# EXPLORING EFFECTS OF TRAINING SET SELECTION ON THE SHALLOW LANDSLIDE SUSCEPTIBILITY MODELING AT REGIONAL SCALE USING MAXENT MACHINE LEARNING TECHNIQUE

**Massimo Conforti, Luigi Borrelli and Giovanni Gullà**

*CNR IRPI, via Cavour 4/6, 87036 Rende (CS)*

e-mail: *massimo.coforti@irpi.cnr.it*, *luigi.borrelli@irpi.cnr.it*, *giovanni.gulla@irpi.cnr.it*

# OBJECTIVES

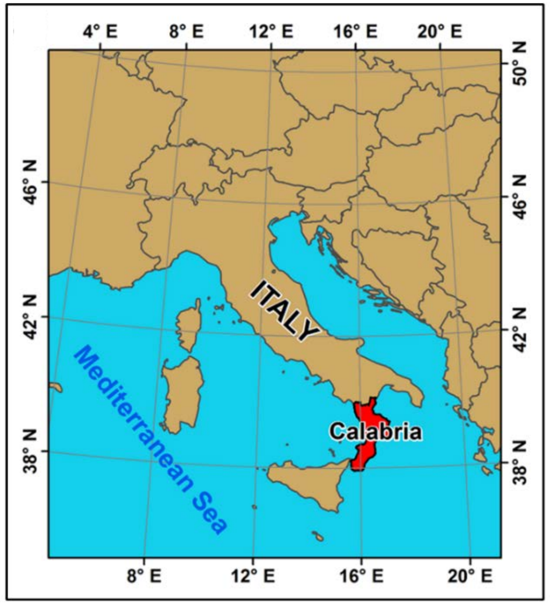
The aims of this study are:

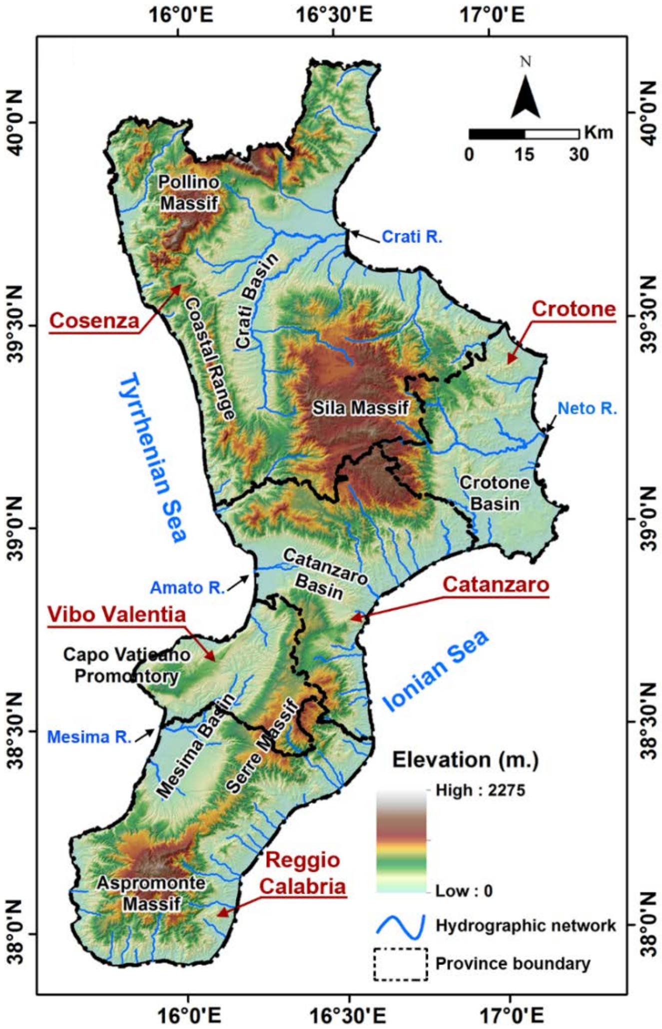
- to evaluate the spatial prediction of shallow landslides at regional scale using Maximum Entropy (MAXENT) method;
- to evaluate training landslides random selection effects on the shallow landslide susceptibility modelling;
- to understand which are the main predisposing factors that control the spatial prediction of shallow landslides.

