LAND-SUITE

(LANDslide - SUsceptibility Inferential Tool Evaluator)
User Guide - Version 1.0

Mauro Rossi[1*], Txomin Bornaetxea[2], Paola Reichenbach[1]

(Mauro.Rossi@irpi.cnr.it)

(1) CNR-IRPI, via Madonna Alta 126, 06128 Perugia, Italy; (2) Euskal Herriko Unibertsitatea (UPV/EHU), Barrio Sarriena s/n, 48940 Leioa, Spain

Consiglio Nazionale delle Ricerche

Istituto di Ricerca per la Protezione Idrogeologica

Index

1. Brief LAND-SUITE description

LAND-SUITE (LANDslide - SUsceptibility Inferential Tool Evaluator) is a software designed to support the landslide susceptibility assessment process, and integrate, extend and complete LAND-SE (Rossi et al., 2010; Rossi & Reichenbach, 2016). The software is implemented in R, a free software environment for statistical computing and graphics (R Core Team, 2021), and allows to: i) facilitate data preparation; ii) perform preliminary and exploratory analysis of the available data; iii) test different combinations of thematic/explanatory variables and select the optimal set; iv) test different model types and their combinations; and v) evaluate the models' performance and uncertainty. In synthesis, LAND-SUITE provides an expert user with the possibility to perform more easily, more flexible and more informed statistically-based landslide susceptibility applications and zonations.

LAND-SUITE is composed of three modules:

- LAND-SIP: LANDslide - Susceptibility Input Preparation

- LAND-SVA: LANDslide - Susceptibility Variable Analysis

- LAND-SE: LANDslide - Susceptibility Evaluation

The modules are coded as separate R scripts and can be executed under different operating systems. Each module requires a configuration depending on the application chosen by the user. The parameters and outputs of LAND-SIP, LAND-SVA and LAND-SE are listed and described in this guide.

The common LAND-SUITE run starts with LAND-SIP, which executes in cascade LAND-SVA and successively LAND-SE. Alternatively, only one of the last two modules can be executed after LAND-SIP, depending on the user needs and the type of application. The three modules can also be executed separately, as long as the user provides the appropriate data input.

LAND-SUITE can use data in raster (GeoTiff) or vector (ESRI shapefile) format, depending on the selected type of application. The input data can be provided with different resolutions (for rasters) and attributes (for vectors), using different reference systems. The software accepts data stored in a local system folder, or alternatively, the data can be stored and imported from a GRASS GIS Mapset. The use of GRASS Mapset is preferred for large datasets because it facilitates the massive code execution via command line interface.

2. Software and hardware requirements

LAND-SUITE requires an R (R Core Team, 2021) installation and some functionalities of LAND-SIP require GRASS GIS installation (GRASS Development Team, 2017). LAND-SUITE was tested successfully with R version 4.1.1 (64bit) and GRASS GIS version 7. The software can be executed under WindowsOS and LinuxOS. MacOS has not been fully tested and problems may arise when using GRASS GIS binding. To run LAND-SUITE under WindowsOS with GRASS GIS binding, it is necessary to install GRASS GIS through OSGeo4W64, which is a binary distribution of a broad set of open source geospatial software for Windows environments (https://trac.osgeo.org/osgeo4w/). We generally recommend using LinuxOS, due to the better software integration at a bash scripting level.

The R packages dependencies include:

- **caret**: Max Kuhn (2021). caret: Classification and Regression Training. R package version 6.0-88. https://CRAN.R-project.org/package=caret
- **corrplot**: Taiyun Wei and Viliam Simko (2021). R package 'corrplot': Visualization of a Correlation Matrix (Version 0.90). Available from https://github.com/taiyun/corrplot
- **data.table**: Matt Dowle and Arun Srinivasan (2021). data.table: Extension of data.frame`. R package version 1.14.0. https://CRAN.R-project.org/package=data.table
- **ggmosaic**: Haley Jeppson, Heike Hofmann and Di Cook (2021). ggmosaic: Mosaic Plots in the 'ggplot2' Framework. R package version 0.3.3. https://CRAN.R-project.org/package=ggmosaic
- **ggplot2**: H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
- **Hmisc**: Frank E Harrell Jr, with contributions from Charles Dupont and many others. (2021). Hmisc: Harrell Miscellaneous. R package version 4.5-0. https://CRAN.R-project.org/package=Hmisc
- **MASS**: Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- **nnet**: Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- **perturb** (need the installation of an old package version from source, since the package is not available for R 4.1.1): John Hendrickx (2019). perturb: Tools for

Evaluating Collinearity. R package version 2.10. https://CRAN.R-project.org/package=perturb

- **raster**: Robert J. Hijmans (2021). raster: Geographic Data Analysis and Modeling. R package version 3.4-13. https://CRAN.R-project.org/package=raster

- **RColorBrewer**: Erich Neuwirth (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer

- **rgdal**: Roger Bivand, Tim Keitt and Barry Rowlingson (2021). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.5-23. https://CRAN.R-project.org/package=rgdal

- **rgrass7**: Roger Bivand (2021). rgrass7: Interface Between GRASS 7 Geographical Information System and R. R package version 0.2-5. https://CRAN.R-project.org/package=rgrass7

- **sp**: Pebesma, E.J., R.S. Bivand, 2005. Classes and methods for spatial data in R. R News 5(2), https://cran.r-project.org/doc/Rnews/.

- **vcd**: David Meyer, Achim Zeileis, and Kurt Hornik (2020). vcd: Visualizing Categorical Data. R package version 1.4-8.

- **verification**: NCAR - Research Applications Laboratory (2015). verification: Weather Forecast Verification Utilities. R package version 1.42. https://CRAN.R-project.org/package=verification

LAND-SUITE does not require specific hardware requirements, however, the CPU clock, RAM dimension, and disk space availability constrain the maximum extension of the study area, data resolution and calculation time.

3. User skills

LAND-SUITE is designed to be properly used by experienced geomorphologists. Experienced practitioners are expected to use the code, with the support of experts in the field of environmental planning and management for a correct and reliable interpretation and exploitation of the results.

Moreover, a appropriate LAND-SUITE execution requires: i) basics knowledge of the R language to run the scripts; ii) GIS skills to prepare and handle spatial input and output data; iii) comprehension of descriptive statistics (statistical distribution, probability density estimation, etc.) and correlation analysis to use properly LAND-SVA functionalities; iv) experience on multivariate statistical models and their evaluation skills/metrics (ROC plot,

contingency table and plots, success/prediction rate curves, etc.) to evaluate model performances; and v) specific geomorphological expertise for a correct interpretation of the results. Advanced R programming language skills are required to customize and tailor the code and to adapt it to different needs.

4. Data requirememnts and specification

The selection and the preparation of input data is an important preliminary phase for the landslide susceptibility assessment and for the execution of LAND-SUITE.

LAND-SUITE, requires different input data:

(i) a landslide inventory map (e.g., historical, geomorphological, event and multi-temporal landslide inventories) used as dependent or grouping variable in the susceptibility analysis, and

(ii) a set of thematic maps to be used as independent explanatory variables that can be continuous (e.g., slope, elevation, etc.) or categorical (e.g., lithology, land use, etc.).

