



# Automatic delineation of geomorphological slope units with `r.slopeunits v1.0` and their optimization for landslide susceptibility modeling

Massimiliano Alvioli, Ivan Marchesini, Paola Reichenbach, Mauro Rossi, Francesca Ardizzone, Federica Fiorucci, and Fausto Guzzetti

CNR IRPI, via Madonna Alta 126, 06128 Perugia, Italy

Correspondence to: Massimiliano Alvioli (massimiliano.alvioli@irpi.cnr.it)

Received: 13 May 2016 – Published in Geosci. Model Dev. Discuss.: 21 June 2016

Revised: 17 October 2016 – Accepted: 19 October 2016 – Published: 9 November 2016

**Abstract.** Automatic subdivision of landscapes into terrain units remains a challenge. Slope units are terrain units bounded by drainage and divide lines, but their use in hydrological and geomorphological studies is limited because of the lack of reliable software for their automatic delineation. We present the `r.slopeunits` software for the automatic delineation of slope units, given a digital elevation model and a few input parameters. We further propose an approach for the selection of optimal parameters controlling the terrain subdivision for landslide susceptibility modeling. We tested the software and the optimization approach in central Italy, where terrain, landslide, and geo-environmental information was available. The software was capable of capturing the variability of the landscape and partitioning the study area into slope units suited for landslide susceptibility modeling and zonation. We expect `r.slopeunits` to be used in different physiographical settings for the production of reliable and reproducible landslide susceptibility zonation.

2012; Saito et al., 2011; Sharma and Mehta, 2012; Fall et al., 2006; Li et al., 2012; Erener and Düzgün, 2012; Schaetzl et al., 2013; Mashimbye et al., 2014). A slope unit (SU) is a type of morphological TU bounded by drainage and divide lines (Carrara, 1988; Carrara et al., 1991, 1995; Guzzetti et al., 1999), and corresponds to what a geomorphologist or an hydrologist would recognize as a single slope, a combination of adjacent slopes, or a small catchment. This makes SUs easily recognizable in the field, and in topographic base maps. Compared to other terrain subdivisions, including grid cells or unique-condition units (Guzzetti et al., 1999; Guzzetti, 2006), SUs are related to the hydrological and geomorphological conditions and processes that shape natural landscapes. For this reason, SUs are well suited for hydrological and geomorphological studies, and for landslide susceptibility (LS) modeling and zonation (Carrara et al., 1991, 1995; Guzzetti et al., 1999; Guzzetti, 2006).

SUs can be drawn manually from topographic maps of adequate scale and quality (Carrara, 1988). However, the manual delineation of SUs is time-consuming and error-prone, limiting the applicability to very small areas. Manual delineation of SUs is also intrinsically subjective. This reduces the reproducibility – and hence the usefulness – of the terrain subdivision. Alternatively, SUs can be delineated automatically using specialized software. The latter exploits digital representations of the terrain, typically in the form of a digital elevation model (DEM) (Carrara, 1988), or adopts image segmentation approaches (Flanders et al., 2003; Dragut and Blaschke, 2006; Aplin and Smith, 2008; Zhao et al., 2012). In both cases, the result is a geomorphological subdivision of the terrain into mapping units bounded by drainage and di-

## 1 Introduction

The automatic subdivision of large and complex geographical areas, or even entire landscapes, into reproducible, geomorphologically coherent terrain units remains a conceptual problem and an operational challenge. Terrain units (TUs) are subdivisions of the terrain that maximize the within-unit (internal) homogeneity and the between-unit (external) heterogeneity across distinct physical or geographical boundaries (Guzzetti et al., 1999; Guzzetti, 2006; Komac, 2006,

vide lines, which can be represented by polygons (in vector format) or groups of grid cells (in raster format).