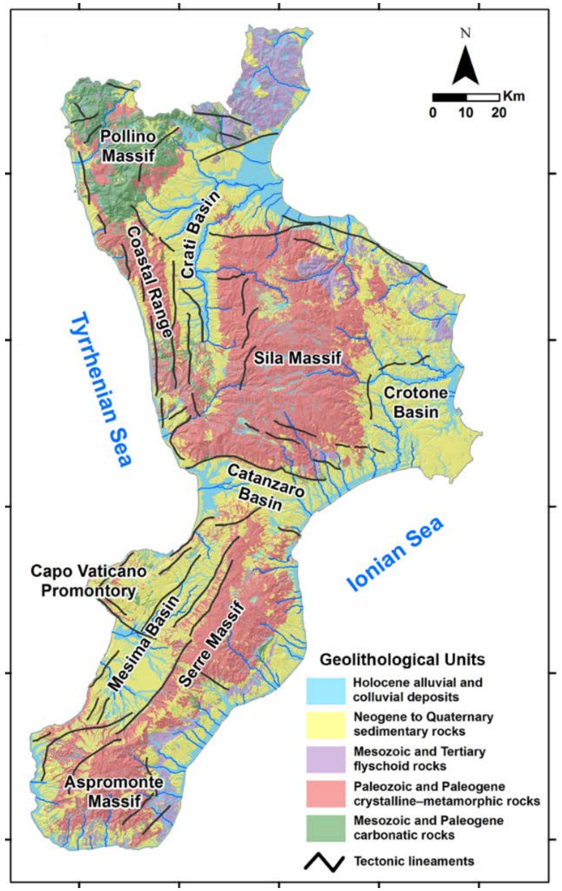
# STUDY AREA



## Calabria Region

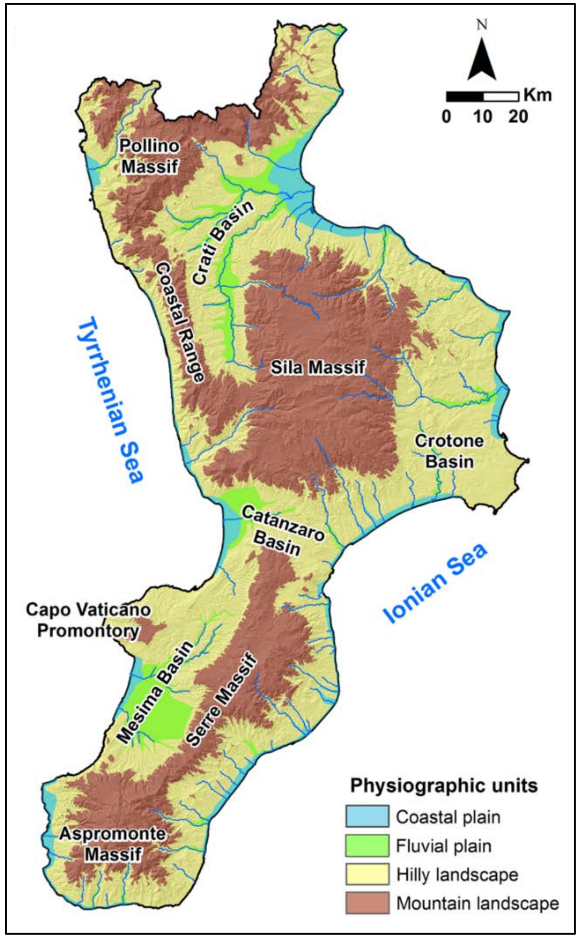


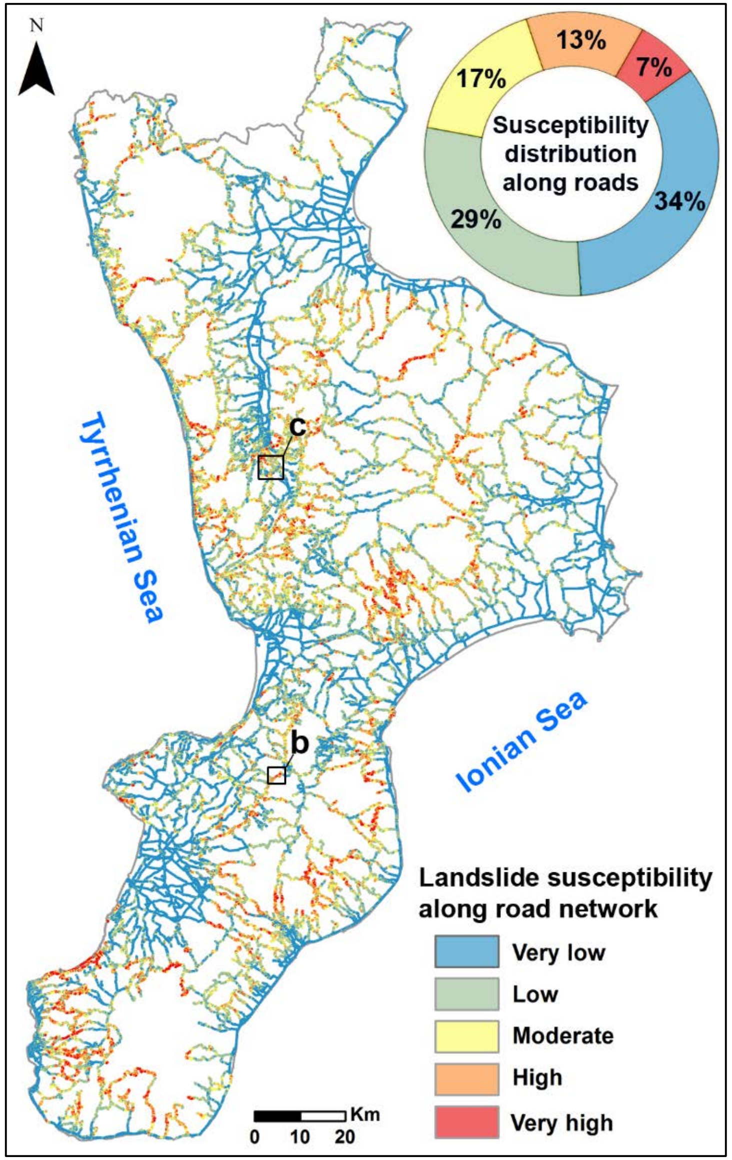




Schematic geological map

Physiographic units map



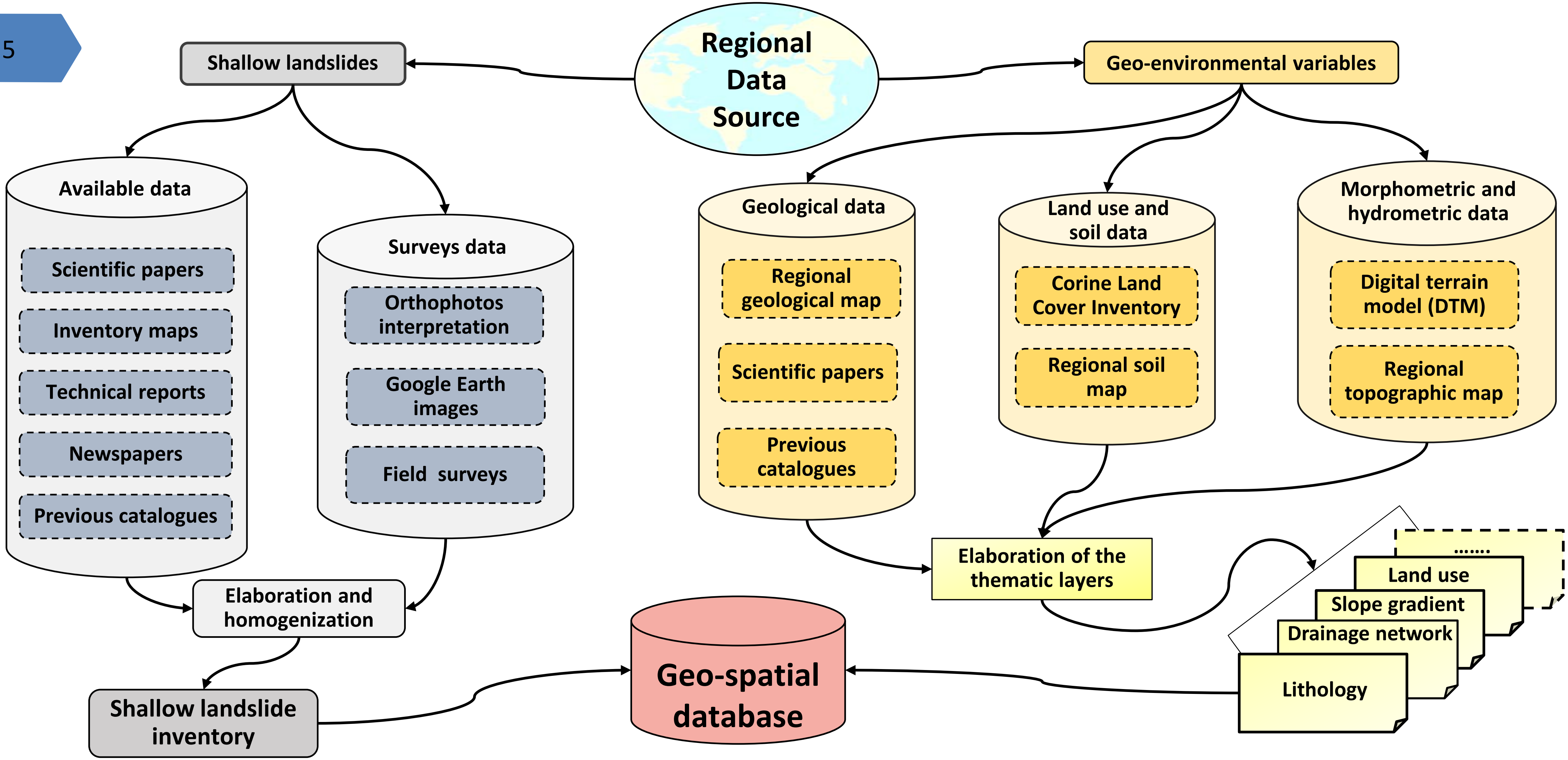
# STUDY AREA




Da Gulla et al., 2021

## Examples of landslides occurring in the Region

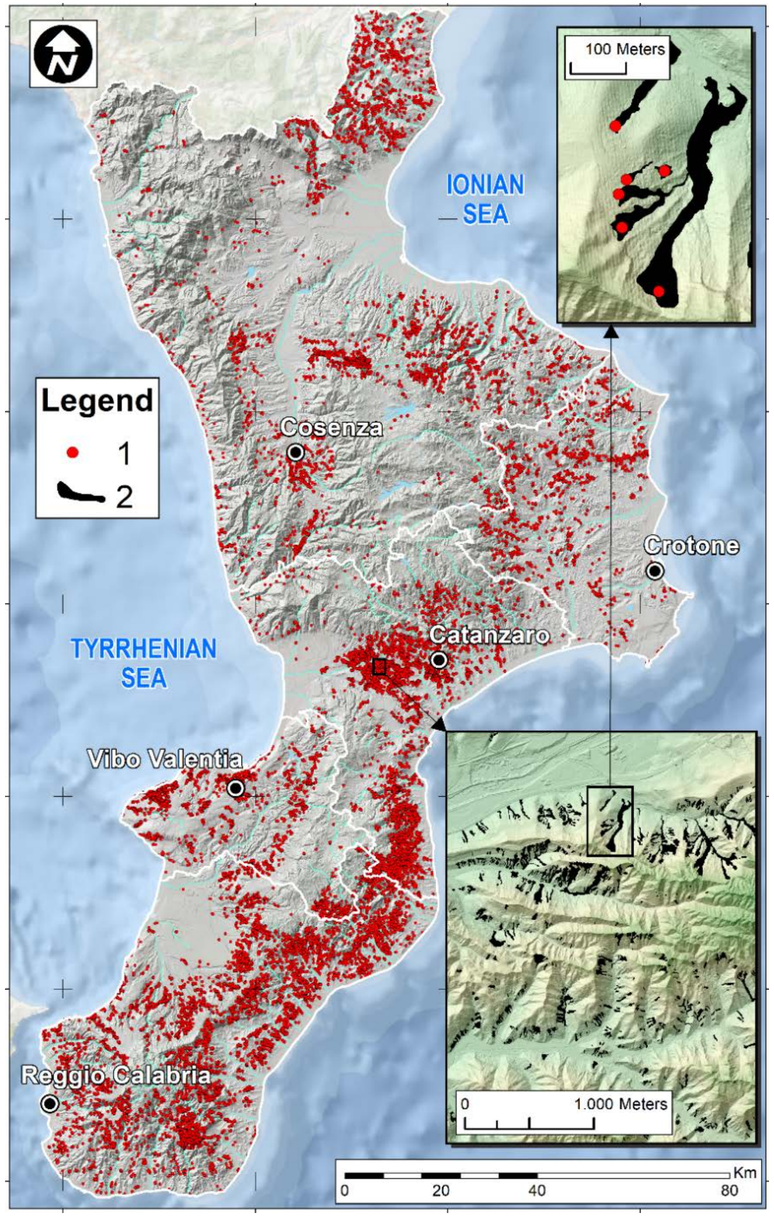

# WORKFLOW OF DATABASE PREPARATION



# Shallow landslide inventory

Source data used to create the shallow landslides database for the Calabria region. (Da Gullà et al., 2021)

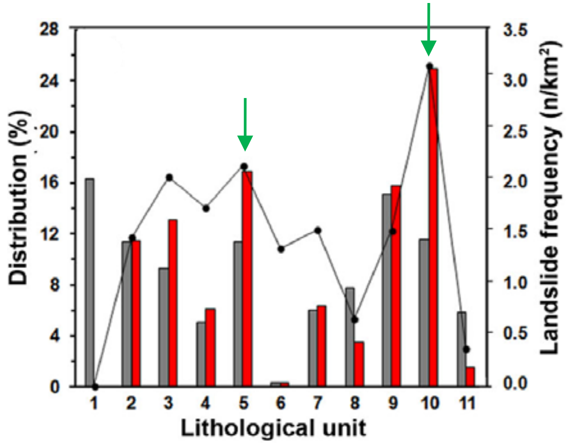
| Source data | References | Period/Rainfall event | Number of landslides | Contribution to inventory (%) |
|---|---|---|---|---|
| | (Gullà et al. 2008) | Rainfall events 1951 and 1953 | 7977 | 36.2 |
| Scientific papers | (Conforti and Ietto 2020) | From 1990 to 2018 | 1358 | 6.2 |
| | (Borrelli et al. 2018) | From 2001 to 2011 | 750 | 3.4 |
| | (Conforti et al. 2016) | From 1998 to 2014 | 266 | 1.2 |
| | (Rago et al. 2013) | From 2000 to 2012 | 210 | 1.0 |
| | (Conforti and Critelli 2012) | From 1991 to 2004 | 127 | 0.6 |
| | (Conforti, Robustelli et al. 2012) | From 1991 to 2004 | 72 | 0.3 |
| | (Conforti and Ietto 2019) | From 1998 to 2011 | 59 | 0.3 |
| | (Borrelli, Gioffrè et al. 2012) | Rainfall event, winter 2000 | 33 | 0.1 |
| | (Conforti, Filomena et al. 2012) | From 1991 to 2011 | 29 | 0.1 |
| Inventory map | (Borrelli et al. 2015) | From 2008 to 2010 | 3399 | 15.4 |
| | (Sorriso-Valvo et al. 2004) | Rainfall event, september 2000 | 2084 | 9.5 |
| | (Borrelli, Critelli et al. 2012) | From 2000 to 2006 | 616 | 2.8 |
| | (Lucà et al. 2011) | From 1990 to 2004 | 423 | 1.9 |
| | (Rago et al. 2017) | Rainfall event 30 October–01 November 2015 | 133 | 0.6 |
| | (Iovine and Merenda 1996) | Rainfall event 1972–73 | 117 | 0.5 |
