Landslide Susceptibility Assessment Tools v1.0.0b - Project Manager Suite: A new modular toolkit for landslide susceptibility assessment.

Jewgenij Torizin[1], Nick Schüßler[1], Michael Fuchs[1]

[1]Federal Institute for Geosciences and Natural Resources (BGR), Hanover, 30655, Germany

Correspondence to: Jewgenij Torizin (jewgenij.torizin@bgr.de)

Abstract. This paper introduces the Landslide Susceptibility Assessment Tools – Project Manager Suite (LSAT PM), an open-source, easy-to-use software written in Python. Primarily developed to conduct landslide susceptibility analyses (LSA), it is not limited to this issue and applies to any other research dealing with supervised spatial binary classification. With its standardized project framework, LSAT PM provides efficient interactive data management supported by handy tools. The application utilizes standard data formats ensuring data transferability to all geographic information systems. LSAT PM has a modular structure allowing to extend the existing toolkit by additional analyses. The LSAT PM v1.0.0b implements heuristic and data-driven methods such as the analytical hierarchy process, weights of evidence, logistic regression, and artificial neural networks. The software was developed and tested over the years in different projects dealing with landslide susceptibility assessment. The emphasis on model uncertainties and statistical model evaluation makes the software a practical modeling tool. Also, it provides the possibility to explore and evaluate different LSA models, even those not created with LSAT PM. The software distribution package includes comprehensive documentation. A dataset for testing purposes of the software is available. LSAT PM is subject to continuous further development.

1 Introduction

Gravitational mass movements or, more general, landslides occur in all mountainous parts of the world mainly in association with other natural hazards such as severe rainfalls, floods, and earthquakes (e.g., Polemio and Petrucci, 2000; Crozier, 2010; Keefer, 2002; Kamp et al., 2008). Landslides contribute to disastrous socio-economic consequences and yearly claim thousands of casualties worldwide (Petley, 2012; Froude and Petley, 2018).

Per definition, a comprehensive assessment of landslide hazards addresses the spatial and temporal landslide occurrence based on three questions: "where?", "when?" and "how large?" (e.g., Varnes, 1984, Guzzeti et al., 1999, Tanyaş et al., 2018; Reichenbach et al., 2018). The landslide susceptibility analysis (LSA) depicts the probability of spatial landslide occurrence (e.g., Brabb, 1985; Guzzetti et al., 2005) covering the spatial domain of the hazard analysis. Addressing the temporal domain of landslide hazard assessment is much more challenging (e.g., Aleotti and Chowdhury, 1999; Van Westen et al., 2006). Therefore, most case studies at regional scales focus on LSA as the most feasible part of the landslide hazard analysis.

30 Regional LSA is usually done based on qualitative (heuristic or knowledge-driven) and quantitative methods (e.g., Reichenbach et al. 2018). The quantitative techniques comprise physically-based and data-driven statistical, and machine learning (ML) approaches. Based on the desired analysis scale and available data, the application of some methods might be limited (e.g., Van Westen et al., 2008; Balzer et al., 2020).

With the advances in remote sensing, the data-driven methods for LSA have gained popularity. These methods belong mainly

35 to supervised binary classification (e.g., Torizin, 2016). In supervised classification, we use a set of recorded observations (labels) and independent explanatory factors (features) such as different geomorphologic, hydrologic, and geological conditions to train a statistic function, the so-called classifier. The classifier's task is, in a narrower sense, to distinguish based on available features whether a specific countable element in a study area (e.g., raster pixel or unique condition unit) is a landslide or non-landslide.

40 Because classification is one of the fundamental tasks in statistics and ML, many different classifiers exist. Consequently, in numerous case studies, the academic community continuously applies and compares classification algorithms and their variations, which were initially developed for other purposes but are sufficiently general to be used for LSA. While some classifiers might outperform others, the drawn conclusions are often valid only for particular settings and study designs (e.g., Balzer et al. 2020). Under other circumstances, such as different data quality or distribution, it is very likely that some of the

45 other classifiers perform on par or better. Reviewing the LSA research of the past 30 years, Reichenbach et al. (2018) counted about 163 different data-driven methods, emphasizing the problem of excessive experimentation with statistical classifiers rather than focusing on the LSA's reliability. Many of these methods have never been adopted or seriously considered by practitioners skeptically following the academic research at their own pace and utilizing a comparably small part of it. Thus, despite the many academic publications dealing with regional LSA, there are few practical solutions. Also, in the software

50 development in this field, compared to the available software applications in geo-engineering, user-friendly standalone applications dealing with regional LSA and especially with its data-driven methods are rare. Most available tools exist as academic code, implying that the user has the necessary programming or scripting skills to set up and run the tools. Thus, the user must program pipes or deal with several apps or packages at different assessment steps. Despite considerable efforts to adopt the earth science curricula to digital transformation (e.g., Hall-Wallace, 1999; Makkawi et al., 2006), the required

55 computational literacy to deal with current data-driven LSA is still not a broad standard in geosciences. However, there is a positive trend. Rising education possibilities on e-learning platforms with exhaustive offers in programming narrow the gap between geosciences and data sciences. So Bouziat et al. (2020) noted that the geoscience community increasingly uses Python for data processing, R for statistical analysis, and custom web services for sharing results.

Since 2010, most LSA tools have been available as plugins or extensions for ESRI ArcGIS, QGIS, and GRASS GIS (e.g.,

60 Polat, 2021; Torizin, 2012; Jebur et al., 2015; Bragagnolo et al., 2020a), scripts in R statistical package, e.g., LAND-SE (Rossi and Reichenbach, 2016), and very few standalone applications, e.g., GeoFIS (Osna et al., 2014).

With the Landslide Susceptibility Assessment Tools – Project Manager Suite (LSAT PM), we introduce an open-source (under GNU General Public License v3), standalone, easy-to-use tool that supports scientific principles of openness, knowledge

integrity, and replicability. Doing so, we want to share our experience in the implementation of heuristic and data-driven LSA

65   methods. Our primary goal is not to introduce as many algorithms as possible for LSA but to provide easy access to a selection of state-of-the-art methods representing groups of different approaches. Providing a convenient framework for model evaluation and uncertainty assessment, we want to highlight such methods' capabilities and limitations, making them more transparent.

2 LSAT PM Software

70   LSAT PM is a desktop application designed to support decision-makers and the scientific community in generating and evaluating landslide susceptibility models based on heuristic and data-driven approaches. LSAT PM's core is the weights of evidence (WoE) method (e.g., Bonham-Carter et al., 1989) widely applied in LSA in the past decades (e.g., Mathew et al., 2007; Moghaddam et al., 2007; Thiery et al., 2007; Neuhäuser and Terhorst, 2012; Teerarungsigul et al., 2015). It is simple to understand and provides a transparent computation algorithm. Enhanced by assessing uncertainties (e.g., Torizin et al., 2018;

75   Torizin et al., 2021), WoE becomes a robust tool for rapid analysis providing a good reference model to test against when exploring new methods. For example, it can be beneficial to investigate the data's dependencies or run several sensitivity analyses based on transparent WoE before applying more sophisticated multivariate statistical analysis techniques, e.g., logistic regression (LR) or ML algorithms such as artificial neural networks (ANN).

Besides WoE, LSAT PM offers two other popular data-driven LSA methods: LR (e.g., Lee, 2005; Budimir et al., 2015;

80   Lombardo & Mai, 2018;) and ANN (e.g., Lee and Evangelista, 2006; Pradhan and Lee, 2010; Bragagnolo et al., 2020b), both based on the *sklearn* library (Pedregosa et al., 2011; Buitinck et al., 2013). For heuristic analyses, LSAT PM embeds a module for the analytical hierarchy process (AHP). This decision support method finds application primarily for areas with insufficient observational data (e.g., Balzer et al., 2020; Panchal & Shrivastava, 2020).

2.1 Development history

85   The development of Landslide Susceptibility Assessment Tools (LSAT) started in 2011 with Python scripting within ESRI ArcGIS 10.0 Toolbox to support technical cooperation (TC) projects (Torizin, 2012). TC projects are usually not aimed at cutting-edge research but summarise, adapt, and implement scientific outcomes by following the *best-practice* approach. Since then, LSAT was continuously improved and tested at different development stages in case studies in Indonesia (Torizin et al., 2013), Thailand (Teerarungsigul et al., 2015), Pakistan (Torizin et al., 2017), China (Torizin et al., 2018) and Germany (Balzer

90   et al., 2020). Working with different data of varying quality helped us to develop efficient methodical workflows. It also enabled us to better understand the limitations of some methods and design practical approaches to assess model uncertainties (e.g., Torizin et al., 2021).

In the beginning, like many other tools, LSAT relied on a host GIS (ESRI ArcGIS) (Torizin, 2012). In 2017, we started to prototype LSAT as a standalone application with the "Project Manager Suite" extension in Python 2 and later in Python 3.

95 This development began within the sino-german scientific cooperation project (Tian et al., 2017) and continued in a cooperation project with german geological surveys (Balzer et al., 2020).

2.2 Software capabilities

LSAT PM v1.0.0a is written in Python 3 (Van Rossum & Drake, 2009). PyQt5 provides the basis for the Graphical User Interface (GUI). Geospatial Data Abstraction software Library (GDAL) (GDAL/OGR Contributors, 2021) and its Python
100 bindings are the basis of the core functions of LSAT PM. Currently, LSAT PM can utilize raster data and vector data formats, such as Tagged Image File Format (GeoTiff), ESRI shapefiles, Keyhole Markup Language (KML), and JavaScript Object Notation (GeoJSON). Further GDAL-supported formats are incorporable on demand.

