



LAND-SE

(LANDslide Susceptibility Evaluation)

Version 1.0

User Guide

Mauro Rossi

(Mauro.Rossi@irpi.cnr.it)

Consiglio Nazionale delle Ricerche

Istituto di Ricerca per la Protezione Idrogeologica

Index

<u>S1. User skills</u>	<u>2</u>
<u>S2. Configuration parameters</u>	<u>2</u>
<u>S3. Input data specifications</u>	<u>5</u>
<u>S4. List of output</u>	<u>7</u>
<u>S5. Installation and running</u>	<u>10</u>
<u>S6. Example dataset</u>	<u>12</u>
<u>S7. Code licence</u>	<u>12</u>
<u>Reference of R software and packages</u>	<u>13</u>

S1. User skills

LAND-SE (LANDslide Susceptibility Evaluation) is a software developed to prepare landslide susceptibility models and zonation at basin and regional scale, with specific functions focused on results evaluation and uncertainty estimation. The software is implemented in R, a free software environment for statistical computing and graphics (R Core Team, 2015). In the field of landslide susceptibility zonation, LAND-SE is designed to be properly and productively 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.

A proper LAND-SE execution requires: (i) a basic knowledge of R language to run the script; (ii) experience on multivariate statistical models and on their evaluation skills/metrics (ROC plot, contingency table and plots, success/prediction rate curves, etc.); (iii) GIS skills to prepare and handle input data; and (iv) specific expertise for a correct and reliable interpretations of the results.

All the modelling types implemented in LAND-SE are basically statistical classification techniques applicable to any multivariate analysis with a binary grouping (dependent or response) variable. This makes the code flexible and appropriate to other scientific fields and usable, with minor customization and tailoring, by user with different expertise.

S1.S2. Configuration parameters

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. To run LAND-SE, the following files are needed:

- 1 *LAND-SE_vX_YYYYMMDD.R*, R script file of the source code (X is the software version and YYYYMMDD the release date);
- 2 *LAND-SE_configuration.txt*, file of the parameters conditioning the susceptibility analysis;
- 3 *LAND-SE_configuration_spatial_data.txt*, file of the parameters and configuration for the spatial output.

The *LAND-SE_configuration.txt* file (Table S1), defines specifications for the susceptibility analysis. In the current version, four single multivariate classification techniques (LDA: linear discriminant analysis, QDA: quadratic discriminant analysis, LR: logistic regression, NN: neural network) and one combined model (CM) can be executed. The user can specify, for each model, the following parameters.

1. RUN (YES/NO): specify to run or not the model.
2. BOOTSTRAP_SAMPLES_ROC_CURVE: indicates the number of bootstrap samples used by the `roc.plot{verification}` R command to evaluate the ROC variability.
3. ANALYSIS_PARAMETER: controls characteristics for some models. In the current version, the user can specify the following:
 - For the QDA model: SEL to eliminate dummy variables from the analysis; DUM to maintain dummy variables, converting them in numerical format but introducing a random noise;
 - For the NN model: NOR to use neural network default weights and a number of nodes in the hidden layer corresponding to half the number of explanatory variables; OPT to perform an auto optimization of the neural network structure (slower and with tendency to over fit data);
4. BOOTSTRAP_MODEL_VARIABILITY_RUN (YES/NO): specify to evaluate or not the model uncertainty, required for the generation of the uncertainty plots and maps;
5. BOOTSTRAP_SAMPLES_MODEL_VARIABILITY: specify the number of bootstrap samples selected in the model uncertainty evaluation.

Table S1. Parameters in the *LAND-SE_configuration.txt* file. (LDA: linear discriminant analysis, QDA: quadratic discriminant analysis, LR: logistic regression, NN: neural network and CM: combined model)

MODEL	RUN	BOOTSTRAP SAMPLES NUMBER ROC CURVE	ANALYSIS PARAMETER	BOOTSTRAP MODEL VARIABILITY RUN	BOOTSTRAP SAMPLES MODEL VARIABILITY
LDA	YES/NO	# (e.g. 10)	na	YES/NO	# (e.g. 100)
QDA	YES/NO	# (e.g. 10)	SEL/DUM	YES/NO	# (e.g. 100)
LR	YES/NO	# (e.g. 10)	na	YES/NO	# (e.g. 100)
NN	YES/NO	# (e.g. 10)	OPT/NOR	YES/NO	# (e.g. 100)
CM	YES/NO	# (e.g. 10)	na	YES/NO	# (e.g. 100)

The *LAND-SE_configuration_spatial_data.txt*, specifies the parameters to run LAND-SE in *geomode* (Table S2):

1. TYPE: in the current version is possible to select only SHAPEFILE;
2. PRESENCE (YES/NO): YES enables the use of spatial data and *training.shp* and *validation.shp* files are needed;
3. ID_FIELD: Name of the column (field name) in the shapefile attribute table containing the univocal code of the mapping unit identification;
4. EPSG_CODE: Coordinate Reference System EPSG Code of the geographical data as defined by IOGP (International Association of Oil & Gas Producers), listed in <http://www.spatialreference.org/>;
5. AREA_SU_FIELD: Name of the column (field name) in the shapefile attribute table containing the mapping unit area;
6. AREA_LANDSLIDE_FIELD: Name of the column (field name) in the shapefile attribute table containing the landslide area in the mapping unit;
7. GEOMETRY: POINTS for pixel-based analysis, POLYGONS for polygon-based analysis;
8. RASTER_RES: Dimension of the grid cell (only for pixel-based analysis);
9. RASTER_EXPORT (TRUE/FALSE): TRUE enable the export of output map also in GeoTIFF format (only for pixel-based analysis).