| | (Tansi et al. 2016) | From 2008 to 2012 | 85 | 0.4 |
| | (Biondino et al. 2018) | From 2014 to 2017 | 92 | 0.4 |
| Catalogues | Calabria Basin Authorities | From 2000 to 2016 | 934 | 4.2 |
| PhD thesis | (Conforti 2009) | From 2000 to 2008 | 95 | 0.4 |
| Degree thesis | (Vigliarolo 2009) | From 2000 to 2006 | 418 | 1.9 |
| Thecnical reports | Cosenza province | Rainfall event winter 2009–2010 | 55 | 0.2 |
| | Calabria region | Rainfall event winter 2009–2010 | 153 | 0.7 |
| | Calabria region | Rainfall event winter 2008–2009 | 371 | 1.7 |
| | CNR-IRPI (Cosenza) | Rainfall events, september 2000 and winter 2008–2009 | 418 | 1.9 |
| Newspapers | Regional newspapers | Rainfall events, winter 2010–2011 and 2013 | 64 | 0.3 |
| Photo interpretation and field survey | Orthophotos dated 2008, Google Earth satellite images dated, 2010, 2011, 2014, 2015 and 2016. | From 2008 to 2017 | 1690 | 7.7 |
| Total shallow landslide cataloged | | From 1951 to 2017 | 22,028 | 100 |




Areal distribution and frequency of shallow landslides in the provinces.

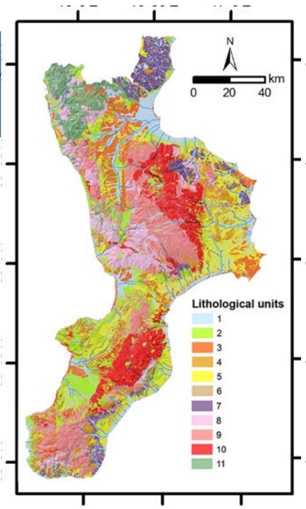
| Province | Shallow landslides | | Landslide frequency |
|---|---|---|---|
| | count | % | (count/km$^2$) |
| Catanzaro | 6410 | 29.1 | 2.7 |
| Cosenza | 4083 | 18.5 | 0.6 |
| Crotone | 1367 | 6.2 | 0.8 |
| Reggio Calabria | 7136 | 32.4 | 2.2 |
| Vibo Valentia | 3032 | 13.8 | 2.7 |
| Whole Region | **22028** | 100 | **1.5** |

## Lithology *vs* shallow landslides

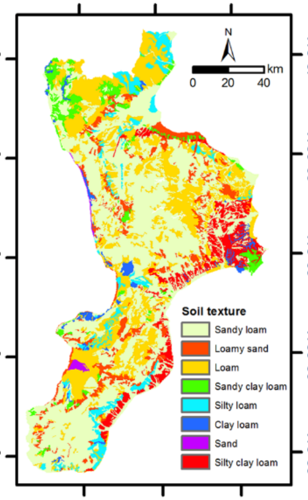



1) alluvial deposits, 2) gravel and sand deposits, 3) sandstone rocks, 4) conglomerate rocks, 5) **clay and marl rocks**, 6) evaporitic rocks, 7) flyschoid rocks, 8) low-grade metamorphic rocks, 9) **middle-high-grade metamorphic rocks**, 10) **intrusive rocks**, 11) carbonate rocks

