Optimizing and validating the Gravitational Process Path model for regional debris-flow runout modelling

Jason Goetz[1], Robin Kohrs[1], Eric Parra Hormazábal[2], Manuel Bustos Morales[3], María Belén Araneda Riquelme[3], Cristián Henríquez[3], Alexander Brenning[1]

[1] Department of Geography, Friedrich Schiller University Jena, Germany
[2] Institute of Geosciences, University of Potsdam, Germany
[3] Instituto de Geografía, Pontificia Universidad Católica de Chile, Chile. Centre for Sustainable Urban Development, CEDEUS. Centro de Cambio Global UC.

Correspondence to: Jason Goetz (jason.goetz@uni-jena.de)

Abstract. Knowing the source and runout of debris-flows can help in planning strategies aimed at mitigating these hazards. Our research in this paper focuses on developing a novel approach for optimizing runout models for regional susceptibility modelling, with a case study in the upper Maipo river basin in the Andes of Santiago, Chile. We propose a two-stage optimization approach for automatically selecting parameters for estimating runout path and distance. This approach optimizes the random walk and Perla et al.'s (PCM) two-parameter friction model components of the open-source Gravitational Process Path (GPP) modelling framework. To validate model performance, we assess the spatial transferability of the optimized runout model using spatial cross-validation, including exploring the model's sensitivity to sample size. We also present diagnostic tools for visualizing uncertainties in parameter selection and model performance. Although there was considerable variation in optimal parameters for individual events, we found our runout modelling approach performed well at regional prediction of potential runout areas. We also found that although a relatively small sample size was sufficient to achieve generally good runout modelling performance; larger samples sizes (i.e. ≥80) had higher model performances and lower uncertainties for estimating runout distances at unknown locations. We anticipate that this automated approach using open-source software R and SAGA-GIS will make process-based debris-flow models more readily accessible and thus enable researchers and spatial planners to improve regional-scale hazard assessments.

1 Introduction

Knowledge of where debris flows are initiated and how far they travel is crucial for assessing their impact over large regions (Aleotti and Chowdhury, 1999; van Westen et al., 2006). Commonly, debris-flow runout modelling for large areas is performed by first delineating source areas and then applying empirical-statistical or process-based numerical methods for simulating the runout characteristics (Blahut et al., 2010a; Horton et al., 2013; Mergili et al., 2019). There is a wide selection of heuristic, statistical and machine learning methods suitable for predicting source areas in large regions (Chung and Fabbri, 1999; Carrara

et al., 1999; Brenning, 2005; Goetz et al., 2015b; Lombardo et al., 2018). There are also many empirical-statistical and numerical methods available to model runout patterns – see McDougall (2017) for an overview.

35   Not all runout methods are suitable for application to large areas. Many of the physically based methods require event-specific geotechnical and rheological parameters, such as material composition (e.g. bulk density and source depths) and flow characteristics (e.g., flow discharge rates). These parameters, such as debris-flow volume, can be extremely difficult to obtain for large areas, let alone single unobserved events (Marchi and D'Agostino, 2004; Dong et al., 2009). Consequently, runout modelling at larger scales has been progressing towards applying simplified conceptual models to simulate debris-flow patterns
40   across different environmental conditions. These models combine spreading algorithms to control runout path with empirical-statistical or numerical friction models to simulate likely runout paths and distances (Guthrie et al., 2008; Horton et al., 2013; Wichmann, 2017; Mergili et al., 2019). Many of the spreading algorithms, including multiple flow direction models (Holmgren, 1994), cellular automata (Guthrie et al., 2008) and random walk (Gamma, 2000; Mergili et al., 2015), simulate runout paths using only topographic data.

45   Calibration of model parameters continues to be one of the main challenges in runout modelling for single-events and over large areas (Hungr, 1995; van Westen et al., 2006; Schraml et al., 2015; McDougall, 2017; Mergili et al., 2019). The objective of model calibration is to determine parameter values that best capture main debris-flow characteristics, such as runout distance, velocity, and distribution of deposits (Hungr, 1995; McDougall, 2017). Approaches for model calibration include adjusting parameters based on visual inspection (i.e. trial and error; Hungr, 1995; Mergili et al., 2012); expert knowledge
50   (Horton et al 2013); posterior analysis (Mergili et al., 2019; Aaron et al., 2019); and optimization algorithms that aim to minimize a cost function, i.e. a quantitative measure of runout model performance. Some measures of performance include estimates of the intersection-over-union (Galas et al., 2007), area under the receiver operating characteristic curve (AUROC; Cepeda et al., 2010; Mergili et al., 2015) and depth error (Aaron et al., 2019) of simulated and observed debris flows. Since most of these calibration approaches are for single observed events, they rarely consider how transferable tuned parameter sets
55   are from local to regional applications.

Assessing spatial transferability is essential for testing the assumption a trained model based on a sample of events captures the generic debris-flow characteristics across a region (Fabbri et al., 2003). The distribution of training data and the sample size can have a strong influence on the model calibration and performance of regional models (Heckmann et al., 2014; Petschko et al., 2014; Rudy et al., 2016). For spatially distributed models, spatial transferability can be assessed by exploring model
60   parameter selection and performance under different spatial partitioning scenarios of training and test data (Wenger and Olden, 2012; Brenning, 2012; Schratz et al., 2019; Mergili et al., 2019). Although spatial transferability has been well explored for regional landslide susceptibility models (Brenning, 2005; Lombardo et al., 2014; Petschko et al., 2014; Goetz et al., 2015b; Cama et al., 2017; Knevels et al., 2020), such analysis is not common for regional runout modelling.

In this study, we developed an optimization procedure for process-based models applied for regional simulation of debris-flow runout patterns. The performance evaluation focuses on the spatial transferability and sensitivity to sample size of an optimized regional debris-flow runout model. We achieve this by utilizing the open-source statistical software R to add optimization and evaluation functionality to the open-source Gravitational Process Path (GPP) modelling framework (Wichmann, 2017). Additionally, this study demonstrates the use of spatial cross-validation and visualization techniques to diagnose uncertainties in the prediction of source areas, runout paths and runout distances, including the sensitivity of optimized parameter selection. The aim of this research is to contribute to improving the development of quantitative techniques for runout model calibration and uncertainty estimation (McDougall, 2017). This is especially important in large and inaccessible mountainous areas where various types of mass movements pose unique challenges to the safety of the local population, the integrity of transportation infrastructure, and the reliability of drinking water supplies.

2 Materials and Methods

2.1.1 Study area

Our study area is the upper Maipo river basin (33°03′ to 34°18′ S, 800 m to 6108 m a.s.l.), located in the semi-arid Andes of central Chile. Debris-flow activity in remote and populated areas of the Maipo river basin have caused many deaths and severe disruptions to critical transportation and water supply infrastructure supporting Chile's capital city Santiago (Hauser, 2002; Sepúlveda et al., 2015; Moreiras and Sepúlveda, 2015).

High-intensity rainfall (Sepúlveda et al., 2015), rapid snowmelt (Moreiras et al., 2012) and seismic activity (Serey et al., 2019) are the main triggers of debris flows in this region. They occur in steep gullies and talus slopes consisting of gravel, small boulders and a fine sandy-silty matrix. Much of this material is from weathered volcanic and sedimentary rocks of the Abanico and Farellones formations in the western Main Cordillera (Sepúlveda et al., 2006). A typical runout track will cut through previously formed debris flow channels and alluvial fans, resulting in new erosion and deposition paths (Sepúlveda et al., 2015). Rainfall-triggered runout distances in this area have been observed up to 5.5 km, and the thickness of deposits varies from 1 to 2 m in deposition areas (Sepúlveda et al., 2015).

2.1.2 The debris-flow inventory

Debris flow polygons and source points were mapped based on photointerpretation of high-spatial resolution (0.50 m) satellite imagery (2000 to 2019) from CNES/Airbus and Maxar Technologies available through Google Earth Pro software, field observations, reviewing news articles and the compilation of data collected by public authorities. Each mapped polygon represents a debris-flow track that includes source, runout and deposition area. In total, 541 source points and 521 debris flow polygons were mapped (**Table 1**). Manually mapping all debris flows across the upper Maipo basin is a challenging task due

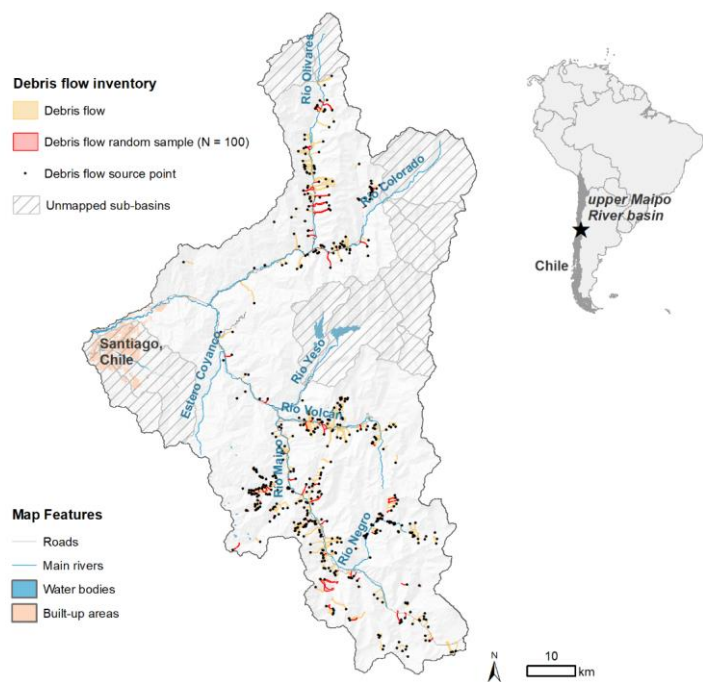
to its large geographical extent and its high abundance of mass movements. Therefore, a mapping strategy was employed that divided the basin into 58 sub-drainage basins (5,439 km^2), 45 (3,936 km^2) of which were selected for mapping (**Figure 1**).

100   **Figure 1: Map providing an overview of the debris flow polygons and source points mapped in the upper Maipo river basin.**

Table 1: Runout characteristics of the debris flow inventory

	Median	IQR	Minimum	Maximum
Runout surface area (ha)	3.7	7.5	0.1	358.1
Runout distance (m)	729	823	54	5668
Max. elevation (m.a.s.l.)	2695	714	1256	4768

Note. IQR = interquartile range.

