviate the time-consuming calibration process of these methods. To investigate this, the study compares the calibration of the model TRIGRS in two different study areas. The first study area was taken from a previous study where the dynamic physically based model TRIGRS was calibrated for the Laternser valley in Vorarlberg, Austria. The calibrated parameter ensemble and its performance from this previous study are compared with a calibrated parameter ensemble of the model TRIGRS for the Passeier valley in South Tyrol, Italy. The comparison showed very similar model performance and large similarities in the calibrated geotechnical parameter values of the best model runs in both study areas. There is a subset of calibrated geotechnical parameter values that can be used successfully in both study areas and potentially other study areas with similar lithological characteristics. For the hydraulic parameters, the study did not find a transferable parameter subset. These parameters seem to be more sensitive to different soil types. Additionally, the results of the study also showed the importance of the inclusion of detailed information on the timing of landslide initiation in the calibration of the model.

KEYWORDS
HPC, shallow landslides, South Tyrol, spatial transferability, TRIGRS

1 | INTRODUCTION

Shallow landslides pose a worldwide threat to infrastructure and the population located in mountainous regions (Petley, 2012). With climate change, this threat will increase even further, and it is becoming increasingly important that reliable and efficient landslide susceptibility assessments are developed and applied in threatened areas (Alvioli et al., 2018; Gariano & Guzzetti, 2016). Physically based methods are well-suited for this, since their high predictive capability and process-based nature make them better suited than statistical approaches for modelling landslide susceptibility under different triggering scenarios (Chae et al., 2017; Schilirò et al., 2018).

Many different physically based slope stability models for shallow landslide susceptibility assessment have been developed over the years. These can be divided into (i) steady-state models (Dietrich & Montgomery, 1998; Pack et al., 1998) and (ii) dynamic models (Baum et al., 2010; Medina et al., 2021; Rossi et al., 2013; Simoni et al., 2008; van Beek, 2002; Wu & Sidle, 1995). Dynamic models differ from steady-state models due to their ability to assess the temporal susceptibility of a study area, such as the slope stability response to a

specific storm event. An issue that still remains with the use of physically based methods is their time consuming and costly nature, due to their requirement for large and detailed input datasets (Gariano & Guzzetti, 2016; Kuriakose et al., 2009; Raia et al., 2014). These high-resolution datasets are often insufficiently available in the areas requiring landslide susceptibility assessments (Kuriakose et al., 2009). Since soil properties can vary significantly within a study area (Herbst et al., 2006; Zhao et al., 2013), it is difficult to construct complete datasets even if samples from fieldwork are available. This means that in physically based modelling, the input datasets often have high uncertainties, which propagate in the model output.

Recent studies have tried to solve this problem with the implementation of parameter ensembles (i.e., ranges of calibrated parameter values) as input for physically based models, instead of single parameter combinations (e.g., de Lima Neves Seefelder et al., 2017; Medina et al., 2021; Park et al., 2013; Raia et al., 2014; Rossi et al., 2013; Zieher et al., 2017). This implementation aimed to improve the model performance and predictive capability of physically based approaches. The hypothesis was that using parameter ranges as input could partially account for the uncertainty in the parameter values and cover their natural variability, which is neglected in approaches using single parameter combinations. The studies showed a significant increase in model performance (i.e., the sensitivity and specificity) when parameter ensembles were used as input for physically based models (Raia et al., 2014).

The demonstrated capability of parameter ensembles to account for parameter variability raises the question if a parameter ensemble can also account for the variability of soil parameters over multiple study areas. This could be used to calibrate a parameter ensemble for a physically based model in one site and then transfer this calibration to another study site without loss of model performance, alleviating the time-consuming process of model calibration. Only a couple of studies using the model TRIGRS (Baum et al., 2008) focus on the values of the calibrated parameters (Ciurleo et al., 2019; de Lima Neves Seefelder et al., 2017; Zieher et al., 2017), and no studies could be found that compare the values of the calibrated parameters from different study areas. However, this comparison could give insight into if and how a spatially transferable parameter ensemble can be created.

In this study, the performance and the similarities in parameter values of two calibrated parameter ensembles for the model TRIGRS in two different study areas are compared with each other. To truly investigate the spatial transferability, the chosen study areas were taken from two different regions. However, to define the extent of the transferability, it was also important that these study areas had distinct similarities which also occur in other study areas. In addition to this, the study will make use of a homogeneous approach. This is done to ensure the applicability of the transferable parameter ensemble to data-poor study areas where insufficient data are available on the lithology or soil properties for the construction of distributed input data. In a previous study by Zieher et al. (2017), the model TRIGRS was calibrated with a large parameter space (e.g., 10.000 parameter value combinations) in the Laternser valley, Vorarlberg (Austria). A second study area was found in the Passeier valley, South Tyrol (Italy) which showed some distinct similarities with the Laternser valley. However, the areas also showed large differences which indicated that the calibrated ensemble from Zieher et al. (2017) cannot be used directly for the calibration of TRIGRS in the Passeier valley. The following research questions have been formulated to investigate if a spatially transferable parameter ensemble could be calibrated for both study areas:

- How well does the parameter space derived for the Laternser valley (Vorarlberg, Austria) perform in the Passeier valley (South Tyrol, Italy)?
- How does the resulting model calibration for the Passeier valley compare or differ from the calibration in the Laternser valley?
- Is there a subset of the parameter space that can be used in both study areas?

2 | STUDY AREA

Two study areas were used to investigate the spatial transferability of TRIGRS. The first study area was a subcatchment of the Passeier valley located in South Tyrol (Italy). The second study area was the Laternser valley located in Vorarlberg (Austria), which was investigated by Zieher et al. (2017). An overview of both study areas is given in Figures 1 and 2. The areas show similarities in their extensive lodgement till deposits and extensive forest cover (see also Figures 1 and 2). However, they also show distinct differences in the storm events that caused landslides in the catchments (see Figure 3) and their geological settings (Figures 1 and 2). For the Passeier valley, the surfacing lithologies, besides to the lodgment till deposits, are mainly metamorphic, while for the Laternser valley, these are mainly sedimentary deposits.

2.1 | Passeier valley

The studied 53 km^2 subbasin of the Passeier valley is located in the southern section of the Passeier valley (Vorderpasseier) and includes the southeast facing slope of the valley. It stretches 14 km in SSW to NNE direction from Meran to Sankt Leonhard. With regard to the topography, the elevation ranges from 350 to 2800 m (a.s.l.) and the slopes are generally steep with an average slope angle of 33°. In terms of land cover, 35% of the study area is covered by forest, which is composed of a mixture of fir (*Albies alba*) and spruce (*Picea Abies*) trees and oak (*Quercus petraea*) trees below 1000 m (a.s.l.) (Geokatalog, 2022). The remaining area is mainly used as either hay pastures or as meadows.

2.1.1 | Geology

Geologically, the Passeier valley is located in the lower central Austroalpine Nappes. The main geological units in the area are the Texel unit, with surfacing orthogneiss, paragneiss and micaschist lithologies, and glacial till deposits from the Last Glacial Maximum. Besides this, the other main surfacing lithologies consist of debris deposits and old landslide bodies, which have not been dated. Previous studies have not found a link between the lithological units and landslide susceptibility in South Tyrol (Amato et al., 2019). According to Piacentini et al. (2012), the occurrence of shallow landslides in South Tyrol is more closely linked to the soil type, where most shallow landslides occur on slopes with high-permeable soils overlaying low-permeable soils or layers.

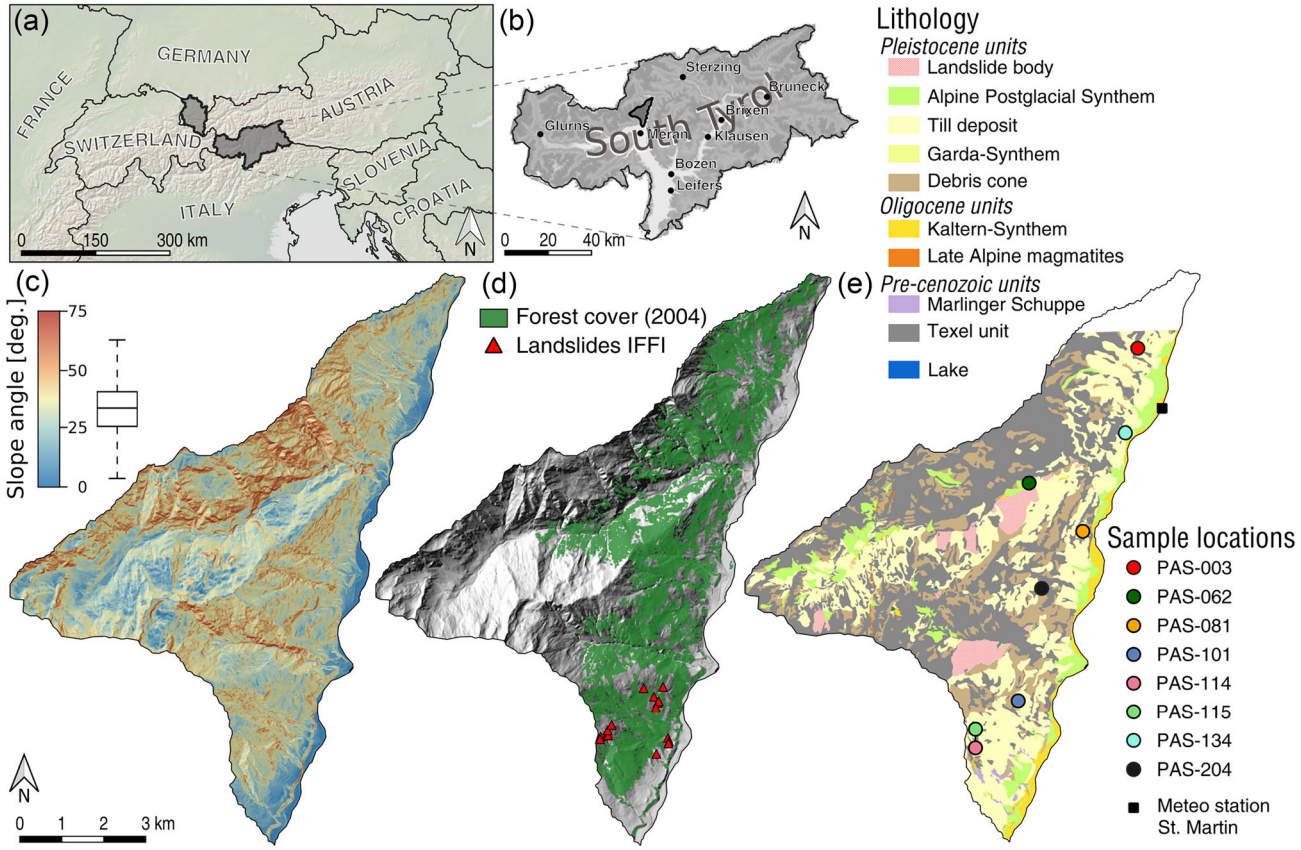

FIGURE 1 Overview of Passeier valley study area, with an overview of the location given in (a) and (b). The slope map of the study area is given in (c), with a boxplot of the slope angle distribution in the entire area. Panel (d) shows the used forest cover map (derived from airborne laser scanning (ALS) data from 2004) and the landslides recorded on 05.08.2016 in the Italian national landslide inventory (IFFI: Inventario dei Fenomeni Franosi in Italia). Panel (e) shows the geological map of the study area with an overlay of the sample locations of this study and the location of the meteorological station in St. Martin.