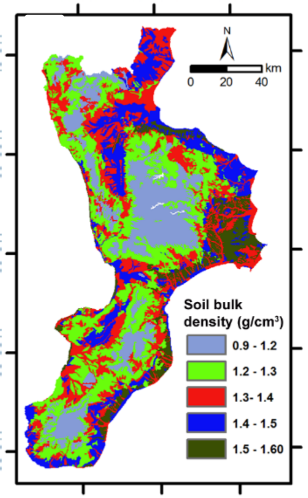
# Geo-environmental variables maps



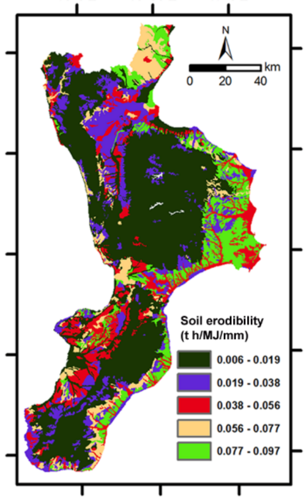

Lithology



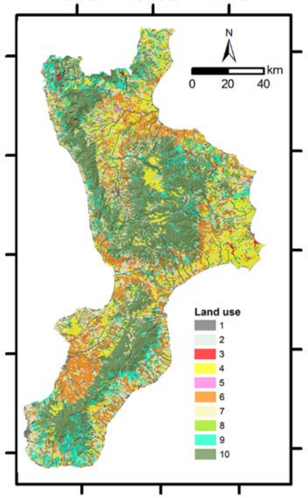

Soil texture



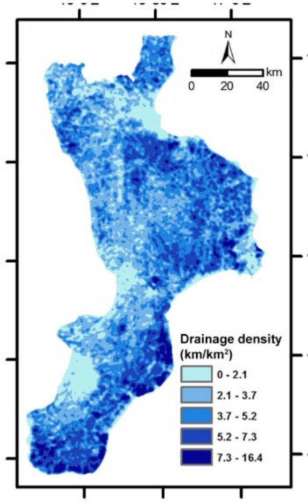

Soil bulk density



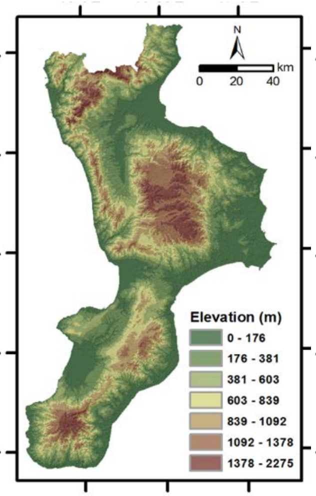

Soil erodibility



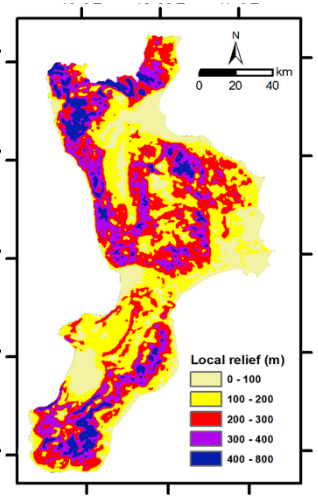

Land use



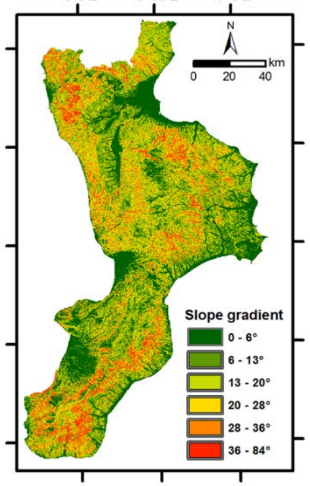

Drainage density



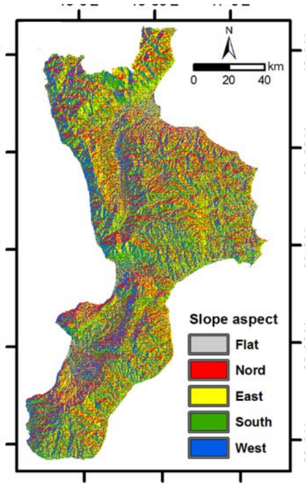

Elevation



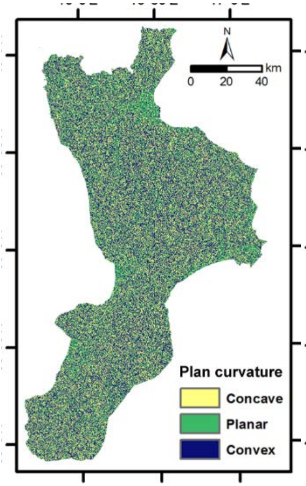

Local relief



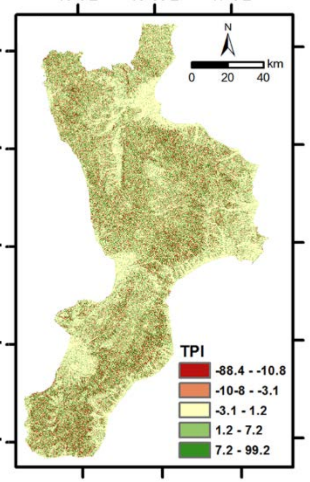

Slope gradient



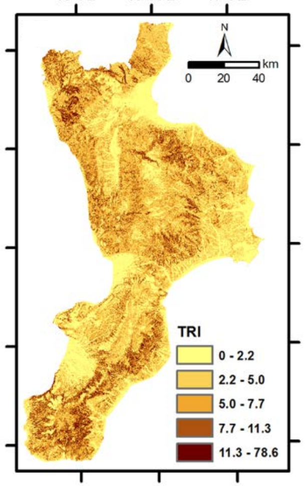

Slope aspect



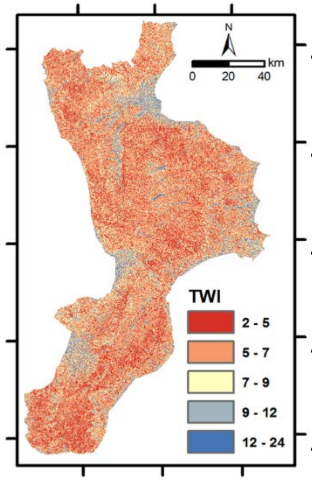

Plan curvature



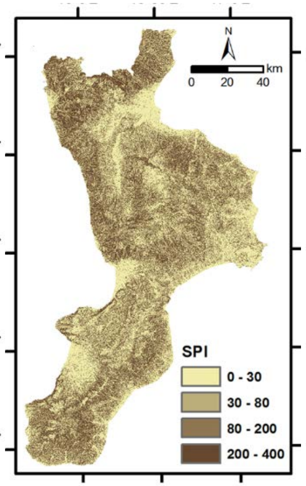

TPI



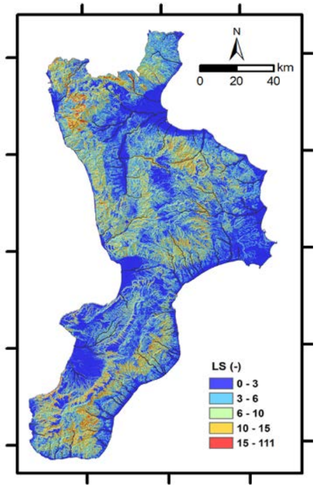

TRI




TWI

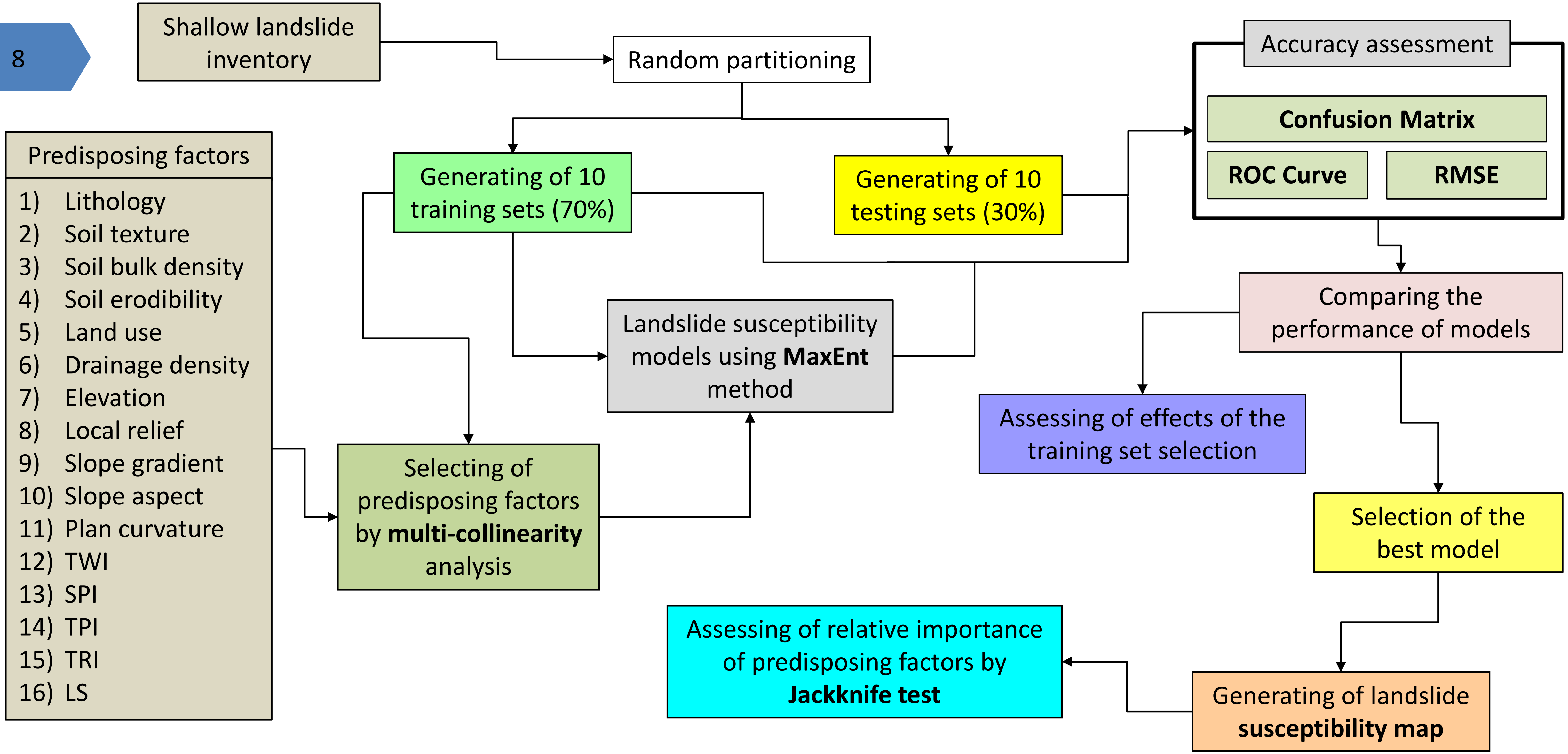



SPI




LS

# Flow chart of the landslide susceptibility analysis



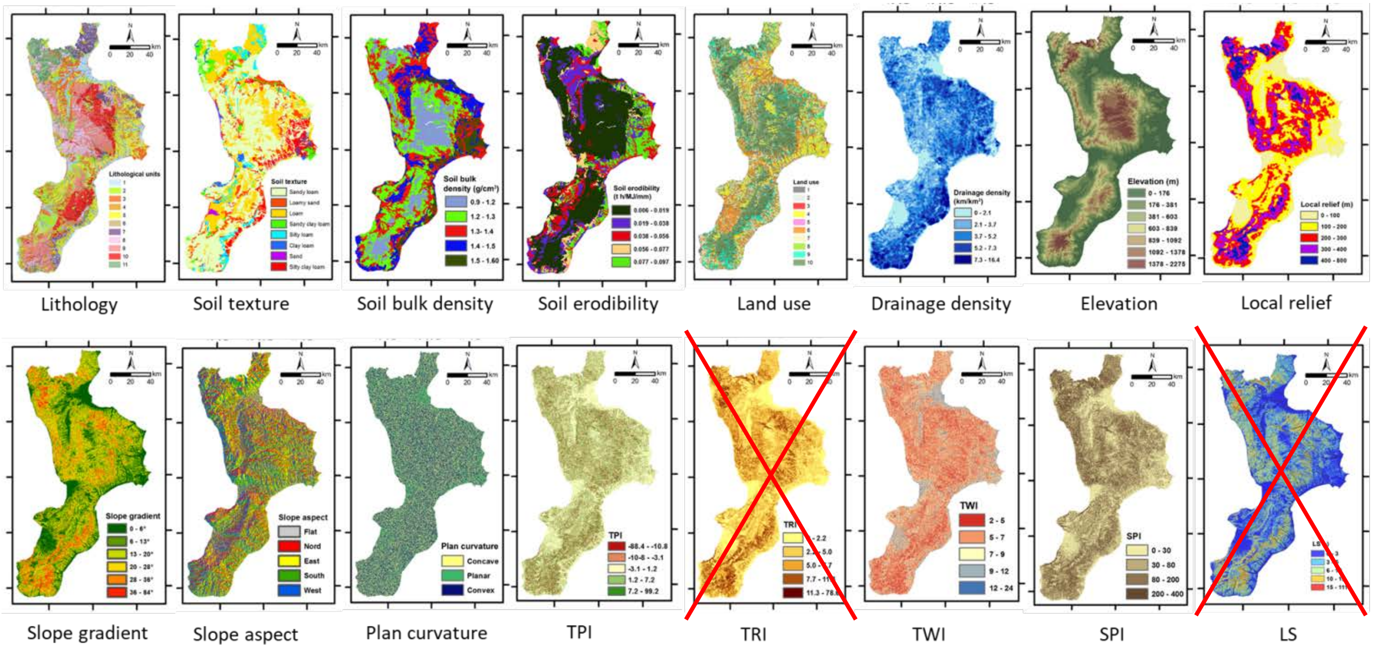
# RESULTS

## Multi-collinearity analysis

| Predisposing factor | Before | | After | |
