Grant Agreement No. 718679
Safety – Sentinel for geohazard prevention and forecasting

Deliverable D.E1: Report on tailoring existing knowledge and tools

A deliverable of Task E:

Geohazard impact assessment

Due date of deliverable: 31/08/2016
Actual submission date: 31/08/2016

Lead contractor for this deliverable: CNR.
Partners: IGME, UNIFI

Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the Consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	
TN	Technical Note, not a deliverable, only internal for members of the Consortium	x

Table of Content

EXECUTIVE SUMMARY

SAFETY is a two years research project (1 January 2016 – 31 December 2018) funded under the ECHO (European Commission's Humanitarian aid and Civil Protection department call "Prevention and preparedness projects in Civil Protection and marine pollution". The mission of the project is to improve the efforts in detecting and mapping geohazards (i.e. landslides and subsidence), by assessing their activity and evaluating their impact on built-up areas and infrastructures' networks. SAFETY will enhance ground deformation risk prevention and mitigation efforts in highly vulnerable geographic and geologic regions. The outcomes of the project will provide Civil Protection Authorities (CPA) with the capability of periodically evaluating and assessing the potential impact of geohazards on the selected sites.

D.E1 "Report on tailoring existing knowledge and tools" is the first deliverable of Task E "Geohazard impact assessment". The main goal of action E.1 is to exploit the existing knowledge and tools developed within projects completed in the framework of different European Funding programs (e.g. LAMPRE, DORIS). Particular attention will be paid to: i) the software for regional landslide susceptibility modelling that exploits statistical methods for the susceptibility zonation, and ii) the software for the determination of landslide statistics from inventory maps that consists of algorithm for the determination of statistics of landslide size (area) derived from inventory maps. The report will describe the existing knowledge and tools, and the strategies used to adapt them to the SAFETY project.

REFERENCE DOCUMENTS

N°	Title
RD1	DoW – FormT3a

CONTRIBUTORS

Contributor(s)	Company	Contributor(s)	Company
Paola Reichenbach	CNR		
Mauro Rossi	CNR		

REVIEW: CORE TEAM

Reviewed by	Company	Date	Signature
			Please add here a scanned image of your sign

1 INTRODUCTION

The project Safety aims to provide to the Civil Protection Authorities (CPA) the capability of periodically monitor and assess the impact of geohazards (landslides and subsidence, volcanos, earthquakes) on urban areas. The project's objectives are to improve the ability to detect and map landslides, to assess and forecast the impact of triggered landslide events on vulnerable elements, and to model landscape changes caused by slope failures. Safety is mainly addressed to the CPAs at different administrative levels.
Landslide hazard is defined by Varnes and his co-workers (1984) as “the probability of occurrence within a specified period of time and within a given area of a potentially damaging phenomenon”. Guzzetti et al. (1999) amended the definition to include the magnitude of the event. The definition of landslide hazard incorporates the concepts of location, time and size. To complete a hazard assessment one has to predict (quantitatively) “where” a landslide will occur, “when” or how frequently it will occur, and “how large” the landslide will be. Landslide hazard can be evaluated as a the joint probability of

$$HL = P_{AL} * P_N * S$$

Where P_{AL} expresses the probability of landslide size; P_N the probability of landslide occurrence in an established period of time; and S the spatial probability given the local environmental setting.

This report on “Tailoring existing knowledge and tools” is the first deliverable of Task E “Geohazard impact assessment” and is prepared as main goal of action E.1 finalized to exploit the existing knowledge and tools developed within projects completed in the framework of different European Funding programs (e.g. LAMPRE, DORIS).
In the task activity and report, particular attention was focused to:

i) the software for regional landslide susceptibility modelling that exploits statistical methods for the susceptibility zonation;
ii) the software for the determination of landslide statistics from inventory maps that consists of algorithm for the determination of statistics of landslide size (area) derived from inventory maps.

The document is divided into two main chapters:
Chapter 2 describes LAND-SE, the software for regional landslide susceptibility modelling;
Chapter 3 describes LAND-STAT, the software for the determination of landslide statistics from inventory maps.

2 LAND-SE

Landslide susceptibility is defined as the likelihood of a landslide occurring in an area on the basis of local terrain conditions (Brabb 1984). It is the degree to which an area can be affected by future slope movements, i.e. an estimate of "where" landslides are likely to occur (Guzzetti *et al.*, 1999, 2005, 2006a, b). In mathematical language, it can be expressed as the probability of spatial (geographical) occurrence of slope failures, given a set of geo-environmental conditions (Chung and Fabbri 1999, Guzzetti *et al.* 2005, 2006a).

The aim of this chapter is to describe LAND-SE, software for regional landslide susceptibility modelling. The software was presented in the article published by Rossi et al. (2010), updated and implemented in the framework of the LAMPRE Project (http://www.lampre-project.eu/). This report will describe the existing knowledge and tools, and the strategies used to adapt them to the SAFETY project end users and test sites.

This chapter is structured as follows:
- Sub-section 2.1: describes the software and its modelling approaches
- Sub-section 2.2: describes the software interface
- Sub-section 2.3: describes LAND-SE implementation in the SAFETY project

2.1 LAND-SE description

LAND-SE, software for regional landslide susceptibility modelling and zonation, follows the procedure proposed by Rossi et al (2010). The software logical scheme for regional landslide susceptibility modelling and zonation is shown in Figure 1. The structure can be subdivided in five parts:

1. Data input preparation
2. Single susceptibility models estimation (single susceptibility maps)
3. Combined model using a logistic regression approach (combined susceptibility maps)
4. Susceptibility model evaluation
5. Uncertainty evaluation (single and combined susceptibility zonations)

In the supervised multivariate statistical approaches, the dependent and the independent variables are defined as follows:

- The dependent variable (or grouping variable) is the presence or absence of landslides in the mapping units (derived from landslide inventory)
- The independent variables are the explanatory variables obtained from thematic and environmental information (morphometry, land cover/use, lithology, etc.).

The software is designed to use different mapping units, reducible to point-like units (pixels) or to polygon-like subdivisions (e.g. geomorphological, administrative, etc.). For each cartographic unit, the analysis requires two input variables: a binary grouping variable (i.e. dependent) showing the absence/presence (respectively 0 and 1) of a landslide, and a set of explanatory variables (i.e. independent). The software exploits different methods of multivariate statistical classification, in particular:

- Linear discriminant analysis model (LDA)
- Quadratic linear discriminant analysis model (QDA)
- Logistic regression model (LRM)
- Self-optimizing neural network model (NNM).