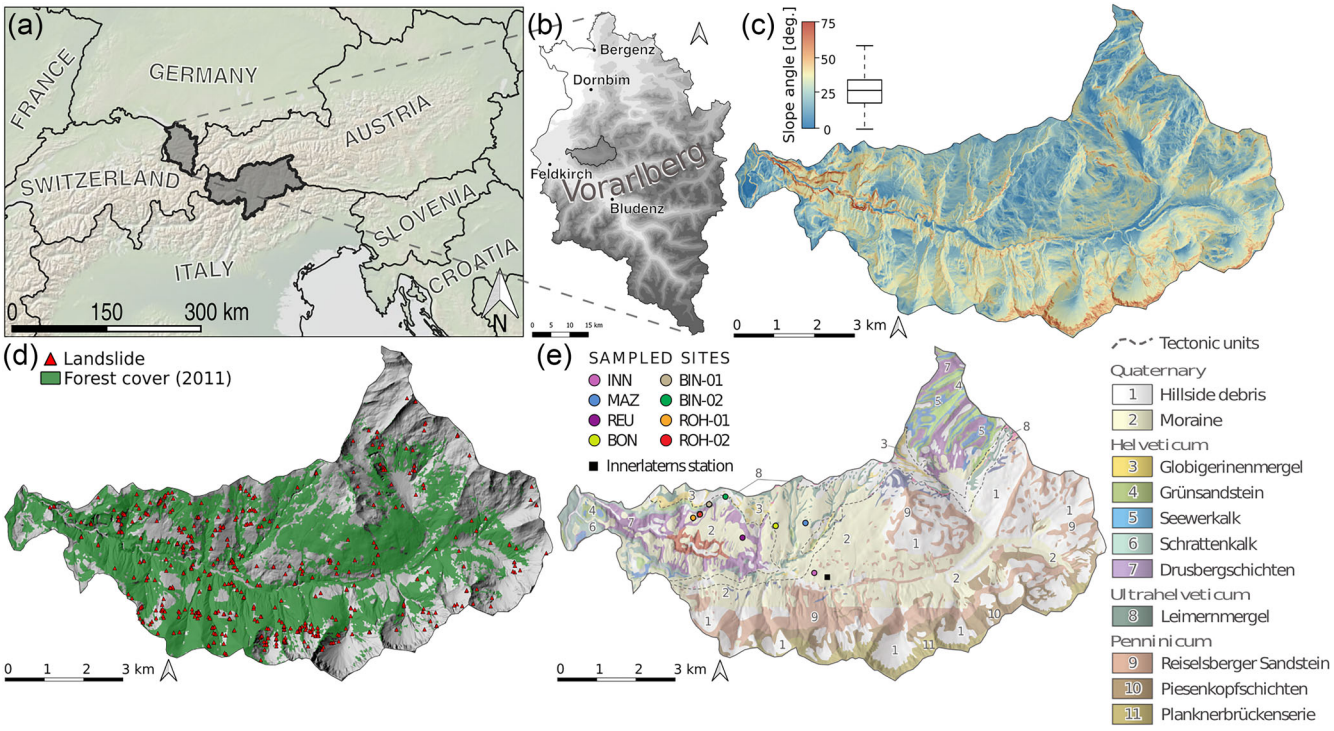

FIGURE 2 Overview of Laternser valley study area (data taken from Zieher *et al.* (2017), Zieher *et al.* (2016)), with an overview of the location given in (a) and (b). The slope map of the study area is given in (c), with a boxplot of the slope angle distribution in the entire area. Panel (d) shows the used forest cover map (from 2011) and the landslides mapped by Zieher *et al.* (2016) for the August 2005 event. Panel (e) shows the geological map of the study area with an overlay of the sample locations (map taken from Zieher *et al.* (2017)).

2.1.2 | Climate and landslide triggering events

The region has a dry marine climate with a cumulative yearly precipitation of 774 mm. The peaks in precipitation occur in summer and later in November. The investigated precipitation event occurred on 5 August 2016 (see Figure 3b). The event was an intense rainfall event that lasted 15 h, where most of the precipitation (accumulating to 120 mm) fell in the first 5 h. The maximum precipitation intensity occurred at 06:00 AM CEST, with an hourly intensity of 36 mm·h^{-1}. Figure 3a also shows that there was no significant precipitation registered at the St. Martin meteorological station in the month prior to the event. However, climate reports show that the months June and July in 2016 were on average wetter months than the climatological mean (1981–2010) (Munari, Peterlin, Tollardo, Geier, & Tartarotti, 2016; Munari, Peterlin, Tollardo, Geier, Tartarotti, & Rastner, 2016) (see also Figure 3a). A further analysis of the full meteorological record since 1981 showed that the event of August 2016 was the second most extreme event (in terms of cumulative daily precipitation) recorded by the meteorological station in St. Martin. The Italian national landslide inventory (IFFI: Inventario dei Fenomeni Franosi in Italia) registered 16 landslides to the event of August 2016. The location of these landslides is shown in Figure 1d.

2.2 | Laternser valley

The Laternser valley, located in Vorarlberg (Austria), covers the catchment of the river Frutz. The valley lies in an east to west direction and is about 52 km^2. The elevation ranges from 500 to 2000 m (a.s.l.). Like the Passeier valley, the slopes are generally steep with a median slope angle of 27°. In terms of land cover, approximately 50% of the area is covered by forest, and similarly, to the Passeier valley, the remaining area is mainly used as hay pastures or meadows. The forest in the Laternser is composed of fir (*Abies alba*) and spruce (*Picea abies*) trees, with beech (*Fagus sylvatica*) trees occurring below 1300 m (a.s.l.) (Zieher et al., 2017).

2.2.1 | Geology

The geological setting includes several different tectonic nappes. In the western and northern parts of the valley lie the Helvetic nappes with limestone lithologies (e.g. Schrattenkalk, Seewerkalk). These lithologies are superimposed to the south-east by the Ultrahelvetic nappes with clayey marls and shales. In the south eastern part of the area lie the Penninic nappes, which mainly consist of sandstone and thinly layered marls. The sandstone from the Penninic nappes together with the Ultrahelvetic marls are known for their landslide susceptibility (Zieher et al., 2016). Similar to the Passeier valley, the Laternser valley has an extensive cover (more than 57%) of till and hillside debris deposits. Due to their often very high compaction values, these deposits are very susceptible to forming impermeable layers within the deposit and can act as the slip surface for the overlying material (Zieher et al., 2016; Zieher et al., 2017). It is expected that a similar high landslide susceptibility of the lodgement till deposits can also be found in the Passeier valley.

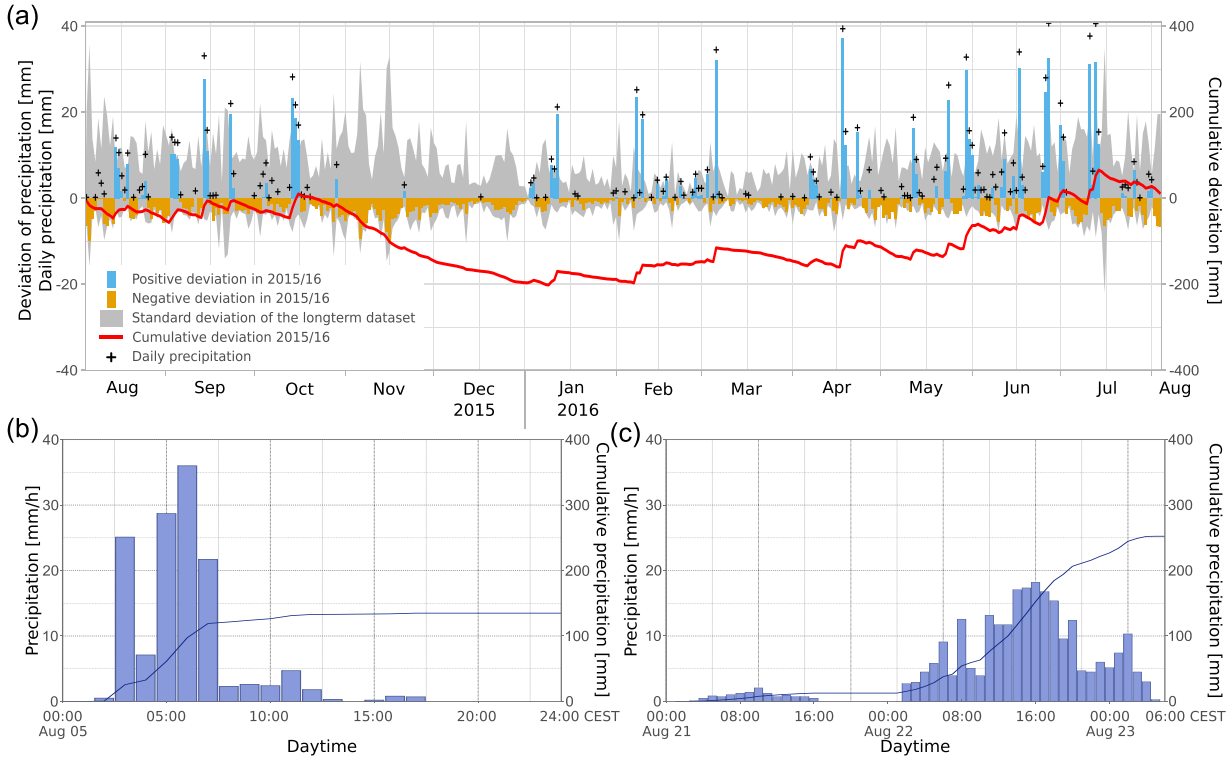

FIGURE 3 Daily precipitation registered at the St. Martin, Passeier valley Meteorological station and its deviation from the long-term mean (1990–2020) for the period of 1 year before the studied event of 5 August 2016 (a). The hourly precipitation intensity of the studied event is given in (b). Panel (c) shows the precipitation event from the Laternser valley on the 21–23 August 2005 (data taken from Zieher *et al.* (2017)).

2.2.2 | Climate and landslide triggering events

The climate of the Laternser valley is warm and temperate. At the Innerlaterns station, the reported annual precipitation is more than 1700 mm. The investigated landslide triggering event occurred from 21–23 August 2005. A graph of the precipitation intensity during this event is shown in Figure 3c. This event was much longer and less intense than the event in the Passeier valley. The maximum hourly precipitation intensity occurred after 40 h with 18 $mm \cdot h^{-1}$. The cumulative precipitation was 252 mm of rainfall after 54 h. There were 356 landslides found to be related to this event (Zieher et al., 2016). Their locations are given in Figure 2d.

3 | MATERIALS AND METHODS

An overview of the approach used in this study is given in Figure 4. A large parameter space, consisting of 10 000 unique parameter value combinations was used to calibrate the model TRIGRS (Baum et al., 2008) in the two study areas. To ensure comparability with the existing calibration for the Laternser valley, the calibration workflow with area-wide homogeneous parameter values developed by Zieher et al. (2017) was reproduced in the Passeier valley. The resulting 10 000 model runs were tested with the landslide inventories of the events that occurred in the two study areas and a ruleset developed by Zieher et al. (2017) to include the predicted timing of landslide initiation in the parameter calibration process. In this calibration, the 10 000 parameter combinations were filtered down to the 25 best performing parameter combinations based on the specificity and sensitivity of their predictions. The 25 best performing parameter combinations from both study areas, also defined as the calibrated parameter ensembles in this study, were then compared with each other in terms of their corresponding model performance and the parameter values occurring within the parameter ensembles. The goal of this comparison was to find an overlapping subset of the parameter space that performs well in both study areas.

3.1 | Landslide inventory

Multispectral satellite remote sensing scenes, that is, imagery by PlanetScope (3 m resolution) and RapidEye (5 m resolution), were used to investigate the event of 5 August 2016. The analysis was done visually with the before and after imagery taken from 05/07/2016 till 25/08/2016 (Planet Team, 2022). Since no elevation data from after the event and no stereographic information were available, the visual mapping method mainly focussed on the change in photographic colour, mottling texture, shape, and size of a signature to determine if the signature originated from a landslide (Guzzetti et al., 2012). It was assumed that this method could identify all landslides in open terrain with a full signature (i.e. both the scarp and depositional area) of more than 81 m^2. In total, 55 landslides were mapped on the Planet imagery. An overview of the mapped landslides and a more detailed subset are given in Figure 5a,c.

The inventory was constructed by outlining the full signature seen on the satellite imagery. However, the full signature also includes the runout area of the landslide. Since physically based models only predict the actual failure surface, the inclusion of the runout area in the validation dataset will lead to overestimation of the predicted failure areas in the final calibration. For this reason, instead, only the

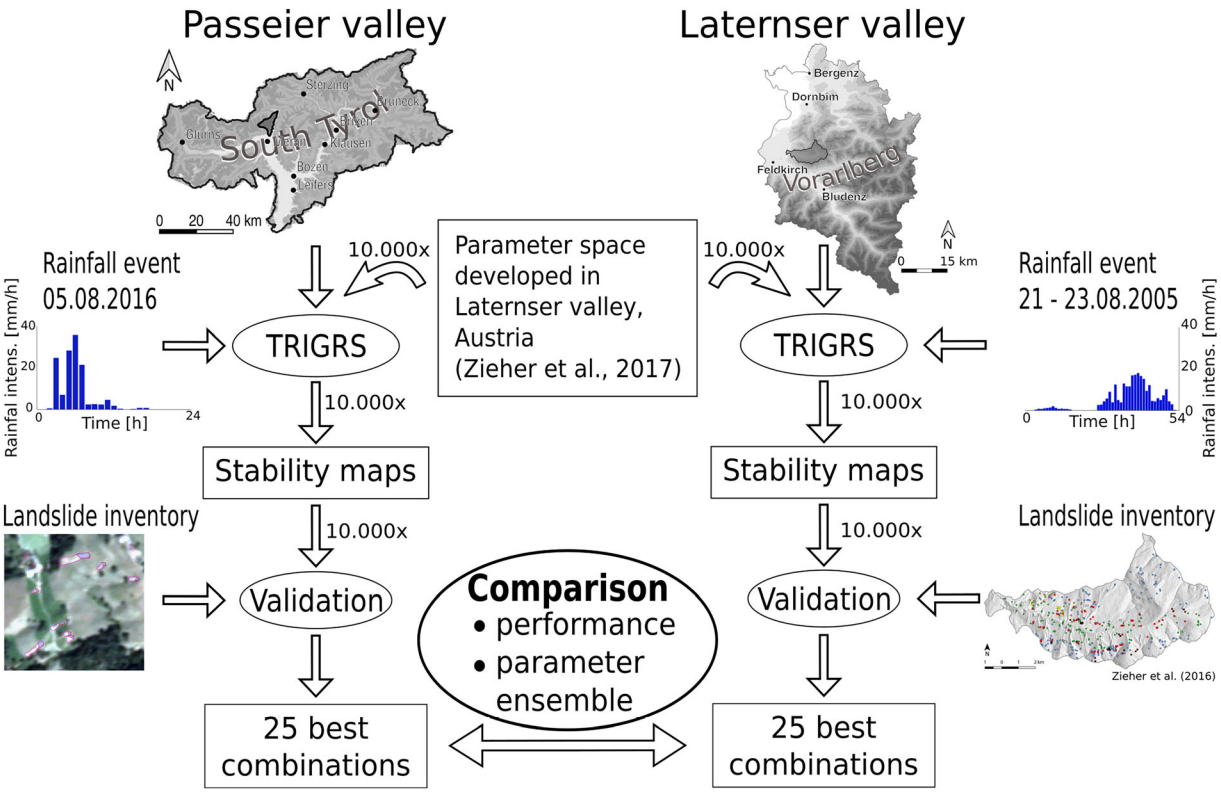

FIGURE 4 Flow diagram of the used workflow.