2.1.3 Modelling debris-flow source areas

105 Potential debris-flow source areas were spatially predicted using a generalized additive model (GAM). In general, GAMs demonstrate good performance for susceptibility modeling compared to other commonly used physically-based and machine-learning techniques (Goetz et al., 2011; Goetz et al., 2015b). To improve model generality and avoid overfitting, the GAM smoothing spline parameters were allowed a maximum 5 effective degrees of freedom (Wenger and Olden, 2012; Goetz et al., 2015a). The training and test data were based on the common 1:1 sampling strategy (Heckmann et al., 2014) of presence to

110 absence of source points. The non-source (i.e. absence) points were randomly sampled within the mapped sub-basins outside of the mapped debris flow polygons. The resulting training and test data contained 541 source points and 541 non-source points.

The predictor variables of source areas included hillslope angle, elevation, catchment area, plan curvature and distance to faults. These predictor variables generally have a high importance for modeling debris-flow initiation susceptibility as observed

115 in previous works (Blahut et al., 2010b; Goetz et al., 2015b; Angillieri, 2020). The publicly available ALOS PALSAR Radiometrically Terrain Corrected (RTC) high-resolution (12.5 m) digital elevation model (DEM; ASF DAAC) was used to derive terrain attributes. Before deriving the terrain attributes, mesh denoising was applied to the DEM to mitigate the propagation of artifacts such as high-frequency noise (Brock et al., 2020) in the prediction of source areas (Sun et al., 2007; Stevenson et al., 2010). We used the implementation of this algorithm in SAGA-GIS. After denoising, an algorithm to fill

120 sinks (Planchon and Darboux, 2002) was applied to the DEM, and the terrain attributes were processed. Distance to faults was calculated as the Euclidean distance from the fault lines (Servicio Nacional de Geología y Minería de Chile, 2003; scale 1:1,000,000).

The performance of the source area prediction model was assessed using repeated k-fold spatial cross-validation (Brenning, 2012). Like cross-validation, spatial cross-validation randomly splits the data (e.g., source points and non-source points) into

125 k disjoint subsets, where the model is trained using k-1 sets and tested with the remaining set during each cross-validation iteration. For *spatial* cross-validation, the data is divided into spatially disjoint sub-areas, in our case using the k-means clustering algorithm (Ruß and Brenning, 2010). This approach should provide a rigorous estimate of the spatial transferability of a model by attempting to reduce spatial autocorrelation between test and training data (Brenning, 2005; Wenger and Olden, 2012; Schratz et al., 2019). We estimated model performance by repeating 5-fold spatial cross-validation 1,000 times. Model

130 performance was measured using the AUROC, which is an overall measure of goodness-of-fit that is independent of any particular decision threshold.

2.2 Modelling debris flow runout

The Gravitational Process Path (GPP; Wichmann, 2017) model was used to regionally model runout. The GPP model is an open-source framework in the SAGA-GIS software that provides users with various model components to simulate runout

135 path, distance, velocity and deposition of material of mass movements (e.g. snow avalanches, rock falls, and debris flows).

Due to the extent and remoteness of the study area, we focus on modelling the likely spatial patterns of runout. That is, we are not modelling flow velocity and depth.

140 Runout path was modelled using the random walk process path component of the GPP model (Gamma, 2000). It is a common approach for debris-flow runout path modelling at medium scales (Mergili et al., 2012; Heckmann and Schwanghart, 2013; Mergili et al., 2015). Random walk models the potential path of runout by iteratively simulating (via Monte Carlo simulation) the downslope movement of debris flows originating from source-area grid cells. These simulations results in a grid with runout frequencies that indicate how many times a grid cell is traversed, this is a cumulative frequency based on simulations 145 from all source areas. There are three parameters that need to be calibrated to obtain a desired runout path: (1) a slope threshold (°) defining where divergent flow is allowed; (2) the exponent of divergence that controls the amount of divergence or lateral spreading in areas below the slope threshold; and (3) a persistence factor that controls the direction of movement (Wichmann, 2017). Flow path is determined using a 3×3 window that first controls the path of a central cell by considering only neighboring cells with lower elevation. If the neighbouring cells are below the slope threshold, the neighbouring cell with the steepest descent 150 is selected; otherwise, neighbours are assigned transition probabilities based on slope. These probabilities are adjusted using the exponent of divergence and the persistence factor. A higher exponent of divergence will result in more even probabilities across the neighbouring cells, allowing for a higher likelihood of not selecting the steepest descent path. The persistence factor considers the previous flow direction in weighting the probabilities. A higher persistence factor increases the probability that the selected neighbor will follow the direction of the previous cell. Based on these transition probabilities, a pseudo-random 155 number generator selects a cell to define the flow path (see Wichmann, 2017 for a more detailed description). With this random-walk implementation, the flow path stops when the neighboring cells have a higher or equal elevation compared to the central cell.

Runout distance was constrained using the two-parameter friction model (PCM; Perla et al., 1980) component of the GPP model. The PCM model, which is also a component of the Flow-R model (Horton et al., 2013), has also been used for modelling 160 debris-flow behaviour at medium scales (Mergili et al., 2012; Heckmann and Schwanghart, 2013; Mergili et al., 2015). It is a centre-of-mass model where motion is mainly controlled by (1) the sliding friction coefficient μ and (2) the mass-to-drag ratio M/D. In the GPP implementation of the PCM model, the velocity (ms^{-1}) at any grid cell v_i along a runout path can be characterized by the velocity of the previous grid cell v_{i-1} (ms^{-1}), the local slope θ (°), distance between grid cells L_i (m), the acceleration due to gravity g (ms^{-2}) and the friction parameters μ and M/D (m),

$$v_i = \sqrt{\alpha_i \cdot (M/D) \cdot (1 - e^{\beta_i}) + (v_{(i-1)})^2 \cdot e^{\beta_i} \cdot \cos(\Delta\theta_i)} \qquad (1)$$

165 where,

$$\alpha_i = g(\sin\theta_i - \mu_i \cos\theta_i),$$ (2)

$$\beta_i = \frac{-2L_i}{(M/D)_i},$$ (3)

and

$$\Delta\theta_i = \begin{cases} \theta_{(i-1)} - \theta_i & \text{if } \theta_{(i-1)} > \theta_i \\ 0 & \text{if } \theta_{(i-1)} \leq \theta_i \end{cases}.$$ (4)

For the case of a concave transition in the slope direction between grid cells, a velocity correction (Eq. 4) based on the conservation of linear momentum is applied (Wichmann, 2017). The μ controls the velocity of movement through the runout path and M/D controls velocity movement over steep terrain. The conditions for acceleration along the runout path can be described by,

$$\tan\theta_i > \mu_i,$$ (5)

and deceleration by

$$\tan\theta_i < \mu_i.$$ (6)

We can use Eq. 6 to help interpret μ, as it can be used to characterize the slope angle under which deposition begins and termination of runout occurs (Perla et al., 1980).

2.2.1 Optimizing model parameters

For regionally applying the runout model we needed to determine the combination of model parameters that result in the best match to our debris-flow inventory. Determining optimal parameters was based on two criteria: (1) the ability of the model to capture our observed runout paths, and (2) its ability to match the observed runout distances. Therefore, we performed this optimization task using a two-stage approach that first optimizes the random walk model and then the PCM model parameters.

A random sample of 100 debris-flow tracks and corresponding source points was used for optimizing the runout models. This sample of the inventory was chosen to facilitate quality control and reduce the computational complexity of the optimization. Source areas were determined by buffering each source point by 50 m and masking away the buffered area that exceeds the runout trimline; this ensures the source area is contained within a mapped debris flow polygon. A sink-filled version of the original DEM was used for the runout modelling.

For each model component, an exhaustive grid search in parameter space was used to find parameter sets that achieve optimal model performance across all sampled debris flows. The search ranges were similar to Wichmann's (2017) suggested parameter limits for debris flows (see **Table 2** for value ranges). We additionally tested the use of a spatially varying sliding

friction coefficient. The value for this spatially varying sliding friction coefficient μ_i was calculated as a function of catchment area a (km²) for maximum runout (Gamma, 2000; Wichmann and Becht, 2006; Mergili et al., 2012; Wichmann, 2017):

$$\mu_i = 0.13a^{-0.25}. \tag{7}$$

Similar to Mergili et al. (2012) and suggested by Wichmann (2017), the spatially varying μ_i was set to a maximum of 0.3 and minimum of 0.045. Runout was computed using 1,000 model iterations.

The AUROC was used as a performance measure for the random walk model. The receiver operating characteristic (ROC) is a plot of the true positive rate versus the false positive rate. AUROC values range from 0.5 (random discrimination between classes) and 1.0 (a perfect classifier; Zweig and Campbell, 1993). Model performance was rated higher if the random walk model contained observed debris-flow tracks within its simulated paths. After optimizing the random walk model, we fixed these parameters for the PCM model, and optimized the μ and M/D parameters for determining runout distance. The performance of the PCM model was measured using the relative error of runout distance. Relative error was used so that each debris flow was weighted equally regardless of its magnitude. The AUROC was used to break any ties in relative error performance between multiple optimal parameter sets.

Runout distance was measured in terms of horizontal length of the debris-flow track. This distance was measured as the length of a minimum area bounding box containing the observed debris-flow track (**Figure 2**; Niculiță, 2016; Taylor et al., 2018). Estimated debris-flow tracks were defined as grid cells with values greater than a median runout frequency (**Figure 2**). In this case, the median value represents the most typical simulated debris-flow track. It also provides a conservative estimate of runout distance, which helps mitigate the chance that the optimized model regionally underestimates runout distances.

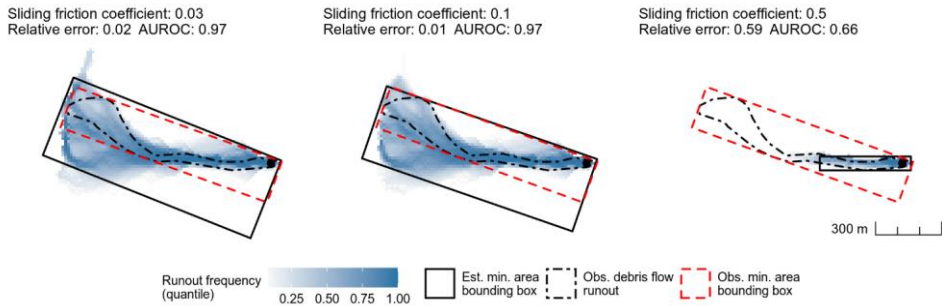



Figure 2: Illustration of runout distance optimization of the sliding friction coefficient μ using a minimum-area bounding box as a measure of travel distance – mass-to-drag ratio M/D fixed at 40 m.