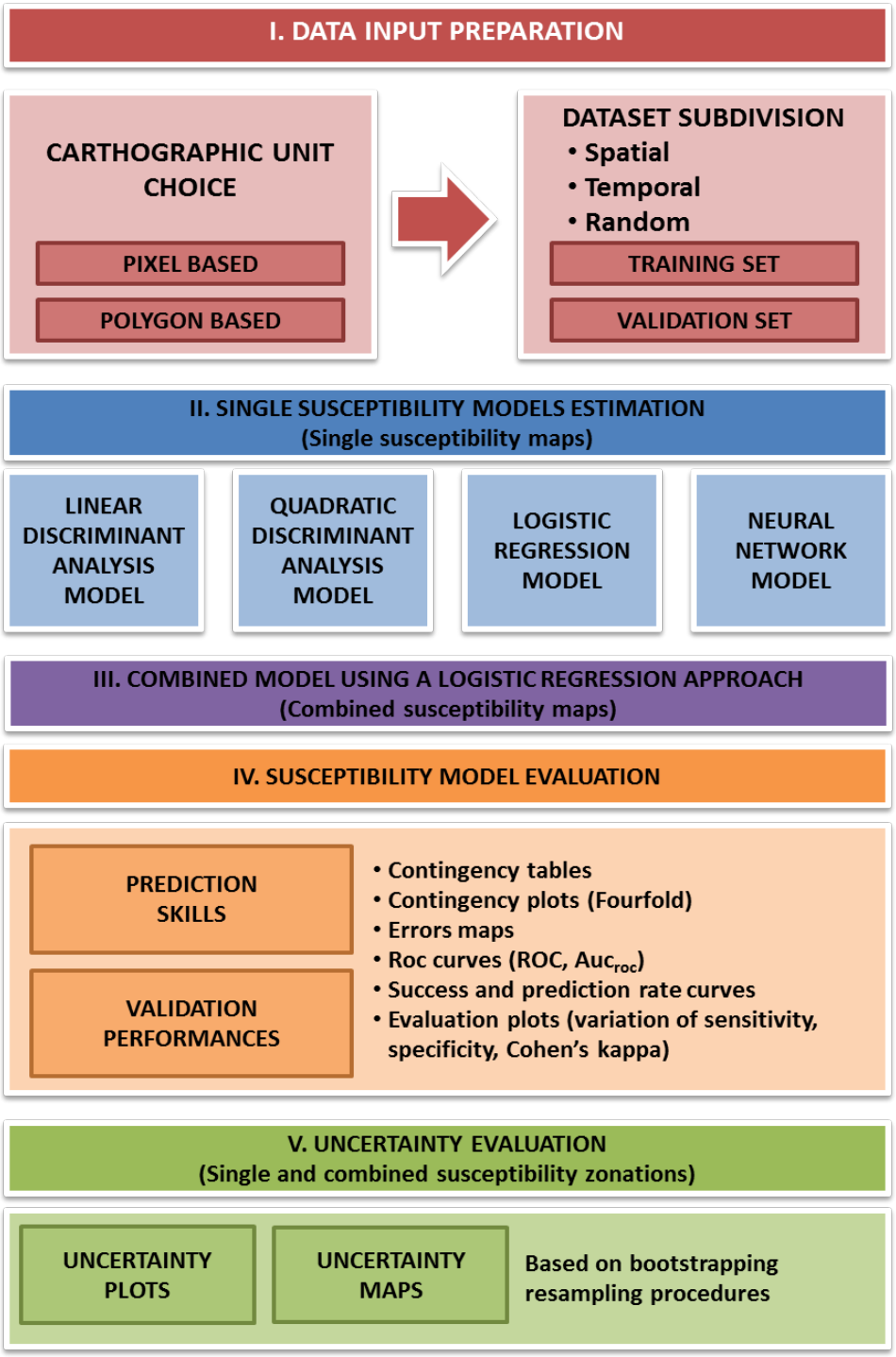

Figure 1. Software logical scheme for regional landslide susceptibility modelling and zonation.

Moreover, LAND-SE considers the combination (CFM) of the single modelling susceptibility results (LDA, QDA, LRM, NNM) or some of them using a logistic regression approach. The logistic regression approach was selected for the combination scheme, where the grouping variable is the presence or absence of landslides in the mapping units (the same of single susceptibility models) and the explanatory variables are the susceptibility prediction obtained from the single models.

The model application follows two phases: a training phase, and a validation phase. In the training phase the model reconstructs the relationships between the two variables (dependent and independent) whereas in the validation phase, these relationships are verified in different conditions.

The software integrates different procedures and measures/indices for evaluating:

- the dependence among explanatory variables;
- the model prediction skill (capability of the model to predict the original data)
- the temporal or spatial model validation performance (capability of the model to predict independent data);
- the model errors evaluation
- the uncertainty evaluation (measuring the uncertainty related to the susceptibility estimate).

The model outputs are stored in .txt format (textual results), in .pdf format (plots and graphics) and in geographical formats (shapefiles .shp, GeoTIFF .tif).

2.2 Software specifications

LAND-SE can be executed using two different modes: *standard* that prepares output in textual and graphical format and *geomode* that requires input spatial data and provides results also in standard geographical formats. The differences are mainly related to the different output types. In the advanced mode, or geographical mode (*geomode*), LAND-SE produces geographical outputs in .shp (shapefile) or in tif (GeoTIFF). In the same mode also .pdf of the geographical data are produced as output. A relevant difference in the *geomode,* is the production of the success rate and prediction rate curve plots, as additional tools to measure the prediction skill and validation performances.

This section describes the specifications (i.e., an R script) required to run LAND-SE. **Section 2.2.1** describes the input data specifications, including the files required for the susceptibility analysis (Table 1). **Section 2.2.2** illustrates the two configuration files containing the parameters controlling directly the susceptibility analysis ("configuration.txt"), and the parameters for the spatial data configuration ("configuration_spatial_data.txt"). **Section 2.2.3,** explains the software output.

2.2.1 Input data specifications

As shown in Figure 1, the software is designed to use different mapping units, reducible to point-like units (pixels) or to polygon-like subdivisions (e.g. geomorphological, administrative, etc.). The software cannot use directly input raster files and hence, a workaround is needed to perform the pixel based analysis. A preliminary operation is needed to convert the raster format into a list format. This operation can be performed in different GIS environment, two possible functions are the "gdal2xyz" function (http://svn.osgeo.org/gdal/trunk/gdal/swig/python/scripts/gdal2xyz.py) integrated in different GIS clients (e.g. QGIS, http://www.qgis.org/en/site/) or the "raster2xyz" function in the ArcGIS platform (http://www.esri.com/software/arcgis).

In the basic mode the software needs the training.txt and the validation.txt files (see Table 1) containing the information on the mapping units for the training and validation phase respectively. The two files are tab-separated .txt file, with named columns (without spaces) containing in order: (i) the identification for the mapping units, (ii) the binary grouping variable (i.e. dependent variable) showing the absence or presence (respectively 0 and 1) of a landslide in the mapping unit, and (iii) and a set of *n* explanatory variables (i.e. independent variables).