The inventory maps are used to classify the mapping units choosen for the analysis (e.g. pixels, polygonal boundaries) as 0 or 1, a largely accepted method to indicate the absence or presence of landslides.

The thematic maps used as explanatory variables require different pre-processing phases to be properly used in the statistical models. In particular, the continuous variables can be used directly in LAND-SUITE, while the categorical need to be converted into dummy variables. Dummy variables represent categorical data using distinct numerical categories and can be derived using the following approaches:

- Heuristical (based on personal judgment/experience), which is performed assigning to different categories an index value that increases/decreases with the expected propensity to instability (Carrara, 1983);

- Numerical, calculating the relative landslide incidence in each category (Lee and Min, 2001; Yilmaz, 2009; Trigila et al., 2015; Van Westen et al., 1997);

- Numerical, calculating the percentage of each category in the mapping unit, generally adopted when using polygon-like terrain partitions (Carrara, 1983; Carrara et al., 2008).

Another common pre-elaboration step for the thematic data is the variable scaling (Becker et al., 1988; Han et al., 2011). The data scaling may hold a strong statistical meaning, particularly

in the case of normally distributed variables (i.e. following a Gaussian distribution), transforming their distribution into a standard normal distribution (i.e., with mean value equal to 0 and standard deviation equal to 1). In all the other cases, such transformation chiefly aims at making variable values ranging on a similar scale, improving their comparability. In many cases, the scaling does not lead to significantly different susceptibility results, but in some conditions, it may avoid problems on numerical model convergence. LAND-SUITE includes a specific option to perform scaling that will be discussed later, but this preliminary transformation can be also done in a GIS or other computational environments.

LAND-SUITE can use data in raster (GeoTiff) or vector (ESRI shapefile) format, depending on the selected type of application. The input data can be provided with different resolutions (for rasters) and attributes (for vectors) and using different reference systems. The software can accept data stored in a local system folder, or alternatively, the data can be stored and imported from a GRASS GIS Mapset. The use of GRASS Mapset is preferred for large datasets because it facilitates the massive code execution via command line interface. To execute the tools in raster mode, the user needs to provide all the inputs as raster layers, while in the vector mode the software requires a vector layer (e.g., a polygon-based slope unit partition layer) with the attribute table containing both the dependent (i.e. derived from the landslide inventory map) and the independent variables (i.e. the thematic information).

5. LAND-SIP: parameters and outputs

LAND-SIP is designed for input preparation and its main purpose is the subdivision and identification of the training and validation datasets that will be used by the other two modules. The dataset partition is controlled and customized by the use, that can select the type of mapping unit (i.e., raster or polygons), choose the appropriate combination of variables, define the extent (i.e., using a mask) of the training and validation areas, and choose the output types. One of the main parameters required by LAND-SIP is the selection of the location and type of the input spatial (or geographical) layers. These can be system files in a local folder or layers in a Grass GIS mapset. This allows the user to use his preferred spatial (GIS or alternative) computational environments. Table S1 lists and briefly describes the variables and the parameters that need to be specified in LAND-SIP. The table groups: i) general parameters; ii) parameters for applications with data in vector format [mapping_unit_type <- "vector"]; and iii) parameters for applications with data in raster format [mapping_unit_type <- "raster"].

Table S1. Variables and parameters to be specified in the *LAND-SIP_vX_YYYYMMDD.R* file.

PARAMETER	Type[1] - Description	EXAMPLE VALUES
General parameters		
main_working_directory	***chr*** Name of the main working directory or folder where the output will be stored	"X:/R/test" (under Windows) "/home/R/test" (under Linux) "/Users/R/test" (under Mac OS)
use_single_working_directory	***logi*** Define if the necessary script and configuration files are located in the same working directory	TRUE / FALSE
name_results_directory	***chr*** A folder with the desired name will be created under the main directory to store all the results	"myresults"
path_to_variable_analysis_script	***chr*** Name of the working directory or folder where LAND-SVA script is located. Only needed if use_single_working_directory = FALSE	"X:/R/test" (under Windows) "/home/R/test" (under Linux) "/Users/R/test" (under Mac OS)
path_to_susceptibility_script	***chr*** Name of the working directory or folder where LAND-SE script is located. Only needed if use_single_working_directory = FALSE	"X:/R/test" (under Windows) "/home/R/test" (under Linux) "/Users/R/test" (under Mac OS)
path_to_susceptibility_script_config_files	***chr*** Name of the working directory or folder where configuration files are located. Only needed if use_single_working_directory = FALSE	"X:/R/test" (under Windows) "/home/R/test" (under Linux) "/Users/R/test" (under Mac OS)
executing_LAND_SVA	***logi*** Define if the analysis of the variables will be executed or not	TRUE / FALSE
LAND_SVA_script_version	***chr*** Specify the version of the LAND-SVA script file	LAND-SVA_*vX_YYYYMMDD*.R
executing_LAND_SE	***logi*** Define if the susceptibility analysis will be executed or not	TRUE / FALSE
LAND_SE_script_version	***chr*** Specify the version of the LAND-SE script file	LAND-SE_*vX_YYYYMMDD*.R
path_to_Rscript_exe	***chr*** Specify the version of the R package used for the analysis. This is only needed under Windows OS.	"C:/PROGRA~1/R/R-3.6.3/bin/x64/Rcmd"

export_shapefiles	*logi* Enables export the *.shp files containing the data of training and validation subsets.	TRUE / FALSE
export_txtfiles	*logi* Enables to export the *.txt files containing the data of training and validation subsets.	TRUE / FALSE
mapping_unit_type	*chr* Specify the type of mapping unit used for the analysis	raster / vector
enable_grass_import	*logi* Enables to read the input data directly from a GRASS location	TRUE / FALSE
myGISDbase	*chr* The system path containing the GRASS Location	"X:/R/test" (under Windows) "/home/R/test" (under Linux) "/Users/R/test" (under Mac OS)
myLocation	*chr* The name of the Location folder	"ELG_Location"
myMapset	*chr* The name of the Mapset folder	"VECTOR_cross_validation"
grass_layer_excluded	*chr* Layers in the GRASS mapset to be excluded from the analysis	("dem","mask_map","shade","geoforms","layer1","layer2","layer3")
myGRASS	*chr* The system path containing the GRASS package	"/usr/lib/grass78"
osgeo4w.root	*chr* Path to OSGEO4W64 (needed to execute LAND-SIP with GRASS under Windows)[2]	"C:\\OSGEO4W64"
Parameters for applications with data in vector format [mapping_unit_type <- "vector"]		
vector_name	*chr* The name of the vector map with the data. It can be in a GRASS mapset or in the main working directory	"SlopeUnits_with_data"
name_id_column	*chr* The name of the column indicating the single identifier of each feature	"cat"
name_dependent_colums	*chr* The name of the column indicating the presence or the absence of landslides with 1 or 0 values.	"landslide_presence"
use_separate_vector_validation_files	*logi* Enable to carry out the validation in a different vector map.	TRUE / FALSE
validation_vector_name	*chr* The name of the vector map to be used for validation. It can be in a GRASS mapset or in the main working	"SlopeUnits_validation"