In addition to determining a global optimal parameter set, the best-performing parameters for individual debris-flow events were also explored. We similarly applied a grid search to each event and determined optimal parameter sets based on performances (AUROC and relative error) of each runout model component.

2.2.2 Assessing spatial transferability

Based on our random sample of 100 debris-flow tracks, we assessed the transferability of optimized runout model (random walk and PCM) parameters by performing 5-fold spatial cross-validation with 1,000 repetitions (**Figure 3**). This approach allows us to explore the sensitivity of grid-search optimized parameter combinations to spatial variation in training and test data. To do this, we observed the frequency of variations in optimized parameter combinations within all cross-validated iterations. Optimal parameter combinations that occurred more frequently were considered to have a higher degree of transferability; thus, being considered more reliable for application to the entire study area.

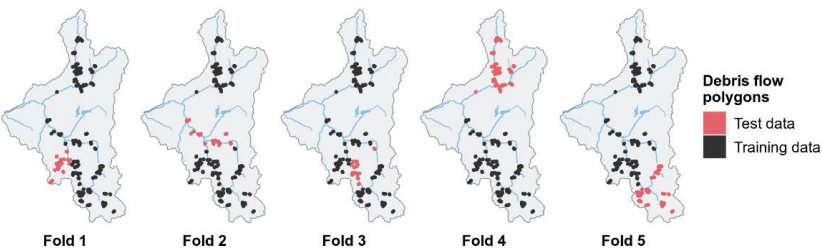



Figure 3: An example realization of random partitions based on k-means clustering of debris-flow polygons for a single repetition of 5-fold spatial cross-validation. The selection of these debris flow polygons was based on a random sample.

We also assessed if there were any spatial patterns in the optimized performance for each model component. That is, were there any spatial trends in model performance that may indicate our model is locally overfitting? We explored such spatial trends by mapping the distribution of individual debris-flow runout model performances based on the optimized parameters. Additionally, we were concerned if the optimized parameters had a stronger fit to debris flows of a certain magnitude or initiating conditions. Therefore, the potential to overfit to certain debris flow characteristics was assessed by determining Spearman's rank correlation (ρ) of individual debris-flow performance (for each model component) with observed runout distance and the elevation, catchment area and hillslope angle of the corresponding source points.

2.2.3 Testing sample-size dependence of performance

We explored how runout model parameter selection, performance and robustness were affected by the number of debris flows used for optimization. Spatial cross-validation was applied to data sets of varying training sample sizes using the random

sample of 100 debris flows used for model optimization. To ensure a fair comparison, the size of the test data for each cross-validation iteration was set to 20 debris flows, which is the maximum test sample size when performing 5-fold spatial cross-validation with a sample of 100 debris flows. We tested training samples sizes from 10 to 80. Model performance for each runout modelling components was summarized using the median and IQR (AUROC and relative error). The optimal parameter

260   set for a given sample size were determined as the parameter combinations that were most frequent.

2.2.4 Finding a suitable source area prediction threshold

For regionally applying the runout model for susceptibility mapping or exposure analysis, a dichotomous classification of the predicted source areas is required to define the grid cells where simulated debris flows initiate. In the case of basing the source areas on a susceptibility model, a suitable threshold of the prediction values needs to be selected. In this study, we determine

265   a suitable prediction threshold to classify source areas by searching for the threshold that results in the best-performing runout model for the entire area. Using the optimized model parameters, we tested runout models based on source areas that were delineated using prediction thresholds from 0.5 to 0.95 with a step of 0.05. The performance of each these models was measured using the AUROC. The AUROC was calculated using a sample of 1,000 debris-flow runout locations and 1,000 locations outside of the debris-flow polygons. The source-area prediction threshold resulting in the highest AUROC values

270   was selected for regionally computing a debris-flow runout map for our study area.

> **Deleted:** non-debris flow locations

2.3 Geocomputing and visualization software

The methods for runout modelling optimization, validation and visualization of the source area prediction and runout modelling were implemented using the open-source statistical software R (ver. 3.6.2; R Core Team, 2019) and SAGA GIS (version 6.1; (Conrad et al., 2015) with its Gravitational Process Path model tool (Wichmann, 2017). Coupling SAGA GIS with R was done

275   using a combination of the 'RSAGA' (Brenning et al., 2018) and 'Rsagacmd' (Pawley, 2019) packages. The GAM was implemented using the 'mgcv' package (Wood, 2011). General handling of spatial data in R used the 'sf' (Pebesma, 2018), 'sp' (Pebesma and Bivand, 2005),'rgeos' (Bivand and Rundel, 2019), 'rgdal' (Bivand et al., 2019) and 'raster' (Hijmans, 2020) packages; spatial cross-validation was applied using the 'sperrorest' (Brenning, 2012) and 'ROCR' (Sing et al., 2005) packages. Parallelization of the optimization and validation procedure used 'foreach' (Microsoft and Weston, 2020).

280   Visualization was done using R's 'ggplot2' (Wickham, 2009) and 'metR' (Campitelli, 2020) packages and ESRI's ArcMap (ver. 10.5).

3 Results

3.1 Source area model performance

The overall performance of the source area prediction based on the GAM was good with a spatially cross-validated median

285   AUROC of 0.80 and an interquartile range (IQR) of 0.001. We found the source-area prediction map was also geomorphologically plausible. Locations most likely to be source areas were within steep terrain associated with channels, gullies and scree slopes. Shallow-flat terrain and areas along ridgelines were modelled as least likely to be source areas (**Figure**

4). This geomorphological knowledge was also expressed in the plots of the GAM spline transformations (**Figure 5**).

290   Relatively steep terrain, slightly concave plan curvature, and areas near faults were modelled as more likely being source areas.

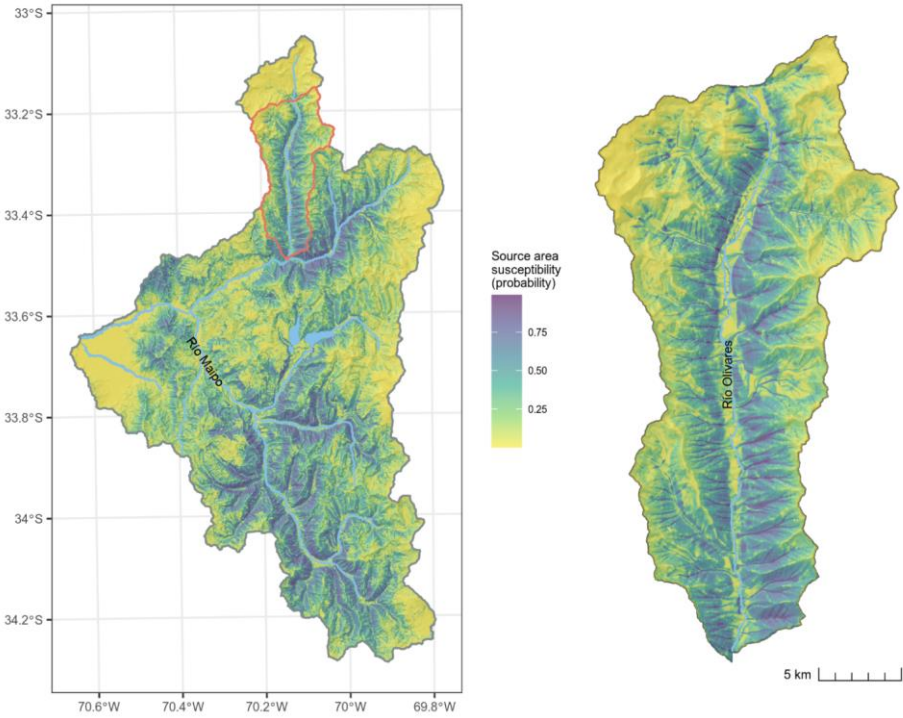



Figure 4: Map of the debris-flow source area prediction based on a generalized additive model (GAM).

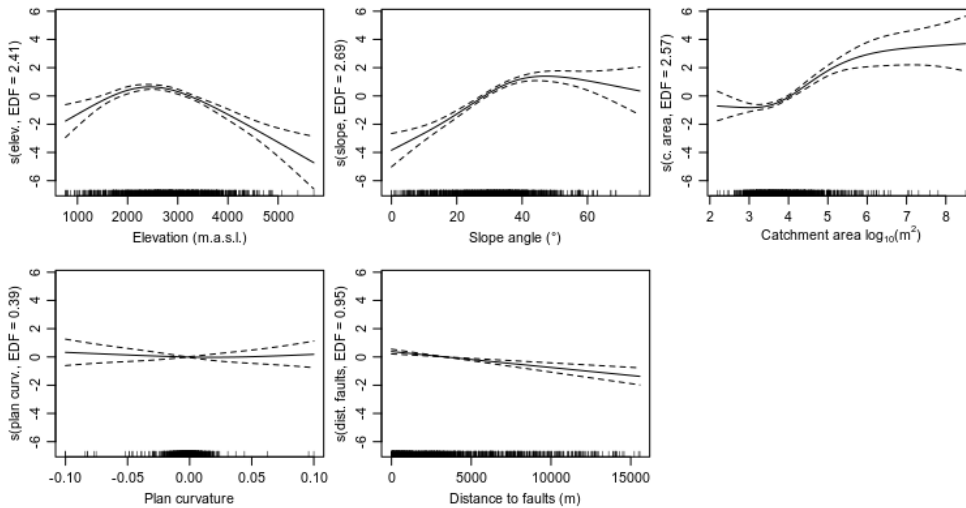



Figure 5: Transformation of predictor variables in the generalized additive model, where the y-axis can be interpreted as the associated likelihood (log-odds) of being a source area. Terms of the form s(predictor) indicate a nonlinear smoothing spline transformation. The effective degrees of freedom (EDF) refer to the flexibility of the smoothers. The dashed lines represent confidence bands at a 95 % level.