Depending of the user main objective, the subdivision of the training and validation dataset (Figure 1) and hence the type of the validation performed by the model can be done in different

way: temporal, spatial and random. In case of a temporal subdivision, the two dataset files will contain the same mapping units, with the same number, the same identifications and the same values of the explanatory variables, but with different values of the grouping variable obtained using a different landslide inventory (possibly successive to that used in the training phase). In the spatial and random subdivisions, the two dataset files will contain different mapping units, with different identification, grouping variable explanatory variable values. The main difference between the two subdivision methods consists in the different method to subdivide the training and the validation datasets: in the spatial case the dataset are relative to two areas separated spatially (contiguous or not), while in the random case the subdivision is based on a random sampling method.

In the advanced mode (*geomode*) the software needs two additional files: training.shp and validation.shp (see Table 1). These can be point or polygon shapefiles containing geographical data for each mapping unit in the training.txt and validation.txt file. The two shapefile attribute table must contains the following fields: (i) the identification values, (ii) the area, and (ii) the landslide area for each mapping units reported in the two corresponding .txt files.

2.2.2 Configuration parameters

To be executed LAND-SE requires two configuration files containing parameters (i) for the control of the type of the susceptibility analysis to be performed ("configuration.txt", Table 2), and (ii) for the spatial data configuration ("configuration_spatial_data.txt", Table 3).

In the "configuration.txt" file the user can specify which type of susceptibility models have to be used (RUN column) in the analysis (LDA, QDA, LRM, NNM in Table 2), and if the combination model have to be used (CFM in Table 2). The BOOTSTRAP_SAMPLES_ROC_CURVE parameter indicates the number of samples to be used in the evaluation of ROC Curve variability. The ANALYSIS_PARAMETER can be different for the different type of models:

- QDA can be: SEL to eliminate dummy variables from the analysis or DUM to maintain dummy variables transforming them in numerical format introducing a random noise;
- NNM can be: NOR to use neural network default weights and a number of nodes in the hidden layer corresponding to half the explanatory variables, or OPT to perform an auto optimization of the neural network structure (slower and with tendency to overfit data).

The BOOTSTRAP_MODEL_VARIABILITY_RUN and BOOTSTRAP_SAMPLES_MODEL_-_VARIABILITY parameters specify if the uncertainty evaluation must be performed and how many bootstrap samples to use in this analysis.

In the "configuration_spatial_data.txt" (Table 3) the user can specify if the advanced analysis (geo mode) have to be performed specifying YES in the "PRESENCE" column. If enabled this analysis requires to specify:

- the name of the shapefile attribute filed containing the identification value of each mapping unit (ID_FIELD);
- the EPSG code identifying the Coordinate Reference System of the geographical data provided in the analysis (EPSG_CODE);
- the name of the shapefile attribute filed containing the area of each mapping unit (AREA_SU_FIELD);
- the name of the shapefile attribute filed containing landslide area in each mapping unit (AREA_LANDSLIDE_FIELD);
- the type of shapefile feature used in the analysis (GEOMETRY);
- the dimension of cell considered in the analysis, used only when pixel-based analysis is performed (RASTER_RES); and
- if the output in geographical raster format have to be produced when the pixel-based analysis is enabled (RASTER_EXPORT).

SOFTWARE FILES	DESCRIPTION
SusceptibilityAnalysis_vX_YYYYMMDD.R	R script file containing the susceptibility analysis source code
configuration_spatial_data.txt	File containing the parameters for the spatial data configuration
configuration.txt	File containing the parameters for the susceptibility model configuration
training.txt	Tab-separated textual input file with named columns (without spaces) containing in order (1) an ID for the mapping units, (2) the grouping variable with 0 or 1 values, (3 to N) explanatory numerical variables. Rows contain these values for the each mapping unit of the training dataset
validation.txt	Tab-separated textual input file with named columns (without spaces) containing in order (1) an ID for the mapping units, (2) the grouping variable with 0 or 1 values, (3 to N) explanatory numerical variables. Rows contain these values for the each mapping unit of the validation dataset
training.shp	Points or polygons shapefile containing the geographical data for each mapping unit in the training.txt file. The shapefile attribute table must contains the following fields (1) the ID, (2) the area, and (3) the landslide area of each mapping units (***needed only in the advanced or geo mode for the spatial data output restitution and for the success rate calculation***)
validation.shp	Points or polygons shapefile containing the geographical data for each mapping unit in the validation.txt file. The shapefile attribute table must contains the following fields (1) the ID, (2) the area, and (3) the landslide area of each mapping units (***needed only in the advanced or geo mode for the spatial data output restitution and for the prediction rate calculation***)

Table 1. List of the files required by LAND-SE for the susceptibility analysis

MODEL	RUN	BOOTSTRAP SAMPLES ROC CURVE	ANALYSIS PARAMETER	BOOTSTRAP MODEL VARIABILITY RUN	BOOTSTRAP SAMPLES MODEL VARIABILITY
LDA	YES/NO	10		YES/NO	10
QDA	YES/NO	10	DUM	YES/NO	10
LRM	YES/NO	10		YES/NO	10
NNM	YES/NO	10	NOR	YES/NO	10
CFM	YES/NO	10		YES/NO	10

Table 2. List of parameter for the susceptibility analysis in the "configuration.txt" file.

PARAMETER	VALUES DESCRIPTION
TYPE	**SHAPEFILE** At present only this value is admitted by the software
PRESENCE	**YES/NO** YES enable the use of spatial data (training.shp and validation.shp file have to be provided)
ID_FIELD	**ID** Name of the shapefile attribute filed containing the identification value of each mapping unit
EPSG_CODE	**4326** EPSG code identifying the Coordinate Reference System of the geographical data provided in the analysis
AREA_SU_FIELD	**area_mapping** Name of the shapefile attribute filed containing the area of each mapping unit
AREA_LANDSLIDE_FIELD	**area_landslide** Name of the shapefile attribute filed containing landslide area in each mapping unit
GEOMETRY	**POINTS/POLYGONS** POINTS for pixel-based analysis POLYGONS when using other polygon based mapping units
RASTER_RES	**30** Dimension of cell considered in the analysis (only when pixel-based analysis is performed)
RASTER_EXPORT	**TRUE/FALSE** TRUE when pixel-based analysis is performed enable the export of map also in GeoTIFF format

Table 3. List of parameter for the spatial data configuration in the "configuration_spatial_data.txt" file.

2.2.3 Software output

Table 4 provides a complete overview of LAND-SE outputs. Moreover, the table gives a description of each output indicating which one are produced by the geo mode (advanced) analysis.

For simplicity in the table the outputs are grouped in graphical (saved in .pdf format), textual (saved in .txt format), and geographical outputs (saved in .shp or .tif format). Inside each group, the outputs are subdivided considering the type of the landslide susceptibility model. The output for a specific model type is generated only if the corresponding model has been enabled in the "configuration.txt" file (see Table 2).

Following the software logical scheme in Figure 1, the software output can be further roughly grouped in outputs that:

- highlight the dependence among explanatory variables;
- measure the model prediction skill;
- measure the temporal or spatial model validation performance;
- evaluate the model errors;
- estimate the model prediction uncertainties.