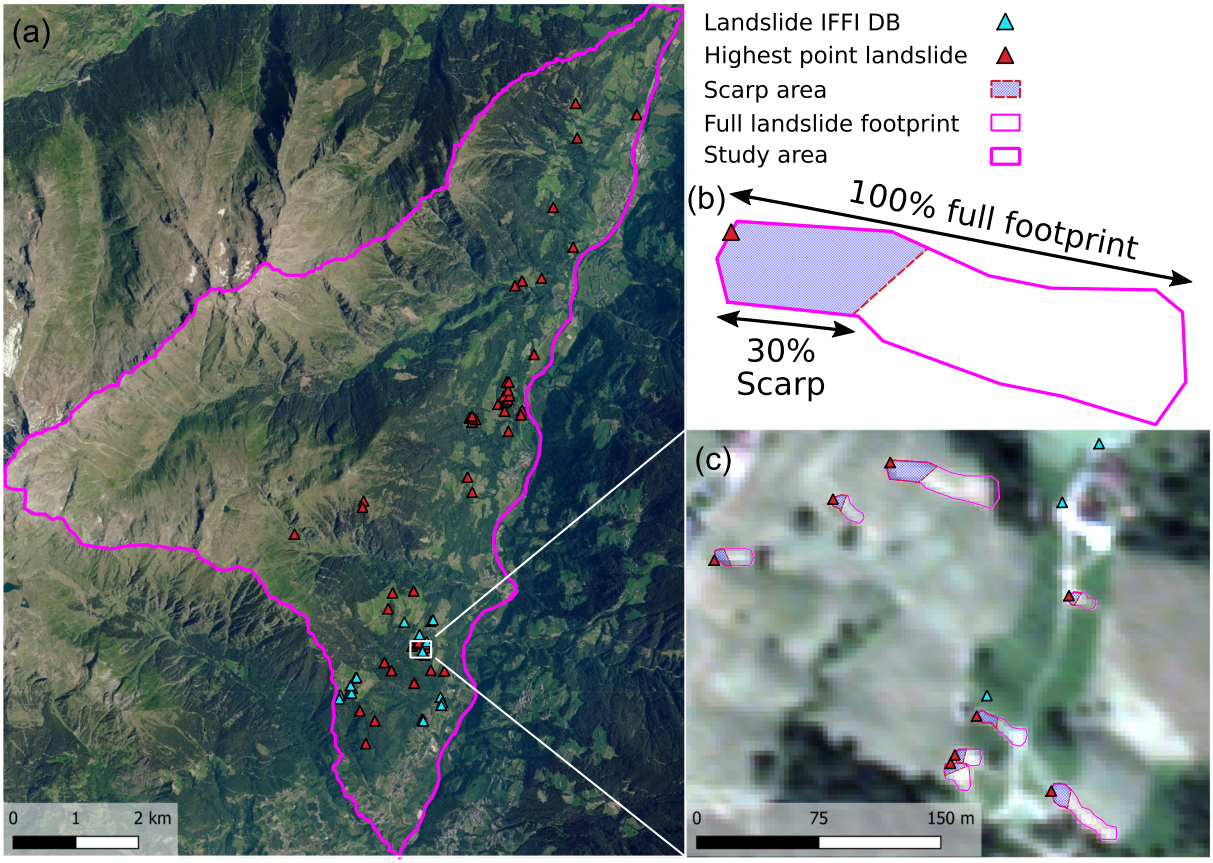

FIGURE 5 Comparison of the mapped landslides on Planet imagery and the original landslides from the Inventario dei Fenomeni Franosi in Italia (IFFI) database (a) (background orthophoto from Geokatalog (2022)), a schematic description of the scarp selection (b) and a detailed view of the mapped landslide with the used Planetscope scene from 25/08/2016 as background (c).

upper 30% of the full footprint was selected, which was assumed to represent the area of failure initiation. A schematic representation of this method is given in Figure 5b. Since the landslides from the IFFI database were mapped as points and contained no information on the landslide scar area, only the manually mapped landslides on the satellite imagery were used in the validation of the model runs.

3.2 | TRIGRS model

The model TRIGRS 2.0 (Transient Rainfall Infiltration Grid-based Regional Slope stability) was developed by Baum et al. (2008). A newer, parallelised version of the model (TRIGRS 2.1) was developed by Alvioli and Baum (2016). However, the study by Zieher et al. (2017) implemented several functions to the TRIGRS 2.0 model that added root cohesion and tree surcharge dynamics. This extended version was also used to model the slope stability in the Passeier valley, to ensure that the results could be compared with the calibration in the Laternser valley.

The original model (TRIGRS 2.0) consists of two components that are combined to model the slope stability for a given precipitation event. First, the model uses a simplified version of Iverson's one-dimensional linearized diffusion model (Iverson, 2000) to model pore pressure as response to infiltration for a specified set of timesteps during the precipitation event. With this, infiltration is modelled in the vertical direction and the resulting pore pressure values are calculated at different depth increments within the soil column of each grid cell. Slope parallel sub-surface flow is not included in the model, since it is considered of minor importance for slope infiltration at shorter timescales (Berti & Simoni, 2010; Iverson, 2000). In the second component, the output of pore pressure values ($\psi(d,t)$) is used as part of the input for the geotechnical component of the model, which uses the infinite slope method (Equation (1)) to model the slope stability for the supplied depth increments in each grid cell. In the final output of the specified timesteps, only the min*FOS* (minimum Factor Of Safety) found in the soil column is written to the output map. Grid cells where the min*FOS* < 1 in one of the timesteps are assumed to have failed during the modelled event and grid cells where the min*FOS* ≥ 1 are assumed to have remained stable during the modelled event.

$$FOS(d,t) = \frac{\tan\varphi'}{\tan\beta} + \frac{c - \psi(d,t)\cdot\gamma_w\cdot\tan\varphi'}{(\gamma_s\cdot d)\cdot\sin\beta\cdot\cos\beta}, \quad (1)$$

where d is the depth increment in the soil column (m), t is the time (s), φ' is the angle of internal friction (deg), β is the slope angle (deg), c' is the cohesion (Pa), γ_w is the unit weight of water (N·m^{-3}), and γ_s the unit weight of soil (N·m^{-3}).

The extended version developed by Zieher et al. (2017) includes root cohesion (c_r) and tree surcharge (s_t) components by integrating the effects directly in the infinite slope method using Equation (2). A parameter for the rooting depth is used to linearly decrease the root cohesion over depth until the rooting depth is reached. If the rooting depth exceeds the regolith depth, the effect of root cohesion is cut-off at the bedrock boundary.

$$FOS(d,t) = \frac{\tan\varphi'}{\tan\beta} + \frac{c' + c_r - \psi(d,t)\cdot\gamma_w\cdot\tan\varphi'}{(s_t + \gamma_s\cdot d)\cdot\sin\beta\cdot\cos\beta} \quad (2)$$

The initial model conditions were set to those used for the calibration in the Laternser valley (Zieher et al., 2017). With these settings, the hydrological model assumes initially wet conditions, a steady background infiltration rate of 0.01 times the used saturated hydraulic conductivity and the initial depth of the water table was assumed to be located at the bedrock boundary. These initial conditions were assumed to be valid for the August 2015 event in the Passeier valley, since the pre-event precipitation showed wetter than average conditions for the 2 months leading up to the event (see also Figure 3a). The values for the root cohesion (2.5 kPa), rooting depth (1 m) and tree surcharge (2.5 kPa) were also taken from Zieher et al. (2017). These values were chosen based on the high occurrence of fir and spruce trees in both the Laternser valley and the Passeier valley.

3.3 | Model input parameterization

Table 1 gives an overview of the required model input for the extended version of TRIGRS. The topographic model input was prepared for the Passeier valley like for the Laternser valley in the study by Zieher et al. (2017). Publicly available airborne laser scanning (ALS) point clouds covering the province of South Tyrol (Geokatalog, 2022) were automatically classified into ground and nonground points using the algorithm proposed by Axelsson (2000). Based on the classified ground points, a digital terrain model (DTM) with a spatial resolution of 1 m was computed, aggregating the mean elevation per raster cell. A corresponding digital surface model (DSM) was computed by aggregating the maximum elevation considering all points. Based on the normalised DSM (DSM-DTM), a forest cover map was computed for the Passeier valley similar as for the Laternser valley. For the model runs, the topographic data and the forest cover map were resampled to a spatial resolution of 10 m which was previously considered feasible for reproducing shallow landslides (Milledge et al., 2012; Zieher et al., 2016; Zieher et al., 2017).

The precipitation intensity data were provided by the meteorological office of South Tyrol (Autonomous Province of Bolzano, 2016), at a 5 min temporal resolution from the meteorological station located in St. Martin in Passeier (46.7824°N, 11.2296°E) (see also Figure 1e). These data were aggregated to a 1 h temporal resolution. Since no precipitation was reported after 12:00 in the afternoon, only the first 12 h of 5 August was modelled. A graph of the registered precipitation is given in Figure 3b.

The regolith depth was modelled using Equation (3). This equation was statistically derived from the relationship of slope values (β) (°) and regolith depth (d_{max}) (m) measurements from DCPTs (dynamic cone penetration tests) by Zieher et al. (2017) for the Laternser valley. This specific model was chosen, since it was in terms of its location the closest validated regolith depth model to the Passeier valley, and no information about the regolith depth in the Passeier valley was available.

$$d_{max} = \begin{cases} 3.028 - 0.049\beta & \text{for } 0.0^\circ \le \beta < 61.8^\circ \\ 0.0 & \text{for } \beta \ge 61.8^\circ \end{cases} \quad (3)$$

To validate the used parameter space and resulting best fitting subset, the study also collected several field samples to construct a ground truths dataset. The samples were collected at eight field locations, as shown in Figure 1e. The sample locations were chosen based on their proximity to the landslides that occurred during the event

TABLE 1 List of required input and the used sources.

	Source	Range or constant	Unit
Topographic data			
Elevation model	ALS data (2004)	370–2840	m
Flow direction	DTM from ALS data (2004)	(-)	(-)
Slope	DTM from ALS data (2004)	0–74	°
General soil parameters			
Regolith depth	Stat. model Zieher et al. (2017)	0 - 3.028	m
Background infiltration rate	Zieher et al. (2017)	0.01 x K_s	$m{\cdot}s^{-1}$
Dry bulk density soil	Zieher et al. (2017)	1.84	$gr{\cdot}cm^{-3}$
Geotechnical parameters			
Cohesion	Parameter space/Triaxial shear test	0 - 18	kPa
Internal friction angle	Parameter space/Triaxial shear test	21 - 39	°
Hydrological parameters			
Precipitation intensity	Meteorological station	0 - 36	$mm{\cdot}h^{-1}$
Saturated hydraulic conductivity	Parameter space	1.0×10^{-6} - 1.0×10^{-3}	$m{\cdot}s^{-1}$
Hydraulic diffusivity	Parameter space/Oedometer test	1.0×10^{-5} - 1.0	$m^2{\cdot}s^{-1}$
Groundwater depth	Zieher et al. (2017)	Set to regolith depth	m
Additional parameters			
Root cohesion	Zieher et al. (2017)	2.5	kPa
Tree surcharge	Zieher et al. (2017)	2.5	kPa
Rooting depth	Zieher et al. (2017)	1	m

and the goal to capture varying lithologies. The final sample locations cover the lodgement till, the Texel unit and the debris cone deposits. At the sample locations, disturbed soil samples were collected from 1-m-deep soil pits. This material was analysed in the laboratory to determine the grain size distribution with the use of sieving and sedimentation methods and the shear strength of the soil (cohesion and friction angle) with consolidated drained triaxial shear tests on reconstructed samples.