3.2 Runout model parameter optimization

The parameter optimization produced runout models with a good spatially cross-validated performance. The optimal parameters for the runout-path model were a slope threshold of 40°, an exponent of divergence of 3.0, and a persistence factor of 1.9 with a median AUROC of 0.94 (IQR = 0.02; **Table 2**). Using these values as plug-in estimates for the PCM runout-distance model component, the optimal μ and M/D were 0.11 and 40 m, respectively. The median spatially cross-validated relative length error of the runout-distance model was 0.11, or 11% (IQR = 0.09). We also found that the model based on a global μ estimate performed better than the runout-distance model using the spatially varying μ_i (optimal M/D = 95 m; **Table 2**), which had a median relative error of 0.19 (IQR 0.09).

Deleted: sliding friction coefficient

Deleted: *mass-to-drag ratio*

Deleted: sliding friction parameter

Deleted: a

Deleted: sliding friction coefficient

Deleted: mass-to-drag

Deleted: -

Deleted: ratio

Table 2: Runout model grid-search optimization setup and results. Optimization performance was assessed using spatial cross-validation (CV).

Model component	Model Parameter	Grid search value range	Grid search steps	Optimal value	Spatial CV performance
Runout path (Random walk)	Slope threshold	20-40°	2°	40°	**AUROC**
	Exponent of divergence	1.3-3.0	0.17	3.0	Median: 0.94
	Persistence factor	1.5-2.0	0.05	1.9	IQR: 0.02
Runout distance (PCM model)	Sliding friction coefficient μ	0.04-0.6	0.01	0.11	**Relative error**
					Median: 0.11
	Mass-to-drag ratio M/D	20-150 m	5 m	40 m	IQR: 0.09
Runout distance (Spatially varying friction PCM model)	Sliding friction coefficient μ_i	$0.13a^{-0.25}$	-	-	**Relative error**
					Median: 0.19
	Mass-to-drag ratio M/D	20-150 m	5 m	95 m	IQR: 0.06

Note. IQR = interquartile range; a = catchment area (km^2)

320 By visualizing the runout-distance optimization results across grid search space, we can observe model performance and sensitivity to different parameter combinations. In this case, we observed only slight model performance differences for μ values just under 0.2 to 0.04 and M/D values from 20 to 150 m (**Figure 6a**). In general, the values in this band across grid search space would result in good performances with median relative errors ≤ 0.15. However, in terms of controlling the spread of error (**Figure 6b**), μ values from about 0.05 to 0.15 and M/D values from 20 to 95 m had the lowest IQRs (≤ 0.2).

325 Exploring the optimal combination of parameter values using spatial cross-validation provides insights into performance reliability of the optimized model. Given different spatial combinations of testing and training data, we found that the optimal combinations of parameters were associated with high performance values (AUROC = 0.94, relative error = 0.11) and high relative frequencies (61-69%) of occurring in each cross-validation iteration (**Figure 7**). Additionally, the spatial cross-validated optimal parameter sets were generally clustered in grid search space, showing there was little variation in optimized 330 parameters depending on the spatial partitions (**Figure 7**). In terms of obtaining multiple optimal solutions, only 5% of the total 5000 cross-validation iterations had ties in relative error performance.

Deleted: ¶
(

Deleted: sliding friction coefficients

Deleted: mass-to-drag ratios

Deleted: sliding coefficients

Deleted: mass-to-drag ratios

Deleted: This result illustrates a generally low sensitivity in runout distance model for a wide range of parameter combinations.

Deleted: ly

Deleted:

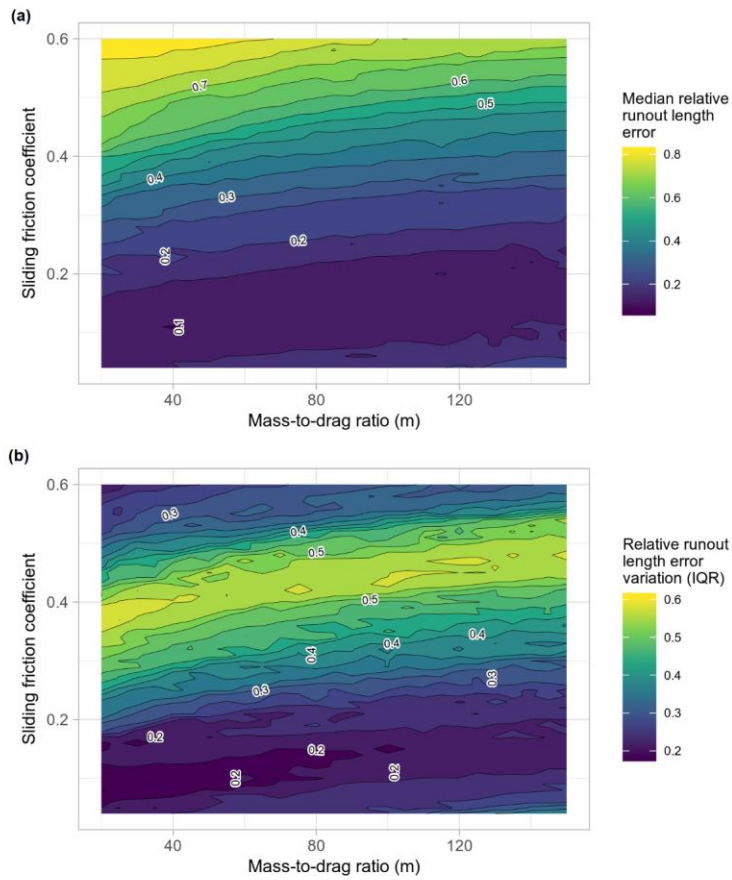



Figure 6: Density contour plots of parameter optimization of sliding friction coefficient μ and mass-to-drag ratio M/D in the PCM model, illustrating median relative runout length errors (a) and the associated variations in relative error (b) using the interquartile range (IQR).

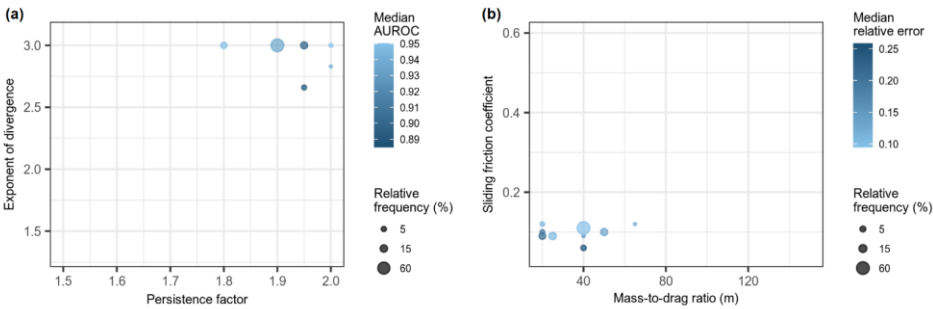



Figure 7: Model performance and frequency of optimal parameters for the runout path (a; given an optimized slope threshold = 40) and runout distance models (b) estimated using 1,000-repeated 5-fold spatial-cross-validation. Relative frequency is the percentage of all repeated spatial-cross validation iterations where a given parameter combination was optimal.

3.3 Thresholding source areas for runout analysis

The best threshold for delineating source areas from the GAM prediction for runout modelling was 0.7, which results in runout affecting 22% of the study area. This threshold had the peak AUROC value of 0.83. The performance of the runout model drastically decreased with thresholds > 0.8 and gradually decreased towards a threshold of 0.5: lower thresholds spatially predicted more runout area **(Figure 8)**. The runout prediction map resulting from the best threshold **(Figure 9)** appears to be geomorphologically plausible. The locations of the runout deposit areas are similar to what we have observed in the satellite imagery. Additionally, lateral spreading is generally low in narrow gullies and much broader on talus slopes, which is what we would expect for debris flows occurring in our study area.

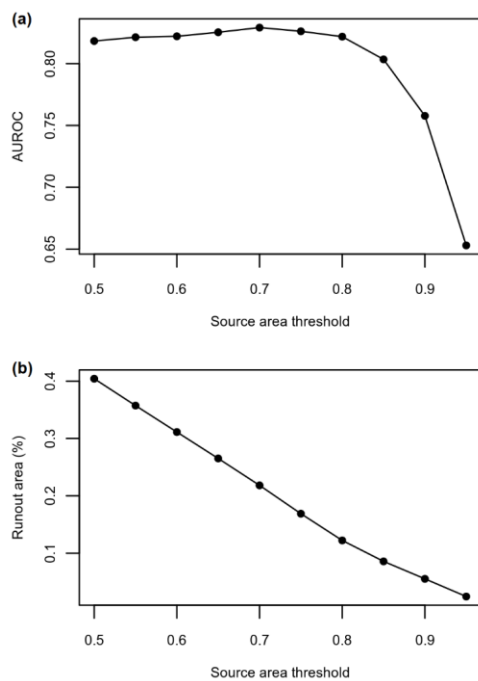



Figure 8: Performance of debris flow runout model for different source area thresholds (a) and percentage of study area impacted by runout (b). Optimal threshold was a threshold of 0.70 (AUROC = 0.83) with runout affecting 22% of the study area.

365

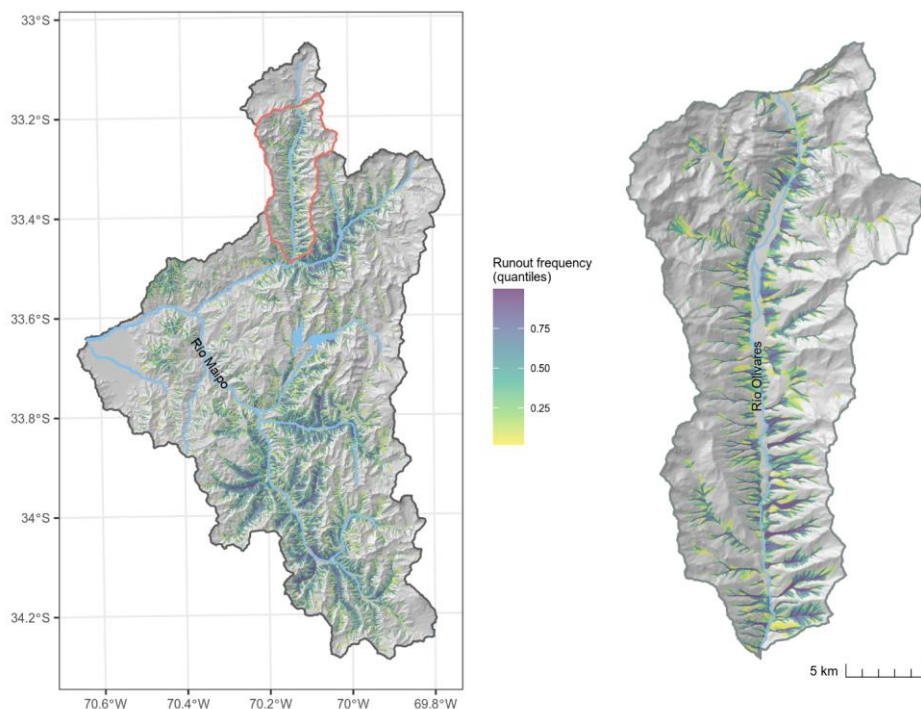



Figure 9: Map of optimized debris-flow runout model based on global parameters and source area threshold of 0.70.