A more detailed description and use of the different software outputs, is given in the article by Rossi *et al.* (2010).

SOFTWARE OUPUT	DESCRIPTION	GEO MODE
GRAPHICAL OUTPUTS		
GroupingVariable_Histogram.pdf	Histogram of the grouping variable	
GroupingVariable_Histogram_Validation.pdf	Histogram of the validation variable	
result_XX_BootstrapMeansComparison.pdf	Comparison of the uncertainty plots	
result_XX_BootstrapPredictionVariability.pdf	Uncertainty plot estimated for the XX model using a resampling approach	
result_XX_BootstrapProbabilityVariability.pdf	Uncertainty plot estimated for the XX model using a sampling approach	
result_XX_FourfoldPlot.pdf	Fourfold (contingency) plot comparing observed vs predicted data (XX model)	
result_XX_FourfoldPlot_Validation.pdf	Fourfold (contingency) plot comparing observed vs validation data (XX model)	
result_XX_Histogram.pdf	Histogram of susceptibility values calculated in the XX model training	
result_XX_ModelEvaluationPlot.pdf	Sensitivity, specificity and Cohen's kappa comparing observed and XX model predicted data classified using different probability thresholds	
result_XX_Model_MatchingCode_Map.pdf	Map of the XX model training errors derived from the contingency table	×
result_XX_Model_Susceptibility_Map.pdf	Map of the XX model susceptibility values obtained in the training phase	×

SOFTWARE OUPUT	DESCRIPTION	GEO MODE
result_XX_PredictionRateCurve.pdf	Prediction rate curve obtained in the validation phase	
result_XX_ROCPlot_bootstrap.pdf	ROC plot comparing observed and predicted data for the XX model	
result_XX_ROCPlot_bootstrap_Validation.pdf	ROC plot comparing observed and validation data for the XX model	
result_XX_SuccessRateCurve.pdf	Success rate curve obtained in the training phase	
result_XX_Validation_MatchingCode_Map.pdf	Map of the XX model validation errors derived from the contingency table	×
result_XX_Validation_Susceptibility_Map.pdf	Map of the XX model susceptibility values obtained in the validation phase	×
result_ModelComparison_LDA_LRM.pdf	LDA vs LRM susceptibility comparison	
result_ModelComparison_LDA_NNM.pdf	LDA vs NNM susceptibility comparison	
result_ModelComparison_LDA_QDA.pdf	LDA vs QDA susceptibility comparison	
result_ModelComparison_LRM_LDA.pdf	LRM vs LDA susceptibility comparison	
result_ModelComparison_LRM_NNM.pdf	LRM vs NNM susceptibility comparison	
result_ModelComparison_LRM_QDA.pdf	LRM vs QDA susceptibility comparison	
result_ModelComparison_NNM_LDA.pdf	NNM vs LDA susceptibility comparison	
result_ModelComparison_NNM_LRM.pdf	NNM vs LRM susceptibility comparison	
result_ModelComparison_NNM_QDA.pdf	NNM vs QDA susceptibility comparison	
result_ModelComparison_QDA_LDA.pdf	QDA vs LDA susceptibility comparison	
result_ModelComparison_QDA_LRM.pdf	QDA vs LRM susceptibility comparison	
result_ModelComparison_QDA_NNM.pdf	QDA vs NNM susceptibility comparison	
TEXTUAL OUTPUTS		
result_Collinearity_Analysis.txt	Results of the test of the collinearity evaluation	
result_XX_BootstrapSamples.txt	XX model susceptibility values for the samples used in the uncertainty evaluation	
result_XX_BootstrapStatistics.txt	Statistics of the XX model susceptibility values for the samples used in the uncertainty evaluation	
result_XX.txt	Summary of the results obtained using the XX model	
GEOGRAPHICAL OUTPUTS FOLDERS		
result_XX_training/training.shp	Folder containing the shapefile of XX model results obtained in the training phase	×
result_XX_validation/validation.shp	Folder containing the shapefile of XX model	×

SOFTWARE OUPUT	DESCRIPTION	GEO MODE
	results obtained in the training phase	
result_XX_Model_MatchingCode_Map.tif	Map of the XX model errors derived from the contingency table in GeoTIFF format (*only for pixel-based analysis*)	×
result_XX_Model_Susceptibility_Map.tif	Map of the XX model susceptibility values obtained in the training phase in GeoTIFF format (*only for pixel-based analysis*)	×
result_XX_Model_Uncertainty_Map.tif	Map of the XX model uncertainty values obtained in the training phase in GeoTIFF format (*only for pixel-based analysis*)	×
result_XX_Validation_MatchingCode_Map.tif	Map of the XX model validation errors derived from the contingency table in GeoTIFF format (*only for pixel-based analysis*)	×
result_XX_Validation_Susceptibility_Map.tif	Map of the XX model susceptibility values obtained in the validation phase in GeoTIFF format (*only for pixel-based analysis*)	×
result_XX_Validation_Uncertainty_Map.tif	Map of the XX model uncertainty values obtained in the validation phase in GeoTIFF format (*only for pixel-based analysis*)	×

Table 4. List of LAND-SE outputs. In the table XX is used in placed of the models available in the software: LDA-linear discriminant analysis, QDA-quadratic discriminant analysis, LR-logistic regression, NN-neural network and CM-combined model. A column specifies output provided by the geomode.

2.3 LAND-SE implementation in the SAFETY project

Following the "User needs and requirements" (Deliverable D.B1), susceptibility maps will be delivered associated with information describing the methodology used to produce them. The maps will be integrated in the project geodatabase and will be prepared in a format suitable for being distributed using web services. The list of data used for the production will to be reported in the metadata together with the accuracy and the resolution of the map.

Landslide susceptibility map will be prepared using LAND-SE, software for landslide statistically-based susceptibility zonation. The LAND-SE code has been improved and new functionalities will be tested in the Volterra site (Tuscany region, Italy). In particular the software implements the following functionalities:

- Procedure to optimize the susceptibility classes selected to prepare maps and plots (additional optimized maps and plots produced as output);
- Procedure to evaluate the interactions between explanatory variables using the logistic regression modelling in the single (LRM) and the combined (CFM) models.

Moreover, a new software aimed to facilitate LAND-SE input data preparation and analysis execution was implemented and will be tested in SAFETY. The tool (LAND-SEAST, LANDslide Susceptibilty Evaluation Analysis Support Tool) provides the following functionalities:

- Procedure to prepare LAND-SE input data (in tabular format) starting from data in ESRI ASCII raster format;
- Possibility to select using different masking approaches, an area of interest to be used as input subset;
- Generation of multiple training and validation samples using different sampling schema (random or spatial);
- Reduction (decreasing sample size) and balancing (modifying the proportion between stable and unstable pixels) of training and validation samples;
- Possibility to execute multiple run of LAND-SE via command line.