Table S2. List of the parameters in the *LAND-SE_configuration_spatial_data.txt* file.

PARAMETER	VALUES DESCRIPTION
TYPE	SHAPEFILE In the current version only this value is possible
PRESENCE	YES/NO YES enable the use of spatial data (<i>training.shp</i> and <i>validation.shp</i> file are required)
ID_FIELD	ID Name of the column (field name) in the shapefile attribute table containing the univocal code of the mapping unit identification

PARAMETER	VALUES DESCRIPTION
EPSG_CODE	4326 Coordinate Reference System EPSG Code
AREA_SU_FIELD	area_mapping Name of the column (field name) in the shapefile attribute table containing the mapping unit area
AREA_LANDSLIDE_FIELD	area_landslide Name of the column (field name) in the shapefile attribute table containing the landslide area in the mapping unit
GEOMETRY	POINTS/POLYGONS POINTS for the pixel-based analysis POLYGONS for the polygon-based analysis
RASTER_RES	30 Dimension of the raster cell (only pixel-based analysis)
RASTER_EXPORT	TRUE/FALSE TRUE enable to export the output map also in GeoTIFF format (only pixel-based analysis)

S2.S3. Input data specifications

LAND-SE can be executed using two different modes: *standard* that provides output only in textual and graphical format and *geomode* that handles also geographical data.

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 1st column contains the univocal code of the mapping unit identification; ii) the 2nd 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 3rd to the n -th columns, values of the n explanatory variables (i.e. independent variables).

Temporal, spatial or random subdivision criteria can be chosen to identify the training and the validation dataset. 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 (this is usually achieved using two different landslide inventory maps). In this case usually the analysis is performed using the same explanatory variables associated to the same set of mapping units. In the spatial and random subdivision, the training and the validation dataset are composed by different mapping units,

characterized by different grouping and explanatory variables. The main difference between the spatial and the random validation relies upon the method chosen to separate the training and the validation dataset: in the first case, the dataset are spatially different (the two areas can be contiguous or not), in the second the subdivision is performed using a random sampling.

LAND-SE is designed to use different mapping units, reducible to point-like units (pixels) or to polygon-like subdivisions (e.g. geomorphological, administrative, etc.). Since raster data cannot be used directly as input, a preliminary conversion is required to perform the pixel based analysis. The conversion from a raster to a tabular format can be completed using standard GIS analysis or a dedicated tool (available on request). Two possible functions are: i) the “*gdal2xyz*” function

(<http://svn.osgeo.org/gdal/trunk/gdal/swig/python/scripts/gdal2xyz.py>) available in different GIS clients (e.g. QGIS, <http://www.qgis.org/en/site/>); and ii) the “*raster2xyz*” function in the ArcGIS platform (<http://www.esri.com/software/arcgis>).

It's important to highlight that when using the point-like units all the thematic data used by LAND-SE must have the same geographical extent and the same resolution. Numbers of columns and rows have to be the same for all the grid data.

LAND-SE is highly demanding in terms of RAM, mainly for the pixel-based approach. The demand of RAM depends on: i) the size of the study area and the pixel resolution; ii) the number of explanatory variables; and iii) the number and type of model selected. When LAND-SE is applied to very large areas, calculations may require very long computational time. A significant improvement in the script execution could be obtained using fast CPUs. A more efficient and advanced solution, that might be improved in the future, consider a code parallelization.

To execute LAND-SE in *geomode*, two additional files are required in shapefile format (.shp) with the geographical information of the training and the validation dataset. The attribute table (Table S2 and S3) of the shapefiles contains, for each mapping unit, the following fields: i) the univocal code of the mapping unit identification (**ID_FIELD**), ii) the area of the mapping unit (**AREA_SU_FIELD**), and iii) the area of landslides in each mapping unit (**AREA_LANDSLIDE_FIELD**).

Table S3. List of the input data files required to run LAND-SE.

SOFTWARE FILES	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 variables. 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 variables. Each row of the file is one mapping unit of the validation dataset.
training.shp Required only for the <i>geomode</i> , 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 <i>geomode</i> , to prepare the spatial output and the calculation of the prediction 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.

S3.S4. List of output

In table S4 is shown the complete list of the LAND-SE outputs with a short description. Output are grouped in graphical (.pdf format), textual (.txt format), and geographical (.shp or .tif format). Outputs for a specific model are generated only when the model has been enabled in the “*LAND-SE_configuration.txt*” file (Table S1).