3.4 Exploring patterns in model performance

370 We observed no clear spatial pattern of individual debris-flow performances of the runout model components (**Figure 10**), which is evidence that there was no local overfitting to a particular region of the study area. The distribution of individual AUROC performances of the runout path model were generally high for most of the region, showing that optimization of the path model performs well across the study area. One debris flow was not captured within the runout path model (AUROC < 0.5; **Figure 10**; **Figure 11**). Having a closer look at this occurrence, we found that the modelled runout paths failed to follow

375 the flow direction of the observed debris flow and source point.

The runout distance model had generally more spatial heterogeneity in performance (**Figure 10**), but again no clear spatial patterns in the distribution of relative errors were observed. We investigated the debris-flows that did not perform well for

modelling runout distance (relative error > 0.8) by looking at the satellite imagery used for mapping and the DEM derived hillslope angle. It was found that these cases were related to misclassifying stream erosion in relatively shallow and long-
380   hillslope channels as debris flows.

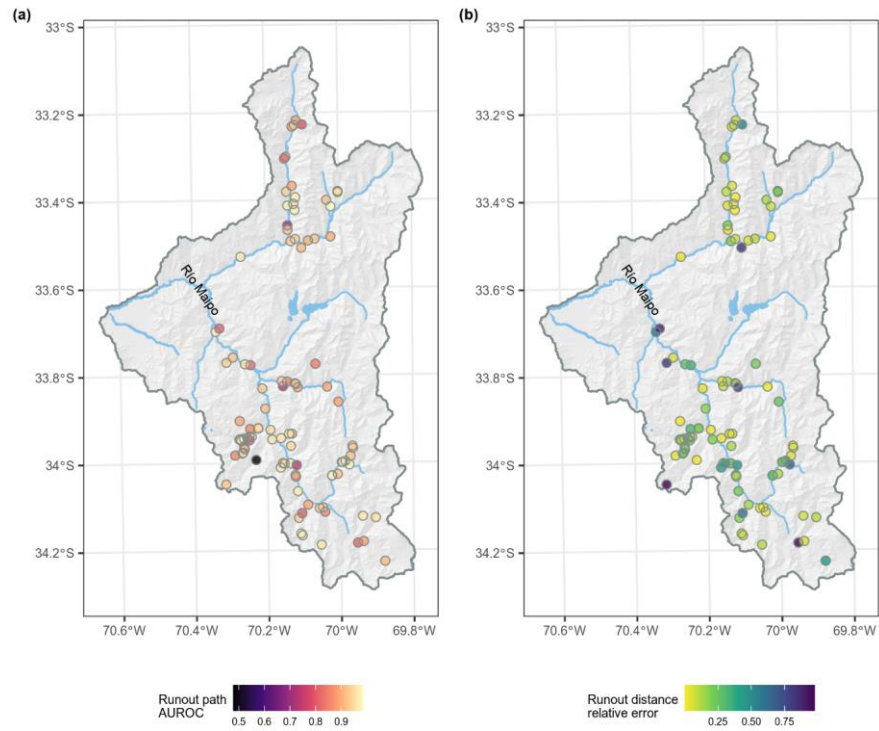



Figure 10: Maps of the performances of the runout path (a) and runout distance models (b) for individual debris-flows based on the global optimal parameters.

Overall, the regionally optimized runout modelled distances fit well with our mapped observations (**Figure 11**). It tended to
385   slightly overestimate debris-flow travel lengths with a median runout distance error of 16.6 m (**Figure 11**). This error is just slightly over a single grid cell size of the DEM. The highest runout distance errors were due to misclassified debris-flows, as previously mentioned. The optimization of the runout model avoided overfitting to debris-flow tracks of a certain magnitude and general terrain conditions. That is, we did not observe a strong correlation between runout distance performance to length

of observed debris flow (ρ = -0.36), starting elevations (-0.21), catchment area (0.11) or hillslope angle (0.29) of source points used for model training.

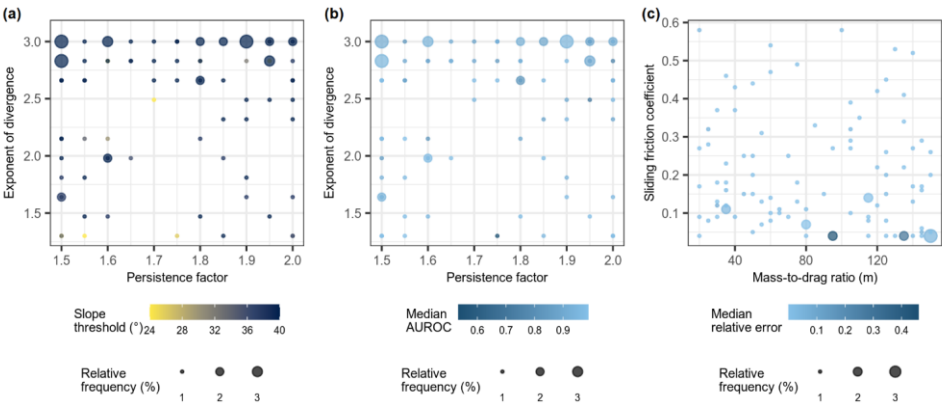
Figure 11: Histogram of performance of the runout path (a), runout distance models based on relative error (b), actual runout length error (c), and a plot of observed runout lengths versus modelled runout lengths (d).

3.5 Optimization for individual debris-flows

The optimal parameters for individual debris-flows were also computed for general comparison to the regional model. We found that the optimized runout model parameters were highly variable for individual debris flows (**Figure 12**). The relative errors were low (median = 0.002, IQR = 0.007), except for a few debris flows that failed to optimize (**Figure 11**). Multiple optimal solutions for the PCM model occurred in 56% of the events. However, after tie breaking using the AUROC, the number of optimal solutions was down to 3%.

Most individual events optimized runout paths with parameter sets leading to high lateral spreading. The optimal-path parameters for most of the individual events had a 40° slope threshold, high exponent of divergence and low persistence values (**Figure 12a**). By individually examining the optimal simulated paths for each training event, we observed that ~ 60% of the

observed debris-flow tracks did occur within the most frequent simulated paths. The other events were typically located on the fringes of the most frequent paths.

420 There was no clear spatial pattern in optimal μ and M/D parameter combinations across the study area, which shows that runout characteristics were quite diverse for individual debris-flows across this large study area (**Figure 13**). We also observed no strong relationships between terrain attributes and optimal parameters. Spearman's correlations of μ and M/D with elevation, slope and catchment area were ≤|0.29|.

425 **Figure 12**: **Performance and frequency of runout path (a, b) and distance (c) optimal parameters determined for individual debris flows. Relative frequency is the percentage of all repeated spatial-cross validation iterations where a given parameter combination was optimal.**

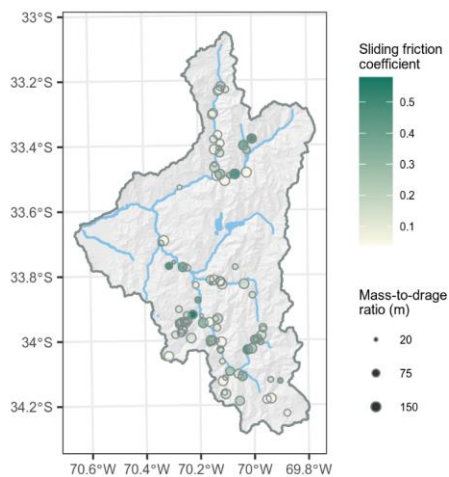



Figure 13: Map of runout distance model optimal parameters determined for individual debris flows.

3.6 Runout model robustness to sample size

Runout model performance and variation tends to improve slightly when using larger sample sizes (**Figure 14**). Large sample sizes also resulted in more consistently selected model parameters across spatial cross-validation iterations **(Figure 15)**. The larger spread across grid search space of the relative frequency of optimal parameter combinations for smaller sample sizes illustrates that we may be less likely to find the best model parameters that generalize well for a large region (**Figure 15**). In general, as we increased sample size, we reduced sampling variability and narrowed the number of optimal combinations of parameters in grid search space, meaning we became more confident that the selected parameters would transfer better to estimate runout distance in adjacent areas.

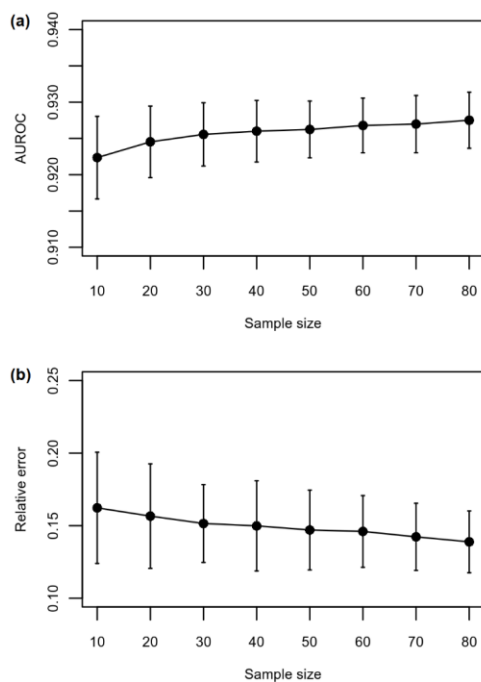



Figure 14: Comparison of runout path (a) and distance (b) performances for different model training samples sizes assessed using spatial cross-validation. The error bars indicate the standard deviation in performances.