The model also requires information about the hydraulic diffusivity (D_0). However, measuring the diffusivity in the lab or in the field is very difficult and comes with high uncertainties (Berti & Simoni, 2010). Instead, the hydraulic diffusivity can be determined using the specific storage (S_s) and the saturated hydraulic conductivity (K_{sat}) using Equation (4).

$$D_0 = \frac{K_{sat}}{S_s} \tag{4}$$

The specific storage is easier to measure in the laboratory and can be determined from the soil compressibility using Equation (5).

$$S_s = \rho_w \cdot g \cdot (\alpha_s + n \cdot \beta_w), \tag{5}$$

where ρ_w is the density of water (kg·m^{-3}), g is the gravitational acceleration (9.81 m·s^{-2}), n is the soil porosity (-), β_w is the compressibility of water (4.4 $\times 10^{-10}$ m^2·N^{-1}) and α_s is the compressibility of the soil (m^2·N^{-1}) (Rowe & Barden, 1966). The disturbed samples from the soil pits were therefore also analysed with oedometer tests to determine their compressibility (Berti & Simoni, 2010; Zieher et al., 2017). Determining the saturated hydraulic conductivity in the laboratory comes with similar uncertainty issues as the hydraulic diffusivity (Berti & Simoni, 2010) and was therefore not measured in this study but calibrated. However, the calibration of the saturated hydraulic conductivity was tested against the predictions of the saturated hydraulic conductivity with several pedotransfer functions from the "LWFBrook90R" R library package (Puhlmann & von Wilpert, 2012; Schmidt-Walter et al., 2023; Wösten et al., 1999) and the Rosetta3 model (Zhang & Schaap, 2017).

Besides the disturbed samples from the soil pits, several ring samples were also taken at different depths in the back-wall of the soil pits (at depths of 50 to 100 cm below the surface). These samples were used to determine the dry bulk density and soil moisture content of the soil using the oven-drying method.

3.4 | Parameter space and calibration methods

The parameter space varies the internal friction angle, soil cohesion, saturated hydraulic conductivity and specific storage. These parameters were chosen to be included in the space after a sensitivity analysis of TRIGRS in the Laternser valley showed that these parameters were the most sensitive parameters in the model. The ranges of the parameter values were determined using ranges described in literature for the soils occurring in the Laternser valley and after analysis of soil samples taken in the valley (Zieher et al., 2017). The used values are given in Table 2.

To test the model runs of the full parameter space and to select the best fitting model runs, Zieher et al. (2017) developed the following ruleset to validate the timing of landslide initiation in the model runs:

1. The initial conditions at the start of the model run should result in predicted stable slopes ($FOS \geq 1.0$).
2. Most shallow landslides should be triggered after the highest precipitation intensity occurred ($FOS < 1.0$).
3. The 25 best fitting model runs are selected by taking the 25 model runs with the lowest Distance to the Perfect Classification (D2PC) values.

Formetta et al. (2016) showed that the D2PC is a good indicator for the performance of a model in the ROC (receiver operator curve) space. The D2PC was, therefore, used as main performance indicator in both study areas. The D2PC gives equal weight to the true positive rate (TPR) and the true negative rate (TNR) and is calculated using Equation (6).

$$D2PC = \sqrt{(1 - TNR)^2 + (1 - TPR)^2} \tag{6}$$

To further investigate the similarities between the best fitting parameter combinations for the Passeier valley and the Laternser valley, the study also looked at the 100 best fitting combinations for each study area.

4 | RESULTS

4.1 | Results of the soil sample analysis

The analysis of the soil samples comprised determining the grain size distribution, the shear strength, the soil compressibility with oedometer tests, the dry bulk density and soil moisture content. Figure 6 shows the results of the grain size distribution analysis and the oedometer tests. The graph from the grain size distribution analysis (Figure 6a) shows that the variance in soil type in the study area is low. The samples are all well-distributed and show a typical grain size distribution for a lodgement till soil (Bell, 2002). The results of the grain size distribution were used with the found dry bulk density values of the soil samples to predict the saturated hydraulic conductivity with several pedotransfer functions. The results of this analysis are given in Figure 7. The graph in Figure 6b shows the measured compressibilities of the soil samples with the oedometer tests. These compressibility values were used with Equation (5) to determine the specific storage of the soil. When the compressibility of the soil samples is extrapolated to a lower pore pressure value of 2 kPa that naturally occurs in an aquifer (i.e. a fully saturated soil) at 2 m depth (Berti & Simoni, 2010), the estimated specific storage ranges from 0.003 to 0.012 m^{-1}. The results from the triaxial shear tests on the soil samples are given in Table 3. The cohesion values of the soil samples range between 0 and 1.5 kPa, while the internal friction angles of the samples range between 31.5° and 35°. Table 3 also lists the values found for the dry bulk density and soil moisture content of the soil samples.

TABLE 2 Parameter values used in the full parameter space.

Cohesion	(kPa)	0.0	2.0	4.0	6.0	8.0	10.0	12.0	14.0	16.0	18.0
Internal friction angle	(°)	21	23	25	27	29	31	33	35	37	39
Sat. hydraulic conductivity	($m \cdot s^{-1}$)	1.0×10^{-6}	2.2×10^{-6}	4.6×10^{-6}	1.0×10^{-5}	2.2×10^{-5}	4.6×10^{-5}	1.0×10^{-4}	2.2×10^{-4}	4.6×10^{-4}	1.0×10^{-3}
Specific storage	(m^{-1})	1.0×10^{-3}	1.7×10^{-3}	2.8×10^{-3}	4.6×10^{-3}	7.7×10^{-3}	1.3×10^{-2}	2.2×10^{-2}	3.6×10^{-2}	6.0×10^{-2}	1.0×10^{-1}

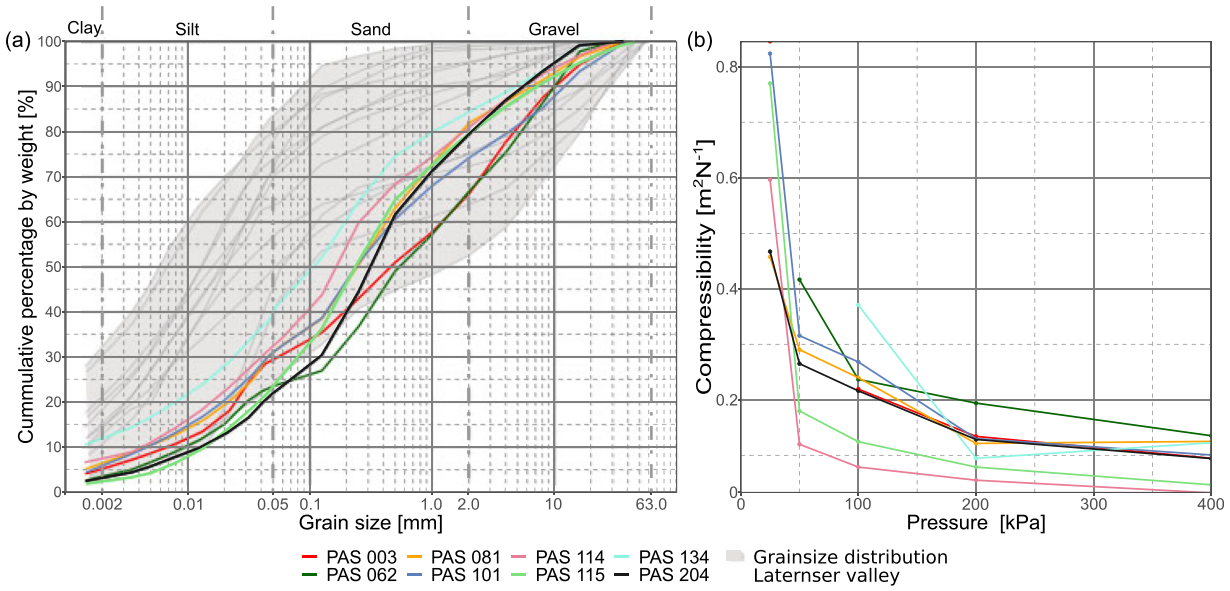

FIGURE 6 Grainsize analysis and oedometer results of the soil sample analysis in the Passeier valley. The grain size distribution of the samples is given in (a) with the grain size distributions found in the Laternser valley in grey. The found compressibility with the oedometer tests is given in (b).

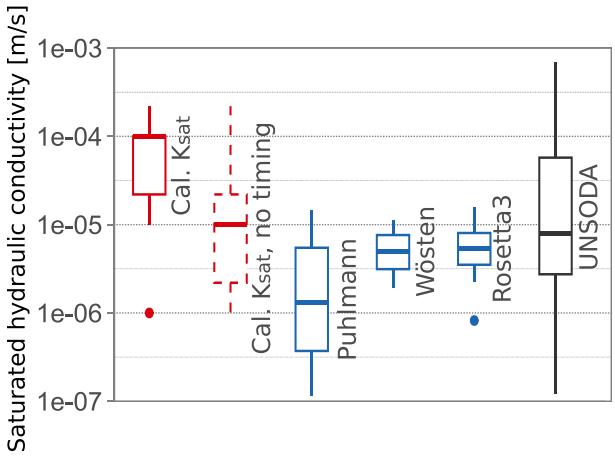

FIGURE 7 Distribution of the saturated hydraulic conductivity in the 25 best performing parameter combinations (in red) and the resulting distribution when timing is not enforced (in dotted red). The blue boxplots show the distributions derived with the pedotransfer functions of the Puhlmann (Puhlmann & von Wilpert, 2012), the Wösten (Wösten *et al.*, 1999) and the Rossetta3 (Zhang & Schaap, 2017) models. The black boxplot shows the distribution for sandy loam soils from the UNSODA soil database (Nemes *et al.*, 2001).

4.2 | Performance of the full parameter space

Figure 8 shows the performance of all parameter combinations in the parameter space and the performance of the remaining parameter combinations after each consecutive filtering rule (see Section 3.4). After the first filtering rule, where all model runs predicting unstable conditions at the first timestep are filtered out, only 7900 parameter combinations remain (Figure 8b). With the second filtering rule, where all remaining model runs where most of the landslides occur before the peak of the precipitation event are filtered out, 7175 parameter combinations remain (Figure 8c).

The performance of the 25 best fitting parameter combinations is good, as is shown by their position in the ROC space (Figure 8d) and their D2PC values (Table 4). A boxplot of the distribution of the D2PC values within the 25 best fitting combinations is given in Figure 11c. The best model run has a D2PC value of 0.37, with a TPR of 71.9% and a TNR of 77.6%. Table 4 shows that the maximum TPR decreases with the consecutive filtering rules, while the minimum TNR increases. The minimum D2PC remains constant.

TABLE 3 Found parameter values from the laboratory tests on the soil samples from the Passeier valley.

Parameter	Unit	PAS 003	PAS 062	PAS 081	PAS 101	PAS 114	PAS 115	PAS 134	PAS 204
Soil moisture	-	0.11	-[a]-[a]	0.21	-[a]	0.07	0.03	0.18	0.11
Dry bulk density	$gr \cdot cm^{-3}$	1.16	-[a]	1.66	-[a]	1.46	1.50	1.71	1.25
Cohesion	kPa	0.0	0.0	0.0	1.5	1.0	0.0	0.0	0.0
Internal friction angle	°	32.5	34.5	35	32.5	31.5	32.5	31.5	33.5
Compressibility (As)	$10^{-4}\ m^2 \cdot kN^{-1}$	8.62	12.22	9.96	9.90	3.45	5.13	10.72	8.87
Specific storage (Ss)	$10^{-3}\ m^{-1}$	8.46	12.01	9.78	9.72	3.39	5.03	10.52	8.71

[a]No ring samples were taken at this location.