|---|---|---|---|---|
| | *Tolerance* | *VIF* | *Tolerance* | *VIF* |
| Lithology | 0.827 | 1.210 | 0.830 | 1.205 |
| Soil texture | 0.797 | 1.255 | 0.802 | 1.247 |
| Soil bulk density | 0.380 | 2.635 | 0.391 | 2.557 |
| Soil erodibility | 0.460 | 2.174 | 0.471 | 2.124 |
| Land use | 0.854 | 1.171 | 0.855 | 1.169 |
| Drainage density | 0.749 | 1.335 | 0.749 | 1.334 |
| Elevation | 0.340 | 2.938 | 0.343 | 2.914 |
| Local relief | 0.386 | 2.588 | 0.395 | 2.530 |
| Slope gradient | **0.074** | **13.484** | 0.431 | 2.321 |
| Slope aspect | 0.992 | 1.008 | 0.993 | 1.008 |
| Plan curvature | 0.998 | 1.002 | 0.998 | 1.002 |
| TPI | 0.710 | 1.408 | 0.735 | 1.360 |
| TRI | **0.063** | **15.847** | - | - |
| TWI | 0.462 | 2.163 | 0.484 | 2.067 |
| SPI | 0.944 | 1.059 | 0.966 | 1.035 |
| LS | **0.151** | **6.610** | - | - |

*VIF*= variance inflation factor

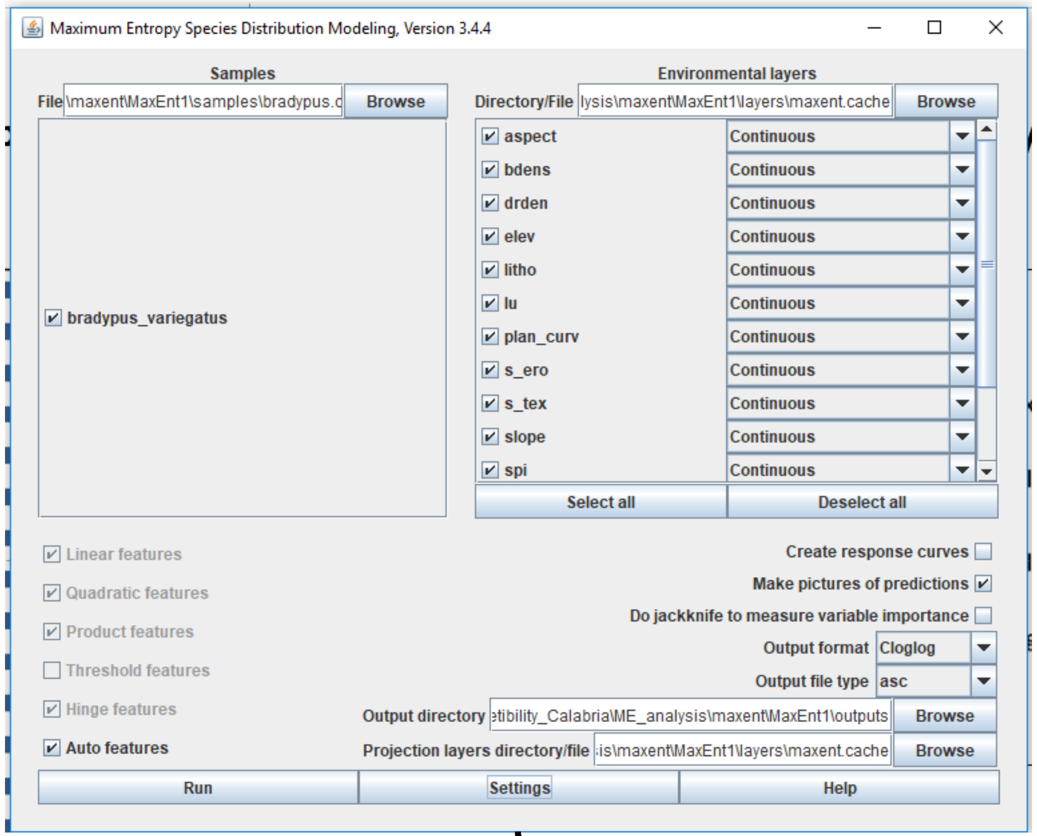
With *Tolerance* > 0.2 and *VIF* < 5, the variables are independent of each other.