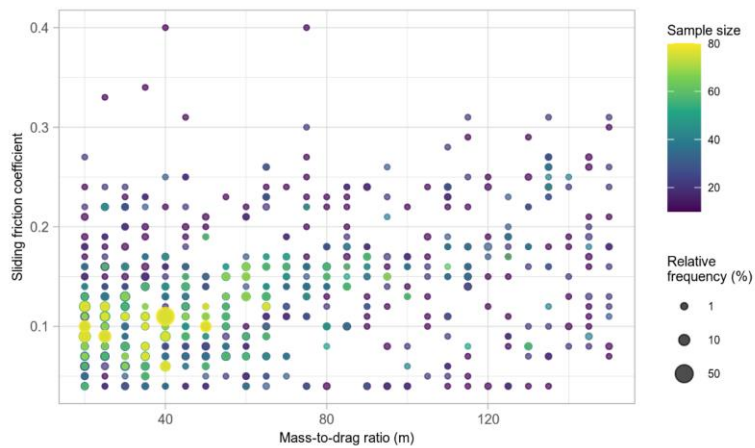



Figure 15: Relative frequencies of optimal parameter combinations for different training sample sizes computed using spatial cross-validation.

4 Discussion

4.1 Model performance and transferability

Assessing the spatial transferability of runout models is essential when extending their use from single or local events to regionally modelling runout across unknown space. Overall, this study demonstrates that our novel optimization approach performed well at regionally modelling the spatial distribution of runout path and distances across the upper Maipo river valley basin. A key component of the success of our modelling approach was its ability to generalize. The transferability of a regional runout model can be affected by the generalization ability of both the source area prediction and the optimized process-based runout model.

4.1.1 Source area modelling

In the development of our source area prediction model, we aimed to produce a simplified empirical model that would result in good performance and transferability. Wenger and Olden (2012) recommended to control the flexibility of the GAM smoothing parameters to avoid overfitting that would lead to poorer model transferability. Therefore, similarly to previous landslide susceptibility studies using the GAM (Goetz et al., 2011; Goetz et al., 2015a; Bordoni et al., 2020) we limited the degrees of freedom for smoothing-spline fitting.

A detailed model based on a large set of predictors can impede its ability to transfer to other locations (Tuanmu et al., 2011). As we and others have demonstrated, good predictive performance of susceptibility models of debris flows can be achieved

with a relatively small set of predictors, which are primarily DEM-derived terrain attributes (Blahut et al., 2010b; Heckmann et al., 2014; Goetz et al., 2015b). Additionally, by fitting our models with multi-temporal data, we may be more likely to achieve better transferability in time (Tuanmu et al., 2011; Knevels et al., 2020). Event-specific inventories may not be large enough for regional optimization and risk the potential of overfitting source conditions, to spatially varying conditions of that event (e.g., precipitation and snowmelt patterns). A smaller sample may also lead to not capturing the range of terrain conditions across the study area required for robust empirical modelling of source area locations (Petschko et al., 2014; Rudy et al., 2016).

The interpretability of the GAM allows us to explore modelled behaviour. For our study, the GAM did well at representing the general geomorphic characteristics of source areas. Some of the relationships between predictors, such as elevation, and debris flow activity can be complex. In the upper Maipo river basin elevation can be a proxy for vegetation, snow cover duration, terrain ruggedness, permafrost and glacial bodies, and geology. It is therefore difficult to discern any direct relationships between elevation and likelihood of being debris source areas. However, we suspect that lower elevations were predicted to be less prone to be source areas due to increased vegetation cover and less rugged terrain. The decrease observed at the highest elevations may relate to permafrost and glacial bodies holding potentially mobilized sediment (e.g., Sattler et al., 2011). A decrease in predicted likelihood of source areas occurring at high slope angles (~ >45°) may be associated with steep rock faces that were more likely sources of rock falls than debris flows (Loye et al., 2009).

4.1.2 Runout analysis

The best-performing regional random-walk parameters allowed for maximum lateral spreading of the runout path given the range of parameters for optimization. Individual events tended to also optimize for high lateral spreading, but not as strongly as the regional model. We believe this high lateral spreading may be due to the location of the observed debris-flows relative to simulated paths and the quality of the DEM. A large proportion of the observed debris flow tracks were located at the fringe of the most frequent simulated paths. Thus, a higher slope threshold and exponent of divergence are required to capture these fringe debris flows. Additionally, the surface of DEMs with resolutions greater than 20 m can be too general to capture minor gullies that may have high flow accumulation (Blahut et al 2010b). The 12.5 m resolution ALOS DEM used in this study is derived from downsampled SRTM data, and would likely contain some of the topographic generalizations of the original DEM (~ 30 m spatial resolution). Despite potential issues with DEM quality, similarly to Horton et al. (2013), we illustrated valuable results can still be achieved.

By optimizing the runout distance model using the median relative error as a metric, we managed to reduce the impact of possible outliers in our training and test data. We additionally reduced our chances of overfitting the regional model to larger debris flow events, which was crucial for a model to maintain a generalization that makes it transferable across large areas.

Plotting the individual optimized models and exploring correlations between terrain attributes and optimal parameters allowed us to see if there were any broad trends in the parameter selection. In our study, we observed high variability in optimal parameters of the PCM model. While applying a trial-and-error approach, Mergili et al. (2012) also observed different optimal parameter combinations for individual events when modelling runout for a couple of debris flows just north of our study area. It seems apparent that determining runout parameter values for regional modelling without an optimization procedure would be difficult. It is also no surprise that the errors of the individually optimized debris-flows were low (**Figure 12**). Whether meaningful or not, the optimization approach should be able to match the runout distance with very low error. Poorly individually optimized events could be attributed to locally poor DEM quality (Horton et al., 2013) and mapping uncertainties (Ardizzone et al., 2002).

The two-parameter PCM model has a uniqueness problem (Perla et al., 1980). Possibly infinitely many pairs of μ and M/D can result in the same runout distances. When optimizing individual events, we did observe this phenomenon. The majority of individual events had more than one optimal combination of parameters. Obtaining a unique solution was not an issue for the regional optimization in our study for the given grid search space. Likely this is due to having to satisfy the runout distances for a variety of hillslope conditions and lengths across the study area. The observed reduced variability in optimal solutions for larger sample sizes (**Figure 15**) provides some evidence for this conjecture.

Although we obtained a unique regional model solution, runout-distance relative errors were only slightly higher than the best performer for combinations of μ and M/D across a band in grid-search space of lower μ values (**Figure 6**). We believe this model performance insensitivity to μ is due to decrease in slope being one of the main factors controlling runout distance. The lowest relative errors tended to occur when the slope condition for deceleration was in the range of 2.3° to 11.3°, and the optimal being 6° ($\mu > 0.04$, $\mu > 0.20$ and $\mu > 0.11$; **Eq. 6**; **Figure 6**). This range matches well to the slope values observed at or near the stopping locations of the debris flows used to train the PCM model. These results are also very similar to modelled observations of debris flow runout in the semi-arid Andes by Mergili et al. (2012). They observed that best PCM runout modelling results for individual events occurred when deposition began at slope angles ranging from 2.6° to 14.0° ($\mu > 0.045$ and $\mu > 0.25$). Additionally, this range fits within other global observations of debris flow deposition occurring on slopes smaller than 6° to 17° (Hungr et al., 1984; Ikeya, 1989; Rickenmann and Zimmermann, 1993; Bathurst et al., 1997; Lorente et al., 2003). The insensitivity across M/D values (**Figure 6**) is due to many possible solutions for obtaining similar runout distances with different combinations of μ and M/D (i.e. the uniqueness problem). The range of μ values resulting in low relative error only slightly increased with higher M/D value, which indicates that μ had a much stronger role than M/D in governing runout distances.

In theory, it is possible that the minimum area-bounding box could contribute to parameter insensitivity. Abrupt changes in flow perpendicular to the initial flow direction, such as a flow meeting a channel, may only slightly increase the length of the

535 bounding box for several iterations of decreasing μ (or increasing M/D). However, we did not observe this to be an issue within our training data for our given parameter ranges of μ and M/D.

4.2 Regionally optimizing runout models

As with any optimization problem, using a suitable cost function is critical to ensure the model parameters are optimized to solve a very specific problem. It can be difficult to define a single metric that simultaneously measures both performance of

540 path and distance simulations. As shown in this study, it may also not be necessary. The modular framework of the GPP model provides the ability to optimize two distinct runout components, the runout path including lateral spreading and the runout distance. In our study, we used the random walk and PCM components of the GPP model to simulate spatial extent of runout. By using the two-stage approach to optimization, where we first optimize the runout path model and then plugin those values to optimize the runout-distance model component, we can considerably reduce computational complexity, while ensuring that

545 our final model result explicitly optimizes for runout distance – a key characteristic for spatially predicting areas potentially impacted by debris flow runout. In our case, we only needed to solve two separate problems with 3 (random walk) and 2 (PCM) unknown parameters, as opposed to solving simultaneously for 5 unknowns. We used an exhaustive grid search to optimize the runout model components because it allowed us to visualize and assess model performance across parameter space. However, a drawback of this optimization method is that it can be computationally slow to explore all candidate

550 parameter combinations. If speed is a requirement for regional runout analysis, then a faster method like random search (Bergstra and Bengio, 2012) may be preferred (Schratz et al., 2019).

In terms of model improvement, analysis of the spatial distribution of optimal parameters may lead to better parameter optimization based on terrain or geological characteristics. The spatially varying values used in this study were based on modelling for alpine regions (Gamma, 2000), and do not account for the potential variability in sliding conditions between

555 different catchments (Guthrie, 2002). This challenge may be overcome by using regional modelling strategies already applied for landslide initiation susceptibility modelling, where the study area is divided into geologically similar regions, runout model optimization is performed for each region and then combined (e.g., Petschko et al., 2014).

5 Conclusions

Modelling the spatial pattern of debris flow runout in large, mainly remote areas, with sparse data, requires model calibration

560 and validation methods that ensure spatial transferability. In this study, we demonstrated that the combination of the statistical prediction (GAM) of source areas and our regional optimization of the GPP runout model (random walk and PCM) performed well at generalizing runout patterns across the upper Maipo river basin. In addition to high model performance, the transparency and interpretability of the GAM provided further confidence in the prediction of source areas by illustrating regionally geomorphically plausible modelled behavior. The optimized runout model parameters sets were consistently similar

565 within grid search space when assessing transferability using spatial cross-validation. We believe this strong transferability of our runout model was due to the hillslope gradient of the deposition area being one of the major controls of runout distance in

the PCM model. The regionally optimized runout model also resulted in geomorphically plausible results, with best performing μ and *M/D* combinations occurring when simulated debris flow deposition and termination occurred on slopes less than 11°. Although obtaining unique PCM parameter solutions for individual events can be an issue, we were able to obtain a unique PCM model solution for our regional model. In general, we found unique regional-optimal PCM model solutions were more prone with larger sample sizes, as well as higher model performance and lower uncertainties. Future improvements to our approach may include building a model that allows for more spatial variation in optimized parameters, especially in regions with better availability of soil physical information. Overall, our open-source modelling approach, which enhanced the GPP model by adding support for automated model calibration and transferability assessment, provides an accessible and extendible modelling framework for such advances in regional runout modelling.