The number and width of landslide susceptibility classes used to prepare maps and histograms can be modified by the user. This can be done changing the default values of the following variables in the script:

- `breaks.histogram.values<-c(0,0.2,0.45,0.55,0.8,1)`
- `breaks.map.susceptibility<-c(0,0.2,0.45,0.55,0.8,1.0001)`

Table S4. 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 output provided by the *geomode*.

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	×
result_XX_Model_Susceptibility_Map.pdf	Map of the XX model susceptibility values obtained in the training phase	×
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	

SOFTWARE OUTPUT	DESCRIPTION	GEO MODE
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 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)	×

S4.S5. Installation and running

The source code of LAND-SE, *LAND-SE_vX_YYYYMMDD.R*, is 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 to use RStudio IDE (<http://www.rstudio.com/>), a free and open source powerful and productive user interface available in many operating systems. For a more programming oriented IDE, we suggest StatET (<http://www.walware.de/goto/statet>), a plug-in for the Eclipse Platform IDE (<http://www.eclipse.org/>). Like R and Eclipse, StatET is open source software, accessible for many operating systems. To run the LAND-SE script, the R packages “vcd”, “verification”, “perturb”, “nnet”, “raster”, “rasterVis”, “RColorBrewer” need to be installed. This can be accomplished in the R console using with the following command:

```
install.package("vcd", "verification", "perturb", "nnet", "raster",
, "rasterVis", "RColorBrewer")
```

Table S5. Parameters to be specified in the *LAND-SE_vX_YYYYMMDD.R* file.

PARAMETER	DESCRIPTION	VALUES
wd_selected	Character Name of the working directory or folder where output will be stored	“X:/R/test” (under Windows) “/home/R/test” (under Linux) “/Users/R/test” (under Mac OS)
cd_selected	Character Name of the working directory or folder containing configurations and input data files	“X:/R/test” (under Windows) “/home/R/test” (under Linux) “/Users/R/test” (under Mac OS)

To run LAND-SE, the user should prepare a local folder containing the *LAND-SE_vX_YYYYMMDD.R*, the *LAND-SE_configuration.txt* and the *LAND-SE_configuration_spatial_data.txt* files. In the same folder, the user should copy the textual (.txt file) and the geographical (.shp file) input files. The user should specify in *LAND-*

SE_vX_YYYYMMDD.R two additional parameters that define the path of the working directory containing the input data and output files (Table S5).

The *LAND-SE_configuration.txt*, the *LAND-SE_configuration_spatial_data.txt* and *LAND-SE_vX_YYYYMMDD.R*, can be edited, using the text editor in the R GUI or IDE environment, or any other text editor. The R script can be executed opening it on different R GUI or IDE environment and using 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 script can be executed via command line, using the following syntax and options:

Windows OS

```
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

```
R CMD BATCH --no-save --no-restore --args -cd PATH-TO-CD-DIR -wd PATH-TO-WD-DIR LAND-SE_vX_YYYYMMDD.R susceptibility.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 (*LAND-SE_configuration.txt*, the *LAND-SE_configuration_spatial_data.txt*)

PATH-TO-WD-DIR is the path (Windows and Linux) of the directory containing the input files (*training.txt*, *validation.txt*, *training.shp*, *validation.shp*) where output files will be stored.

S5,S6. Example dataset

To test LAND-SE script and to have a complete overview of the output types, we provide three different example dataset organized in the following directories:

- pixel_Random_validation_entirearea (pixel-based landslide susceptibility zonation with random selection of the training data and validation on the entire area);
- polygon_NS_validation (polygon-based landslide susceptibility zonation with selection of the training and validation data on contiguous areas);
- polygon_Random_validation_entirearea (polygon-based landslide susceptibility zonation with random selection of the training data and the validation on the entire area).

Each directory contains the complete set of files required to run LAND-SE:

- LAND-SE_vX_YYYYMMDD.R;
- LAND-SE_configuration.txt;
- LAND-SE_configuration_spatial_data.txt;
- training.shp;
- validation.shp;
- training.txt;
- validation.txt.

Spatial location of the provided datasets and relative results obtained running LAND-SE should be considered only as example.

S6,S7. Code licence

LAND-SE Copyright © Mauro Rossi. LAND-SE 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.

Reference of R software and packages

- Hendrickx, J.: perturb: Tools for evaluating collinearity. R package version 2.05. <http://CRAN.R-project.org/package=perturb>, 2012.
- Hijmans, R.J., and van Etten, J.: raster: Geographic analysis and modeling with raster data. R package version 2.0-12. <http://CRAN.R-project.org/package=raster>, 2012.
- Meyer, D., Zeileis, A., Hornik, K.: vcd: Visualizing Categorical Data. R package version 1.4-1, 2015.
- NCAR - Research Applications Laboratory (2014). verification: Weather Forecast Verification Utilities. R package version 1.41. <http://CRAN.R-project.org/package=verification>.
- Neuwirth, E.: RColorBrewer: ColorBrewer palettes. R package version 1.0-2, 2007.
- Perpinan Lamigueiro, O., and Hijmans, R.: rasterVis. R package version 0.31. doi:10.5281/zenodo.12394, 2014.
- R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>, 2015.
- Venables, W.N., and Ripley, B.D.: Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0, 2002.