Complementary to the spatial data output in the same formats as input, LSAT PM supports table exports to Microsoft Excel-Sheets, graphs to portable network graphic files (PNG), and automated analysis reports to Microsoft Word-Documents.

105 For spatial analysis, LSAT PM implements basic geoprocessing functionalities for data preparation. Morphological analyses, such as slope, aspect, topographic position index (TPI), and many others, can be performed based on raster datasets in GeoTiff format. Functions such as euclidean distance and raster classification are also available. A simple data viewer visualizes raster data. Table 1 provides an overview of implemented tools.

However, although LSAT PM provides some GIS capabilities for geoprocessing, it cannot be characterized as a solid GIS
110 application and was never supposed to become one. It has a slim structure tailored to manage and prepare the data for the targeted binary spatial classification.

Table 1: Overview of LSAT PM's tools.

Category	Tool domain	Tools (with GUI support)
PROJECT	Project	• Create project • Open project • Project info
	Settings	• Language
	Context Menu	• Attribute table (raster, vector) • Data properties • View data
DATA	Import	• Import raster • Import inventory
	Vector Tools	• Random sampling • Subset by attribute • Geoprocessing tools
	DEM Tools	• Slope • Aspect • Hillshade • Topographic positional index (TPI) • Roughness

		• Topographic ruggedness index (TRI)
		• Euclidean distance
		• Contingency
	Raster Tools	• Combine
		• Reclassify
		• Lookup
		• Sensitivity reclassification
		• Reproject
	Visualization	• Viewer
		• Weights of Evidence (WoE)
	Analysis	• Logistic Regression (ANN)
		• Artificial Neural Networks (ANN)
ANALYSIS		• Analytic Hierarchy Process (AHP)
		• Model Builder
	Model Management	• Model Info
		• Zoning

3 Test data

115 A dataset to test the functionalities of the software is available separately from https://doi.org/10.5281/zenodo.5109620. The example or test dataset is an excerpt of the data collected in the German-Pakistani technical cooperation project Georisk Assessment Northern Pakistan (GANP), carried out by the Federal Institute for Geosciences and Natural Resources and Geological Survey of Pakistan (e.g., Torizin et al., 2017). The dataset covers about 664 square kilometers, including parts of the Kaghan and Siran Valleys in Khyber Pakhtunkhwa, Northern Pakistan. Table 2 and Fig.1 provide an overview of the test

120 data.

A severe Kashmir Earthquake struck the test data area on October 8th, 2005, with a moment magnitude of 7.8, triggering thousands of landslides. The collected landslide inventory results from the visual interpretation of optical satellite images available through Google Earth (GE), which, during the data acquisition, consisted mainly of imagery from Quickbird (up to 0.60 m ground resolution), IKONOS (~4 m ground resolution), SPOT (SPOT5 about 5 m ground resolution), and Landsat (15

125 m ground resolution) (Torizin et al., 2017). In total, the landslide inventory includes 3819 events for the test area depicted as polygons. Landslide sizes range from about 12 square meters to about 88444 square meters representing the depletion area of the landslides (as far as it was possible to determine by visual interpretation of imagery).

The digital elevation model is the ALOS Global Digital Surface Model (AW3D30) (JAXA, 2017) with a ground resolution of approximately 30 m (Fig. 1a). The geological information and the tectonic features (faults) were derived from the geological

130 map of Calkins et al. (1975) (Fig. 1b). The landcover results from the supervised classification on Landsat imagery performed by Fuchs et al. (2015) (Fig. 1c).

5

Geoscientific
Model Development
Discussions

Table 2: Overview of the datasets in the test data.

Data layer	Data Type	Source
Geology	raster, shapefile (.tif, .shp)	Calkins et al. (1975)
Landcover	raster, shapefile (.tif, .shp)	Fuchs et al. (2015)
AW3D30	raster (.tif)	JAXA (2017)
Confirmed faults	shapefile (.shp)	Calkins et al. (1975)
Landslides	shapefile (.shp)	Torizin et al. (2017)
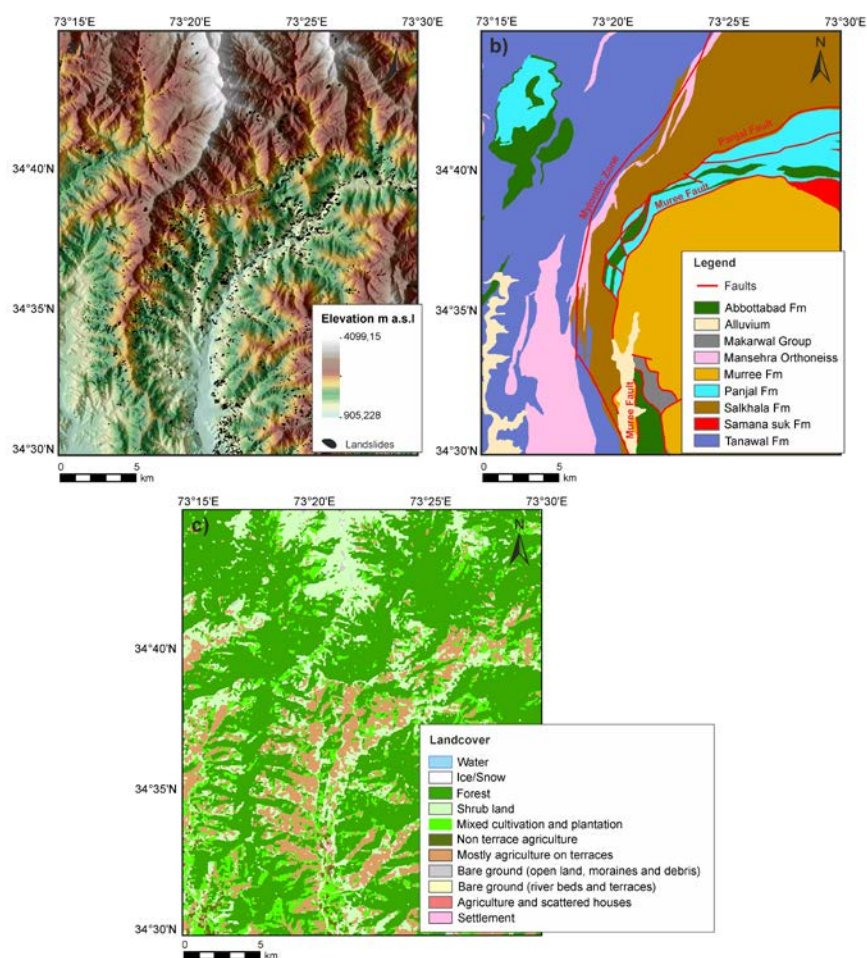
135 **Figure 1: Example dataset within LSAT PM. a) AW3D30 with overlaid by landslide inventory; b) Geology with prominent (confirmed) faults; c) Landcover.**

4 Working with LSAT PM

LSAT PM provides handy tools to set up the model through data exploration, preprocessing, analysis, model evaluation, and postprocessing.

140 The logical workflow in Fig. 2 schematically sketches the working process with LSAT PM. The following sections briefly introduce the single steps of this workflow and their corresponding modules. For more details, we refer to the software documentation distributed standalone or as part of the installer package (see also Sect. 8).
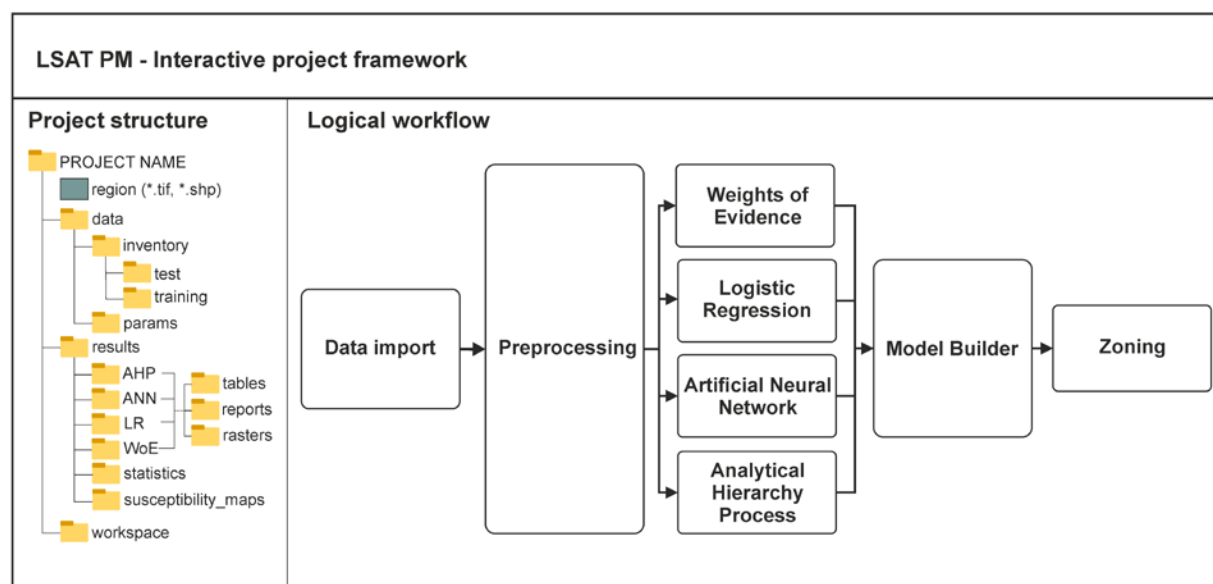

Figure 2: Standardized project structure and the logical workflow of LSAT PM.