3 LAND-Stat

The aim of this session is to describe LAND-Stat, the software to analyse landslide statistics. The statistical distribution of landslide size is a proxy for the landslide magnitude, relevant for landslide hazard assessment. The session describes the software input and output.

Statistical models are used to estimate the probability of occurrence of landslides of a given size as part of a triggered event or geomorphic inventory. In the literature, two frequency-area landslide distribution probability models have been proposed: the Double Pareto distribution (Stark & Hovius, 2001) and the Inverse Gamma distribution (Malamud et al., 2004). Both models are able to describe the right tail (i.e. the medium and largest landslides) of the landslide size distribution with an inverse power-law, and both models reproduce the fact that as the landslide area decreases, occurrence increases, until (for triggered events) a landslide area of several hundred m^2 where one arrives at a maximum probability and there is then a "rollover" (i.e. the modal or the most frequent landslide size value), and then as the landslide area continues to be smaller, the occurrence probability decreases. This has been shown by observed data and described in the literature (Malamud et al., 2004).

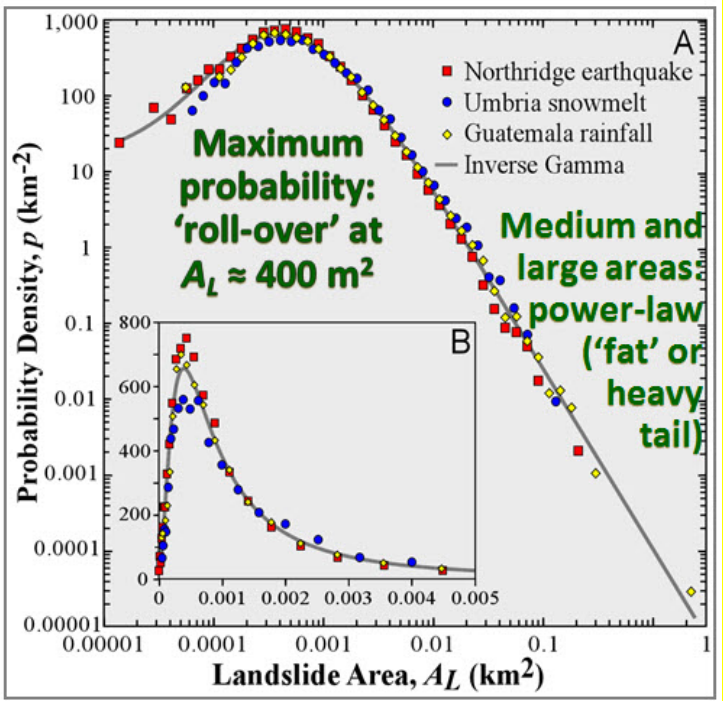

Figure 2. Inverse gamma probability density function fit to frequency densities from three substantially complete triggered event inventories (Malamud et al., 2004)

An inverse gamma fit for three substantially complete landslide inventory areas is given in Figure 2. The main difference between the two statistical models is the description of the left tail (i.e., smaller landslide areas). In the case of the Double Pareto, this left tail basically consists of another power law function (this time not inverse) that gradually censors the upper tail Double Pareto from the "rollover" to the smaller landslide area values. In the case of the Inverse Gamma, the lower tail is described by an exponential rollover. While the Inverse Gamma must have a maximum (i.e., assumed by definition a rollover), the Double Pareto does not. These two different mathematical representations of the landslide size distribution can converge to similar probability density function estimates.

The statistical characterisation of the underlying complete (or in some cases not so complete) event, geomorphic, and multi-temporal inventories has been widely used in the literature as a way of characterisation for individual regions, but also as a way of comparison (e.g., for completeness), and can lead to conclusions of potential magnitude of a given event (and corresponding erosion) and probabilistic risk for the future (Ardizzone *et al.*, 2013; Guzzetti *et al.*, 2005; Fiorucci *et al.*, 2011). This report here will not detail these studies and their conclusions, but rather, as often the actual calculation of the best-fit underlying distribution is difficult for the user, describe a software tool that is designed for the user to input (via a text file or shapefile) a set of landslide areas, with the software tool giving the best estimates for both statistical models, the Double Pareto and the Inverse Gamma.

This section is structured into the following sections:

- **Sub-section 3.1**: Describes the background to the landslide statistical models
- **Sub-section 3.2**: Describes the software and its interfaces
- **Sub-section 3.3**: LAND-Stat implementation in the SAFETY project

3.1 Landslide statistical models

3.1.1 Double Pareto Distribution

The Double Pareto distribution can be written as one of the following (Stark & Hovius, 2001):

Double Pareto Simplified (DPS) [Three parameter]:

$$\mathrm{pdf}(\mathrm{x}|\alpha,\beta,t) = \left[\frac{\beta(t^{\alpha})}{\left(1+\left(\frac{x}{t}\right)^{-\alpha}\right)^{\left(1+\left(\frac{\beta}{\alpha}\right)\right)}(x^{(\alpha+1)})}\right] \qquad \text{Eq. 1}$$

Double Pareto (DP) [Five parameters]:

$$\mathrm{Pdf}(x|\alpha,\beta,t,c,m) = \left[\frac{\beta}{t}\left(1-\left(\frac{1+\left(\frac{m}{t}\right)^{-\alpha}}{1+\left(\frac{c}{t}\right)^{-\alpha}}\right)^{\left(\frac{\beta}{\alpha}\right)}\right)^{-1}\right]\left[\frac{\left(1+\left(\frac{m}{t}\right)^{-\alpha}\right)^{\left(\frac{\beta}{\alpha}\right)}}{\left(1+\left(\frac{x}{t}\right)^{-\alpha}\right)^{\left(1+\frac{\beta}{\alpha}\right)}}\left(\frac{x}{t}\right)^{(-\alpha-1)}\right] \qquad \text{Eq. 2}$$

with the following parameters:

- α controls the exponent of the inverse power-law for the right tail
- β controls the exponent of the power-law for the left tail
- t constrains rollover position (but does not correspond to exact rollover position)
- *c*, *m*: two additional parameters for DP (vs. DPS) that constrain estimation of the pdf inside a specific landslide size range from c (minimum size value) to m (maximum size value)

Eq. 1 corresponds to the Double Pareto Simplified (DPS) function and has three parameters. Eq. 2 corresponds to the Double Pareto (DP) function and is similar to Eq. 1, but has two additional parameters, *c* and *m*. Eq. 1 can be derived from Eq. 2 with simple mathematical passages posing $m = \infty$ and $c = 0$.