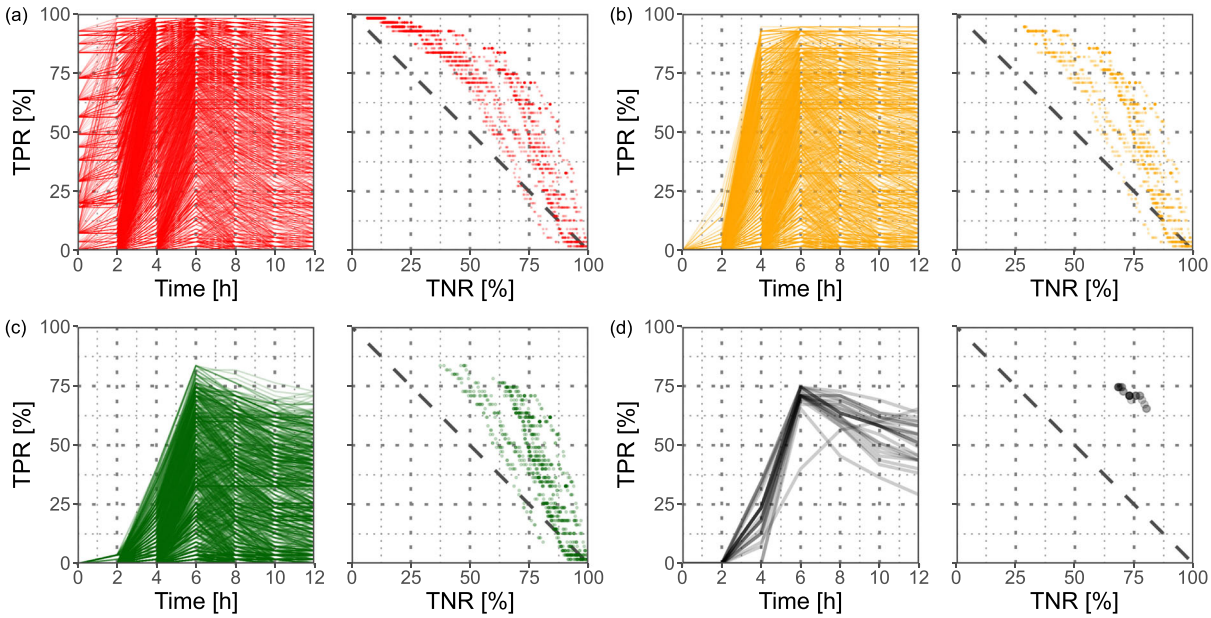

FIGURE 8 Performance of the model runs in the parameter space over time and in the receiver operator curve (ROC) space for the Passeier valley, where (a) shows the performance of all parameter combinations, (b) shows the parameter combinations remaining after filtering rule 1 (see Section 3.4), (c) shows the parameter combinations remaining after filtering rule 2 (see Section 3.4) and (d) shows the 25 best performing parameter combinations out of the model runs in (c).

TABLE 4 Performance metrics for the model runs in the Passeier valley of the timestep with the maximum predicted failure in each model run.

	Full ensemble		Stable at $t = 0$		Peak TPR after $t = 6$ h		25 best combinations	
	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
TPR	0.0%	98.2%	0.0%	94.5%	0.0%	83.6%	65.5%	74.5%
TNR	6.6%	100%	28.6%	100%	37.4%	100%	68.2%	80.4%
FPR	0.0%	93.4%	0.0%	71.4%	0.0%	62.6%	19.6%	31.8%
FNR	1.8%	100%	5.5%	100%	16.4%	100%	25.5%	34.5%
D2PC	0.37	1.00	0.37	1.00	0.37	1.00	0.37	0.41

Abbreviations: D2PC, Distance to the Perfect Classification; TNR, true negative rate; TPR, true positive rate.

4.3 | Best performing parameter combinations

The parameter values occurring in the 25 best fitting parameter combinations are visualised using frequency scatter plots in Figure 9a,b. The figure shows that the best fitting combinations contain cohesion values of 6 and 8 kPa and internal friction angle values of 21° to 31°. With regard to the calibrated hydraulic parameters, the specific storage ranges between 0.001 and 0.013 m^{-1} and the saturated hydraulic conductivity between 1.0 $\times 10^{-6}$ and 2.2 $\times 10^{-4}$ $m \cdot s^{-1}$ (see also Figure 7).

The predicted slope failures in the 25 best fitting parameter combinations have also been used to create an average predicted stability map. The result is shown in Figure 10. This map shows how often each raster cell was predicted as unstable in each of the 25 best fitting

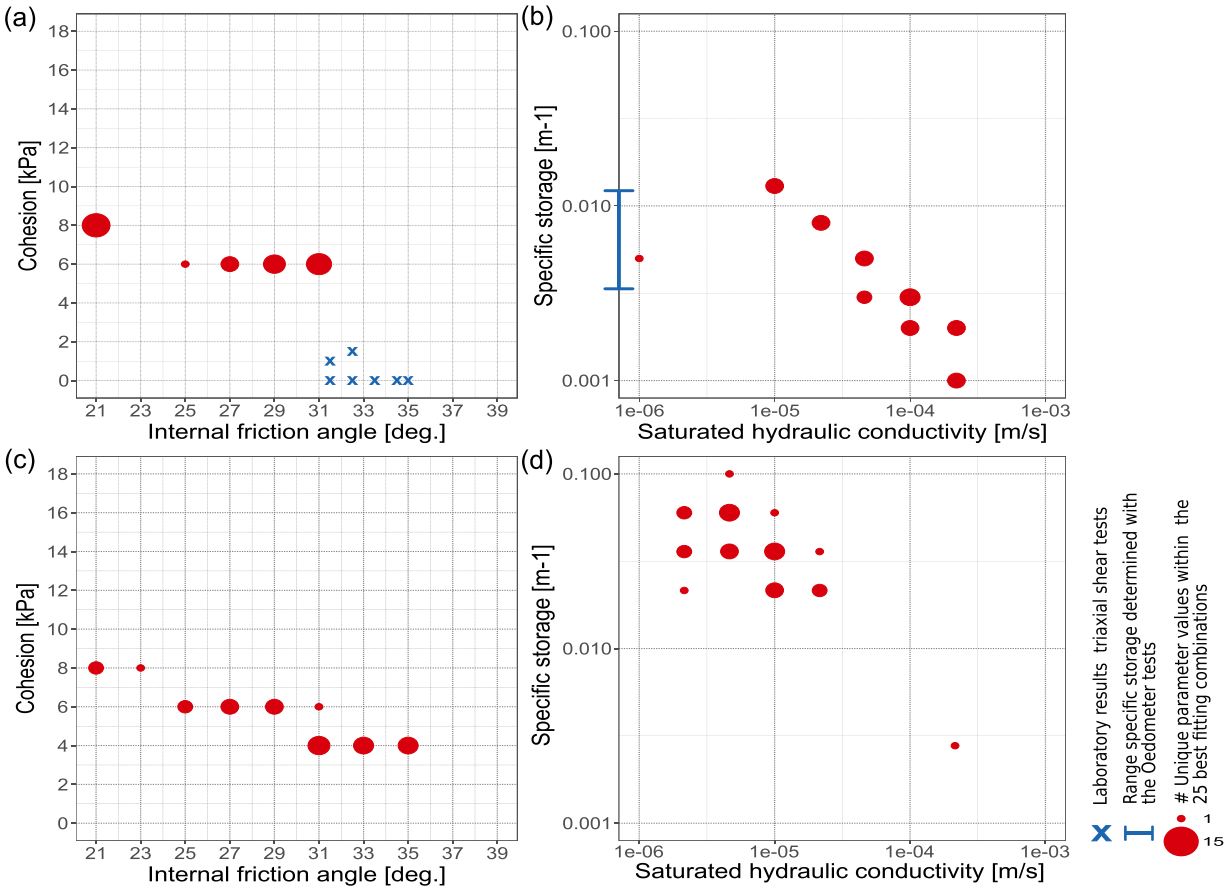

FIGURE 9 Frequency scatter plots of the unique geotechnical (a, c) and hydraulic parameter values (b, d) occurring within the 25 best fitting parameter combinations for the Passeier valley (a, b) and the Laternser valley (c, d). The blue crosses show the internal friction angle and cohesion values found with the triaxial shear tests, and blue range bar indicates the range of the specific storage that was derived with the oedometer tests on the soil samples in the Passeier valley.

model runs. The map shows a distinct spatial pattern in the model prediction, where the eastern lower part of the catchment is less susceptible to slope failure in the model and the upper western part of the catchment shows a high susceptibility to failure. This pattern coincides with the forest mask (Figure 1d), which covers the eastern area with lower predicted susceptibility to slope failure. The locations with a higher susceptibility coincide with the slope values of around 26° to 44° or regolith depth values of around 0.9 to 1.8 m; see also Figure 1c for the used slope angle map. Besides this, the map also shows high average stability values for the cells that were predicted as unstable. This indicates that the individual model runs within the calibrated parameter ensemble showed very similar spatial patterns of predicted slope failure.

4.4 | Comparison parameter ensembles of both study areas

The performance of the parameter ensemble in the Laternser valley is given in Table 5. Comparing the performances of the parameter ensemble in the Laternser valley and the ensemble derived in the Passeier valley shows that the calibration of the Passeier valley performs better than the calibration in the Laternser valley. The D2PC values of all the best model runs of the Passeier valley are lower and thus better than the lowest D2PC value of the model runs of the Laternser valley (see also Figures 11c,f). The TNR for the best model runs in the Passeier valley is lower than the best model runs in the Laternser valley. However, the TPR is generally higher for the best model runs in the Passeier valley, which causes the lower D2PC values. Both the TNR and TPR values for the best model runs in the Laternser valley show a larger spread.

A comparison of the frequency scatter plots of the best fitting parameter combinations for both study areas in Figure 9a,b and 9c,d shows very similar calibrations of the geotechnical parameter values. The best fitting cohesion and internal friction angle values for the Passeier valley are a subset of the calibrated geotechnical parameter values in the Laternser valley. The spread of the geotechnical parameters is slightly smaller in the best fitting combinations for the Passeier valley. In contrast, the comparison of the calibrated hydraulic parameters shows a very distinct difference in the best fitting combinations for both study areas. The calibrated saturated hydraulic conductivity values are on average higher in the Passeier valley, with a median saturated hydraulic conductivity of 1.0 $\times 10^{-4}$ m·s^{-1} against a median value in the Laternser valley of 4.64 $\times 10^{-6}$ m·s^{-1}. The spread of the best fitting hydraulic conductivity for the Passeier valley is also larger than in the calibration of the Laternser valley. The standard deviation of the calibrated saturated hydraulic conductivity in the Passeier valley is 7.97 $\times 10^{-5}$ m·s^{-1} and 4.18 $\times 10^{-5}$ m·s^{-1} for the calibration in the Laternser valley. The values of the specific storage are generally much lower in the best fitting combinations for the Passeier valley.

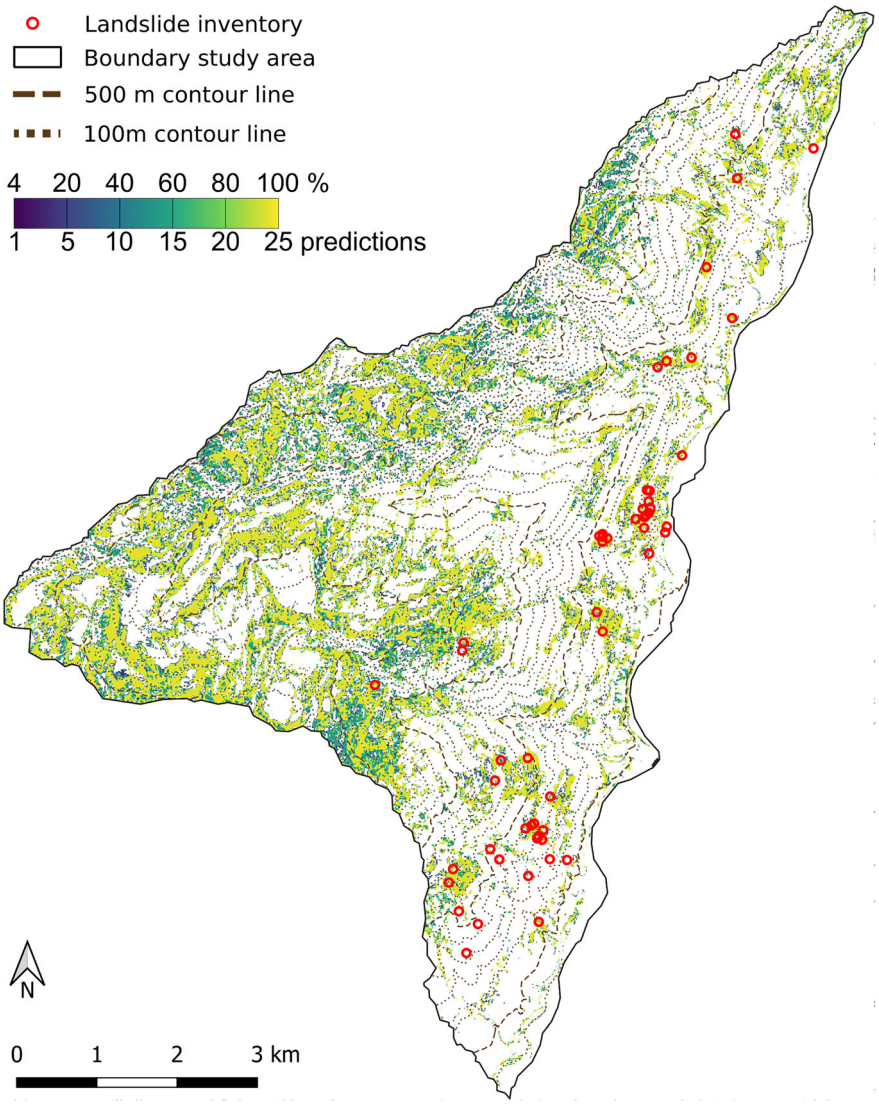

FIGURE 10 Map of average stability constructed with the 25 best fitting parameter combinations for the Passeier valley.