145 **4.1 Project**

After launching the software, the user can create a new or open an existing project. The LSAT PM project has a standardized predefined structure, as shown in Fig. 2. The project structure includes a specified spatial reference that needs to be assigned manually or by selecting an already existing dataset as a project reference.

4.2 Data import

150 Data import distinguishes between *Import Raster* and *Import Inventory* (Fig. 3). The first tool imports datasets considered to represent independent exploratory factors or features, the second imports vector-based datasets for observational data, the so-called inventory or labels.

The raster import tool ensures the data's consistency by validating imported datasets against the project reference datasets (region). If not consistent with the project reference, the data gets warped accordingly. This procedure is comparable to the 155 concept of regions found in GRASS GIS (GRASS Development Team, 2021) and helps avoid processing errors due to specific

7

Geoscientific
Model Development
Discussions
Open Access
EGU

resolution and spatial reference inconsistencies. During the import, LSAT PM generates copies; thus, original files remain unchanged.

Inventory import supports splitting the dataset into training and test datasets (optional). The data split is an essential step in modern data-driven approaches (e.g., Xu and Goodacre, 2018), which is employed to evaluate the data-driven model by a test
160   dataset not involved in the model's training.
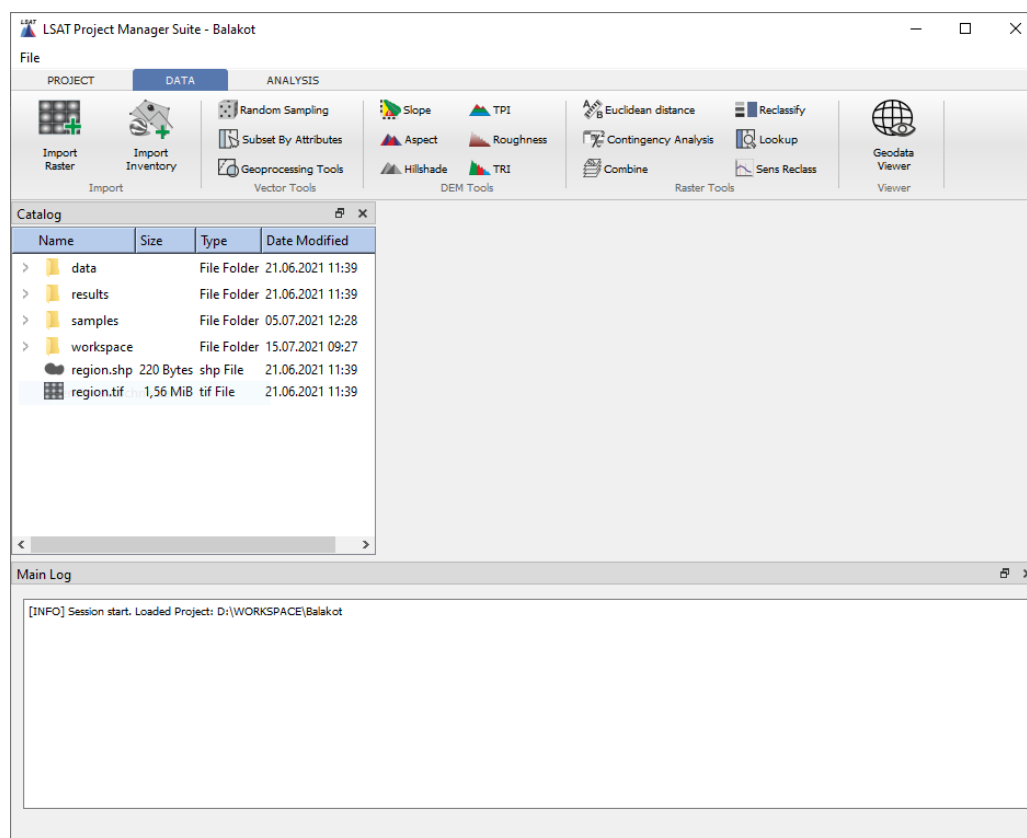
Figure 3: Main GUI with activated *Data* tab. The *Data* tab exhibits tools for data import, vector data processing, DEM tools, tools for raster processing, and a simple data viewer.

4.3 Data exploration and preprocessing

165   In the preprocessing step, we derive parameters and prepare the data according to the requirements of the upcoming analysis. So the DEM in the test dataset serves as a basis for morphometric parameters such as e.g., slope, topographic position index, or aspect. Using the combine tool, the user can combine several discrete datasets to generate a new dataset exhibiting unique conditions of higher complexity. Alternatively, the user can convert vector data unsuitable for spatial analysis such as tectonic features or roads into meaningful parameters depicting their spatial effect, e.g., over the distance, by generating a euclidean
170   distance raster. In this step, the user also can generate datasets more suitable for the analysis type. For example, the implemented WoE analysis utilizes discrete data only. Thus, continuous raster data such as slope or distance rasters need to be

Geoscientific
Model Development
Discussions

classified. In LSAT PM, the user can perform the raster classification task by the tools *Reclassify* or the so-called *Sensitivity reclassification*, which serves as a sensitivity analysis to find optimal class boundaries (e.g., Torizin et al., 2017).

Also, preprocessing of inventory data is available. Depending on the study design, it might be necessary to subset the data.

175 LSAT PM offers tools for random sampling and subsetting by attributes allowing subdividing datasets by spatial or temporal criteria. Other vector geoprocessing tools allow operations such as, e.g., erase, intersect, or union datasets.

These preprocessing tools apply to datasets in any location. Thus, it is not obligatory to import all datasets into the project before preprocessing. However, the crucial point is to use the LSAT PM's built-in import tools to import preprocessed data afterward.

180 Contingency analysis is the only tool in the tool domain *Raster data*, which produces an output file that stores the calculated contingency tables in the subfolder *statistics* of the projects' results folder. This tool represents an analysis helping to explore the dependencies (associations) in the discrete datasets rather than a preparation step.

4.4 Analyses

As already introduced in Sect. 2, LSAT PM implements heuristic and data-driven methods for LSA representing different
185 categories (e.g., bivariate and multivariate). All of the methods have different levels of complexity, which need to be accounted for when choosing a specific analysis method. Table 3 briefly summarizes the corner points of the approaches such as category, supported data types, and complexity. The introduced complexity is a subjective measure that we assigned based on our experience evaluating criteria such as needed mathematical background, the complexity of data preparation, model structure and computation algorithm (transparency), and interpretability of the results. The more complex methods might appear easier
190 to apply depending on the automation degree because the user has to do only a few steps. However, this deception will vanish once the user encounters the advanced settings of the methods.

Table 3 Analyses included in LSAT PM and their specifications

Analysis name	Category	Supported data types	Complexity
Weights of Evidence	data-driven, bivariate	discrete	moderate
Logistic regression	data-driven, multivariate	continuous, discrete	high
Artificial Neural Network	data-driven, multivariate	continuous, discrete	very high
Analytic Hierarchy Process	heuristic	discrete	low

4.4.1 Weights of Evidence

195 WoE is a bivariate statistical approach estimating the association between the observational data (dependent variable represented by the training landslide inventory) and potential controlling factors (independent variables represented by, e.g.,

9

geological conditions), analyzing them separately. The analysis output is a raster of the specific controlling factor containing logarithmic log-likelihood weights, which characterize the relationship of discrete factor classes with a landslide occurrence. Afterward, individually weighted factors are overlain into a linear model to obtain the overall landslide susceptibility pattern (Torizin et al., 2017).

LSAT PM offers three different analysis modes. The default option is simple cross-validation. With this option, the analysis runs once with the entire training dataset. The two other options are the so-called on-the-fly subsampling and subsampling with predefined samples. For on-the-fly subsampling, the weight estimation runs for several user-defined iterations taking random samples (without replacement) of user-defined size. The estimated weights are mean values from all iterations. The analysis with predefined samples utilizes predefined sample datasets in a specified folder location. The predefined samples must be created beforehand by any subsetting algorithm introduced above. Likewise to the random subsampling, the computed weights are mean values.

After the training, the result table and the weighted raster are automatically exported into the corresponding result folders.

The model generation process is performed in the next step using the LSAT PM model builder module (see Sect. 4.5).

4.4.2 Logistic regression

Logistic regression (LR) is a multivariate statistical classification method to estimate relationships between the dependent variable and independent controlling factors. In contrast to WoE, LR analyzes the relationships for all controlling factors at once and can utilize both continuous and discrete data for independent variables.

In LSAT PM, LR runs more automated than WoE. For the parameters, the user has to determine if the parameter is a discrete or continuous variable. The data preparation process runs automatically. The continuous datasets are scaled using a so-called min-max-scaler to the value range between 0 and 1; discrete datasets are transformed into binary dummy variables. All setting options for the logistic regression, e.g., regularisation or solver algorithm, implemented in the *sklearn* library, are adjustable upon the user's needs in the advanced settings GUI. After the training, the result table and the prediction raster are automatically exported into the corresponding result folders.