Large and complex geographical areas or landscapes can be partitioned by different SU subdivisions. Unique (i.e., universal) subdivisions do not exist, and optimal (best) terrain subdivisions depend on multiple factors, including the size and complexity of the study area, the quality and resolution of the available terrain elevation data, and – most importantly – the purpose of the terrain subdivision (e.g., geomorphological or hydrological modeling, landslide detection from remote-sensing images, landslide susceptibility, hazard or risk modeling). An open problem is that an optimal SU subdivision for LS modeling cannot be decided unequivocally, a priori, or in an objective way, and the quality and usefulness of a LS zonation depends on the SU subdivision (Carrara et al., 1995).

In this work, we propose an innovative modeling framework to determine an optimal terrain subdivision based on SUs best suited for LS modeling. For the purpose, we also present the `r.slopeunits` software for the automatic delineation of SUs, and we propose a method to optimize the terrain subdivision in SUs performed by the software. The `r.slopeunits` software is written in Python for the GRASS GIS (Neteler and Mitasova, 2007), and automates the delineation of SUs, given a DEM and a set of user-defined input parameters. We tested the `r.slopeunits` software and the proposed optimization procedure for LS modeling in a large area in central Italy, where sufficient landslide and thematic information was available to us (Cardinali et al., 2001, 2002).

The paper is organized as follows. First, we present the proposed approach for the delineation of an optimal terrain subdivision into SUs best suited for LS modeling, based on an optimization method (Sect. 2). Next (Sect. 3), we present the method for the automatic delineation of SUs, which we have implemented in the `r.slopeunits` software for the GRASS GIS, and (in Sect. 4) we describe a segmentation metric useful for the evaluation of the SU internal homogeneity. Next, we introduce landslide susceptibility modeling (Sect. 5) and our optimization approach to the SU partitioning (Sect. 6). This is followed (in Sect. 7) by a description of the study area, in central Italy, and of the data used for LS modeling. In Sect. 8, we present the results obtained in our study area in central Italy; then we summarize the obtained results (Sect. 9), and outline possible uses of the `r.slopeunits` software and of the optimization procedure (Sect. 10).

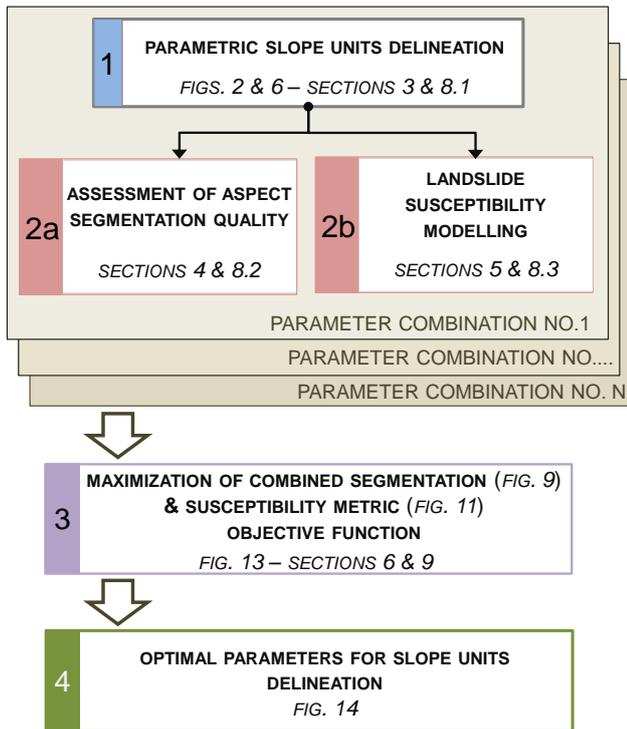
## 2 Modeling framework

We propose a new modeling framework for the parametric delineation of SUs and their optimization, as a function of a few input parameters, for the specific purpose of determining landslide susceptibility (LS) adopting statistically based

classification methods (Guzzetti et al., 1999, 2005, 2006; Guzzetti, 2006; Rossi et al., 2010). The framework is exemplified in Fig. 1, and consists of the following steps:

1. First, the `r.slopeunits` SU delineation software – described in Sect. 3, and whose flowchart is shown in Fig. 2 – is run multiple times with different combinations of the input modeling parameters. Each software run results in a different subdivision of the landscape into a different set of SUs. In each run, the number and size of the SUs depend on the input modeling parameters.
- 2a. Second, the internal homogeneity and external inhomogeneity of each SU subdivision – required by any meaningful terrain subdivision – are defined in terms of terrain aspect, measured by the circular variance of the unit vectors perpendicular to the local topography represented by all grid cells in a slope unit. For each set of SUs obtained in step 1 using different input parameters, the quality of the aspect segmentation is evaluated adopting a general-purpose segmentation objective function, presented in Sect. 4.
- 2b. At the same time, for each set of SUs obtained in step 1, a LS model is calibrated adopting a logistic regression model (LRM) for SU classification (Sect. 5). In the LRM, each SU is classified as stable (i.e., free of landslides) or unstable (i.e., having landslides) depending on a (in our case, linear) combination of the local terrain conditions (i.e., the geo-environmental variables). The performance of the model calibration is evaluated using the area under the curve (AUC) of receiver operating characteristic (ROC), referred to as the  $AUC_{ROC}$  metric, a standard and objective metric commonly adopted in the literature to evaluate the performance of LS models (Rossi et al., 2010).
3. Lastly, an overall (combined) objective function is defined by properly combining the segmentation (step 2a) and the  $AUC_{ROC}$  (step 2b) objective functions, as described in Sect. 6. Maximization of this quantity allows to single out the optimal set of SUs (i.e., the optimal terrain subdivision) that, simultaneously, (i) provides a good aspect segmentation and (ii) results in an effective calibration of the LS model.
4. Maximization of the combined objective function allows selecting objectively the optimal combination of the input terrain modeling parameters best suited for LS modeling (step 4 in Fig. 1) and the corresponding SU subdivision.

In summary, the proposed modelling framework relies on an optimization procedure that maximizes a proper, specific function that contains information on (i) the morphology of the study area, represented by the aspect segmentation metric



**Figure 1.** Logical framework for the proposed method for (1) the parametric delineation of SUs, (2a) the assessment of the quality of terrain aspect segmentation using of a proper segmentation objective function, (2b) the calculation and assessment of the quality of the LS modeling using a standard  $AUC_{ROC}$  metric, and (3) the definition and maximization of a combined objective function for (4) the determination of optimal parameters for the delineations of SUs best suited for LS modeling and zonation.

(step 2a), and on (ii) the specific landslide processes under investigation (in our case, slow- to very slow-moving shallow slides, deep-seated slides, and earth flows), represented by the LS model performance and the associated  $AUC_{ROC}$  metric (step 2b). The optimization approach removes subjectivity from the SU delineation algorithm, and produces a result that is objective and completely reproducible. This is a significant advantage over manual methods (Carrara, 1988), or specialized software that needs multiple parameters and specific calibration procedures.

### 3 Automatic delineation of slope units

Automatic delineation of SUs can be performed adopting two strategies. The first strategy defines a large number of small homogeneous areas, and enlarges or aggregates them progressively, maximizing the aspect homogeneity of the SUs (Zhao et al., 2012). Following this approach, the size of the initial polygons representing the small homogeneous areas is significantly smaller than the size of the desired (final) SUs, which results from the aggregation of multiple areas per-