	directory. It is only used if use_separate_vector_validation_files is TRUE	
validation_dim	*num* The percentage of mapping units to be used during the validation	70
change_dependent_proportion	*logi* Enable to change the proportion between mapping units with values 1 and 0.	TRUE / FALSE
nodependent_percentage	*num* Value to change the proportion between mapping units with values 1 and 0. For balanced amount set to 50	50
validation_entire_area	*logi* Enable to carry out the validation in the complete study area, i.e. considering all the mapping units.	TRUE / FALSE
reducing_sample_size	*logi* Enable to reduce the size of the sample creating an aleatory subset.	TRUE / FALSE
sample_size_proportion	*num* Value to reduce the size of the sample.	10
vector_column_mappingunitarea	*chr* The name of the column in the vector map containing the area of the mapping unit	"area_slu"
vector_column_landslidearea	*chr* The name of the column of the vector map containing the area of the mapping unit covered by landslides	"area_lnd"
Parameters for applications with data in raster format [mapping_unit_type <- "raster"]		
raster_dependent_name	*chr* Name of the landslide data in raster format	"landslides"
mask_name	*chr* Name of the mask to be used in the analysis	"mask"
EPSG_code	*num* Coordinate Reference System EPSG Code	32633
use_mask	*logi* Enable to use a mask for the analysis	TRUE / FALSE
use_separate_raster_validation_files	*logi* Enable to use different input data for training and validation. This is used for spatial validation approaches. If TRUE use_separate_raster_validation_dependent must be FALSE	TRUE / FALSE

use_separate_raster_valida tion_dependent	*logi* Enable to use different dependent variables for training and validation, but the same explanatory variables. This is used for temporal validation approaches. If TRUE use_separate_raster_validation_files must be FALSE	TRUE / FALSE
raster_dependent_name_v alidation	*chr* The name of the landslide data in raster format for validation	"landslides_validation"
mask_name_validation	*chr* Name of the mask to be used in the validation analysis	"mask_validation"
validation_filename_string	*chr* This is a suffix to identify the separate layers for the validation. This requires use_separate_raster_validation_depend ent=TRUE	"_validation"
generate_random_validati on_dependent	*logi* Enable a random validation dependent variable generation. Useful when there isn't a different validation dependent	TRUE / FALSE
use_mask_validation	*logi* Enable the use of a specific mask for the validation	TRUE / FALSE
original_data_training	*logi* Enable training only in a portion of the study area. When TRUE, this requires reducing_sample_size=FALSE and change_dependent_proportion=FALSE	TRUE / FALSE
validation_dim	*num* **The** percentage of mapping units to be used during the validation	70
change_dependent_proport ion	*logi* Enable to change the proportion between mapping units with value 1 and 0.	TRUE / FALSE
nodependent_percentage	*num* Value to change the proportion between mapping units with value 1 and 0. For balanced amount set to 50	50
validation_entire_area	*logi* Enable to carry out the validation in the complete study area, i.e. considering all the mapping units.	TRUE / FALSE
reducing_sample_size	*logi* Enable to reduce the size of the sample creating an aleatory subset.	TRUE / FALSE
sample_size_proportion	*num* Value to reduce the size of the sample.	10

use_polygon_based_landslide_random_selection	*logi* Enable the random selection of landslides as single units instead of selecting at random the landslide raster cells	TRUE / FALSE
vector_dependent_name	*chr* The name of the polygon shapefile containing the landslide inventory	"landslide_inventory.shp"
tolerance	*num* Tolerance to check the areas covered by training and validation samples when the landslide sampling criteria is selected	0.02
load_training_rawdata	*logi* Enable the skip of the raw training data preparation and the load of an existing .RData containing a training set.	TRUE / FALSE
training_raw_rdata	*chr* The name of the .RData file containing a training set. It is only used if load_training_rawdata is TRUE	"raw_data_binded.RData"

[1] *num*: numerical; *logi*: logical; *chr*: character
[2] OSGeo4W64 is a binary distribution of a broad set of open source geospatial software for windows environments (https://trac.osgeo.org/osgeo4w/); the actual LAND-SIP configuration works with GRASS GIS 7.8.

The main output of LAND-SIP is a binary R data file (Table S2) containing the training and validation sets in a tabular structure (data.frame, Spatial*DataFrame). When executing LAND-SUITE in cascade mode (i.e., when "executing_LAND_SVA = TRUE" or "executing_LAND_SE = TRUE" or both equal TRUE) the file is used as input.

Table S2. List of LAND-SIP outputs.

SOFTWARE OUTPUT	DESCRIPTION
R OBJECT OUTPUT	
datatable_inventory.RData	RData file containing data in tabular format to be used as input in LAND-SVA and LAND-SE

6. LAND-SVA: parameters and outputs

LAND-SVA is designed for the explorative analysis of the LAND-SE training and validation input datasets and facilitates the selection of the optimal combination of variables. The tool automatically detects continuous or dummy variables (i.e., derived from categorical data and normally represented with numerical discrete values) and selects the outputs accordingly. All the analyses are performed separately for the training and validation datasets, with the main purpose to provide the possibility to analyse and control the dataset differences. Table S3 lists and briefly describes the variables and the parameters that need to be specified in LAND-SVA.

Table S4 lists the outputs that are explained in detail in the text of the LAND-SUITE article.

Table S3. Variables and parameters to be specified in the *LAND-SVA_vX_YYYYMMDD.R* file.

PARAMETER	Type[1] - Description	EXAMPLE VALUES
wd_selected *	*chr* Name of the working directory or folder where the output will be stored. * *Only needed if LAND-SVA is run on its own.*	"X:/R/test" (under Windows) "/home/R/test" (under Linux) "/Users/R/test" (under Mac OS)
rdata_file *	*chr* Name of the .RData file containing input data. * *Only needed if LAND-SVA is run on its own.*	"datatable_inventory.RData" (default value)
enable_NA_removal	*logi* Define if the possible NA values will be removed or not. If TRUE rows with at least an NA value will be removed.	TRUE / FALSE
enable_variable_scaling	*logi* Define if the variables will be scaled or not.	TRUE / FALSE
max_categories	*num* The max number of unique values for a variable to be considered as categorical.	10
enable_multicollinearity_test	*logi* Enable multicollinearity test execution.	TRUE / FALSE
type_correlation	*chr* Specifies the type of correlations to compute in the correlation test.	"pearson" / "spearman"
exclude_insignificant_correlation_coefficients	*logi* If TRUE in the correlation plot only the significant correlations are shown. * *Use this option only when fully confident with its functioning.*	TRUE / FALSE
sig_level_value	*num* Significance level for discriminating correlated variable: normally 0.001, 0.01, 0.05	0.01
export_shapefiles	*logi* Define if the shapefile of points corresponding to the data tables will be exported or not	TRUE / FALSE
export_txtfiles	*logi* Define if the raining and validation tables in tab-separated .txt format will be exported or not	TRUE / FALSE

[1] *num*: numerical; *logi*: logical; *chr*: character.

Table S4. List of LAND-SVA outputs. In the table, XX is used in place of the continuous variable names, while YY for the categorical variables.