Unlike WoE, where an additional modeling step is required to overlay the weighted parameter raster to a model, the analysis output from multivariate LR already provides the multi-parameter landslide susceptibility model.

4.4.3 Artificial Neural Network

ANNs are computer models inspired by biological neural networks. They consist of a network of artificial neurons to simulate information processing, information storage, and learning. The structure of an ANN usually consists of an input layer, one or more hidden layers, and an output layer. The number of hidden layers determines the depth of an ANN (e.g. Schmidhuber, 2015, Hernández-Blanco et al., 2019). This structure is also known as the Multi-Layer Perceptron (MLP). The layers are composed of neurons in which the information processing takes place. Theoretically, the number of hidden layers and their neurons is unlimited. Thus, the network design is strongly dependent on the complexity and non-linearity of the task and the

available processing capacity. Most ANNs utilized in LSA are the so-called feedforward networks, which have a multi-layer

230   perceptron (MLP) structure with usually one hidden layer (e.g., Ermini et al., 2005; Lee and Evangelista, 2006; Alimohammadlou et al., 2014).

The implemented module for artificial neural network (ANN) has an experimental status and runs comparable to the LR. The routine allows the user to flag the variables as discrete or continuous. Additionally, the user can specify the ANN properties. Among others, these settings comprise the number of neurons in the layers, the number of hidden layers, the activation function,

235   and the solver to use. The design of the network and the activation function selection is one of the most sensitive steps of the analysis with ANN. Despite reviewing numerous studies, we do not have a straightforward, practical recipe for designing the network yet. Therefore, the default settings do not represent the best practical approach but rather the defaults delivered with the *sklearn* library.

Nevertheless, the GUI and included data preprocessing provides easy and fast access to the capabilities given by the *sklearn*

240   library and allow exploring the ANN performance for LSA. For more information, we refer to the LSAT PM documentation and corresponding references for the *sklearn* library.

The data scaling process runs identically to the LR module. After the training process, the output raster defining the spatial probability of landslide occurrence is exported to the corresponding result folder.

Comparable to LR, the resulting model raster can be imported to the model builder tool, compared with other models, or

245   evaluated with other inventories.

4.4.4 Analytical Hierarchy Process

As a heuristic approach, the AHP is different from the data-driven applications explained above. The user must specify the weights, usually based on his general expertise (knowledge of geological processes) and specific knowledge about the investigation area. The weighting process is a pairwise comparison of the inputs at different hierarchical levels. For LSA, the

250   AHP typically takes two hierarchical levels. The first level controls the class priorities inside multi-class parameters (e.g., raster values), the second sets the priorities between the multi-class parameters (e.g., raster datasets).

The pairwise comparison is a complex task when working with parameters exhibiting a large number of classes. The human's ability to compare is limited to approximately seven objects, plus or minus two (Miller, 1956). Saaty (1977) considered this proposing rank values between 1 and 9. Therefore, in the preparation process, it is advisable to reduce the number of classes

255   by generalization or subdivision in different hierarchical groups (e.g., Balzer et al., 2020). The latter will make the hierarchy more complex.

The advantage of pairwise ranking compared to the simple ranking is the ability to verify the logical consistency of the decision mathematically. Therefore, the AHP introduces a consistency ratio CR that indicates whether the introduced ranking is a logical inference or a random guess. Saaty (1980) recommends that the CR value should not exceed 0.1 for consistent assessment.

260   Notable is that some studies applying AHP for LSA use a hybrid approach combining bivariate methods with AHP (e.g., Kamp et al. 2008, 2010). In the hybrid approach, the first hierarchical level is assessed using a data-driven bivariate approach, e.g.,

WoE. Afterward, additional expert-based weights derived from the AHP priority vector are applied to overlay the parameters to the model. Such an approach can preserve the crude generalization of the patterns in the first hierarchical level making the analysis applicable with more detailed datasets. For the AHP part, the method becomes more applicable by involving the expert

265   weights at a higher hierarchical level, which benefits more from general process understanding than detailed local knowledge. Conversely, it also has implications for the bivariate analysis part. Using conditionally dependent parameters becomes less critical since experts adjust the parameter's contribution in the upper hierarchical level. However, the hybrid approach is only possible if a sufficient number of observations for the first data-driven step is available.

Currently, the implemented AHP is a pure expert-based tool supporting two hierarchy levels. The user has to perform the

270   pairwise ranking for single parameter classes in the first step and the parameters in the second step.

The output consists of the landslide susceptibility model. Like LR and ANN, the results raster can be evaluated in the model builder if a landslide inventory is available. In the absence of observational data in urbanized regions, local geotechnical models or existing structural measures such as slope reinforcement may provide appropriate evaluation points (e.g., Balzer et al., 2020).

275   4.5 Model builder

The model builder is the core module for generating models. It can combine WoE outputs to a model in an additive procedure (e.g., Torizin et al., 2017) and evaluate all model types generated within LSAT PM and foreign models in supported data formats.

The model evaluation is basing on the computation of the receiver operating characteristics (ROC) curve. The ROC curve is a

280   technique to visualize and evaluate the classifier's performance (e.g., Fawcett 2006) by depicting the ratio of the True Positive Rate (Sensitivity) and the False Positive Rate (1-Specificity). The area under the ROC curve (AUC or AUROC) provides a quantitative measure to compare the goodness of different models. We extended the evaluation procedure with the ROC curve by subsampling mode to assess sampling errors, as introduced by Torizin et al. (2018, 2021). Exploring the variance of the training dataset in conjunction with the test ROC curve helps to understand whether the uncertainty of the model is the property

285   of the insufficient data or the model accuracy.

The user can also generate hybrid models or model ensembles combining outputs from different classifiers. Therefore, the model builder supports the generation of complex mathematical expressions. The model metadata includes the expression used to build the model, providing the necessary transparency.

4.6 Zoning

290   The zonation procedure applies to all models generated or evaluated with LSAT PM. It uses the model's ROC curve to aggregate the model output with many different landslide susceptibility index (LSI) values to a legible map with few susceptibility zones (e.g., Chung and Fabbri 2003). Therefore, the user must specify the proportions of landslide areas that should fall within the specific zone. In our case studies (e.g., Torizin et al. 2017, Torizin et al. 2018), we usually used the

proportions of 50% of all landslides or landslide areas in the *Very high*, further 30% for the *High*, 15% for *Moderate*, 4% for

295 *Low* and about 1% for the *very low* susceptibility zone (Fig. 4). The implemented tool uses these values as the default table.

However, there are no well-established standards for the number of zones or the definition of zone boundaries. Therefore, the

user can freely specify the number of classes and the proportion of landslides in them.

Alternatively, the user can use the *reclass tool* to aggregate the model to zones with user-specified thresholds on the model's
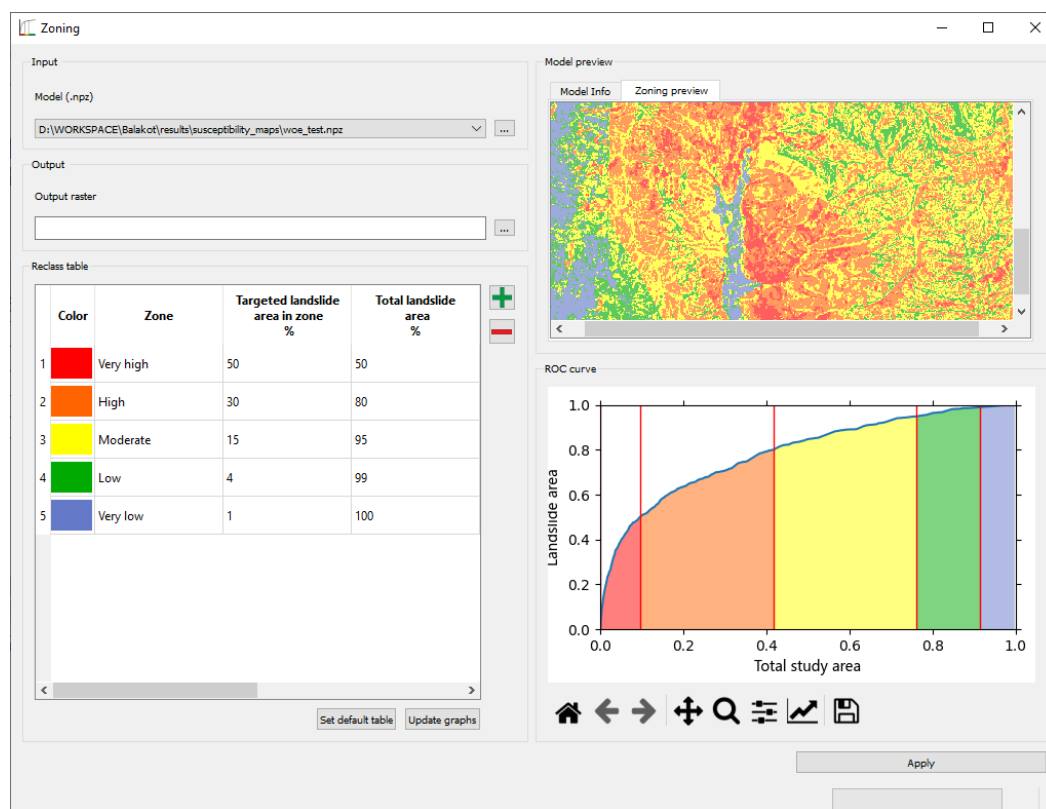
LSI or probability values.