# RESULTS

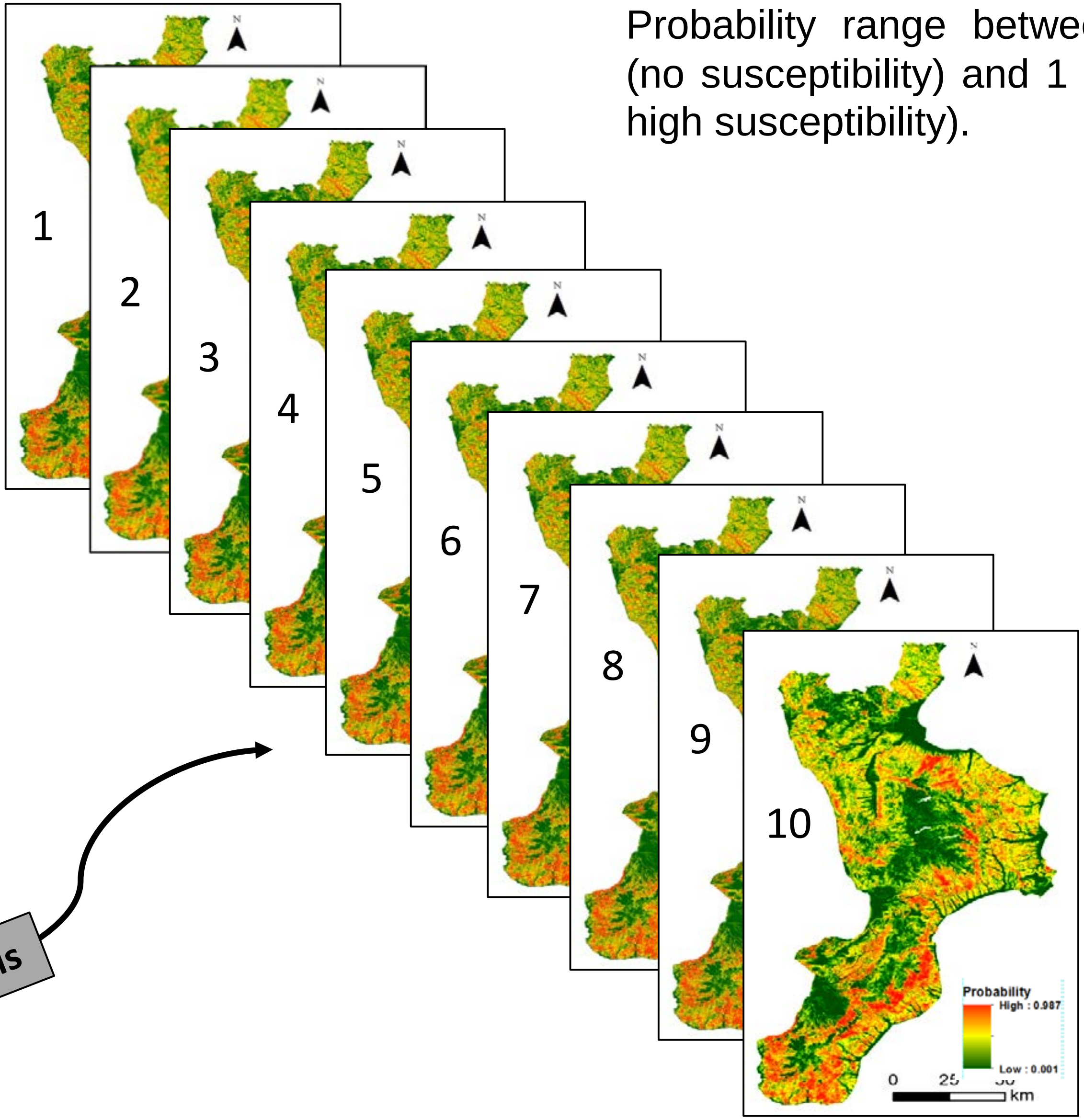
10

Building of the 10 shallow landslide susceptibility models using Maximum Entropy method




MaxEnt 3.4.4 software

Models




Probability range between 0 (no susceptibility) and 1 (very high susceptibility).


Phillips et al. (2006) - Maximum entropy modeling of species geographic distributions. Ecological Modeling 190, 231–259.

# RESULTS

## Accuracy of models obtained by 10 replicates of training dataset

Best model

### Training dataset

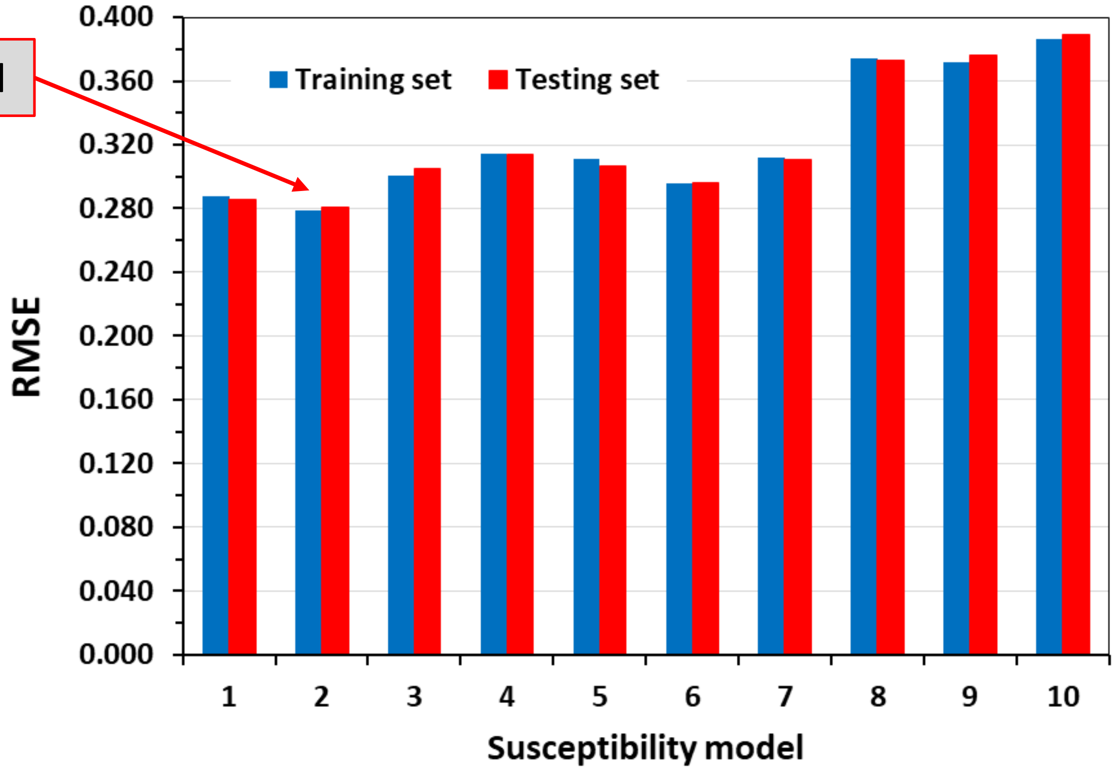
| Model | Sensitivity | Specificity | Accurancy | Kappa index |
|---|---|---|---|---|
| | Accuracy parameters | | | |
| 1 | 0.797 | 0.824 | 0.810 | 0.620 |
| 2 | **0.813** | **0.853** | **0.832** | **0.664** |
| 3 | 0.794 | 0.858 | 0.823 | 0.645 |
| 4 | 0.785 | 0.832 | 0.807 | 0.613 |
| 5 | 0.802 | 0.848 | 0.823 | 0.646 |
| 6 | 0.782 | 0.836 | 0.807 | 0.614 |
| 7 | 0.800 | 0.849 | 0.823 | 0.645 |
| 8 | 0.800 | 0.832 | 0.815 | 0.631 |
| 9 | 0.784 | 0.820 | 0.801 | 0.601 |
| 10 | 0.800 | 0.844 | 0.820 | 0.641 |
| Min | 0.782 | 0.820 | 0.801 | 0.601 |
| Max | 0.813 | 0.858 | 0.832 | 0.664 |
| Mean | 0.796 | 0.840 | 0.816 | 0.632 |
| S.dev | 0.010 | 0.013 | 0.010 | 0.020 |
| C.V. | 1.213 | 1.513 | 1.186 | 3.089 |