SOFTWARE OUTPUT	DESCRIPTION
GRAPHICAL OUTPUTS	
DensityPlot_XX.pdf	Density plot for the variable XX
ConditionalDensityPlot_XX.pdf	Conditional Density Plot for the variable XX
HistogramPlot_YY.pdf	Histogram plot for the variable YY
MosaicPlot_YY.pdf	Mosaic Plot for the variable YY
Variables_Correlogram_matrix_training.pdf	Correlogram matrix for the training data set
Variables_Correlogram_matrix_validation.pdf	Correlogram matrix for the validation data set
TEXTUAL OUTPUTS	
Variables_CollinearityAnalysis_training.txt	Result of the collinearity analysis (training)
Variables_CollinearityAnalysis_validation.txt	Result of the collinearity analysis (validation)
R OBJECT OUTPUT Needed only if enable_NA_removal = TRUE or enable_variable_scaling = TRUE or exclude_insignificant_correlation_coefficients = TRUE	
datatable_inventory.RData	RData file containing data in tabular format to be used as input in LAND-SE

7. LAND-SE: parameters and outputs

LAND-SE is the module for landslide susceptibility modelling that is described in detail in Rossi and Reichenbach (2016). The software allows the user to perform and combine different statistical susceptibility modelling methods, evaluate the results and estimate the associated uncertainty. In particular, the script is focused on: i) selection of different combinations of multivariate approaches; ii) evaluation of the model prediction skills and performances using success contingency matrices and plots, ROC curve and prediction rate curves; iii) estimation of the associated uncertainty and errors; iv) production of results in standard geographical formats (shapefiles, GeoTIFF); and v) usage of additional computational parameters to tune the calculation procedure for the analysis of large data sets. LAND-SE parameters are specified in the *LAND-SE_vX_YYYYMMDD.R* script (Table S5) and in two configuration files, namely the *configuration.txt* and *configuration_spatial_data.txt* (Table S6 and Table S7). *LAND-SE* can be executed in two modes: the *standard mode* that provides output only in textual and graphical format and the *geomode* that also handles geographical data. When running LAND-SE in cascade from LAND-SIP, the *geomode* should be enabled to fully benefit from the additional spatial-related output.

Table S5. Variables and parameters to be specified in the LAND-SE_vX_YYYYMMDD.R file.

PARAMETER	Type[1] - Description	EXAMPLE VALUES
wd_selected *	*chr* Name of the working directory or folder where the output will be stored. * *Only needed if LAND-SE is run on its own.*	"X:/R/test" (under Windows) "/home/R/test" (under Linux) "/Users/R/test" (under Mac OS)
cd_selected *	*chr* Name of the working directory or folder containing configurations and input data files. * *Only needed if LAND-SE is run on its own. Normally this corresponds to wd_selected.*	"X:/R/test" (under Windows) "/home/R/test" (under Linux) "/Users/R/test" (under Mac OS)
load_rdata	*logi* Enable reading input data from a .RData file. * *Need to be TRUE when running LAND-SE in cascade from LAND-SIP.*	TRUE / FALSE
rdata_file	*chr* Name of the .RData file containing input data.	"datatable_inventory.RData" (default value)
enable_screen_plotting	*logi* Enable screen plotting during LAND-SE execution.	TRUE / FALSE
enable_multicollinearity_test	*logi* Enable multicollinearity test execution.	TRUE / FALSE
enable_rocplot_confidence interval	*logi* Enable using confidence interval estimation for ROC plots. * *TRUE requires intensive computational resources.*	TRUE / FALSE
bootstrap_constant_correction	*logi* Enable a procedure to avoid convergence problems when explanatory variables contain mostly 0 value (blocking susceptibility model convergence).	TRUE / FALSE
enable_detailed_data_export	*logi* Enable the final shapefile export. * *Need to be FALSE if the shapefile in output is estimated to be bigger than 2Gb.*	TRUE / FALSE
enable_detailed_data_export_tiff	*logi* Enable the uncertainty and matching code map export in .tiff format.	TRUE / FALSE

[1] *num*: numerical; *logi*: logical; *chr*: character

Table S6. List of the parameters in the *configuration.txt* file. The file is a tab separated .txt and has a tabular structure that allow the user to modify the parameters for each model (i.e., LDA, QDA, LRM, NNM, CFM) listed in the first column.

PARAMETER	Type[1] - Description	EXAMPLE VALUES
RUN	*logi* Enable the execution of the different statistical susceptibility model	YES / NO
BOOTSTRAP_SAMPLES_ROC_CURVE	*num* Number of bootstrap resampling used to estimate ROC uncertainty	100
ANALYSIS_PARAMETER	*chr* Specific analysis parameters applicable for different statistical susceptibility models.	For QDA can be: SEL (eliminate dummy variables), DUM (maintain dummy variables transformed in numeric introducing a random variation between -0.1 and +0.1 and 0 are avoided sampling quantities close to 0). For LRM can be: INT (considering all the explanatory variables interactions). For NNM can be: NOR (default weight, half variables in the hidden layer), OPT (auto optimize the neural structure: slower). For LDA and CFM no option are available.
BOOTSTRAP_MODEL_VARIABILITY_RUN	*logi* Enable the execution of the bootstrap model variability analysis	YES / NO
BOOTSTRAP_SAMPLES_MODEL_VARIABILITY	*num* Number of bootstrap resampling used to estimate model variability	100

[1] *num*: numerical; *logi*: logical; *chr*: character

Table S7. List of the parameters in the *configuration_spatial_data.txt* file. The file is a tab separated .txt and has a tabular structure that allow the user to modify the parameters listed in the first row.

PARAMETER	Type[1] - Description	EXAMPLE VALUES
TYPE	*chr* Specific the input and output data type	SHAPEFILE / RASTER
PRESENCE	*logi* Enable the *geomode* option to use spatial data in input and output. * *Need to be YES when running LAND-SE in cascade from LAND-SIP. If NO, LAND-SE will produce only a limited number of output. See Table S8 for checking the output produced.*	YES / NO
ID_FIELD	*chr* Name of the column (field name) in the shapefile attribute table containing	ID

	the univocal code of the mapping unit identification	
EPSG_CODE	*num* Coordinate Reference System EPSG Code	32633
AREA_SU_FIELD	*chr* Name of the column (field name) in the shapefile attribute table containing the mapping unit area	area_mapping
AREA_LANDSLIDE_FIELD	*chr* Name of the column (field name) in the shapefile attribute table containing the landslide area in the mapping unit	area_landslide
GEOMETRY	*chr* Type of features used in the analysis	POINTS (for the pixel-based analysis) POLYGONS (for the polygon-based analysis)
RASTER_RES	*num* Dimension of the raster cell (only pixel-based analysis)	30
RASTER_EXPORT	*logi* Enable the export of the output map also in GeoTIFF format (only pixel-based analysis)	TRUE / FALSE

[1] *num*: numerical; *logi*: logical; *chr*: character

Table S8 shows the complete list of the LAND-SE outputs for the two execution modes with a short description. Output are grouped in graphical (.pdf format), textual (.txt format), and geographical (.shp or .tif format).

Table S8. List of LAND-SE outputs. In the table, XX is used in place 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 the output provided by the *geomode* that can be enabled in the *configuration_spatial_data.txt* file (see Table S7).