300

Figure 4: Zoning tool with applied default classification on the WoE results of the Kaghan Valley dataset.

5 LSA with the test data

The included Kaghan Valley test dataset serves for exploration of the software. In the following example, we use the Khagan

Valley dataset to showcase how to perform a simple LSA based on WoE, LR, and ANN in LSAT PM and compare the model

305 outputs. We skip the analysis with AHP in this example due to its high subjectivity and our lacking detailed knowledge about

the investigation area.

5.1 Step by step procedure

To get the results, we performed the following steps:

- Create an LSAT PM project utilizing the test dataset AW3D30.tif as a project reference (mask)

310
- Import landslide inventory dataset (landslides.shp), setting the random seed to 42 and the training sample size to 70%
- Import raster datasets AW3D30.tif, geology.tif, and landcover.tif
- Calculate the slope from AW3D30.tif based on Horn (1981) algorithm using the included slope tool.
- Reclass generated slope raster to defined intervals of 5-degree using the reclass tool
- Generate a euclidean distance raster based on the confirmed_faults.shp using the euclidean distance tool

315
- Reclass the euclidean distance raster, based on the sensitivity analysis with ten quantiles and the training inventory, into five classes [0-637, >637-1150, >1150-1831, >1831-2727, >2727] using the sensitivity reclass tool. Note that the reclassify tool currently provides better handling for reproducing the analysis results with given interval values.

The following two steps prepare the data for evaluation of the model uncertainties. Therefore, we compared the size of the landslide training and test datasets. The training dataset has 2674 events; the test dataset has 1146, which is about 43% of the

320 training dataset. To estimate the sample size-dependent model variance, we generated 100 random subsamples from the training dataset with the size of the test dataset. Therefore, we used the random sampling tool. To make the results reproducible also here, we set the random seed to 42. The idea behind it is simple. The training and test inventory are random subsamples from the same dataset with a given spatial distribution. Therefore, they represent the same spatial distribution but with a mean sampling error (MSE) related to the sample size. Torizin et al. (2021) show that evaluation of the model performance based

325 on a test inventory of a smaller size than a training inventory must consider this for correct interpretation of the results. Evaluation with the 100 subsamples of the same size as the test inventory but derived from the training inventory dataset has two implications. First, all events are known to the model and follow the same spatial distribution as training inventory. Variations in model performance on these datasets define the MSE, which is expected to be in a similar range by evaluating the model with a test dataset of the corresponding size. The ROC curve and AUC value should fall within the MSE range if

330 the model generalizes well without significant overfit.

- Generate in the project data/inventory/ subfolder samples.
- Run the random sampling tool to generate 100 subsamples from the training inventory with random seed 42 and a sample size of 43%.

In the analysis part, we calculated the models with WoE, LR, and ANN in different ways to contrast the approaches. The WoE

335 can utilize only discrete data; therefore, we used the classified and initially discrete data to generate the weighted layers. The LR and ANN support usage of both: continuous and discrete data. Thus, we used the capability of both approaches to utilize discrete and continuous data and generated two models for LR and ANN, respectively. The first marked with a _c suffix utilizes both data types, the second marked by _d suffix utilizes only discrete datasets as used in WoE. In detail, the steps are the following:

340
- WoE-analysis with discrete layers: geology, landcover, classified slope, and classified distance to faults

- Build the WoE model in model builder by adding single weighted layers to the model. To better compare the results from WoE with results from ANN and LR, we transformed the log-likelihoods of the WoE model into probability space, applying the logistic function (see also Sect. 7.1 for details).

- LR_c with discrete layers geology and landcover and continuous layers slope and distance to faults (computation with
345   default settings, except the random state, which we set as Integer, 42)

- LR_d with discrete layers geology, landcover, classified slope, and classified distance to faults (computation with default settings, except the random state, which we set as Integer, 42)

- ANN_c with discrete layers geology and landcover and continuous layers slope and distance to faults computation with default settings, except the random state, which we set as 42)

350   - ANN_d with discrete layers geology, landcover, classified slope, and classified distance to faults computation with default settings, except the random state, which we set as 42)

For model comparison, we used the model builder. Therefore, in the advanced settings of the model builder tool, we selected the *predefined subsamples* option. The model performance is evaluated based on subsamples from the folder "samples" generated in the steps above. The step for this is:

355   - Compare the WoE, LR_c, LR_d, ANN_c, and ANN_d models in Model Builder computing the ROC curves with the predefined subsets and the test inventory.

Finally, we used the ROC curve from the test dataset to generate legible susceptibility maps consisting of five susceptibility zones using the *Zoning tool*. For the definition of the susceptibility zones, the default table was used.

5.2 Results and discussion

360   All three approaches deliver comparably good models (Fig. 5). Figure 5 shows the performance of the models based on the predefined subsamples. The greyish error band around the mean ROC curve indicates the MSE. As we can see from the violin plot inserts in Fig. 5a-e, the variance of the corresponding AUC values shows the normal distribution. At first glance, ANN_d and ANN_c models have the best training performance with AUCs of 0.84 and 0.83, followed by LR_c with an AUC of 0.81. WoE and LR_d model show with AUC of 0.80 worst training performance. To see why ANN_d performed better than the

365   other models using discrete data only, we have to look at the test performance of the models.

We see that for the WoE, LR_d, and LR_c, the test ROC curve is within the expected MSE. For ANN_c and ANN_d, however, the test performance is significantly lower. In this case, we can interpret this as an overfit of the models. Using discrete variables in ANN_d, we introduced additional degrees of freedom compared to the ANN_c model; therefore, overfitting is more prominent in the ANN_d model. The reason is the flexibility of the ANNs with multiple neurons to fit also non-linear data

370   relations, which might be an advantage but at the same time a significant uncertainty source. Thus, if we had aimed to optimize the susceptibility map with ANN, we would need to review the network design or number of iterations in the network training process to prevent overfitting. At this point, it is worth noting that imbalanced samples can also cause comparable effects.

When using polygon landslide data, the imbalance may develop from rare large landslides appearing only in the training or the test inventory. Although this affects all model types, flexible ANN might suffer more than, e.g., WoE. The solution would

375   be to check the distributions of the training and test landslide datasets and if imbalances are present to generate balanced samples by, e.g., randomly drawing pixels from landslide areas instead of using them as a whole.

Because the interpretation of ANN results is not intuitive, we would generally recommend a parallel application of a multivariate linear model and ANN to see how much non-linearity is introduced by the ANN and how it affects the model generalization capabilities. LSAT PM serves well for that purpose.

380   Further, looking at the test ROC curves of all models (Fig. 5f), we see that the predictivity of the models is comparable with minor advantages for models utilizing continuous datasets. Thus, given the simple study design and available data, the models are equivalent alternatives from the statistical point of view. However, although the ROC curve provides a quantitative measure for classifier performance, as any statistical measure, it is not suitable for evaluating the model's reliability (e.g., Rossi et al., 2010). As we can demonstrate here, models with equivalent AUCs can exhibit substantially different susceptibility patterns.