TABLE 5 Performance metrics for model runs in the Laternser valley of the timestep with the maximum predicted failure in each model run (data taken from Zieher *et al.*, 2017).

	Full ensemble		Stable at $t = 0$		Peak TPR after $t = 6h$		25 best combinations	
	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
TPR	0.0%	99.4%	0.0%	90.4%	0.3%	70.5%	49.2%	70.5%
TNR	10.5%	100.0%	57.4%	100.0%	71.0%	100.0%	71.0%	90.2%
FPR	0.0%	89.5%	0.0%	42.6%	0.0%	29.0%	9.8%	29.0%
FNR	0.6%	100.0%	9.6%	100.0%	29.5%	99.7%	29.5%	50.8%
D2PC	0.34	1.00	0.34	1.00	0.41	1.00	0.41	0.53

Abbreviations: D2PC, Distance to the Perfect Classification; TNR, true negative rate; TPR, true positive rate.

The median specific storage in the calibration of the Passeier valley is 0.003 m^{-1} against 0.035 m^{-1} in the calibration of the Laternser valley. The spread of the specific storage in the Passeier valley is narrower than in the calibration for the Laternser valley, with a standard deviation of 0.004 m^{-1} for the Passeier valley against 0.020 m^{-1} for the Laternser valley.

The comparison of the 100 best fitting parameter combinations for each study area is shown in Figure 11. For the Passeier valley, enlarging the best fitting ensemble to 100 model runs extends the range of the best fitting cohesion to lower values and the range of the internal fiction angle values to higher values (Figures 11a and 9a). Due to this, the distributions of the calibrated geotechnical parameters for both study areas become even more similar. For the hydraulic parameters, the distribution does not change much from the selection of the 25 best fitting model runs (Figures 11b and 9b). For the Passeier valley, the median of the saturated hydraulic conductivity remains the same. However, the maximum saturated hydraulic conductivity increases to 1.1 $\times 10^{-3}$ $m \cdot s^{-1}$. The range of specific storage is extended with larger values and the median value of the specific storage increases to 0.005 $m \cdot s^{-1}$. For the Laternser valley, increasing

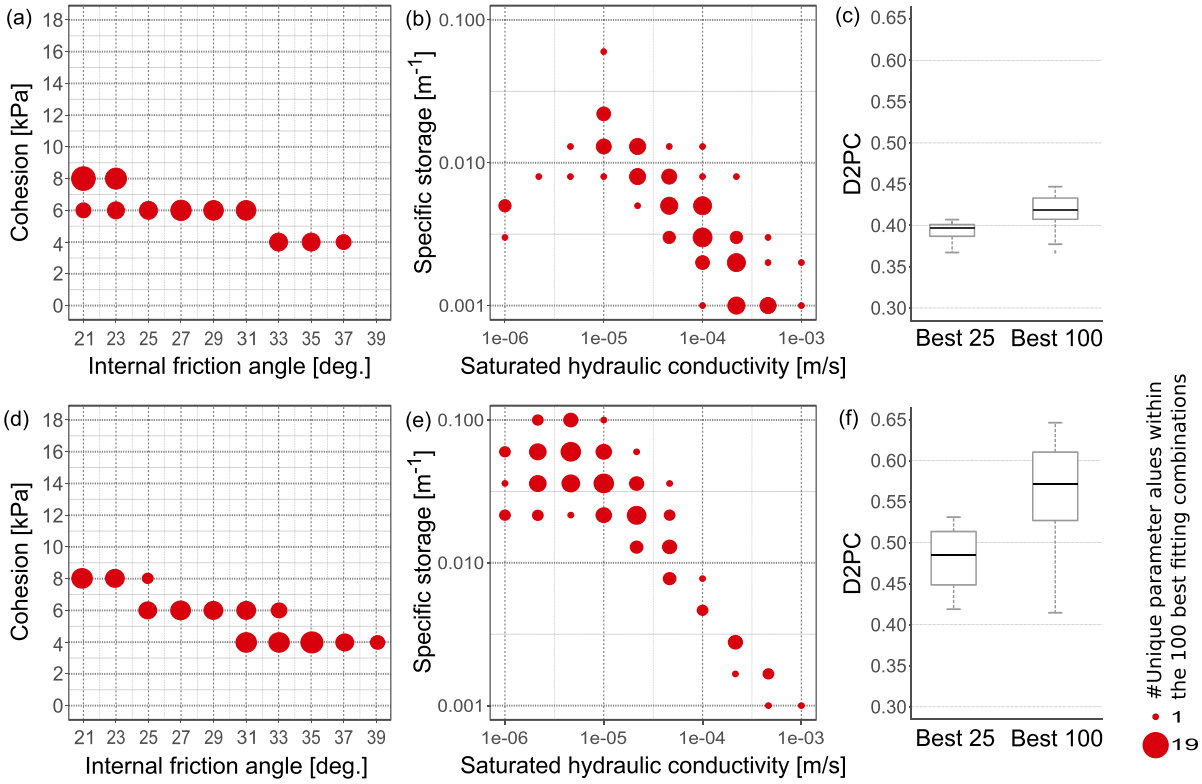

FIGURE 11 Frequency scatter plots of the 100 best fitting parameter combinations for the Passeier valley (a, b) and the Laternser valley (d, e). The boxplots on the right show the distribution of the D2PC values within the 25 and 100 best parameter combinations for the Passeier valley (c) and the Laternser valley (f).

the number of runs for the calibration impacts the distributions of the calibrated parameter combinations less than for the Passeier valley. The median hydraulic conductivity in the calibration for the Laternser valley increases to 1.0 $\times 10^{-5}$ m·s^{-1}. The range of the specific storage is extended to lower values, but the median specific storage stays the same.

Figure 11c shows that the performance of the calibrated ensemble does not change much for the Passeier valley when the number of selected model runs is increased to 100. The maximum D2PC value increases from 0.41 to 0.45 and the median from 0.40 to 0.42. The performance of the model runs in the Laternser valley changes more drastically, with the maximum D2PC value increasing from 0.53 to 0.61 and the median from 0.41 to 0.55.

5 | DISCUSSION

5.1 | Discussion laboratory results

The grain size analysis of the soil samples collected in the Passeier valley showed that all samples fall in the United States Department of Agriculture (USDA) soil texture class sandy loam and showed very little variation in the grain size distributions of the soil samples. The triaxial shear tests and oedometer tests also showed a very narrow distribution in the found cohesion, internal friction angle and specific storage values, even though the samples were taken from several different lithologies. When the found geotechnical soil parameters are compared with values reported in literature, they agree with the values reported in other studies on sandy loam type soils. The reported values of the internal friction angle for these types of soil ranges between 27° and 34° (Carter & Bentley, 2016). The range of the reported soil cohesion for sandy loam type soils is relatively large between 0 and 50 kPa and is heavily dependent on the soil moisture content, fine particle content and the compaction of the soil (Naval Facilities Engineering Command, 1986; Matsushi & Matsukura, 2006). Validating the specific storage was not possible, as the only values reported in literature are on aquifers located at much larger depths than the 2 m considered in this study.

Zieher et al. (2017) report a higher variance within the soil samples taken in the Laternser valley (see Figure 6). The grain size distributions of the samples in the Laternser valley are also well-distributed and comparable with the lodgment till soils reported in Bell (2002). However, almost all samples from the Passeier valley are sandier than those from the Laternser valley and show a lower clay content. With regard to the geotechnical parameters, the found cohesion and internal friction angle values in the Laternser valley again show a higher variation. The geotechnical parameters found in the Passeier valley do fall in the same range; however, the cohesion values for the samples from the Laternser valley are higher and the internal friction angle values are lower. Besides the higher clay content of the soils in the Laternser valley, this difference in shear strength properties in both study areas could also be caused by the uncertainty of shear strength values that are derived with the triaxial shear tests (Schneider-Muntau et al., 2018; Schneider-Muntau et al., 2021). The specific storage reported in Zieher et al. (2017) is a factor 10 higher than the values found in the Passeier valley. This is in agreement with the fact that the sampled soils in the Passeier valley are sandier. The specific storage is inversely related to the hydraulic diffusivity which increases

with grain size (Dingman, 2015). The higher diffusivity values in the Passeier valley mean that the groundwater recharge and in turn the rise of pore pressure respond faster and more intensely to infiltration.

5.2 | Model performance and resulting calibrated parameter ensemble

The D2PC values in Table 4 show that the full parameter space taken from Zieher et al. (2017) does contain a subsample of parameter combinations that performs well for the August 2016 event in the Passeier valley. Compared with other studies, the minimum D2PC value of the best performing combinations in the Passeier valley is higher than what is reported in these studies (Bordoni et al., 2015; Ciurleo et al., 2019; Salciarini et al., 2006). These studies report both higher TPRs and TNRs. However, it should be considered with this comparison that both Bordoni et al. (2015) and Salciarini et al. (2006) used spatially variable geotechnical input for their models, based on pedological or lithological units. Salciarini et al. (2006) showed that the use of spatially variable input significantly improved the performance of TRIGRS. Besides the spatially variable input, Salciarini et al. (2006) and Ciurleo et al. (2019) also include a significant amount of trivial areas, that is, flat areas that are inherently stable. Steger and Glade (2017) showed that this can drastically improve the assessed model performance, since the model will always correctly predict these inherently stable areas as stable, increasing the TNR of the model performance. Considering that the TPR of the best performing model runs (ordered by D2PC) is not higher than 85% in both studies, the minimum D2PC values might be more similar to the performance of TRIGRS in the Passeier valley when the trivial areas are omitted in the validations of the models. For example, for the Passeier valley, excluding the area with a slope ≤ 10° in the model performance would result in a best model run with a TPR of 72.2% and a TNR of 76.4%. As expected, the TNR thus slightly drops. The TPR slightly increases, since there are one or two landslides in the inventory located in cells with a slope angle ≤ 10°. Lastly, only a few studies explicitly consider the timing of landslide initiation in the model calibration of TRIGRS (Bordoni et al., 2015). However, as Figure 8 and the results from Zieher et al. (2017) show, the model calibration requires information on the timing of landslide initiation and is very sensitive to how the timing of landslide initiation is handled. For the Passeier valley, the calibration of the parameter ensemble, while disregarding the timing of landslide initiation, would reduce the range of the geotechnical parameters to a single parameter combination (8 kPa and 21°). The range of the hydraulic parameters would not change, but the median saturated hydraulic conductivity would decrease to 1.0 $\times 10^{-5}$ m·s^{-1} (see also Figure 7). For the Laternser valley, the minimum D2PC value even decreases from 0.41 to 0.34 (see Table 5).

Regarding the parameter ensemble for the Passeier valley, the calibrated hydraulic parameters were tested with values derived using laboratory tests, pedotransfer functions and values reported in literature. The range of the specific storage from the oedometer tests falls within the range of specific storage values in the best fitting parameter ensemble. The range of calibrated saturated hydraulic conductivity values was not validated with a laboratory test but with literature values and values derived using pedotransfer functions. The values from the pedotransfer functions predict saturated hydraulic conductivity values that are approximately a factor 10 lower (ranging from 1.13 $\times 10^{-5}$ m·s^{-1} to 1.17 $\times 10^{-7}$ m·s^{-1}) than the values derived in the calibration of the hydraulic conductivity (see also Figure 7). However, the values from the pedotransfer functions were derived without consideration of the structure of the soil, such as the occurrence of macropores, and could therefore underestimate the *in situ* saturated hydraulic conductivity (Fusco et al., 2021; Fatichi et al., 2020; Zhang & Schaap, 2019). Fatichi et al. (2020) also reported a factor 3–20 difference between saturated hydraulic conductivity values derived with the consideration of soil structural information and values derived without consideration of the soil structure, which would fit the factor 10 difference found in this study. The range of values reported in literature for sandy loam soils also overlaps with the range of calibrated hydraulic conductivity values (2.74 $\times 10^{-6}$ m·s^{-1} to 5.74 $\times 10^{-5}$ m·s^{-1}) (Nemes et al., 2001) (see also Figure 7).