SOFTWARE OUTPUT	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	X
result_XX_Model_Susceptibility_Map.pdf	Map of the XX model susceptibility values obtained in the training phase	X
result_XX_PredictionRateCurve.pdf	Prediction rate curve obtained in the validation phase	X
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	X
result_XX_Validation_MatchingCode_Map.pdf	Map of the XX model validation errors derived from the contingency table	X
result_XX_Validation_Susceptibility_Map.pdf	Map of the XX model susceptibility values obtained in the validation phase	X
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	X
result_XX_validation/validation.shp	Folder containing the shapefile of XX model results obtained in the training phase	X
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*)	X
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*)	X
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*)	X
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*)	X
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*)	X
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*)	X

In the *standard* mode, LAND-SE requires two input files (training.txt and validation.txt) in tab-separated .txt format, storing data in a tabular structure. The two files are organized in columns (column names cannot contain spaces) with the following information: i) the first column contains the univocal code of the mapping unit identification; ii) the second column contains the value of the grouping variable (i.e. dependent variable) showing the absence/presence (respectively 0 and 1) of landslides in the mapping unit, and iii) from the third to the *n*-th columns, values of the *n* explanatory variables (i.e. independent variables). In *geomode*, two additional files are required in shapefile format (.shp) with the geographical information of the training and validation dataset. The LAND-SE input may be also provided directly in an .RData file produced by LAND-SIP or LAND-SVA (i.e., when load_rdata = TRUE).

Table S9. List of the input data files required to run LAND-SE when load_rdata = FALSE. If load_rdata = TRUE these data are provided as R objects in the .RData file, that are generated by LAND-SIP and LAND-SVA.

LAND-SE inputs (when load_rdata = FALSE)	DESCRIPTION
training.txt	Tab-separated textual file with columns (named without spaces) containing in order: 1) the univocal ID of the mapping unit, 2) the grouping variable with 0 or 1 values, 3 to n explanatory numerical variable. Each row of the file is one mapping unit of the training dataset.
validation.txt	Tab-separated textual file with columns (named without spaces) containing in order: 1) the univocal ID of the mapping unit, 2) the grouping variable with 0 or 1 values, 3 to n explanatory numerical variable. Each row of the file is one mapping unit of the validation dataset.
training.shp Required only for the *geomode*, to prepare the spatial output and the calculation of the success rate curve	Shapefile with the geographical data (points or polygons) of the training dataset. The shapefile attribute table contains the following attributes: 1) the univocal ID, 2) the area, and 3) the landslide area in the mapping unit.
validation.shp Required only for the *geomode*, to prepare the spatial output and the calculation of the prediction rate curve	Shapefile with the geographical data (points or polygons) of the validation dataset. The shapefile attribute table contains the following attributes: 1) the univocal ID, 2) the area, and 3) the landslide area in the mapping unit.

8. LAND-SUITE configuration examples

LAND-SUITE can be applied in a flexible way, with different configurations of the parameters of LAND-SIP, LAND-SVA and LAND-SE. In particular, LAND-SIP allows to execute many different susceptibility applications that require differentiated training and validation dataset partitions. As an example, in the following, we list and explain how to configure the software parameters (including LAND-SIP, LAND-SVA and LAND-SE) to run 5 common susceptibility applications, hereafter referred to as "Cases". Figure 1 shows graphically the application differences.

Case A: In this case, the susceptibility modelling is performed applying a regular cross-validation approach. A balanced random sampling is used to select the grouping variable mapping units following the "pixels sampling" selection criteria, with the size of training and validation datasets (e.g., 70% training and 30% validation) selected by the user (Figure 1 Case A). This configuration is usually applied for exploratory analysis focused on the preliminary evaluation of the explanatory variables (see LAND-SVA section), and of the statistical performance of the model. During this phase, the user can select, add or remove explanatory variables before the application of the model to the entire study area (Case C).

Case B: This application considers a cross-validation approach similar to Case A, but the partition of the training and validation datasets apply the "landslides sampling" selection. As in Case A, a balanced random sampling and a specific size of the training and validation datasets (e.g., 70% training and 30% validation) are chosen by the user (Figure 1 Case B). Case B can be used to analyse the explanatory variables and to test the modelling results as well as its dependency from the selection of different landslide samples.

Case C: The training configuration can be similar either to Case A or B, but the validation is applied to the entire study area. This case should be applied when the definitive set of explanatory variables is selected and the statistical performance of the model is satisfactory and acceptable. The validation map will show the susceptibility zonation for the entire extent of the study area (Figure 1 Case C).

Case D: This case performs a temporal validation, applicable when a landslide inventory map is available to train the model and a more recent inventory map (i.e., an event map) can be used for validation. In this case, the map used for the validation may cover the entire study area or only part of it (Figure 1 Case D). If the validation can be done only for a portion of the study area, the configuration requires two mask files, one covering the entire area and the other the area affected by the event. The selection of the explanatory variables and the preliminary evaluation of the model can be done with Case A or B. In the temporal subdivision, the training and the validation set are constituted by the same set of mapping units, but classified as 1 (presence) or 0 (absence) according to different distribution of landslides, usually achieved using two different landslide inventory maps. In this case the analysis is usually performed using the same explanatory variables associated to the same set of mapping units.

Case E: This case performs a spatial validation, with the model calibration performed in a given region of the study area and the validation in a different one. In such a case, the variable selection and the preliminary model testing could be performed only in one area similarly to Case A or B. In this case, the explanatory variables and landslide inventory map should be available for the two areas with the same characteristics (Figure 1 Case E). This configuration requires two different landslide inventory maps, two mask files and two explanatory variables datasets, respectively for the calibration and the validation region. In the spatial and random subdivision, the training and the validation dataset are composed of different mapping units, characterized by different grouping and explanatory variables. The main difference between

the spatial (Case E) and the random validation (Case A, B and C) relies upon the method chosen to separate the training and the validation dataset: in Case E, the dataset are spatially different (the two areas can be contiguous or not), in the other cases the subdivision is performed using a random sampling.
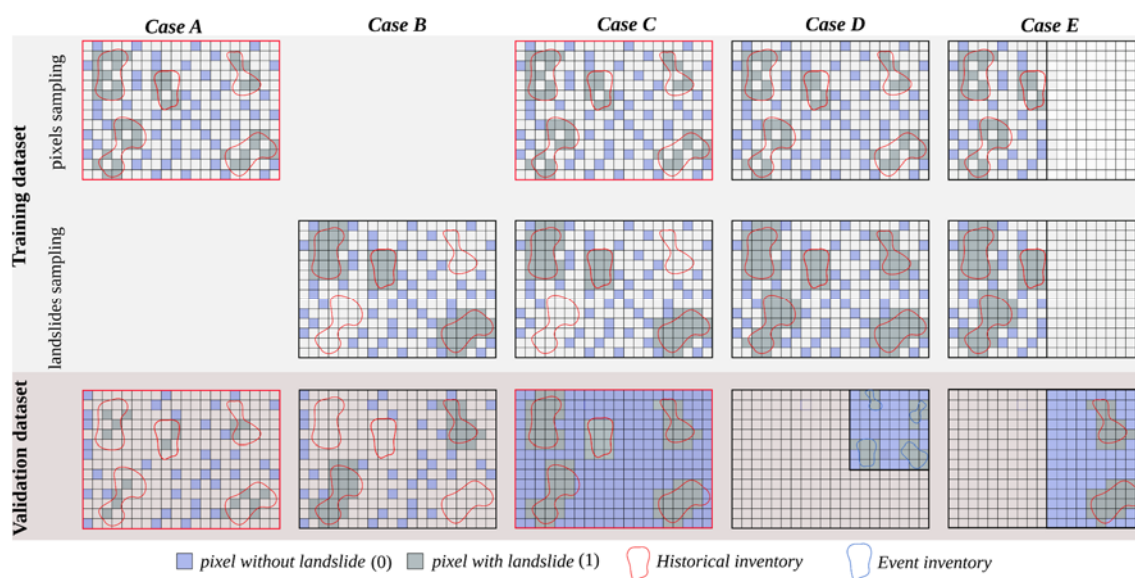

Figure 1. Simplified representations of the five applications, referred to as "Cases" in the figures and text, representing common susceptibility investigations. Red boxes show the cases described in the article.