385   How meaningful those patterns are is beyond the statistical analysis capabilities and has to be verified based on other sources of information.
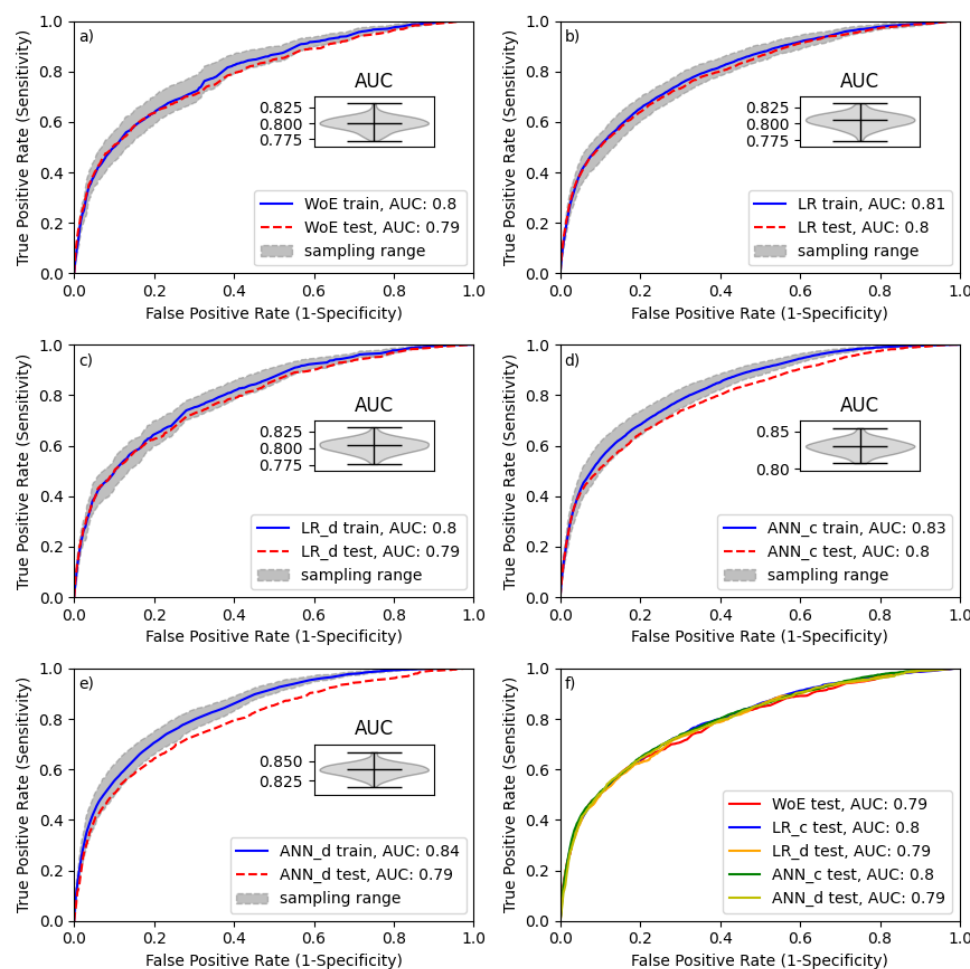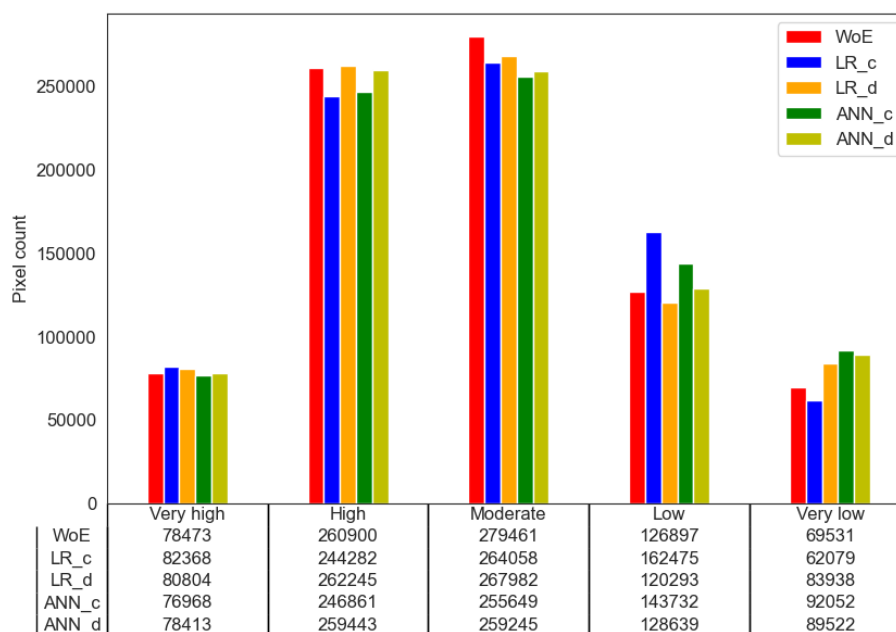
Figure 5 ROC curves for the different models. The greyish band marks the model uncertainty based on the MSE. The insert in a-e shows the corresponding distribution of AUC values for the utilized samples.

390   To evaluate the differences in obtained susceptibility patterns, we compare the susceptibility models in a pair plot. The pair plot in Fig. 6 visualizes the pairwise pixel-by-pixel comparison of the model values. The matrix diagonal shows the distribution of the model values (marginal probabilities). In contrast, the scatter plots in the lower matrix corner of Fig. 6 show the covariance of the value pairs overlain by linear regression to emphasize the trend. The pairwise comparison reveals general linear relation for all models with better comparability for multivariate models ANN and LR but substantial differences in

395   detail. In fact, dependent on the data composition and used approach, we generated five different susceptibility patterns.
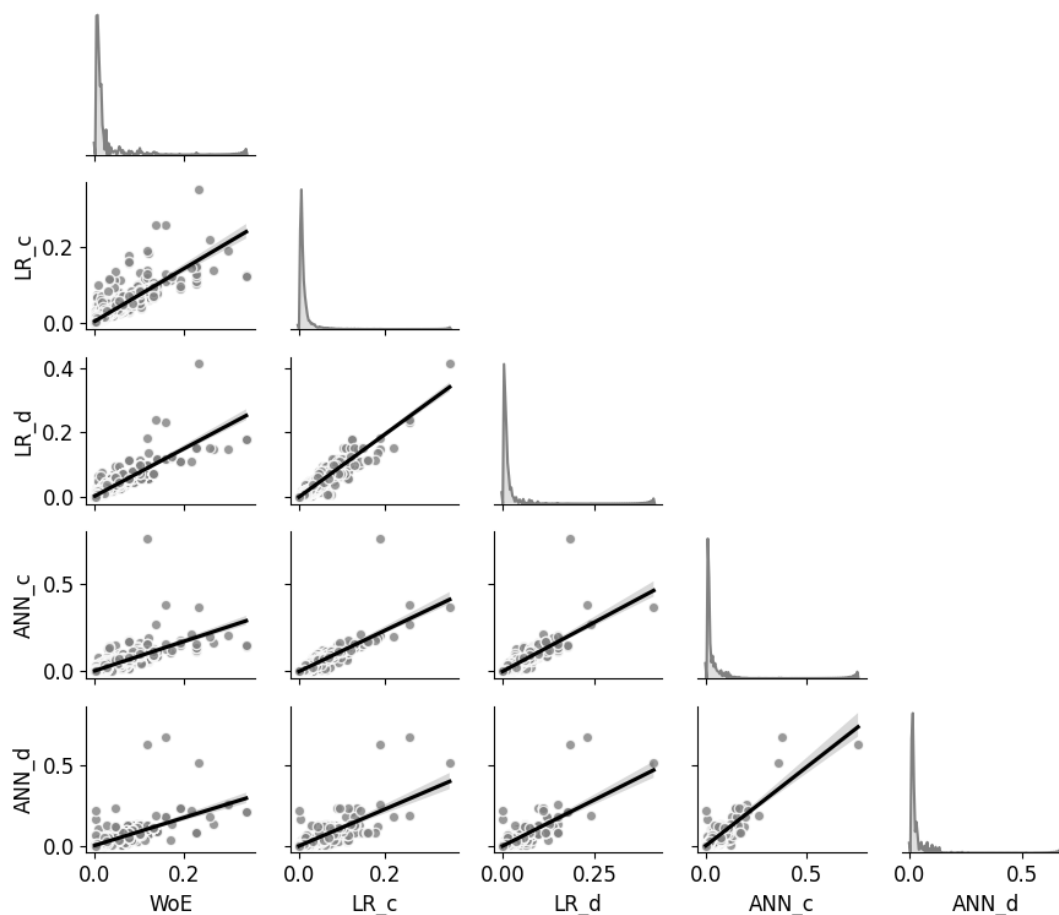
Figure 6 Pairwise comparison of the predicted landslide susceptibility patterns.

We wanted to see how these differences affect the final zonation and compared the models using simple class frequencies statistics after the zonation procedure. In Fig. 7, the classified models are compared regarding their pixel counts within the
400 susceptibility classes. While the highest susceptibility class containing 50% of all landslide areas show minor differences, they become more significant in the lower susceptibility zones.

	Very high	High	Moderate	Low	Very low
WoE	78473	260900	279461	126897	69531
LR_c	82368	244282	264058	162475	62079
LR_d	80804	262245	267982	120293	83938
ANN_c	76968	246861	255649	143732	92052
ANN_d	78413	259443	259245	128639	89522

Figure 7 Susceptibility zones distribution for the susceptibility models. *Very high* **contains about 50%,** *High* **about 30%,** *Moderate* **about 15%,** *Low* **about 4% and** *Very low* **about 1% of all landslide areas. The values in columns show the number of pixels within the zones.**

6 Conclusions and Outlook

LSAT PM streamlines the application and evaluation of popular data- and knowledge-driven spatial binary classification methods. The tool further provides the ability to build and evaluate hybrid models and model ensembles and evaluate any other models. With the provided test dataset, users can explore different analysis techniques at different complexity levels.

Of course, there is always room for improvement. Therefore, the LSAT PM is subject to further development. We intend to implement additional and improve existing methods, e.g., improve the AHP and machine learning workflows. Especially for the latter, we intend to implement additional features to visualize the results and increase the interpretability of the model outputs. To support massive datasets in machine learning methods, we intend to implement TensorFlow (Tensor Flow Team, 2015), adding GPU computation capability. Also, we want to implement new features such as a plugin builder tool that would support the easy implementation of customized plugins into the LSAT PM framework. Software documentation is going to be extended by short video tutorials introducing the work with LSAT PM.

With the open-source approach, we would like to encourage interested scientists to join the development, introducing and discussing new ideas and sharing experience in spatial modeling of landslides and scientific programming in general.

7 Appendices

420 **Appendix A: Supporting Information for Weights of Evidence**

WoE uses the Bayes rule to estimate the conditional probability of an event based on prior knowledge and a set of pieces of evidence. Prior knowledge in terms of prior probability usually represents the average expectation of an event given a study area. For the raster-based analysis, the prior probability calculates as the number of event pixels divided by the total pixels in the study area. Thus, the prior probability follows a uniform distribution over the entire study area. We update the prior

425 probability by weighted evidence factors, which we assume to be conditionally independent. For the spatial analysis, the factors characterize how much the probability value in a specific location is higher or lower than the prior probability. The updated probability is called posterior probability. The performed knowledge update cannot find new events but rather redistributes the probability density patterns conserving the total events. Thus, given the conditional independence of evidential patterns, we should obtain (approximately) the number of initial event pixels when summing up all pixels of the posterior probability raster.