A similar fit was not found for the calibrated geotechnical parameter values and the geotechnical values from the laboratory tests. The calibrated values of the effective cohesion are higher, and the values for the internal friction angle are on average lower than those derived from the triaxial shear tests. An important reason for this is the use of disturbed samples. A similar discrepancy between values of cohesion obtained with laboratory triaxial shear tests and *in situ* bore-hole shear tests was also reported by Rinaldi et al. (2004). In addition to this, the failure envelopes found with the laboratory tests were determined using a limited range of applied loads, due to the technical constraints of the triaxial shear test. The stress loads in this limited range were higher (>50 kPa) than the stresses that naturally occur in shallow landslides (approximately 20 kPa). This results in the extrapolation of lower cohesion and higher internal friction angle values from the triaxial shear tests (Schneider-Muntau et al., 2022; Wood, 1990). Besides this, Schneider-Muntau et al. (2018) and Schneider-Muntau et al. (2021) found that measuring shear strength using a standard triaxial shear test comes with large uncertainties. Lastly, the discrepancy between the laboratory values and calibrated values can also be attributed to the use of spatially homogeneous model input, which leads to generalised values in the calibration.

5.3 | Comparison parameter ensembles both study areas

A comparison of the D2PC values of the calibration in both study areas shows that the performance of the parameter ensembles is very similar (min. D2PC of 0.37 for the Passeier and 0.41 for the Laternser valley). However, from the distributions of the D2PC values within the parameter ensembles (Figure 11c,f), it can be concluded that the TRIGRS setup using initially wet conditions with homogeneous input works better in the Passeier valley. The distributions show that almost all model runs within the final parameter ensemble of the Passeier valley (max. D2PC: 0.41) outperform the model runs in the final parameter ensemble of the Laternser valley (min. D2PC: 0.41). It could be that the use of homogeneous parameter values works better in the Passeier valley, because the soil in the Passeier is more homogeneous than in the Laternser valley. This would be in line with what was found in the comparison of the soil samples from both study areas.

The ensemble of the Passeier valley is in general more conservative than the ensemble of the Laternser valley, as the FPR and TPR values are overall higher (Passeier: min. FPR 20%, max. TPR 75%, Laternser: min. FPR 10%, max. TPR 71%) (Figure 12). Thus, the calibrated model runs of the Laternser valley generally predict more stable scenarios than the model runs of the Passeier valley. Since the slopes in the Laternser valley are less steep than in the Passeier valley, it is expected that the infinite slope method predicts less slope failure in the Laternser valley when the other parameter values are kept constant. Besides this, the Laternser valley has a larger trivial area, where the slope angle is half the minimum internal friction angle in the parameter space (10°). In the Laternser valley, this area amounts to 7.2% of the study area, while the trivial area amounts to 4.9% in the Passeier valley. This larger trivial area means that the model runs will automatically have a higher TNR in the Laternser valley, since these trivial areas are unconditionally stable with the infinite slope method.

For further improvements in model performance in the Passeier valley, the focus should lie on the incorporation of spatial variability in the model input. Several studies using TRIGRS have incorporated the spatial variability using the lithology map as basis for the zonation of the model input (Bordoni et al., 2015; Ciurleo et al., 2019). To account for the variation of the soil properties within the lithological classes and the uncertainties of their values, these approaches could be further improved with application of probabilistic methods (Raia et al., 2014; Rossi et al., 2013; Salciarini et al., 2017). However, a prerequisite for these approaches is detailed information on the spatial distribution of the soil parameters. The homogeneous results from the laboratory indicate that the distribution of soil properties in the Passeier valley is not solely dependent on the underlying lithology. For the Passeier valley, it will, therefore, first be required to derive a zonation of the soil properties for the model input. In addition to this, the model would also benefit from a more detailed assessment of the regolith depth model, since this is the most sensitive parameter in the infinite slope method (Kuriakose et al., 2009; Zieher et al., 2017). However, previous studies have also shown that the construction of an accurate regolith depth map is very difficult, since information on the spatial distribution of regolith depth is often insufficiently available and difficult to acquire (von Ruette et al., 2013; Van den Bout et al., 2021). Efforts into the incorporation of the spatial distribution of other input datasets, such as the soil properties or precipitation, should therefore not be neglected and could prove more effective in improving the model performance.

In the comparison of the calibrated parameter values for the Laternser valley and the Passeier valley, there were large similarities found in the calibrated geotechnical parameter ensembles. These similarities show that this calibrated geotechnical ensemble is transferable across the two valleys and possibly applicable to other study areas. When the lithology maps of the two study areas are compared, the main commonality between the the two valleys is their extensive lodgement till deposits. It can thus be expected that this ensemble (see Figure 9) is also applicable for calibrating TRIGRS in other study areas with extensive lodgement till deposits. de Lima Neves Seefelder et al. (2017), who investigated the calibration of the infinite slope method for a study area in Brazil, also report very similar calibrated parameter combinations for the geotechnical parameter values. It would therefore be interesting to investigate if the calibration of TRIGRS in significantly different study areas also results in similar geotechnical parameter calibrations. The comparison of the calibrated parameter ensembles found no similarities in the calibrated hydraulic parameter values. The calibrations instead reflected the more prominent soil types in the two study areas. It seems that these parameters are more sensitive to different soil types. In addition, the hydraulic parameters mainly force the timing of landslide initiation predicted by the TRIGRS model. Their calibration is, therefore, dependent on the temporal pattern of the modelled precipitation event and how the timing of landslide initiation is handled in the temporal calibration of the model. Since the precipitation events in the Passeier and Laternser valleys are distinctly different, this difference is reflected in the calibrated hydraulic parameter ensembles.

Lastly, the large variances in the calibrated parameter ensembles in combination with the high values in the average stability maps in both study areas indicate equifinality in the calibration of TRIGRS. With this equifinality, it is uncertain if a calibration of TRIGRS will work on different storm events (Brazier et al., 2000). This uncertainty

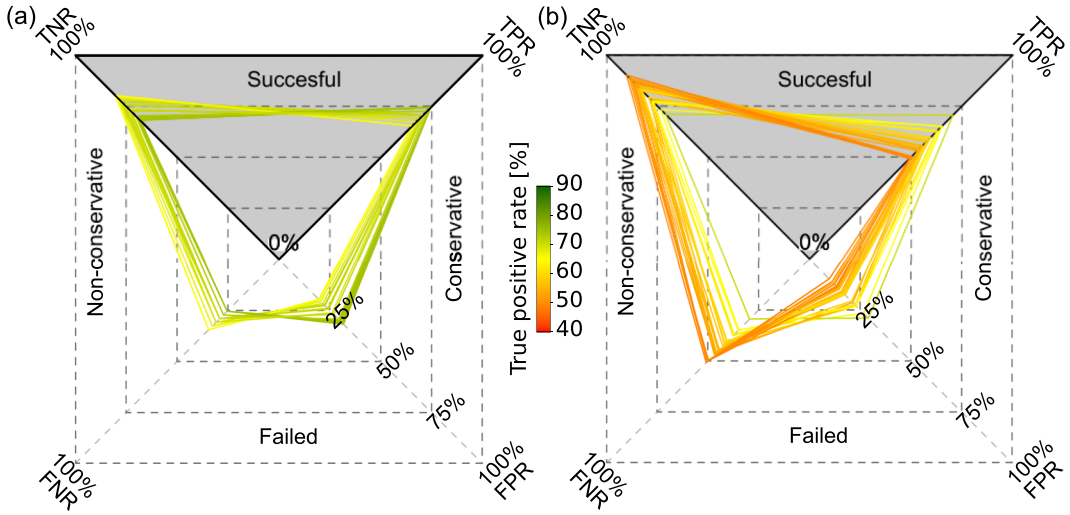

FIGURE 12 Predictive rates of the model runs in the calibrated ensembles of the Passeier valley (a) and the Laternser valley (b). The colours indicate the true positive rate (TPR) of the individual model runs.

adds to the uncertainty from the assumptions made for the timing of landslide initiation and the uncertainties of general assumptions, such as those made for the initial conditions of the model. A different storm event could not be tested for the Passeier valley, since there is no record of other precipitation events causing landslides in the study area. However, the study by Zieher et al. (2017) showed that the calibration of the Laternser valley performed well for another precipitation event. The equifinality and uncertainties in the calibration of the model stress the importance of using of parameter ensembles in the calibration of TRIGRS. Not only to account for the inherent variance of the parameters and uncertainties that come with them but also because of the resulting higher predictive capability of the model.

6 | CONCLUSIONS

In this study, the calibrations of the dynamic physically based slope stability model TRIGRS in two study areas were compared with each other in terms of their model performance and the values occurring within the calibrated parameter ensembles. The goal was to find an overlapping parameter ensemble that could be used in both study areas, and potentially other study areas, without losing model performance. The comparison of the model performances showed that the methods from Zieher et al. (2017) can be used to derive a well-performing parameter ensemble for the Passeier valley for the assessment of landslide susceptibility during the August 2016 event. The calibrated parameter ensemble of the Passeier valley slightly outperforms the calibration of the Laternser valley. Compared with other studies using a setup of TRIGRS with spatially homogeneous input, the calibrated model in the Passeier valley performs similarly. However, compared with studies using spatially variable input, the performance of the model setup in the Passeier valley is slightly worse.

The study also investigated the validity of the parameter values in the calibrated parameter ensemble for the Passeier valley. This investigation showed that the calibrated hydraulic parameters fit with values from laboratory tests on soil samples from the Passeier valley and the values found for these parameters in literature. For the geotechnical parameters, a discrepancy was found between the laboratory values and the calibrated values. This can be explained with the uncertainties of the values found in the triaxial shear tests and the use of a homogeneous model setup. However, this finding does stress the importance of the use of parameter ensembles for the input of physically based methods, since it shows that the values from laboratory tests cannot always be used directly in physically based models.

The comparison of the parameter values within the calibrated parameter ensembles of the Passeier valley and the Laternser valley showed that a small parameter space exists for the geotechnical parameters that can be applied to both study areas without losing model performance. For the hydraulic parameter ensemble, it does not seem possible that a small spatially transferable parameter space can be constructed. The comparison of the calibrated ensemble in the Passeier valley and in the Laternser valley showed that these parameters are more sensitive to different soil types than the geotechnical parameters. In addition, the difference in the calibration of the hydraulic parameters is also caused by the distinct differences in the precipitation events in the Passeier and Laternser valleys. The hydraulic parameters drive the temporal response of the slope stability in the model output. The differences in the precipitation events are, therefore, reflected in the calibrated hydraulic parameter ensemble.

Based on the results from this study, the authors recommend the use of a similar homogenous setup and the found transferable set of strength parameter values (cohesion and internal friction angle) for a faster calibration of the model TRIGRS in other study areas with extensive lodgement till deposits when limited information on the spatial distribution of the soil parameters is available. The hydraulic parameters from this study cannot be transferred and should therefore be calibrated to the specific soil types of the study area. For the specific storage values, the calibration can be determined using oedometer tests. For the calibration of the saturated hydraulic conductivity, the study found a discrepancy between the calibrated values and the values derived from pedotransfer functions. The calibration of the saturated hydraulic conductivity can be derived from the range of values used in the full parameter space of this study and may yield better results than taking parameter values from pedotransfer functions. Lastly, the results from the study also show that it is important that detailed information on the timing of landslide initiation is available for the calibration of the hydraulic parameters and that this information is used to explicitly enforce this timing in the model prediction.

AUTHOR CONTRIBUTIONS

Lotte de Vugt: methodology, resource collection, data collection, coding, data analysis and writing—original draft. **Thomas Zieher:** conceptualization, funding acquisition, methodology, resource collection, data collection, coding, data analysis, supervision and writing—review and editing. **Barbara Schneider-Muntau:** methodology, data analysis and writing—review and editing. **Mateo Moreno:** writing—review and editing. **Stefan Steger:** conceptualization, funding acquisition, resource collection and writing—review and editing. **Martin Rutzinger:** conceptualization, funding acquisition, methodology, supervision and writing—review and editing.

ACKNOWLEDGEMENTS

This study was conducted within the PROSLIDE project that received funding from the research program Research Südtirol/Alto Adige 2019 under D52F19000140003 of the Autonomous Province of Bozen/Bolzano (Südtirol/Alto Adige). For the modelling, the study made use of the High-Performance Computing (HPC) LEO network at the University of Innsbruck. The authors thank the Autonomous Province of Bolzano/Bozen for providing access to the used environmental datasets. The authors also thank the geotechnical laboratory at the university of Innsbruck for providing access to and aid with the conduction of the laboratory tests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Lotte de Vugt https://orcid.org/0009-0003-0221-036X
Barbara Schneider-Muntau https://orcid.org/0000-0001-5947-9181
Mateo Moreno https://orcid.org/0000-0002-9530-3076

REFERENCES

Alvioli, M. & Baum, R.L. (2016) Parallelization of the TRIGRS model for rainfall-induced landslides using the message passing interface. *Environmental Modelling & Software*, 81, 122–135.

Alvioli, M., Melillo, M., Guzzetti, F., Rossi, M., Palazzi, E., von Hardenberg, J., Brunetti, M.T. & Peruccacci, S. (2018) Implications of climate change on landslide hazard in Central Italy. *Science of The Total Environment*, 630, 1528–1543.