Table S10 lists the input files that need to be provided by the user to apply the 5 cases. The file of the landslides spatial distribution, which is used in the modelling as the dependent variable, has to be classified as 1 for the landslide bodies and as 0 for the no-landslide areas. For Case D and Case E, respectively performing a temporal and a spatial validation, some variable layers need to be identified by a string at the end of the names (e.g., "_validation").

Table S10: List of the files (i.e., landslide data, mask and thematic explanatory variables) required to apply the 5 cases.

	Thematic data	Case A	Case B	Case C	Case D	Case E
Training	Landslide data	YES e.g. inventory	YES e.g. inventory	YES e.g. inventory	YES e.g. inventory	YES e.g. inventory
	Mask	YES e.g. mask	YES e.g. mask	YES e.g. mask	YES e.g. mask	YES e.g. mask_1
	Thematic variables	YES e.g. slope, lithology	YES e.g. slope, lithology	YES e.g. slope, lithology	YES e.g. slope, lithology	YES e.g. slope, lithology
Validation	Landslide data	Same as training	Same as training	Same as training	Different from e.g. inventory_validation	Different from e.g. inventory2_validation
	Mask	Same as training	Same as training	Same as training	Different from e.g. mask_validation	Different from e.g. mask2_validation
	Thematic variables	Same as training	Same as training	Same as training	Same as training	Different from e.g. slope_validation lithology_validation

Table S11 lists the parameters required by LAND-SIP to run the five cases in raster mode, while Table S12 shows the parameters required for Cases B, C, D and E using the polygon as mapping unit.

Table S11: The table shows the parameters that should be specified in LAND-SIP and the options required to run the described 5 cases with raster selected as mapping unit. In the table, the type *string* refers to text characters that should be written in the script within double-quotes. The type *logic* can be TRUE or FALSE. The type *value* is a numerical value. The text written in red are examples. The lines with an asterisk (*) are needed only for

windows operating systems. The lines with two asterisks (**) are options needed only if *enable_grass_import option* is TRUE. The lines with three asterisks (***), have to be verified depending on the operation system. The references to the .R scripts may change according to the released version.

	Type	Case A	Case B	Case C	Case D	Case E
GENERAL OPTIONS						
main_working_directory	string	"/home/data "	"/home/data "	"/home/data "	"/home/data "	"/home/data "
use_single_working_directory	logic	TRUE	TRUE	TRUE	TRUE	TRUE
name_results_directory	string	"myresults"	"myresults"	"myresults"	"myresults"	"myresults"
path_to_variable_analysis_script	string	"/home/code "	"/home/code "	"/home/code "	"/home/code "	"/home/code "
path_to_susceptibility_script	string	"/home/code "	"/home/code "	"/home/code "	"/home/code "	"/home/code "
path_to_susceptibility_script_config_files	string	"/home/code "	"/home/code "	"/home/code "	"/home/code "	"/home/code "
executing_LAND_SVA	logic	TRUE	TRUE	TRUE	TRUE	TRUE
LAND_SVA_script_version	string	LAND-SVA.R	LAND-SVA.R	LAND-SVA.R	LAND-SVA.R	LAND-SVA.R
executing_LAND_SE	logic	TRUE	TRUE	TRUE	TRUE	TRUE
LAND_SE_script_version	string	LAND-SE.R	LAND-SE.R	LAND-SE.R	LAND-SE.R	LAND-SE.R
path_to_Rscript_exe *	string	--	--	--	--	--
export_shapefiles	logic	TRUE	TRUE	TRUE	TRUE	TRUE
export_txtfiles	logic	TRUE	TRUE	TRUE	TRUE	TRUE
mapping_unit_type	string	"raster"	"raster"	"raster"	"raster"	"raster"
enable_grass_import	logic	TRUE	TRUE	TRUE	TRUE	TRUE
MyGISDbase **	string	"/home/data "	"/home/data "	"/home/data "	"/home/data "	"/home/data "
MyLocation **	string	"gipu_location"	"gipu_location"	"gipu_location"	"gipu_location"	"gipu_location"
MyMapset **	string	"cross_validation"	"cross_validation"	"cross_validation"	"temporal_validation"	"spatial_validation"
grass_layer_excluded **	string	--	--	--	--	--
MyGRASS ***	string	"/usr/lib/grass78"	"/usr/lib/grass78"	"/usr/lib/grass78"	"/usr/lib/grass78"	"/usr/lib/grass78"
RASTER OPTION						
raster_dependent_name	string	"inventory_rast"	"inventory_rast"	"inventory_rast"	"inventory_rast"	"inventory_1"
mask_name	string	"mask"	"mask"	"mask"	"mask"	"mask_1"
EPSG_code	value	25830	25830	25830	25830	25830
use_mask	logic	TRUE	TRUE	TRUE	TRUE	TRUE
use_separate_raster_validation_files	logic	FALSE	FALSE	FALSE	FALSE	TRUE
use_separate_raster_validation_dependent	logic	FALSE	FALSE	FALSE	TRUE	FALSE
raster_dependent_name_validation	string	--	--	--	"inventory_rast_validation"	"inventory_2_validation"
mask_name_validation	string	--	--	--	"mask_validation"	"mask_2_validation"
validation_filename_string	string	--	--	--	"_validation"	"_validation"
generate_random_validation_dependent	logic	FALSE	FALSE	FALSE	FALSE	FALSE
use_mask_validation	logic	FALSE	FALSE	FALSE	TRUE	TRUE
execute_multiple_validation	logic	FALSE	FALSE	FALSE	FALSE	FALSE
names_multiple_validation_layers	string	--	--	--	--	--
original_data_training	logic	FALSE	FALSE	FALSE	FALSE	FALSE
validation_dim	value	30	30	30	0	0
change_dependent_proportion	logic	TRUE	TRUE	TRUE	TRUE	TRUE
nodependent_percentage	value	50	50	50	50	50
validation_entire_area	logic	FALSE	FALSE	TRUE	TRUE	FALSE/TRUE
reducing_sample_size	logic	FALSE	FALSE	FALSE	FALSE	FALSE
sample_size_proportion	value	--	--	--	--	--
use_polygon_based_landslide_random_selection	logic	FALSE	TRUE	TRUE	FALSE	FALSE
vector_dependent_name	string	--	"landslides.shp"	"landslides.shp"	--	--
tolerance	value	--	0.02	0.02	--	--
load_training_rawdata	logic	FALSE	FALSE	FALSE	FALSE	FALSE
training_raw_rdata	string	--	--	--	--	--

Table S12: The table shows the parameters that should be specified in LAND-SIP and options required to run the described 5 cases with polygon selected as mapping unit. In the table, the type *string* refers to text characters that should be written in the script within double-quotes. The type *logic* can be TRUE or FALSE. The type *value* is a numerical value. The text written in red are examples. The lines with an asterisk (*) are needed only for windows operating systems. The lines with two asterisks (**) are options needed only if *enable_grass_import option* is

TRUE. The lines with three asterisks (***), have to be verified depending on the operation system.