3.1.2 *Inverse Gamma Distribution*

In the software, the Inverse Gamma (IG) distribution is formulated as follows:

$$\mathrm{pdf}(x|\alpha,\eta,\lambda) = \left[\frac{\lambda^{2\alpha}}{\Gamma(\alpha)}\right]\left[\left(\frac{1}{x+\eta^2}\right)^{(\alpha+1)}\right]\exp\left[-\frac{\lambda^2}{x+\eta^2}\right] \qquad \text{Eq. 3}$$

with the following three parameters and $\Gamma(\alpha)$ the gamma function of α:

- α controls the exponent of the inverse power law, i.e. right tail (i.e. corresponding to the Double Pareto α parameter),
- η controls the amount the left tail bends
- λ controls the position of the rollover.

Eq. (3) basically corresponds to the inverse gamma formulation by Malamud *et al.* (2004) given in Eq. (4). In fact, substituting $\alpha = \rho$, $\eta = \sqrt{-s}$ and $\lambda = \sqrt{a}$ the following formulation for the Inverse Gamma probability distribution (Malamud *et al.*, 2004) can be obtained:

$$\mathrm{pdf}(x|\rho,a,s) = \frac{1}{a\Gamma(\rho)}\left[\frac{a}{x-s}\right]^{\rho+1}\exp\left[-\frac{a}{x-s}\right] \qquad \text{Eq. 4}$$

with the following three parameters and $\Gamma(\rho)$ the gamma function of ρ:

- ρ controls the exponent of the inverse power law, i.e. right tail (i.e. corresponding to the Double Pareto α parameter) [unitless]
- s controls amount left tail bends as part of exponential [same units as *x*]
- a controls the position of the rollover [same units as *x*]

The estimation of the distribution parameters is not as straightforward as for other classical distribution parameters (e.g., normal distributions), and at present different statistical approaches are used in the literature. In specific cases, this is a significant problem, for example when analysing landslide data characterized by some bias (e.g., due to sampling problem) results obtained using different statistical parameter estimation approaches can be significantly different. To overcome this limitation, LAND-Stat, software for the determination of landslide statistics from inventory maps, includes different statistical methods to estimate the parameters of the probability density functions mentioned above.

3.2 LAND-Stat Description

3.2.1 *Basic vs. Advanced Version of software*

The LAND-Stat software for the determination of landslide statistics (Figure 3) is realized in R (a free software environment for statistical computing, http://www.r-project.org/) and implements parametric and non-parametric approaches to estimate the parameters of the three probability density function:

1. Histogram Density Estimation (HDE)
2. Kernel Density Estimation (KDE)
3. Maximum Likelihood Estimation (MLE)

Each of these approaches exploits different optimization procedures thus can give slightly different results. Two different versions of the software have been developed, a **Basic Version** and an **Advanced Version**.

In its **Basic Version**, for each parameter (Table 5) the tool gives:

1. an estimate of its value
2. standard error (Std. Err.),
3. the estimated error variance (t_value), and
4. the correlations among the parameters (Pr(>|t|)).

The latter in particular could be useful in cases of difficulty in producing a solution: very high correlations between parameters are indicative of ill-conditioning. The tool estimates the exact rollover position (*r*) a useful information for the landslide inventory comparison.

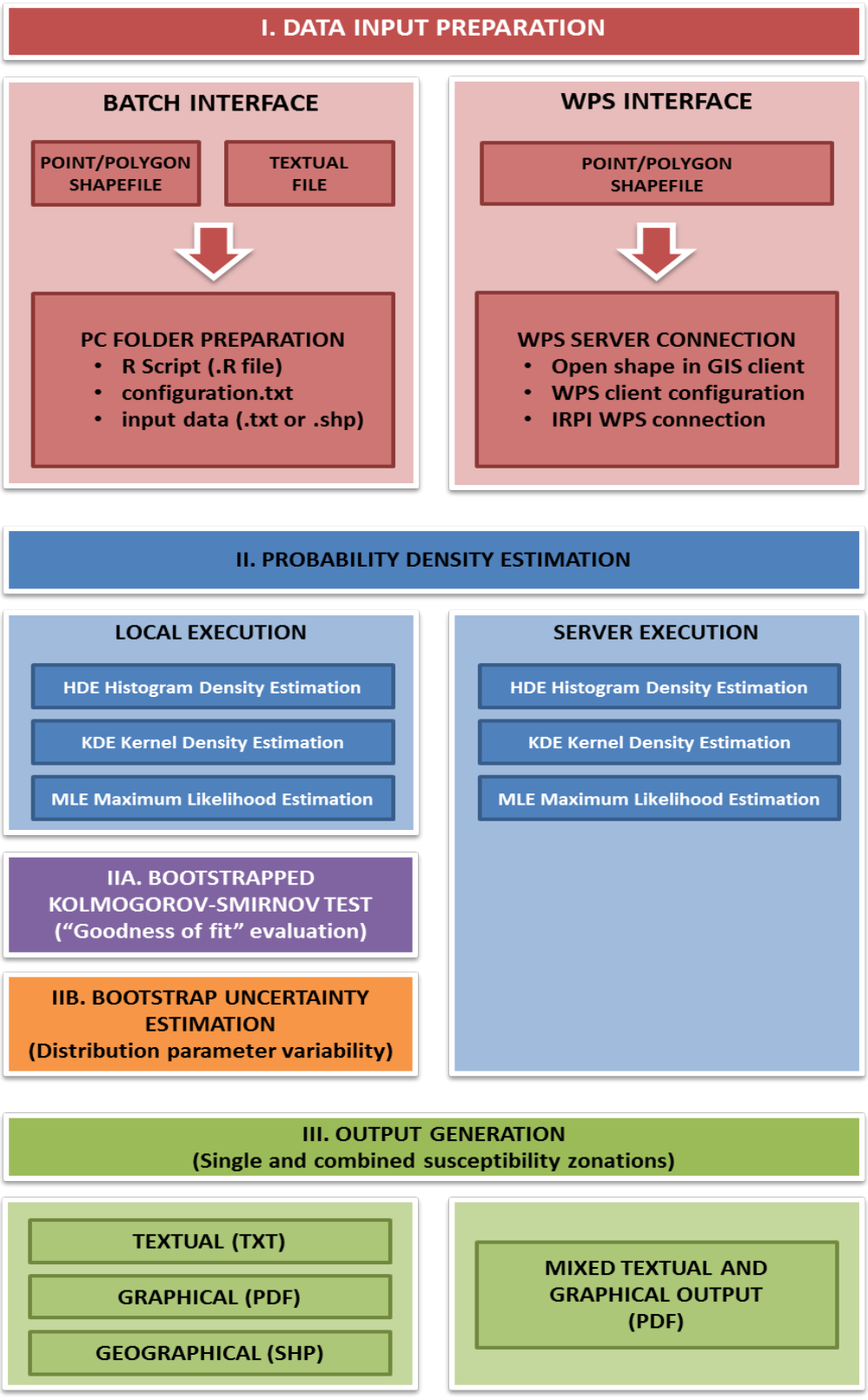

Figure 3. Software logical scheme for the determination of landslide statistics (LAND-Stat) showing two interfaces: a batch interface (Advanced version of the software) and a WPS (Web Processing Service) interface (Basic version of the software).