Code availability

To go with this paper, we developed the runoptGPP R package for optimizing mass movement runout models using the random walk and PCM model components of the GPP tool in SAGA-GIS. It is available for download from the GitHub repository: https://github.com/jngtz/runoptGPP (DOI: 10.5281/zenodo.4428050). This package contains tutorials on how to apply the optimization methods regionally or for single runout events. The GitHub repository also contains the R code used to conduct our analysis.

Data availability

The data is available for download from Zenodo https://doi.org/10.5281/zenodo.4428080.

Acknowledgements

This research was funded by CETAQUA Chile on behalf of Aguas Andinas from the project titled "Mass movement processes in the upper Maipo basin: Modelling the susceptibility and probable sediment transfers", which was done in collaboration between the Friedrich Schiller University Jena and the Institute of Geography at the Pontificia Universidad Católica de Chile.

References

Aaron, J., McDougall, S., and Nolde, N.: Two methodologies to calibrate landslide runout models, Landslides, 16, 907–920, https://doi.org/10.1007/s10346-018-1116-8, 2019.

Aleotti, P. and Chowdhury, R.: Landslide hazard assessment: summary review and new perspectives, Bull Eng Geol Env, 58, 21–44, https://doi.org/10.1007/s100640050066, 1999.

Angillieri, M. Y. E.: Debris flow susceptibility mapping using frequency ratio and seed cells, in a portion of a mountain international route, Dry Central Andes of Argentina, CATENA, 189, 104504, https://doi.org/10.1016/j.catena.2020.104504, 2020.

Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., and Reichenbach, P.: Impact of mapping errors on the reliability of landslide hazard maps, Nat. Hazards Earth Syst. Sci., 2, 3–14, https://doi.org/10.5194/nhess-2-3-2002, 2002.

605   ASF DAAC: ALOS PALSAR Radiometric Terrain Corrected high resolution digital elevation model: Includes Material ©JAXA/METI, https://doi.org/10.5067/Z97HFCNKR6VA, 2011.

Bathurst, J. C., Burton, A., and Ward, T. J.: Debris Flow Run-Out and Landslide Sediment Delivery Model Tests, Journal of Hydraulic Engineering, 123, 410–419, https://doi.org/10.1061/(ASCE)0733-9429(1997)123:5(410), 1997.

Bergstra, J. and Bengio, Y.: Random search for hyper-parameter optimization, Journal of Machine Learning Research, 13, 610   281–305, 2012.

Bivand, R. and Rundel, C.: rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.5-2., https://CRAN.R-project.org/package=rgeos, 2019.

Bivand, R., Keitt T., and Rowlingson B.: rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.4-8., https://CRAN.R-project.org/package=rgdal, 2019.

615   Blahut, J., Horton, P., Sterlacchini, S., and Jaboyedoff, M.: Debris flow hazard modelling on medium scale: Valtellina di Tirano, Italy, Nat. Hazards Earth Syst. Sci., 10, 2379–2390, https://doi.org/10.5194/nhess-10-2379-2010, 2010a.

Blahut, J., van Westen, C. J., and Sterlacchini, S.: Analysis of landslide inventories for accurate prediction of debris-flow source areas, Geomorphology, 119, 36–51, https://doi.org/10.1016/j.geomorph.2010.02.017, 2010b.

Bordoni, M., Galanti, Y., Bartelletti, C., Persichillo, M. G., Barsanti, M., Giannecchini, R., Avanzi, G. D.'A., Cevasco, A., 620   Brandolini, P., Galve, J. P., and Meisina, C.: The influence of the inventory on the determination of the rainfall-induced shallow landslides susceptibility using generalized additive models, CATENA, 193, 104630, https://doi.org/10.1016/j.catena.2020.104630, 2020.

Brenning, A.: Spatial prediction models for landslide hazards: review, comparison and evaluation, Nat. Hazards Earth Syst. Sci., 5, 853–862, https://doi.org/10.5194/nhess-5-853-2005, 2005.

625   Brenning, A., Bangsm D., and Becker, M.: RSAGA: SAGA Geoprocessing and Terrain Analysis. R package version 1.3.0, https://CRAN.R-project.org/package=RSAGA, 2018.

Brenning, A.: Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest, in: 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 2012-07-22 - 2012-07-27, 5372–5375, 2012.

630 Brock, J., Schratz, P., Petschko, H., Muenchow, J., Micu, M., and Brenning, A.: The performance of landslide susceptibility models critically depends on the quality of digital elevation models, Geomatics, Natural Hazards and Risk, 11, 1075–1092, https://doi.org/10.1080/19475705.2020.1776403, 2020.

Cama, M., Lombardo, L., Conoscenti, C., and Rotigliano, E.: Improving transferability strategies for debris flow susceptibility assessment: Application to the Saponara and Itala catchments (Messina, Italy), Geomorphology, 288, 52–65, 635 https://doi.org/10.1016/j.geomorph.2017.03.025, 2017.

Campitelli, E.: metR: Tools for Easier Analysis of Meteorological Fields. R package version 0.7.0, https://CRAN.R-project.org/package=metR, 2020.

Carrara, A., Guzzetti, F., Cardinali, M., and Reichenbach, P.: Use of GIS technology in the prediction and monitoring of landslide hazard, Nat Hazards, 20, 117–135, https://doi.org/10.1023/A:1008097111310, 1999.

640 Cepeda, J., Chávez, J. A., and Cruz Martínez, C.: Procedure for the selection of runout model parameters from landslide back-analyses: application to the Metropolitan Area of San Salvador, El Salvador, Landslides, 7, 105–116, https://doi.org/10.1007/s10346-010-0197-9, 2010.

Chung and Fabbri: Probabilistic prediction models for landslide hazard mapping, Photogrammetric Engineering & Remote Sensing, 65, 1389–1399, 1999.

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

Dong, J.-J., Lee, C.-T., Tung, Y.-H., Liu, C.-N., Lin, K.-P., and Lee, J.-F.: The role of the sediment budget in understanding debris flow susceptibility, Earth Surf. Process. Landforms, 34, 1612–1624, https://doi.org/10.1002/esp.1850, 2009.

650 Fabbri, A. G., Chung, C.-J. F., Cendrero, A., and Remondo, J.: Is Prediction of Future Landslides Possible with a GIS?, Nat Hazards, 30, 487–503, https://doi.org/10.1023/B:NHAZ.0000007282.62071.75, 2003.

Galas, S., Dalbey, K., Kumar, D., Patra, A., and Sheridan, M.: Benchmarking TITAN2D mass flow model against a sand flow experiment and the 1903 Frank Slide, in: Proceedings of the 2007 International Forum on Landslide Disaster Management, Hong Kong, 10-12 December 2007 / edited by Ken Ho, Victor Li, edited by: Ho, K. and Li, V., Geotechnical Division, Hong 655 Kong, 899–918, 2007.

Gamma, P.: dfwalk – Ein Murgang-Simulationsprogramm zur Gefahrenzonierung, Geographica Bernensia, University of Bern, Switzerland, G66, 14pp, 2000.

Goetz, J. N., Guthrie, R. H., and Brenning, A.: Forest harvesting is associated with increased landslide activity during an extreme rainstorm on Vancouver Island, Canada, Nat. Hazards Earth Syst. Sci., 15, 1311–1330, https://doi.org/10.5194/nhess-15-1311-2015, 2015a.

Goetz, J. N., Brenning, A., Petschko, H., and Leopold, P.: Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling, Computers & Geosciences, 81, 1–11, https://doi.org/10.1016/j.cageo.2015.04.007, 2015b.

Goetz, J. N., Guthrie, R. H., and Brenning, A.: Integrating physical and empirical landslide susceptibility models using generalized additive models, Geomorphology, 129, 376–386, https://doi.org/10.1016/j.geomorph.2011.03.001, 2011.

Guthrie, R. H.: The effects of logging on frequency and distribution of landslides in three watersheds on Vancouver Island, British Columbia, Geomorphology, 43, 273–292, https://doi.org/10.1016/S0169-555X(01)00138-6, 2002.

Guthrie, R. H., Deadman, P. J., Cabrera, A. R., and Evans, S. G.: Exploring the magnitude–frequency distribution: a cellular automata model for landslides, Landslides, 5, 151–159, https://doi.org/10.1007/s10346-007-0104-1, 2008.

Hauser, A.: Rock avalanche and resulting debris flow in Estero Parraguirre and Río Colorado, Region Metropolitana, Chile, in: Catastrophic Landslides: Effects, Occurrence, and Mechanisms, Geological Society of America, 135–148, https://doi.org/10.1130/REG15-p135, 2002.

Heckmann, T., Gegg, K., Gegg, A., and Becht, M.: Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows, Nat. Hazards Earth Syst. Sci., 14, 259–278, https://doi.org/10.5194/nhess-14-259-2014, 2014.

Heckmann, T. and Schwanghart, W.: Geomorphic coupling and sediment connectivity in an alpine catchment — Exploring sediment cascades using graph theory, Geomorphology, 182, 89–103, https://doi.org/10.1016/j.geomorph.2012.10.033, 2013.

Hijmans, R. J.: raster: Geographic Data Analysis and Modeling. R package version 3.0-12, https://CRAN.R-project.org/package=raster, 2020.

Holmgren, P.: Multiple flow direction algorithms for runoff modelling in grid based elevation models: An empirical evaluation, Hydrol. Process., 8, 327–334, https://doi.org/10.1002/hyp.3360080405, 1994.

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

685   Hungr, O., Morgan, G. C., and Kellerhals, R.: Quantitative analysis of debris torrent hazards for design of remedial measures, Can. Geotech. J., 21, 663–677, https://doi.org/10.1139/t84-073, 1984.