	Type	Scenario A	Scenario B	Scenario C	Scenario D	Scenario E
GENERAL OPTIONS						
main_working_directory	string	--	"/home/data "	"/home/data "	"/home/data "	"/home/data "
use_single_working_directory	logic	--	TRUE	TRUE	TRUE	TRUE
name_results_directory	string	--	"myresults"	"myresults"	"myresults"	"myresults"
path_to_variable_analysis_script	string	--	--	--	--	--
path_to_susceptibility_script	string	--	--	--	--	--
path_to_susceptibility_script_config_files	string	--	--	--	--	--
executing_LAND_SVA	logic	--	TRUE	TRUE	TRUE	TRUE
LAND_SVA_script_version	string	--	LAND-SVA.R	LAND-SVA.R	--	LAND-SVA.R
executing_LAND_SE	logic	--	TRUE	TRUE	TRUE	TRUE
LAND_SE_script_version	string	--	LAND-SE.R	LAND-SE.R	LAND-SE.R	LAND-SE.R
path_to_Rscript_exe *	string	--	--	--	--	--
export_shapefiles	logic	--	TRUE	TRUE	TRUE	TRUE
export_txtfiles	logic	--	TRUE	TRUE	TRUE	TRUE
mapping_unit_type	string	--	"vector"	"vector"	"vector"	"vector"
enable_grass_import	logic	--	TRUE	TRUE	TRUE	TRUE / FALSE
MyGISDbase **	string	--	"/home/data "	"/home/data "	"/home/data "	"/home/data "
MyLocation **	string	--	"gipu_location"	"gipu_location"	"gipu_location"	"gipu_location"
MyMapset **	string	--	"cross_validation"	"cross_validation"	"temporal_validation"	"spatial_validation"
grass_layer_excluded **	string	--	--	--	--	--
MyGRASS ***	string	--	"/usr/lib/grass78"	"/usr/lib/grass78"	"/usr/lib/grass78"	"/usr/lib/grass78"
VECTOR OPTIONS						
vector_name	string	--	"slope_units"	"slope_units"	"slope_units_old"	"slope_units_1"
name_id_column	string	--	"cat"	"cat"	"cat"	"cat"
name_dependent_colums	string	--	"slide"	"slide"	"slide"	"slide"
use_separate_vector_validation_files	logic	--	FALSE	FALSE	TRUE	TRUE
validation_vector_name	string	--	--	--	"slope_units_new"	"slope_units_2"
validation_dim	value	--	30	30	0	0
change_dependent_proportion	logic	--	TRUE	TRUE	TRUE	TRUE
nodependent_percentage	value	--	50	50	50	50
validation_entire_area	logic	--	FALSE	TRUE	TRUE	FALSE/TRUE
reducing_sample_size	logic	--	FALSE	FALSE	FALSE	FALSE
sample_size_proportion	value	--	--	--	--	--
vector_column_mappingunitarea	string	--	"slope_unit_area"	"slope_unit_area"	"slope_unit_area"	"slope_unit_area"
vector_column_landslidearea	string	--	"slide_area"	"slide_area"	"slide_area"	"slide_area"

Table S13 lists as an example of the configuration of LAND-SVA prepared to run Case A, which can be used also for the other cases. Depending on the selected configuration, LAND-SVA can automatically remove pairwise correlated variables using the command line "exclude_insignificant_correlation_coefficient". We suggest the user to use this option only when fully confident with the tool functioning, preferring at the initial stage to remove the variables manually.

Table S13: List of LAND-SVA configuration parameters, alphanumeric types, possible values and comments, specified for Case A (in red), but also valid for the other cases.

	Type	Case A	Comments
enable_NA_removal	logic	FALSE	If TRUE rows with at least an NA value will be removed. If FALSE and error message will be returned.
enable_variable_scaling	logic	FALSE	Helpful when rank_defiency problem
max_categories	value	10	max number of unique value for a variable to be considered categorical
enable_multicollinearity_test	logic	TRUE	
type_correlation	string	"pearson"	It could be "pearson" or "spearman" and it specifies the type of correlations to compute.
exclude_insignificant_correlation_coefficients	logic	FALSE	If TRUE in the correlation plot only the significant correlations are shown
sig_level_value	value	0.01	Significance level for discriminating correlated variable: normally 0.001, 0.01, 0.05
enable_linearcombos_test	logic	FALSE	Use only for advanced users
export_shapefiles	logic	TRUE	Enable this to export shapefile of points corresponding to the data tables, Usefull to check location of point of training and validation datasets using GIS clients
export_txtfiles	logic	TRUE	Enable this to export the training and validation tables in tab separeted .txt format

Table S14, S15 and S16 list an example of the configuration of LAND-SE prepared to run Case A using data in raster format.

Table S14: List of LAND-SE configuration parameters, alphanumeric types, possible values and comments, for Case A (in red), but also valid for the other cases.

	Type	Case A	Comments
load_rdata	logic	TRUE	Enable reading input data from a .RData file produced by LAND-SIP
rdata_file	string	"datatable_inventory.RData"	Name of the .RData file containing input data.
enable_screen_plotting	value	FALSE	Enable screen plotting during LAND-SE execution.
enable_multicollinearity_test	logic	FALSE	Enable multicollinearity test execution.
enable_rocplot_confidenceinterval	logic	FALSE	TRUE requires intensive computational resources.
bootstrap_constant_correction	logic	TRUE	Enable a procedure to avoid convergence problems when explanatory variables contain mostly 0 value
enable_detailed_data_export	logic	TRUE	Enable the final shapefile export.
enable_detailed_data_export_tiff	logic	TRUE	Enable the uncertainty and matching code map export in .tiff format.

Table S15: List of *configuration_file.txt* parameters specified for Case A, but also valid for the other cases.

MODEL	RUN	BOOTSTRAP_SAMPLES_ROC_CURVE	ANALYSIS_PARAMETER	BOOTSTRAP_MODEL_VARIABILITY_RUN	BOOTSTRAP_SAMPLES_MODEL_VARIABILITY
LDA	YES	100		YES	100
QDA	YES	100	DUM	YES	100
LRM	YES	100		YES	100
NNM	YES	10	NOR	YES	10
CFM	YES	100		YES	100

Table S16: List of *spatial_configuration_file.txt* parameters specified for Case A, but also valid for the other cases.