		HISTOGRAM DENSITY ESTIMATION				KERNEL DENSITY ESTIMATION				MAXIMUM LIKELIHOOD ESTIMATION			
		Estimate	Std. Err.	t_value	Pr(>|t|)	Estimate	Std. Err.	t_value	Pr(>|t|)	Estimate	Std. Err.	t_value	Pr(>|t|)
DPS	α	0.96	0.1	11.35	0.000	0.96	0.1	11.35	0.000	0.96	0.1	11.35	0.000
	β	5.00	2.9	1.71	0.110	5.00	2.9	1.71	0.110	5.00	2.9	1.71	0.110
	t	206	203	1.02	0.327	206	203	1.02	0.327	206	203	1.02	0.327
	r	433	-	-	-	433	-	-	-	433	-	-	-
DP	α	0.96	0.1	11.41	0.000	0.96	0.1	11.41	0.000	0.96	0.1	11.41	0.000
	β	5.00	2.9	1.71	0.109	5.00	2.9	1.71	0.109	5.00	2.9	1.71	0.109
	t	207	203	1.02	0.324	207	203	1.02	0.324	207	203	1.02	0.324
	c	88	-	-	-	88	-	-	-	88	-	-	-
	m	328375	-	-	-	328375	-	-	-	328375	-	-	-
	r	434	-	-	-	434	-	-	-	434	-	-	-
IG	α	0.92	0.09	9.78	0.000	0.92	0.09	9.78	0.000	0.92	0.09	9.78	0.000
	η	9.22	2.71	3.41	0.004	9.22	2.71	3.41	0.004	9.22	2.71	3.41	0.004
	λ	31.73	3.89	8.17	0.000	31.73	3.89	8.17	0.000	31.73	3.89	8.17	0.000
	r	438	-	-	-	438	-	-	-	438	-	-	-

Table 5. Example of parameters estimated by LAND-Stat for three landslide size distribution models and three different estimation methods.

The **Advanced version** of LAND-Stat integrates the following additional features:

1. the bootstrapped parameter uncertainty estimation, for a more significant statistical comparison of the obtained probability distribution parameters;
2. the bootstrapped Kolmogorov–Smirnov test (KS test) to serve as a “goodness of fit” test providing a measure of the suitability of the different distribution types,
3. an improved version of the cumulative density function calculation,
4. an improved use of shapefiles, allowing the automatic landslide size calculation, and
5. an improved output version, allowing the calculation of the probability density.

The software produces textual (.txt format), graphical (in .pdf format) and geographical outputs (in .shp format). The type of output is differentiated in the two software versions. Different interfaces were prepared for the two software versions integrating different specifications: (a) a batch script interface (Advanced Version), and (b) a WPS (Web Processing Service) interface (Basic Version).

3.2.2 LAND-Stat software tool interfaces

As mentioned previously, the software tool has a two-fold interface (Figure 3): the first is a batch interface (i.e. an R script) while the second is a Web Processing Service interface (i.e. a standard WPS as defined by the Open Geospatial Consortium).

The batch interface corresponds to the **LAND-Stat Advanced Version** of the software and requires the download of the source code, consisting of a textual file with extension *.R (available from the LAMPRE web site, see introduction). This must be executed on a local machine.

The WPS interface (**LAND-Stat Basic Version**) integrates the basic software characteristics and can be directly accessed through the Web, using a specific plugin (PyWPS) of the QGIS software (A Free and Open Source Geographic Information System).

The two interfaces have different specifications, in particular the batch interface (LAND-Stat Advanced Version) can be executed using as input either a textual file or a geographical shapefile. This interface produces additional outputs compared to the WPS interface (LAND-Stat Basic Version). Table 6 lists the software outputs with the relative description, produced by the two software interfaces.

SOFTWARE OUPUT	DESCRIPTION	BATCH interface (advanced)	WPS interface (basic)
GRAPHICAL OUTPUTS			
BoxPlot.pdf	Box plot of raw data series	×	×
BoxPlot_Comparison.pdf	Box plot of raw data series (if multiple data series provided)	×	×
Histogram.pdf	Histogram of the log of raw data series (log-binning)	×	×
KDE_Density.pdf	Kernel density of the log of the raw data series (log-bandwidth)	×	×
KDE_FrequencyDensity.pdf	Kernel frequency density of the log of the raw data series (log-bandwidth)	×	×
HDE_Fit_Comparison.pdf	Comparison of DP, DPS and IG probability density functions estimated with HDE	×	×
HDE_Fit_DP.pdf	DP probability density function estimated with HDE	×	×
HDE_Fit_DP_uncertainty.pdf	DP probability density function and relative uncertainty estimated with HDE		×
HDE_Fit_DPS.pdf	DPS probability density function estimated with HDE	×	×
HDE_Fit_DPS_uncertainty.pdf	DPS probability density function and relative uncertainty estimated with HDE		×
HDE_Fit_IG.pdf	IG probability density function estimated with HDE	×	×
HDE_Fit_IG_uncertainty.pdf	IG probability density function and relative uncertainty estimated with HDE		×
KDE_Fit_Comparison.pdf	Comparison of DP, DPS and IG probability density functions estimated with KDE	×	×
KDE_Fit_DP.pdf	DP probability density function estimated with KDE	×	×
KDE_Fit_DP_uncertainty.pdf	DP probability density function and relative uncertainty estimated with KDE		×
KDE_Fit_DPS.pdf	DPS probability density function estimated with KDE	×	×
KDE_Fit_DPS_uncertainty.pdf	DPS probability density function and relative uncertainty estimated with KDE		×
KDE_Fit_IG.pdf	IG probability density function estimated with KDE	×	×
KDE_Fit_IG_uncertainty.pdf	IG probability density function and relative uncertainty estimated with KDE		×