### Testing dataset

| Model | Sensitivity | Specificity | Accurancy | Kappa index |
|---|---|---|---|---|
| | Accuracy parameters | | | |
| 1 | 0.795 | 0.820 | 0.807 | 0.614 |
| 2 | **0.811** | **0.838** | **0.824** | **0.647** |
| 3 | 0.800 | 0.833 | 0.816 | 0.632 |
| 4 | 0.776 | 0.832 | 0.802 | 0.603 |
| 5 | 0.807 | 0.832 | 0.819 | 0.637 |
| 6 | 0.798 | 0.820 | 0.809 | 0.617 |
| 7 | 0.794 | 0.826 | 0.809 | 0.618 |
| 8 | 0.789 | 0.806 | 0.797 | 0.594 |
| 9 | 0.805 | 0.792 | 0.798 | 0.596 |
| 10 | 0.808 | 0.826 | 0.817 | 0.634 |
| Min | 0.776 | 0.792 | 0.797 | 0.594 |
| Max | 0.811 | 0.838 | 0.824 | 0.647 |
| Mean | 0.798 | 0.823 | 0.810 | 0.619 |
| S.dev | 0.010 | 0.014 | 0.009 | 0.018 |
| C.V. | 1.315 | 1.705 | 1.127 | 2.920 |

# RESULTS

## Accuracy of models for each selection of training set and testing set

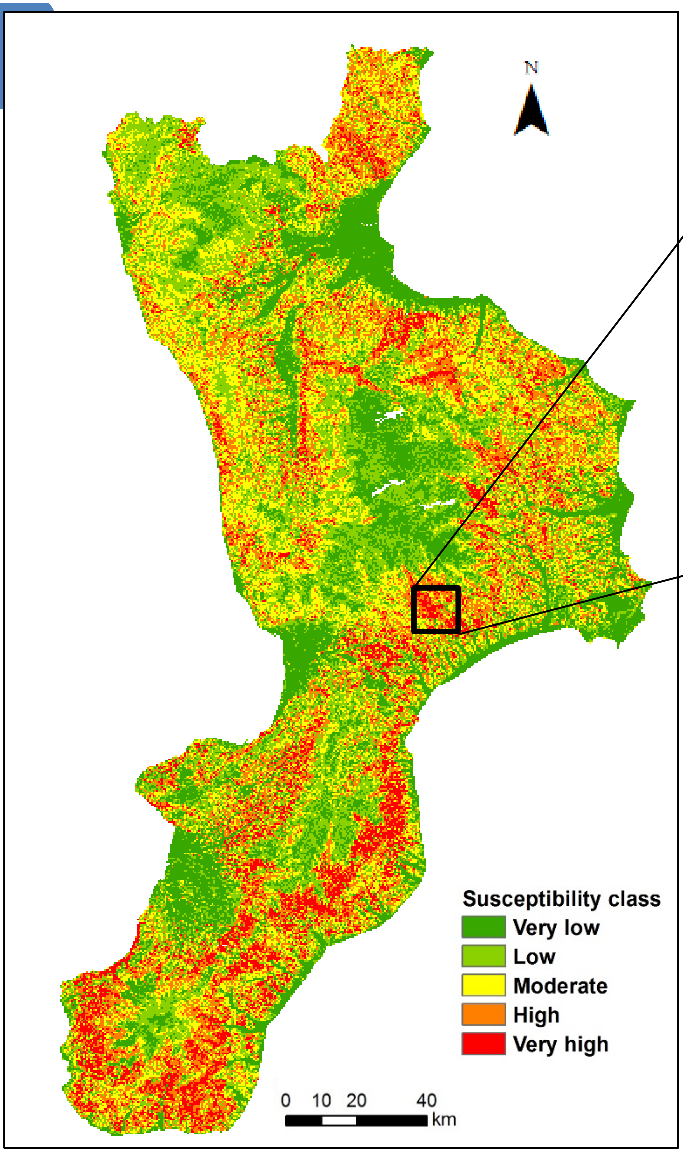



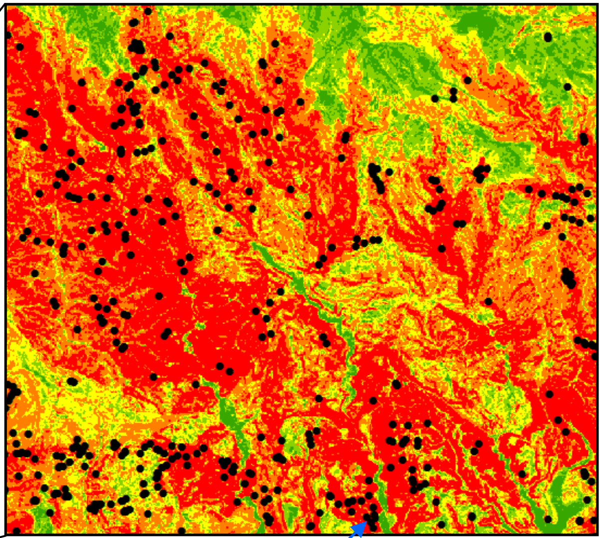

| | Min | Max | Mean | S.dev | C.V. |
|---|---|---|---|---|---|
| Training sets | 0.848 | 0.882 | 0.860 | 0.010 | 1.110 |
| Testing sets | 0.835 | 0.866 | 0.853 | 0.009 | 1.010 |

| | Min | Max | Mean | S.dev | C.V. |
|---|---|---|---|---|---|
| Training sets | 0.279 | 0.386 | 0.323 | 0.039 | 12.064 |
| Testing sets | 0.281 | 0.389 | 0.324 | 0.040 | 12.321 |