Hungr, O.: A model for the runout analysis of rapid flow slides, debris flows, and avalanches, Can. Geotech. J., 32, 610–623, https://doi.org/10.1139/t95-063, 1995.

Ikeya, H.: Debris flow and its countermeasures in Japan, Bull Eng Geol Env, 40, 15–33, https://doi.org/10.1007/BF02590339, 690   1989.

Knevels, R., Petschko, H., Proske, H., Leopold, P., Maraun, D., and Brenning, A.: Event-Based Landslide Modeling in the Styrian Basin, Austria: Accounting for Time-Varying Rainfall and Land Cover, Geosciences, 10, 217, https://doi.org/10.3390/geosciences10060217, 2020.

Lombardo, L., Cama, M., Maerker, M., and Rotigliano, E.: A test of transferability for landslides susceptibility models under 695   extreme climatic events: application to the Messina 2009 disaster, Nat Hazards, 74, 1951–1989, https://doi.org/10.1007/s11069-014-1285-2, 2014.

Lombardo, L., Opitz, T., and Huser, R.: Point process-based modeling of multiple debris flow landslides using INLA: an application to the 2009 Messina disaster, Stoch Environ Res Risk Assess, 32, 2179–2198, https://doi.org/10.1007/s00477-018-1518-0, 2018.

700   Lorente, A., Beguería, S., Bathurst, J. C., and García-Ruiz, J. M.: Debris flow characteristics and relationships in the Central Spanish Pyrenees, Nat. Hazards Earth Syst. Sci., 3, 683–691, https://doi.org/10.5194/nhess-3-683-2003, 2003.

Loye, A., Jaboyedoff, M., and Pedrazzini, A.: Identification of potential rockfall source areas at a regional scale using a DEM-based geomorphometric analysis, Nat. Hazards Earth Syst. Sci., 9, 1643–1653, https://doi.org/10.5194/nhess-9-1643-2009, 2009.

705   Marchi, L. and D'Agostino, V.: Estimation of debris-flow magnitude in the Eastern Italian Alps, Earth Surf. Process. Landforms, 29, 207–220, https://doi.org/10.1002/esp.1027, 2004.

McDougall, S.: 2014 Canadian Geotechnical Colloquium: Landslide runout analysis — current practice and challenges, Can. Geotech. J., 54, 605–620, https://doi.org/10.1139/cgj-2016-0104, 2017.

Mergili, M., Krenn, J., and Chu, H.-J.: r.randomwalk v1, a multi-functional conceptual tool for mass movement routing, 710   Geosci. Model Dev., 8, 4027–4043, https://doi.org/10.5194/gmd-8-4027-2015, 2015.

Mergili, M., Schwarz, L., and Kociu, A.: Combining release and runout in statistical landslide susceptibility modeling, Landslides, 16, 2151–2165, https://doi.org/10.1007/s10346-019-01222-7, 2019.

Mergili, M., Fellin, W., Moreiras, S. M., and Stötter, J.: Simulation of debris flows in the Central Andes based on Open Source GIS: possibilities, limitations, and parameter sensitivity, Nat Hazards, 61, 1051–1081, https://doi.org/10.1007/s11069-011-9965-7, 2012.

Microsoft and Weston, S.: foreach: Provides Foreach Looping Construct. R package version 1.4.8., https://CRAN.R-project.org/package=foreach, 2020.

Moreiras, S., Lisboa, M. S., and Mastrantonio, L.: The role of snow melting upon landslides in the central Argentinean Andes, Earth Surf. Process. Landforms, 37, 1106–1119, https://doi.org/10.1002/esp.3239, 2012.

Moreiras, S. M. and Sepúlveda, S. A.: Megalandslides in the Andes of central Chile and Argentina (32°–34°S) and potential hazards, Geological Society, London, Special Publications, 399, 329–344, https://doi.org/10.1144/SP399.18, 2015.

Niculiță, M.: Automatic landslide length and width estimation based on the geometric processing of the bounding box and the geomorphometric analysis of DEMs, Nat. Hazards Earth Syst. Sci., 16, 2021–2030, https://doi.org/10.5194/nhess-16-2021-2016, 2016.

Pawley, S.: Rsagacmd: Linking R with the Open-Source 'SAGA-GIS' Software. R package version 0.0.2, https://CRAN.R-project.org/package=Rsagacmd, 2019.

Pebesma, E. and Bivand, R.: Classes and methods for spatial data in R, R News, 5, available at: https://cran.r-project.org/doc/Rnews/, 2005.

Pebesma, E.: Simple Features for R: Standardized Support for Spatial Vector Data, The R Journal, 10, 439, https://doi.org/10.32614/RJ-2018-009, 2018.

Perla, R., Cheng, T. T., and McClung, D. M.: A Two–Parameter Model of Snow–Avalanche Motion, J. Glaciol., 26, 197–207, https://doi.org/10.3189/S002214300001073X, 1980.

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

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.

R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/, 2019.

Rickenmann, D. and Zimmermann, M.: The 1987 debris flows in Switzerland: documentation and analysis, Geomorphology, 8, 175–189, https://doi.org/10.1016/0169-555x(93)90036-2, 1993.

Rudy, A. C.A., Lamoureux, S. F., Treitz, P., and van Ewijk, K. Y.: Transferability of regional permafrost disturbance susceptibility modelling using generalized linear and generalized additive models, Geomorphology, 264, 95–108, https://doi.org/10.1016/j.geomorph.2016.04.011, 2016.

Ruß, G. and Brenning, A.: Data Mining in Precision Agriculture: Management of Spatial Information, in: Computational Intelligence for Knowledge-Based Systems Design, edited by: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Hüllermeier, E., Kruse, R., and Hoffmann, F., Springer Berlin Heidelberg, Berlin, Heidelberg, 350–359, https://doi.org/10.1007/978-3-642-14049-5_36, 2010.

Sattler, K., Keiler, M., Zischg, A., and Schrott, L.: On the Connection between Debris Flow Activity and Permafrost Degradation: A Case Study from the Schnalstal, South Tyrolean Alps, Italy, Permafrost and Periglac. Process., 22, 254–265, https://doi.org/10.1002/ppp.730, 2011.

Schraml, K., Thomschitz, B., McArdell, B. W., Graf, C., and Kaitna, R.: Modeling debris-flow runout patterns on two alpine fans with different dynamic simulation models, Nat. Hazards Earth Syst. Sci., 15, 1483–1492, https://doi.org/10.5194/nhess-15-1483-2015, 2015.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., and Brenning, A.: Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data, Ecological Modelling, 406, 109–120, https://doi.org/10.1016/j.ecolmodel.2019.06.002, 2019.

Sepúlveda, S. A., Moreiras, S. M., Lara, M., and Alfaro, A.: Debris flows in the Andean ranges of central Chile and Argentina triggered by 2013 summer storms: characteristics and consequences, Landslides, 12, 115–133, https://doi.org/10.1007/s10346-014-0539-0, 2015.

Sepúlveda, S. A., Rebolledo, S., and Vargas, G.: Recent catastrophic debris flows in Chile: Geological hazard, climatic relationships and human response, Quaternary International, 158, 83–95, https://doi.org/10.1016/j.quaint.2006.05.031, 2006.

Serey, A., Piñero-Feliciangeli, L., Sepúlveda, S. A., Poblete, F., Petley, D. N., and Murphy, W.: Landslides induced by the 2010 Chile megathrust earthquake: a comprehensive inventory and correlations with geological and seismic factors, Landslides, 16, 1153–1165, https://doi.org/10.1007/s10346-019-01150-6, 2019.

Servicio Nacional de Geología y Minería de Chile: Geología de Chile Escala 1:1,000,000 (v.1.0), Publicacion Geologica Digital, No. 4.

Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T.: ROCR: visualizing classifier performance in R, Bioinformatics (Oxford, England), 21, 3940–3941, https://doi.org/10.1093/bioinformatics/bti623, 2005.

770   Stevenson, J. A., Sun, X., and Mitchell, N. C.: Despeckling SRTM and other topographic data with a denoising algorithm, Geomorphology, 114, 238–252, https://doi.org/10.1016/j.geomorph.2009.07.006, 2010.

Sun, X., Rosin, P., Martin, R., and Langbein, F.: Fast and effective feature-preserving mesh denoising, IEEE transactions on visualization and computer graphics, 13, 925–938, https://doi.org/10.1109/TVCG.2007.1065, 2007.

Taylor, F. E., Malamud, B. D., Witt, A., and Guzzetti, F.: Landslide shape, ellipticity and length-to-width ratios, Earth Surf.
775   Process. Landforms, 43, 3164–3189, https://doi.org/10.1002/esp.4479, 2018.

Tuanmu, M.-N., Viña, A., Roloff, G. J., Liu, W., Ouyang, Z., Zhang, H., and Liu, J.: Temporal transferability of wildlife habitat models: implications for habitat monitoring, Journal of Biogeography, 38, 1510–1523, https://doi.org/10.1111/j.1365-2699.2011.02479.x, 2011.

van Westen, C. J., van Asch, T.W.J., and Soeters, R.: Landslide hazard and risk zonation—why is it still so difficult?, Bull
780   Eng Geol Environ, 65, 167–184, https://doi.org/10.1007/s10064-005-0023-0, 2006.

Wenger, S. J. and Olden, J. D.: Assessing transferability of ecological models: an underappreciated aspect of statistical validation, Methods in Ecology and Evolution, 3, 260–267, https://doi.org/10.1111/j.2041-210X.2011.00170.x, 2012.

Wichmann, V. and Becht, M.: Rockfall modelling: methods and model application in an alpine basin (Reintal, Germany), in: SAGA – Analysis and Modelling Applications, edited by: Böhner, J., McCloy, K. R., and Strobl, J., 105–116, 2006.

785   Wichmann, V.: The Gravitational Process Path (GPP) model (v1.0) – a GIS-based simulation framework for gravitational processes, Geosci. Model Dev., 10, 3309–3327, https://doi.org/10.5194/gmd-10-3309-2017, 2017.

Wickham, H.: ggplot2: Elegant graphics for data analysis, Use R!, Springer, New York, viii, 212, 2009.

Wood, S. N.: Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73, 3–36,
790   https://doi.org/10.1111/j.1467-9868.2010.00749.x, 2011.

Zweig, M. H. and Campbell, G.: Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, Clinical Chemistry, 39, 561–577, https://doi.org/10.1093/CLINCHEM/39.4.561, 1993.

795