430 In practice, complete conditional independence of evidential patterns is seldom. Therefore, the usage of many factors may cause inflation of posterior probability by occasional double counting of the effects (e.g., Agterberg and Cheng, 2002). This sets a general requirement to perform a contingency analysis (e.g., based on chi-square metrics) to estimate possible associations between evidence patterns.

In WoE, the weights for specific evidence patterns derive from the Bayes rule's formulation in logarithmic odds notation (e.g.,

435 Bonham-Carter, 1989), considering the evidence as a binary pattern for presence or absence of a specific feature. For multi-class datasets, the computation is done as if the dataset had been disintegrated into binary dummy variables. However, the straightforward analysis allows the weight calculation for multi-class datasets in one table without generating binary dummy variables explicitly. The weight calculation for a particular raster cell distinguishes two cases. First, if the particular feature class C is present, then the logit is given by:

440
$$w^+ = log_e \left(\frac{P\{C|E\}}{P\{C|\bar{E}\}} \right).$$
(1)

else, the logit is given by:

$$w^- = log_e \left(\frac{P\{\bar{C}|E\}}{P\{\bar{C}|\bar{E}\}} \right).$$
(2)

where $P\{C/E\}$ is the conditional probability of C given event E; $P\{C|\bar{E}\}$ is the conditional probability of C given $\bar{E}$ (no event); $P\{\bar{C}/E\}$ is the conditional probability of $\bar{C}$ (no feature) given E; $P\{\bar{C}/\bar{E}\}$ is the conditional probability of $\bar{C}$ (no feature) given

445 $\bar{E}$.

The weight notations w^+ and w^- represent not the mathematical sense of the values but the feature class presence (positive) and absence (negative) in the given raster cell.

With this formulation, positive logit values suggest a positive effect of the given variable, negative logits indicate a negative effect, and logits with a zero value indicate no effect. Thus, the latter does not modify the prior probability.

450   For a specific number of weighted layers, the posterior logit z is obtainable as:

$$z = PriorLogit + \sum_{i=1}^{n} w_i. \tag{3}$$

where n is the number of weighted layers (evidence), and w_i is the *ith* weighted layer, and the *PriorLogit* is:

$$PriorLogit = log_e \left(\frac{\frac{\text{total landslide pixel}}{\text{total area pixel}}}{1 - \frac{\text{total landslide pixel}}{\text{total area pixel}}} \right). \tag{4}$$

To convert the logit formulation to posterior probability, we use the logistic function:

455
$$p_{post} = \frac{1}{1 + e^{-z}}. \tag{5}$$

The sum of the weighted layers from Eq. 3 is the default output in the LSAT PM model builder for WoE models. It is already sufficient to obtain the relative susceptibility pattern needed for evaluation with the ROC curve. To obtain a model with probability values, the user should first compute the prior logit and modify the model generating expression in the model builder according to Eq. 5. Necessary information on the total number of landslide pixels and the total number of pixels in the
460   study area is obtainable from the result table of any layer. We recommend exporting the result table to Excel and conducting the simple side calculation as shown in Eq. 4.

8 Code Availability

The current version of LSAT PM is available from the project website: https://github.com/BGR-EGHA/LSAT under the GNU GPL v3.0 license. The exact version of LSAT PM used to produce the results used in this paper is archived on Zenodo:
465   https://doi.org/10.5281/zenodo.5511768. The LSAT PM documentation is available separately from https://github.com/BGR-EGHA/LSAT-Documentation under the CC BY-SA 4.0 license. The documentation for the exact version is archived on Zenodo: https://doi.org/10.5281/zenodo.5511920.

9 Data availability

The corresponding test dataset is archived on Zenodo: https://doi.org/10.5281/zenodo.5109620 under the CC BY-SA 4.0
470   license.

Geoscientific
Model Development
Discussions

EGU
Open Access

10 Author contribution

Jewgenij Torizin developed the theoretical concept, designed and coded LSAT PM. Nick Schüßler designed and coded parts of LSAT PM and migrated the code from Python 2.7 to Python 3. Michael Fuchs contributed with theoretical concepts and testing of the application through all stages of the development.

11 Competing interests

The authors declare that they have no conflict of interest.

12 Acknowledgements

13 References

Agterberg, F.P. and Cheng, Q.: Conditional independence Test for Weight-of-Evidence Modeling. Natural Resources Research 11 (4): 249-255, https://doi.org/10.1023/A:1021193827501, 2002.

Aleotti, P. and Chowdhury, R.: Landslide hazard assessment: summary review and new perspectives, Bull. Eng. Geol. Envir., 58(1), 21–44, https://doi.org/10.1007/s100640050066, 1999

Alimohammadlou, Y., Najafi, A., & Gokceoglu, C.: Estimation of rainfall-induced landslides using ANN and fuzzy clustering methods: A case study in Saeen Slope, Azerbaijan province, Iran. Catena, 120, 149-162, https://doi.org/10.1016/j.catena.2014.04.009, 2014.

Bonham-Carter, G. F., Agterberg, F. P. and Wright, D. F.: Weights of evidence modelling: a new approach to mapping mineral potential. Stat. Appl. Earth. Sci. Geol. Survey Can. Paper, 89–9, 171–183, 1989.

Bouziat, A., Schmitz, J., Deschamps, R., and Labat, K.: Digital transformation and geoscience education: New tools to learn, new skills to grow. European Geologist, 50, https://doi.org/10.5281/zenodo.4311379, 2020

Brabb, E. E.: Innovative approaches to landslide hazard and risk mapping. Proceedings of the 4th International Symposium on Landslides, Toronto, vol. 1, 307–324, 1985

Bragagnolo, L., da Silva, R. V., and Grzybowski, J. M. V.: Artificial neural network ensembles applied to the mapping of landslide susceptibility. Catena, 184, 104240, https://doi.org/10.1016/j.catena.2019.104240, 2020a

Bragagnolo, L., da Silva, R. V., and Grzybowski, J. M. V.: Landslide susceptibility mapping with r.landslide: A free open-
500   source GIS-integrated tool based on Artificial Neural Networks. Environmental Modelling & Software, 123, 104565, https://doi.org/10.1016/j.envsoft.2019.104565, 2020b

Budimir, M. E. A., Atkinson, P. M., Lewis, H. G.:A systematic review of landslide probability mapping using logistic regression. Landslides, 12, 419–436, https://doi.org/10.1007/s10346-014-0550-5, 2015.

Calkins, J. A., Offield, T. W., Abdullah, S. K. M., Ali, T.: Geology of the Southern Himalaya in Hazara, Pakistan, and Adjacent
505   Areas. Geological Survey Professional Paper 716-C, United States Government Printing Office, Washington, 1975

Chung, A. J. F., Fabbri, A. G.:Validation of spatial prediction models for landslide hazard mapping. Nat Hazards, 30, 451–472, 2003.

Crozier, M. J.: Deciphering the effect of climate change on landslide activity: A review. Geomorphology 124(3-4): 260-268. https://doi.org/10.1016/j.geomorph.2010.04.009, 2010.

510   Ermini, L., Catani, F., and Casagli, N.: Artificial Neural Networks applied to landslide susceptibility assessment. Geomorphology, 66, 327-343, https://doi.org/10.1016/j.geomorph.2004.09.025, 2005.

Fawcett, T.: An introduction to ROC analysis. Pattern Recogn Lett, 27(8): 861–874. doi: 10.1016/j.patrec.2005.10.010, 2006.

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

515   Fuchs, M., Khalid N.: Land Cover Map for the Districts of Mansehra & Torghar, Province Khyber Pakhtunkhwa, Islamic Republic of Pakistan, Final Report, 44 p., Islamabad/Hannover, 2015

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

George A. Miller: The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information.
520   The Psychological Review, 63, 81–97, https://doi.org/10.1037/h0043158, 1956.

GRASS Development Team: Geographic Resources Analysis Support System (GRASS) Software, Version 7.2.1 Open Source Geospatial Foundation. http://grass.osgeo.org, 2021

Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P.: Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology, 31:181-216. https://doi.org/10.1016/s0169-555x(99)00078-
525   1, 1999.

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., Ardizzone, F.: Probabilistic landslide hazard assessment at the basin scale. Geomorphology, 72:272-299, https://doi.org/10.1016/j.geomorph.2005.06.002, 2005.

Hall-Wallace, M. K.: Integrating Computing Across a Geosciences Curriculum Through an Applications Course. Journal of Geoscience Education 47(2): 119-123, https://doi.org/10.5408/1089-9995-47.2.119, 1999.

530   Hernández-Blanco, A., Herrera-Flores, B., Tomás, D., Navarro-Colorado, B.: A Systematic Review of Deep Learning Approaches to Educational Data Mining. Complexity, Article ID 1306039, https://doi.org/10.1155/2019/1306039, 2019.

JAXA: ALOS Global DSM AW3D30 Dataset Product Format Description for V 1.1, available on http://www.eorc.jaxa.jp/ALOS/en/aw3d30/aw3d30v11_format_e.pdf, 2017.