TYPE	PRESENCE	ID_FIELD	EPSG_CODE	AREA_SU_FIELD	AREA_LANDSLIDE_FIELD	GEOMETRY	RASTER_RES	RASTER_EXPORT
SHAPEFILE	YES	id	25830	areaslu	arealds	POINTS	20	TRUE

9. Installation

The source code of LAND-SUITE is composed of 3 R scripts *LAND-SIP_vX_YYYYMMDD.R,
LAND-SVA_vX_YYYYMMDD.R, LAND-SE_vX_YYYYMMDD.R,* that are executed in an R
console environment. R is free software for statistical computing and graphics, which can be
compiled and run on a wide variety of UNIX, Windows and MacOS platforms. R can be
downloaded from a CRAN mirror site (https://cran.r-project.org/mirrors.html). The R base
package includes a basic Graphical User Interface (RGui) that allows loading an R console
where commands and scripts can be executed. Alternatively, other different Integrated
Development Environment (IDE) can be downloaded and used to run the script. For its
simplicity, we suggest using RStudio IDE (http://www.rstudio.com/), a free and open source
powerful and productive user interface available in many operating systems. To run the LAND-
SUITE scripts, the R packages listed in section 2 need to be installed. This can be accomplished
in the R console using the following command
`install.package("package_name")`or using the specific functions of the selected
IDE. In case the user wants to benefit from the GRASS GIS binding, this needs to be installed
into the system. Such installation depends on the type of OS used
(https://grass.osgeo.org/download/). As mentioned before, when running LAND-SUITE under
WindowsOS we suggest installing GRASS GIS through OSGeo4W64, which is a binary
distribution of a broad set of open source geospatial software for Windows environments
(https://trac.osgeo.org/osgeo4w/).

10. Running/Execution

To run LAND-SUITE in a standard way, the user should prepare a local folder containing the
scripts *LAND-SIP_vX_YYYYMMDD.R, LAND-SVA_vX_YYYYMMDD.R, LAND-
SE_vX_YYYYMMDD.R* and the files *configuration.txt* and *configuration_spatial_data.txt*. Such
files can be edited, using classical text editors, or using the editors in the R GUI or in the selected
IDE environment. The three scripts are executed in cascade starting from LAND-SIP.

To use GeoTiff files as input, the user should save them in the main folder as standard .tif
format. Alternatively, to use GRASS GIS inputs, the files should be loaded as layer input files
(.shp file) in the same folder, or provide an RData file containing them as R objects.

The R scripts can be executed using different R GUI or IDE environment and the specific commands:

- Classical RGui IDE: "Run" or "Run line or selection" in the "Edit" menu;
- RStudio IDE: "Source" or "Source with echo" in the "Code" menu.

Alternatively, the scripts can be executed via command line, with the following syntax and options.

Windows OS

To execute LAND-SIP:

```
PATH-TO\R.exe --no-save --args -wd PATH-TO-WD-DIR < LAND-
SIP_vX_YYYYMMDD.R > input_preparation.log
```

To execute LAND-SVA:specific mapset and location, which names need to be specified during LAND-SIP configuration. To run only LAND-SE, the user should copy the textual (.txt file) and the geographical

```
PATH-TO\R.exe --no-save --args -wd PATH-TO-WD-DIR < LAND-
SVA_vX_YYYYMMDD.R > variable_analysis.log
```

To execute LAND-SE:

```
PATH-TO\R.exe --no-save --args -cd PATH-TO-CD-DIR -wd PATH-TO-
WD-DIR < LAND-SE_vX_YYYYMMDD.R > susceptibility.log
```


Linux OS

To execute LAND-SIP:

```
R CMD BATCH --no-save --no-restore '--args -wd PATH-TO-WD-DIR'
LAND-SIP_vX_YYYYMMDD.R input_preparation.log
```

To execute LAND-SVA:

```
R CMD BATCH --no-save --no-restore '--args -wd PATH-TO-WD-DIR'
LAND-SVA_vX_YYYYMMDD.R variable_analysis.log
```

To execute LAND-SE:

```
R CMD BATCH --no-save --no-restore '--args -cd PATH-TO-CD-DIR
-wd PATH-TO-WD-DIR' LAND-SE_vX_YYYYMMDD.R susceptibilty.log
```

where

`PATH-TO\R.exe` is the full windows path of the R.exe file

`PATH-TO-CD-DIR` is the path (Windows and Linux) of the directory containing the configuration files (*configuration.txt*, the *configuration_spatial_data.txt*)

`PATH-TO-WD-DIR` is the path (Windows and Linux) of the directory containing the input files where also the output files will be stored.

11. Code availability and licence

LAND-SUITE is composed of three modules (LAND-SIP, LAND-SVA, LAND-SE), coded as separate .R script files that can be executed under different operating systems. The software was mainly tested under WindowsOS and LinuxOS, with the version of R-4.1.1 (64bit). Some code functionalities of LAND-SIP require GRASS GIS binding. We tested the script using GRASS GIS version 7 under WindowsOS and LinuxOS. We recommend LinuxOS, due to the better software integration at a bash scripting level.

LAND-SUITE is free software; it can be redistributed or modified under the terms of the GNU General Public (either version 2 of the license, or any later version) as published by the Free Software Foundation. The program is distributed in the hope that it will be useful, but without any warranty; without even the implied warranty of merchantability or fitness for a particular purpose. See the GNU General Public License for more details.

LAND-SUITE V1.0 is archived in ZENODO repository with the DOI: 10.5281/zenodo.5650810.

12. References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (Eds.): The New S Language, Wadsworth & Brooks, Cole, 1988.

Carrara, A.: Multivariate models for landslide hazard evaluation, Journal of the International Association for Mathematical Geology, 15(3), 403-426, 1983.

Carrara, A., Crosta, G., and Frattini, P.: Comparing models of debris-flow susceptibility in the alpine environment, Geomorphology, 94(3-4), 353-378, https://doi.org/10.1016/j.geomorph.2006.10.033 2008.

GRASS Development Team: Geographic Resources Analysis Support System (GRASS) Software, Version 7.2. Open Source Geospatial Foundation. Electronic document:. http://grass.osgeo.org, 2017

Han, J., Pei, J., and Kamber, M. (Eds.): Data mining: concepts and techniques, Elsevier, 2011.

Lee, S., & Min, K.: Statistical analysis of landslide susceptibility at Yongin, Korea, Environmental geology, 40(9), 1095-1113, 2001, https://doi.org/10.1007/s002540100310

R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/, 2021.

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

Rossi, M., and Reichenbach, P.: LAND-SE: a software for statistically based landslide susceptibility zonation, version 1.0, Geoscientific Model Development, 9(10), 3533-3543, 2016.

Trigila, A., Iadanza, C., Esposito, C., and Scarascia-Mugnozza, G.: Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy), Geomorphology, 249, 119-136, https://doi.org/10.1016/j.geomorph.2015.06.001, 2015.

Van Westen, C. J.: Statistical landslide hazard analysis. ILWIS 2.1 for Windows application guide, ITC publication, 2, 73-84, 1997.

Yilmaz, I.: Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat-Turkey), Computers & Geosciences, 35(6), 1125-1138, https://doi.org/10.1016/j.cageo.2008.08.007, 2009.