Amato, G., Eisank, C., Castro-Camilo, D. & Lombardo, L. (2019) Accounting for covariate distributions in slope-unit-based landslide susceptibility models. A case study in the alpine environment. *Engineering Geology*, 260, 105237.

Autonomous Province of Bolzano (2016) Precipitation data, St. Martin weather station. Dataset. Available from: http://daten.buergernetz.bz.it/services/meteo/v1/timeseries?station_code=22210MS&output_format=CSV&sensor_code=N&date_from=201608010000&date%_to=201608100000

Axelsson, P. (2000) DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing* 33(4).

Baum, R.L., Godt, J.W. & Savage, W.Z. (2010) Estimating the timing and location of shallow rainfall-induced landslides using a model for transient, unsaturated infiltration. *Journal of Geophysical Research: Earth Surface* 115(F3).

Baum, R.L., Savage, W.Z. & Godt, J.W. (2008) TRIGRS—a fortran program for transient rainfall infiltration and grid-based regional slope-stability analysis, version 2.0. In 1159, Open-File Report.

Bell, F.G. (2002) The geotechnical properties of some till deposits occurring along the coastal areas of eastern England. *Engineering Geology*, 63(1), 49–68.

Berti, M. & Simoni, A. (2010) Field evidence of pore pressure diffusion in clayey soils prone to landsliding. *Journal of Geophysical Research: Earth Surface* 115(F3).

Bordoni, M., Meisina, C., Valentino, R., Bittelli, M. & Chersich, S. (2015) Site-specific to local-scale shallow landslides triggering zones assessment using TRIGRS. *Natural Hazards and Earth System Sciences*, 15(5), 1025–1050.

Brazier, R.E., Beven, K.J., Freer, J. & Rowan, J.S. (2000) Equifinality and uncertainty in physically based soil erosion models: application of the GLUE methodology to WEPP–the Water Erosion Prediction Project–for sites in the UK and USA. *Earth Surface Processes and Landforms*, 25(8), 825–845.

Carter, M. & Bentley, S.P. (2016) Shear strength. In *Soil properties and their correlations*, 2nd edn., John Wiley & Sons, Ltd; 107–137.

Chae, B.-G., Park, H.-J., Catani, F., Simoni, A. & Berti, M. (2017) Landslide prediction, monitoring and early warning: a concise review of state-of-the-art. *Geosciences Journal*, 21(6), 1033–1070.

Ciurleo, M., Mandaglio, M.C. & Moraci, N. (2019) Landslide susceptibility assessment by TRIGRS in a frequently affected shallow instability area. *Landslides*, 16(1), 175–188.

de Lima Neves Seefelder, C., Koide, S. & Mergili, M. (2017) Does parameterization influence the performance of slope stability model results? a case study in Rio de Janeiro, Brazil. *Landslides*, 14(4), 1389–1401.

Dietrich, W.E. & Montgomery, D.R. (1998) SHALSTAB: a digital terrain model for mapping shallow landslide potential, NCASI (National Council of the Paper Industry for Air and Stream Improvement).

Dingman, S.L. (2015) *Physical hydrology: third edition*. Waveland Press.

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M.H., Ghezzehei, T.A., Hengl, T., Kollet, S., Agam, N. & Avissar, R. (2020) Soil structure is an important omission in earth system models. *Nature Communications*, 11(1), 522.

Formetta, G., Capparelli, G. & Versace, P. (2016) Evaluating performance of simplified physically based models for shallow landslide susceptibility. *Hydrology and Earth System Sciences*, 20(11), 4585–4603.

Fusco, F., Mirus, B.B., Baum, R.L., Calcaterra, D. & De Vita, P. (2021) Incorporating the effects of complex soil layering and thickness local variability into distributed landslide susceptibility assessments. *Water*, 13(5), 713.

Gariano, S.L. & Guzzetti, F. (2016) Landslides in a changing climate. *Earth-Science Reviews*, 162, 227–252.

Geokatalog (2022) Open geodatabase of the autonomous province of South Tyrol. Available from: http://geokatalog.buergernetz.bz.it/geokatalog/

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M. & Chang, K.-T. (2012) Landslide inventory maps: new tools for an old problem. *Earth-Science Reviews*, 112(1), 42–66.

Herbst, M., Diekkrüger, B. & Vanderborght, J. (2006) Numerical experiments on the sensitivity of runoff generation to the spatial variation of soil hydraulic properties. *Journal of Hydrology*, 326(1), 43–58.

Iverson, R.M. (2000) Landslide triggering by rain infiltration. *Water Resources Research*, 36(7), 1897–1910.

Kuriakose, S.L., van Beek, L. P. H. & van Westen, C. J. (2009) Parameterizing a physically based shallow landslide model in a data poor region. *Earth Surface Processes and Landforms*, 34(6), 867–881.

Matsushi, Y. & Matsukura, Y. (2006) Cohesion of unsaturated residual soils as a function of volumetric water content. *Bulletin of Engineering Geology and the Environment*, 65(4), 449–455.

Medina, V., Hürlimann, M., Guo, Z., Lloret, A. & Vaunat, J. (2021) Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale. *CATENA*, 201, 105213.

Milledge, D.G., Griffiths, D.V., Lane, S.N. & Warburton, J. (2012) Limits on the validity of infinite length assumptions for modelling shallow landslides: testing the infinite length assumption in landslide models. *Earth Surface Processes and Landforms*, 37(11), 1158–1166.

Munari, M., Peterlin, D., Tollardo, M., Geier, G. & Tartarotti, P. (2016) Climareport – Juni – Giugno 2016. In N. 246, Landeswetterdienst – Autonome Provinz Bozen.

Munari, M., Peterlin, D., Tollardo, M., Geier, G., Tartarotti, P. & Rastner, L. (2016) Climareport – Juli – Luglio 2016. In N. 247, Landeswetterdienst – Autonome Provinz Bozen.

Naval Facilities Engineering Command (1986) Foundations and earth structures: NAVFAC DM 7.02.

Nemes, A., Schaap, M.G., Leij, F.J. & Wösten, J. H.M. (2001) Description of the unsaturated soil hydraulic database UNSODA version 2.0. *Journal of Hydrology*, 251(3), 151–162.

Pack, R.T., Tarboton, D.G. & Goodwin, C.N. (1998) Terrain stability mapping with SINMAP, technical description and users guide for version 1.00. Civil and Environmental Engineering Faculty Publications.

Park, H.J., Lee, J.H. & Woo, I. (2013) Assessment of rainfall-induced shallow landslide susceptibility using a GIS-based probabilistic approach. *Engineering Geology*, 161, 1–15.

Petley, D. (2012) Global patterns of loss of life from landslides. *Geology*, 40(10), 927–930.

Piacentini, D., Troiani, F., Soldati, M., Notarnicola, C., Savelli, D., Schneiderbauer, S. & Strada, C. (2012) Statistical analysis for assessing shallow-landslide susceptibility in South Tyrol (south-eastern Alps, Italy). *Geomorphology* 151–152: 196–206.

Planet Team (2022) Planet application program interface: in space for life on earth. Dataset, San Fransisco, CA.

Puhlmann, H. & von Wilpert, K. (2012) Pedotransfer functions for water retention and unsaturated hydraulic conductivity of forest soils. *Journal of Plant Nutrition and Soil Science*, 175(2), 221–235.

Raia, S., Alvioli, M., Rossi, M., Baum, R.L., Godt, J.W. & Guzzetti, F. (2014) Improving predictive power of physically based rainfall-induced shallow landslide models: a probabilistic approach. *Geoscientific Model Development*, 7(2), 495–514.

Rinaldi, M., Casagli, N., Dapporto, S. & Gargini, A. (2004) Monitoring and modelling of pore water pressure changes and riverbank stability during flow events. *Earth Surface Processes and Landforms*, 29(2), 237–254.

Rossi, G., Catani, F., Leoni, L., Segoni, S. & Tofani, V. (2013) HIRESSS: a physically based slope stability simulator for HPC applications. *Natural Hazards and Earth System Sciences*, 13(1), 151–166.

Rowe, P.W. & Barden, L. (1966) A new consolidation cell. *Géotechnique*, 16(2), 162–170.

Salciarini, D., Fanelli, G. & Tamagnini, C. (2017) A probabilistic model for rainfall—induced shallow landslide prediction at the regional scale. *Landslides*, 14(5), 1731–1746.

Salciarini, D., Godt, J.W., Savage, W.Z., Conversini, P., Baum, R.L. & Michael, J.A. (2006) Modeling regional initiation of rainfall-induced

shallow landslides in the eastern Umbria Region of central Italy. *Landslides*, 3(3), 181.

Schilirò, L., Cevasco, A., Esposito, C. & Mugnozza, G.S. (2018) Shallow landslide initiation on terraced slopes: inferences from a physically based approach. *Geomatics, Natural Hazards and Risk*, 9(1), 295–324.

Schmidt-Walter, P., Trotsiuk, V., Hammel, K., Kennel, M. & Federer, A. (2023) LWFBrook90R: simulate evapotranspiration and soil moisture with the SVAT model LWF-Brook90. Manual.

Schneider-Muntau, B., Dai, X. & Fellin, W. (2022) Sensitivity analyses of the different influencing factors on numerical investigations of landslides. *Geomechanics and Tunnelling*, 15(5), 582–595.

Schneider-Muntau, B., Medicus, G., Desrues, J., Andò, E. & Viggiani, G. (2021) Investigation of uncertainty in strength parameter identification. In: Barla, M., Di Donna, A., Sterpi D. (Eds.) *Challenges and innovations in geomechanics*. Lecture Notes in Civil Engineering. Cham: Springer International Publishing, 277–284.

Schneider-Muntau, B., Schranz, F. & Fellin, W. (2018) The possibility of a statistical determination of characteristic shear parameters from triaxial tests. *Beton- und Stahlbetonbau*, 113(S2), 86–90.

Simoni, S., Zanotti, F., Bertoldi, G. & Rigon, R. (2008) Modelling the probability of occurrence of shallow landslides and channelized debris flows using GEOtop-FS. *Hydrological Processes*, 22(4), 532–545.

Steger, S. & Glade, T. (2017) The challenge of "trivial areas" in statistical landslide susceptibility modelling. In: Mikos, M., Tiwari, B., Yin, Y., Sassa, K. (Eds.) *Advancing culture of living with landslides*. Cham: Springer International Publishing, 803–808.

van Beek, L.P.H. (2002) Assessment of the influence of changes in land use and climate on landslide activity in a mediterranean environment. Ph.D. Thesis, University of Utrecht.

Van den Bout, B., Lombardo, L., Chiyang, M., van Westen, C. & Jetten, V. (2021) Physically-based catchment-scale prediction of slope failure volume and geometry. *Engineering Geology*, 284, 105942.

von Ruette, J., Lehmann, P. & Or, D. (2013) Rainfall-triggered shallow landslides at catchment scale: threshold mechanics-based modeling for abruptness and localization. *Water Resources Research*, 49(10), 6266–6285.

Wösten, J.H.M., Lilly, A., Nemes, A. & Le Bas, C. (1999) Development and use of a database of hydraulic properties of European soils. *Geoderma*, 90(3), 169–185.

Wood, D.M. (1990) *Soil behaviour and critical state soil mechanics*. Cambridge University Press.

Wu, W. & Sidle, R.C. (1995) A distributed slope stability model for steep forested basins. *Water Resources Research*, 31(8), 2097–2110.

Zhang, Y. & Schaap, M.G. (2017) Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *Journal of Hydrology*, 547, 39–53.

Zhang, Y. & Schaap, M.G. (2019) Estimation of saturated hydraulic conductivity with pedotransfer functions: a review. *Journal of Hydrology*, 575, 1011–1030.

Zhao, H.F., Zhang, L.M., Xu, Y. & Chang, D.S. (2013) Variability of geotechnical properties of a fresh landslide soil deposit. *Engineering Geology*, 166, 1–10.

Zieher, T., Perzl, F., Rössel, M., Rutzinger, M., Meißl, G., Markart, G. & Geitner, C. (2016) A multi-annual landslide inventory for the assessment of shallow landslide susceptibility – two test cases in Vorarlberg, Austria. *Geomorphology*, 259, 40–54.

Zieher, T., Rutzinger, M., Schneider-Muntau, B., Perzl, F., Leidinger, D., Formayer, H. & Geitner, C. (2017) Sensitivity analysis and calibration of a dynamic physically based slope stability model. *Natural Hazards and Earth System Sciences*, 17(6), 971–992.

How to cite this article: de Vugt, L., Zieher, T., Schneider-Muntau, B., Moreno, M., Steger, S. & Rutzinger, M. (2024) Spatial transferability of the physically based model TRIGRS using parameter ensembles. *Earth Surface Processes and Landforms*, 1–18. Available from: https://doi.org/10.1002/esp.5770