SOFTWARE OUPUT	DESCRIPTION	BATCH interface (advanced)	WPS interface (basic)
MLE_Fit_Comparison.pdf	Comparison of DP, DPS and IG probability density functions estimated with MLE	×	×
MLE_Fit_DP.pdf	DP probability density function estimated with MLE	×	×
MLE_Fit_DP_uncertainty.pdf	DP probability density function and relative uncertainty estimated with MLE		×
MLE_Fit_DPS.pdf	DPS probability density function estimated with MLE	×	×
MLE_Fit_DPS_uncertainty.pdf	DPS probability density function and relative uncertainty estimated with MLE		×
MLE_Fit_IG.pdf	IG probability density function estimated with MLE	×	×
MLE_Fit_IG_uncertainty.pdf	IG probability density function and relative uncertainty estimated with MLE		×
CDF_MLE_Fit_Comparison.pdf	Comparison of DP, DPS and IG cumulative density functions estimated with MLE	× (discrete calculation)	× (integral calculation)
CDF_MLE_DP.pdf	DP cumulative density function estimated with MLE	× (discrete calculation)	× (integral calculation)
CDF_MLE_DPS.pdf	DPS cumulative density function estimated with MLE	× (discrete calculation)	× (integral calculation)
CDF_MLE_IG.pdf	DPS cumulative density function estimated with MLE	× (discrete calculation)	× (integral calculation)
Model_Comparison_DP.pdf	Comparison of DP cumulative density functions estimated with HDE, KDE and MLE	×	×
Model_Comparison_DPS.pdf	Comparison of DPS cumulative density functions estimated with HDE, KDE and MLE	×	×
Model_Comparison_IG.pdf	Comparison of IG cumulative density functions estimated with HDE, KDE and MLE	×	×
TEXTUAL OUTPUTS			
HDE_Results.txt	Probability Density Functions parameters estimated for DP, DPS and IG probability density functions estimated with HDE	×	×
HDE_Fit_DP_parameter_uncertainty.txt	Percentile values of the parameters of the DP probability density function estimated with HDE		×
HDE_Fit_DPS_parameter_uncertainty.txt	Percentile values of the parameters of the DPS probability density function estimated with HDE		×
HDE_Fit_IG_parameter_uncertainty.txt	Percentile values of the parameters of the IG probability density function estimated with HDE		×

SOFTWARE OUPUT	DESCRIPTION	BATCH interface (advanced)	WPS interface (basic)
KDE_Results.txt	Probability Density Functions parameters estimated for DP, DPS and IG probability density functions estimated with KDE	×	×
KDE_Fit_DP_parameter_uncertainty.txt	Percentile values of the parameters of the DP probability density function estimated with KDE		×
KDE_Fit_DPS_parameter_uncertainty.txt	Percentile values of the parameters of the DPS probability density function estimated with KDE		×
KDE_Fit_IG_parameter_uncertainty.txt	Percentile values of the parameters of the IG probability density function estimated with KDE		×
MLE_Results.txt	Probability Density Functions parameters estimated for DP, DPS and IG probability density functions estimated with MLE	×	×
MLE_Fit_DP_parameter_uncertainty.txt	Percentile values of the parameters of the DP probability density function estimated with MLE		×
MLE_Fit_DPS_parameter_uncertainty.txt	Percentile values of the parameters of the DPS probability density function estimated with MLE		×
MLE_Fit_IG_parameter_uncertainty.txt	Percentile values of the parameters of the IG probability density function estimated with MLE		×
Bootstrapped_KS_Test_Results.txt	Bootstrapped Kolmogorov-Smirnov test results obtained for the DP, DPS and IG probability density functions estimated with HDE, KDE and MLE		×
ProbabilityResults.txt	PDF and CDF values calculated for each landslide area values (specified in the textual input file) for DP, DPS and IG distributions using HDE, KDE and MLE estimation method		×
GEOGRAPHICAL OUTPUTS (only when using a shapefile as input)			
ProbabilityResults.shp	PDF and CDF values calculated for each landslide point/polygon (specified in the input file) for DP, DPS and IG distributions using HDE, KDE and MLE estimation method		×

Table 6. List of the LAND-Stat (Landslide Statistics) software outputs for the batch interface (Advanced Version) and WPS (Basic Version) of the software.

3.3 LAND-Stat implementation in the SAFETY project

LAND-Stat will be exploited to derive landslide size statistics in the two project test areas: Volterra site (Tuscany region, Italy) and the Canary Islands (Spain).

During SAFETY, improvements made to the LAND-Stat software code will be tested and verified. The new functionalities are focused to analyse different landslide size statistics. In addition to the landslide area, LAND-Stat is able to estimate the statistical distribution of

population/samples of other landslide size measures as volume, length, width, length-width ratios, etc. derived from inventories. This makes the software applicable in the Canary Island mainly affected by rock falls.

REFERENCES

Ardizzone F., Fiorucci F., Santangelo M., Cardinali M., Mondini A. C., Rossi M., Guzzetti F. (2013). Very-high resolution stereoscopic satellite images for landslide mapping. In Landslide Science and Practice (pp. 95-101). Springer Berlin Heidelberg.

Ardizzone F., Cardinali M., Carrara A., Guzzetti F., Reichenbach P. (2002). Impact of mapping errors on the reliability of landslide hazard maps. Natural Hazards and Earth System Sciences, 2:1-2, 3-14.

Brabb E.E. 1984. Innovative approaches to landslide hazard mapping. Proceedings 4th International Symposium on Landslides, Toronto, 1, pp 307–324

Chung C.–J.F. and Fabbri A.G. 2003. Validation of Spatial Prediction Models for Landslide Hazard Mapping. Natural Hazards 30:3, 451–472

Fawcelt T. (2006) An Introduction to ROC Analysis. Pattern Recognition Letters 27 (8): 861–874. doi:10.1016/j.patrec.2005.10.010.

Guzzetti F., Carrara A., Cardinali M., Reichenbach P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi–scale study, Central Italy. Geomorphology 31: 181–216

Guzzetti F., Galli M., Reichenbach P., Ardizzone F., Cardinali M. (2006b). Landslide hazard assessment in the Collazzone area, Umbria, Central Italy. Natural Hazards and Earth System Sciences, 6, 115–131.

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

Guzzetti F., Reichenbach P., Ardizzone F., Cardinali M., Galli M. (2006a). Estimating the quality of landslide susceptibility models. Geomorphology, 81:1-2, 166-184

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., & Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale.Geomorphology, 72(1), 272-299.

Fiorucci F., Cardinali M., Carlà R., Rossi M., Mondini A. C., Santurri L., Guzzetti F. (2011). Seasonal landslide mapping and estimation of landslide mobilization rates using aerial and satellite images. Geomorphology, 129(1), 59-70.

Malamud B. D., Turcotte D. L., Guzzetti F., Reichenbach P. (2004). Landslide inventories and their statistical properties. Earth Surface Processes and Landforms, 29(6), 687-711.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Reichenbach P., Busca C., Mondini A. C., Rossi M. (2015) Land Use Change Scenarios and Landslide Susceptibility Zonation: The Briga Catchment Test Area (Messina, Italy). In Engineering Geology for Society and Territory-Volume 1 (pp. 557-561). Springer International Publishing.

Reichenbach P., Busca C., Mondini A. C., Rossi M. (2014) The Influence of Land Use Change on Landslide Susceptibility Zonation: The Briga Catchment Test Site (Messina, Italy). Environmental Management 54:1372–1384. DOI 10.1007/s00267-014-0357-0.

Rossi M., Guzzetti F., Reichenbach P., Mondini A.C., Peruccacci S. 2010. Optimal landslide susceptibility zonation based on multiple forecasts. Geomorphology, 114:3, 129-142.

Stark C. P. and Hovius, N. (2001). The characterization of landslide size distributions. Geophysical Research Letters, 28(6), 1091-1094.

Varnes D.J. IAEG Commission on Landslides (1984) Landslide hazard zonation-a review of principles and practice.