Jebur, M. N., Pradhan, B., Shafri, H. Z. M., Yusoff, Z.M., Tehrany, M.S. An integrated user-friendly ArcMap tool for bivariate
535   statistical modelling in geoscience applications. Geosci Model Dev 8:881–891. https://doi.org/10.5194/gmd-8-881-2015, 2015.

Kamp, U., Growley, B. J., Khattak, G. A., Owen, L. A.: GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. Geomorphology 101: 631-642, https://doi.org/10.1016/j.geomorph.2008.03.003, 2008.

Kamp, U., Owen, L. A., Growley, B. J., Khattak, G. A.: Back analysis of landslide susceptibility zonation mapping for the
540   2005 Kashmir earthquake: an assessment of the reliability of susceptibility zoning maps, Natural Hazards 54: 1-25, 2010.

Keefer, D. K.: Investigating Landslides Caused by Earthquakes – A Historical Review. Surveys in Geophysics, 23:473-510, https://doi.org/10.1023/A:1021274710840, 2002.

Lee, S., Evangelista, D. G.: Earthquake-induced landslide susceptibility mapping using an artificial neural network. Natural Hazards and Earth System Sciences 6:687–695, https://doi.org/10.5194/nhess-6-687-2006, 2006.

545   Lee, S.: Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data International Journal of Remote Sensing, 26:7, 1477-1491. https://doi.org/10.1080/01431160412331331012, 2005.

Lombardo, L., Mai, M. P.: Presenting logistic regression-based landslide susceptibility results. Engineering Geology 244: 14-24, https://doi.org/10.1016/j.enggeo.2018.07.019, 2018.

550   Makkawi, M. H., Hariri, M. M., Ghaleb, A. R.: Computer Utilization in Teaching Earth Sciences: Experience of King Fahd University of Petroleum and Minerals. International Education Journal 4(2): 89-97, 2003.

Mathew, J., Jha V. K, Rawat, G. S.: Weights of evidence modelling for landslide hazard zonation mapping of Bhagirathi Valley, Uttarakhand. Current Science 92 (5): 628-638, 2007.

Moghaddam, M. H. R., Khayyam, M., Ahmadi, M., Farajzadeh, M.: Mapping susceptibility Landslide by using Weight-of
555   Evidence Model: A case study in Merek Valley, Iran. Journal of Applied Sciences 7(22): 3342-3355, https://doi.org/10.3923/jas.2007.3342.3355, 2007.

Neuhäuser, B., Damm, B., Terhorst, B.: GIS-based assessment of landslide susceptibility on the base of the Weights-of-Evidence model. Landslides 9(4): 511-528, https://doi.org/10.1007/s10346-011-0305-5, 2012.

Osna, T., Sezer, E.A., Akgun, A.: GeoFIS: an integrated tool for the assessment of landslide susceptibility. Comput Geosci
560   66:20–30, https://doi.org/10.1016/j.cageo.2013.12.016, 2014.

Panchal, S., Shrivastava, A. K.: Application of analytic hierarchy process in landslide susceptibility mapping at regional scale in GIS environment. Journal of Statistics and Management Systems 23(2):199-206, https://doi.org/10.1080/09720510.2020.1724620, 2020

Petley, D. (2012) Global patterns of loss of life from landslides. Geology 40: 927–930, https://doi.org/10.1130/G33217.1.

565   Polat, A.: An innovative, fast method for landslide susceptibility mapping using GIS-based LSAT toolbox. Environ Earth Sci 80, 217, https://doi.org/10.1007/s12665-021-09511-y, 2021.

Polemio, M., Petrucci, O.: Rainfall as a Landslide Triggering Factor: an overview of recent international research. In Landslides in Research, Theory and Practice: Proceedings of the 8th International Symposium on Landslides; Bromhead, E., Dixon, N., Ibsen, M.-L., Eds.; Thomas Telford: London, UK, pp. 1219–1226, 2020.

570   Pradhan, B., Lee, S. (2010) Landslide susceptibility assessment, and factor effect analysis: back propagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environmental Modelling and Software 25: 747-759, https://doi.org/10.1016/j.envsoft.2009.10.016, 2010.

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

575   Rossi, M., Reichenbach, P.: LAND-SE: a software for statistically-based landslide susceptibility zonation, version 1.0. Geoscientific Model Development 9: 3533-3543, https://doi.org/10.5194/gmd-9-3533-2016, 2016.

Rossi, P. H., Guzzetti, F., Reichenbach, P., Mondini, A. C., Perruccacci, S.: Optimal landslide susceptibility zonation based on multiple forecasts. Geomorphology, 114(3),129–142, https://doi.org/10.1016/j.geomorph.2009.06.020, 2010.

Saaty, T. L.: A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology 15: 234-281, 1977.

580   Saaty, T. L.: The analytic hierarchy process, New York: McGraw, 1980.

Schmidhuber, J.: Deep Learning in neural networks: An overview. Neural Networks, 61, 85-117, 2015.

Tanyaş, H., Allstadt, K. E. and van Westen, C. J.: An updated method for estimating landslide-event magnitude. Earth surface processes and landforms, 43(9), 1836–1847, https://doi.org/10.1002/esp.4359, 2018.

Teerarungsigul, S., Torizin, J., Fuchs, M., et al.: An integrative approach for regional landslide susceptibility assessment using

585   weight of evidence method: a case study of Yom River Basin, Phrae Province, Northern Thailand. Landslides. 13(5): 1151-1165, https://doi.org/10.1007/s10346-015-0659-1, 2015.

Tensor Flow Team: Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jozefowicz, R., Jia, Y., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Schuster, M., Monga, R., Moore, S., Murray, D., Olah, C., Shlens, J., Steiner, B., Sutskever, I.,

590   Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng X.: TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org., 2015

Thiery, Y., Malet, J-P., Sterlacchini, S., Puissant, A., Maquaire, O.: Landslide susceptibility assessment by bivariate methods at large scales: Application to a complex mountainous environment. Geomorphology, 92(1-2), 38-59,

595   ff10.1016/j.geomorph.2007.02.020, 2007.

Tian, T., Balzer, D., Wang, L., Torizin, J., Wan, L., Li, X., Chen, L., Li, A., Kuhn, D., Fuchs, M., Lege, T., Tong, B.: Landslide hazard and risk assessment Lanzhou, province Gansu, China - Project introduction and outlook. In: Mikoš, M., Tiwari, B.,

Yin, Y., Sassa, K. (eds) Advancing culture of living with landslides, pp. 1027–1033. WLF 2017. Springer, Cham. https://doi.org/10.1007/978-3-319-53498-5_116, 2017.

600 Torizin, J., Fuchs, M., Awan, A. A., et al.: Statistical landslide susceptibility assessment of the Mansehra and Thorgar districts, Khyber Pakhtunkhwa Province, Pakistan. Natural Hazards 89(2): 757-784, https://doi.org/10.1007/s11069-017-2992-2, 2017.

Torizin, J., Fuchs, M., Balzer, D., Kuhn, D., Arifianti, Y., Kusnadi: Methods for generation and evaluation of landslide susceptibility maps: a case study of Lombok Island, Indonesia. Proceedings of 19th Conference on Engineering Geology, Munich, pp 253–258, 2013.

605 Torizin, J., Fuchs, M., Kuhn, D., Balzer, D., Wang, L.: Practical Accounting for Uncertainties in Data-Driven Landslide Susceptibility Models. Examples from the Lanzhou Case Study. In: Guzzetti F., Mihalić Arbanas S., Reichenbach P., Sassa K., Bobrowsky P.T., Takara K. (eds) Understanding and Reducing Landslide Disaster Risk. WLF 2020. ICL Contribution to Landslide Disaster Risk Reduction. Springer, Cham. https://doi.org/10.1007/978-3-030-60227-7_27, 2021.

Torizin, J., Wang, L. C., Fuchs, M., et al.: Statistical landslide susceptibility assessment in a dynamic environment: A case
610 study for Lanzhou City, Gansu Province, NW China. Journal of Mountain Science 15(6), https://doi.org/10.1007/s11629-017-4717-0, 2018.

Torizin, J.: Elimination of informational redundancy in the weight of evidence method: An application to landslide susceptibility assessment. Stochastic Environmental Research and Risk Assessment 30(2): 635-651, https://doi.org/10.1007/s00477-015-1077-6, 2016.

615 Torizin, J.: Landslide Susceptibility Assessment Tools for ArcGIS 10 and their Application. In: Proceedings of 34th IGC, Brisbane, 5-10 August 2012, p. 730, 2012.

Van Rossum, G., Drake, F. L.: Python 3 Reference Manual. Scotts Valley, CA: CreateSpace, 2009.

Van Westen, C. J., Castellanos Abella, E. A., & Sekhar, L. K.: Spatial data for landslide susceptibility, hazards and vulnerability assessment: an overview. Engineering Geology, 102(3-4), 112-131,
620 https://doi.org/10.1016/j.enggeo.2008.03.010, 2008.

Van Westen, C., Van Asch, T. W. and Soeters, R.: Landslide hazard and risk zonation – why is it still so difficult? Bulletin of Engineering geology and the Environment, 65(2), 167–184, 2006.

Varnes, D. J.: Landslide hazard zonation: a review of principles and practice, Natural Hazards, 3. UNESCO. Paris. 63 pp., 1984.

625 Xu, Y., Goodacre, R.: On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. J. Anal. Test. 2, 249–262, https://doi.org/10.1007/s41664-018-0068-2, 2018.