# RESULTS

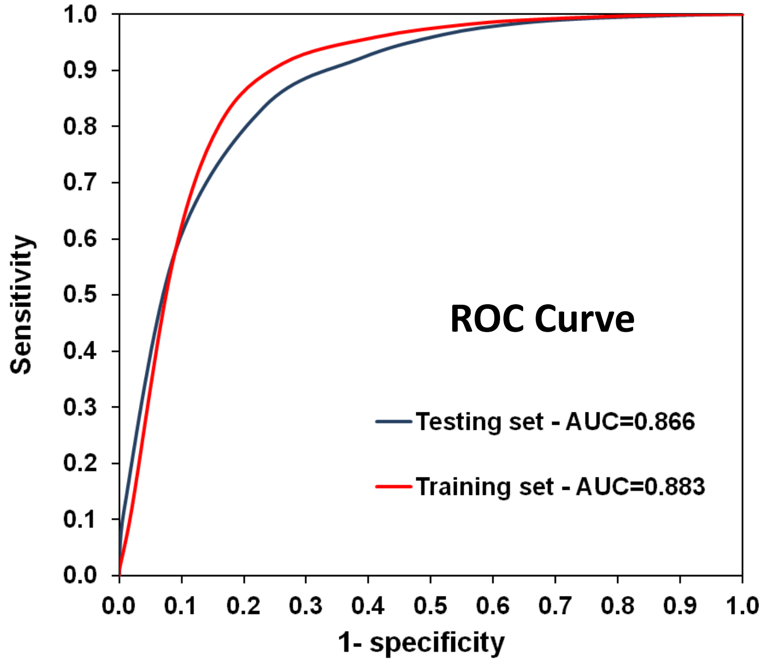
## Shallow landslide susceptibility map



Shallow landslide

## Validation of susceptibility map



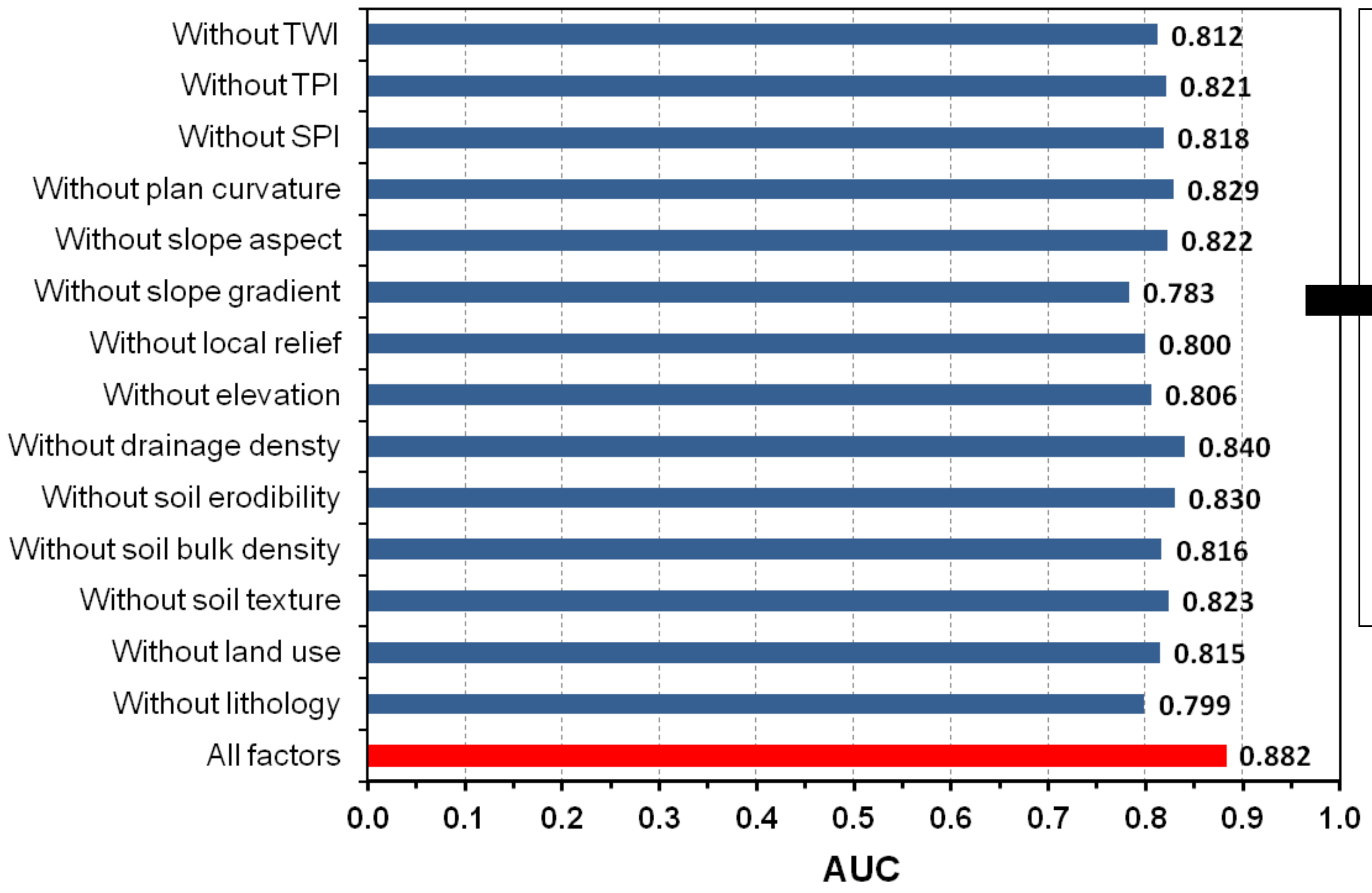

**Percentage area of the susceptibility classes in the map, percentage of shallow landslides distribution and related landslide frequency falling in the susceptibility classes.**

| Probability range | Susceptibility class | Area (%) | Shallow landslides (%) | | | Landslide frequency |
|---|---|---|---|---|---|---|
| | | | Training set | Testing set | All dataset | |
| 0.00 – 0.18 | Very low | 22.8 | 0.7 | 0.7 | 0.7 | 0.03 |
| 0.19 - 0.41 | Low | 25.5 | 3.9 | 3.5 | 3.8 | 0.15 |
| 0.42 - 0.61 | Moderate | 23.6 | 12.1 | 9.8 | 11.4 | 0.48 |
| 0.62 - 0.81 | High | 18.6 | 28.9 | 29.1 | 28.9 | 1.55 |
| 0.81 - 0.98 | Very high | 9.6 | 54.4 | 56.9 | 55.2 | 5.78 |
| | | | 83.3% | 86.0% | 84.1% | |

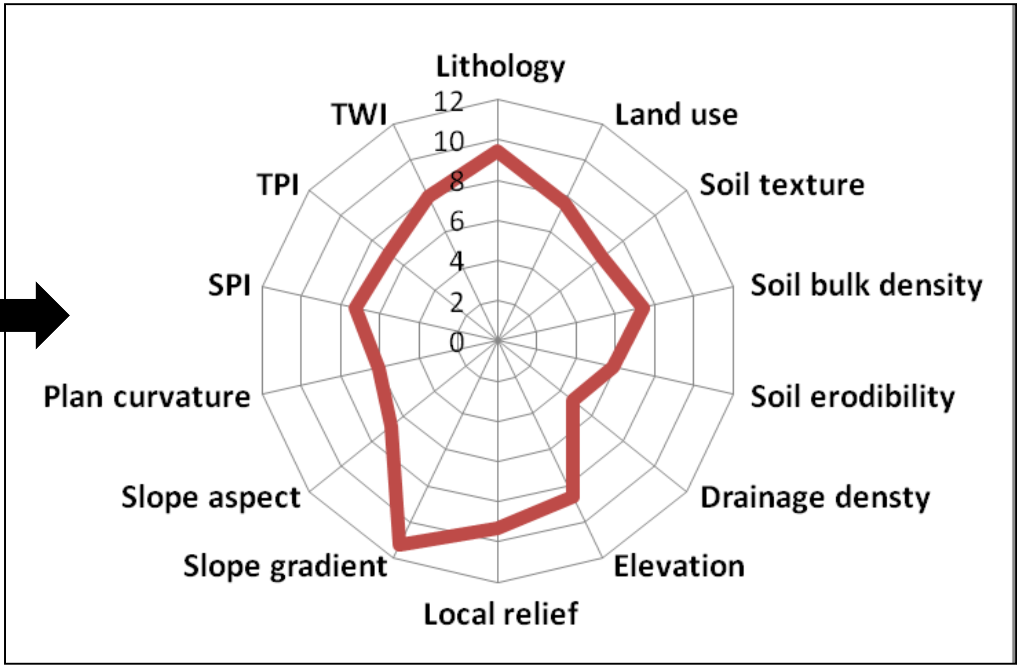
# RESULTS

## Sensitivity of predisposing factors used in the shallow landslide susceptibility model

Jack-knife test




Relative importance (RI)




$$RI = \frac{(AUC_{all} - AUC_i)}{AUC_{all}} \times 100$$

# CONCLUSIONS



- The data collected in this study contribute to build a regional shallow landslide database that can be used both for land use planning and hazard-risk assessment at regional scale;
- The results indicate that MaxEnt method is able for mapping shallow landslide prone-areas within Calabria region;
- The MaxEnt method appeared to be less sensible to the change of random training sets;
- The results highlighted also that the selection and the sensitivity evaluation of the predisposing factors is a key requirement for modeling of landslide susceptibility;
- Others methods should be tested in order to find the most appropriate model to produce landslide susceptibility maps, as well as evaluate the effects of the random selection of the training sets.