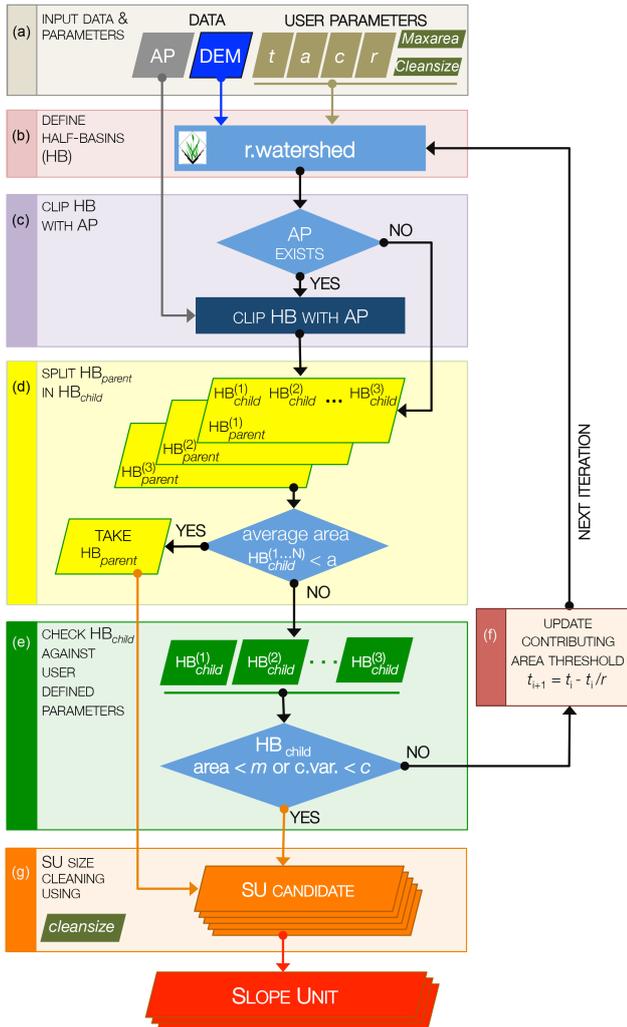
formed maximizing an objective function (Espindola et al., 2006). The second strategy defines an initial small number of large or very large areas, and progressively reduces their size until a satisfactory result is obtained (Carrara, 1988; Carrara et al., 1991, 1995). In the second strategy, the study area is subdivided into large subcatchments, which can be further subdivided into left and right sides (looking downstream with respect to the main drainage), with the resulting two sides named half basins (HBs). The size of the initial HB is much larger than the desired size for the SUs.

For both strategies, the final subdivision of the landscape into SUs does not maintain memory of the terrain partitioning represented by the initial areas or HBs. In both strategies, deciding when to stop the aggregation or the partitioning to obtain a terrain subdivision suitable for a specific use (in our case, LS modeling) is critical. Both strategies are subject to the selection of user-defined modeling parameters, which introduce subjectivity and reduce the reproducibility of the results. These are conceptual and operational problems that hamper the design and the implementation of an automatic procedure for the effective delineation of terrain subdivisions based on SUs (Carrara et al., 1995; Espindola et al., 2006; Dragut et al., 2010, 2014).

#### 3.1 Slope unit delineation algorithm

For the delineation of the SUs, we adopt the second strategy outlined above, i.e., we start from a relatively small number of large HBs, and we gradually reduce their size by subdividing the HBs into smaller TUs. Hydrological conditions and terrain aspect requirements control the subdivision of the large HBs into smaller TUs. The approach is adaptive, and it results in a geomorphological subdivision of the terrain based on SUs of different shapes and sizes that capture the real (natural) subdivisions of the landscape.

We implemented the approach to the delineation of SUs in a specific algorithm, coded in the `r.slopeunits` software (Fig. 2). The algorithm (and the software) uses the hydrological module `r.watershed` (Metz et al., 2011) available in the GRASS GIS. Using a DEM to represent terrain morphology, `r.watershed` produces a map of HBs adopting an advanced flow accumulation (FA) area analysis. Each grid cell in the DEM is attributed the total contributing area FA, based on the number of cells that drain into it. The FA values are low along the divides and increase downstream along the drainage lines. This information is used to single out streams and divides, i.e., the main elements bounding a SU (Carrara, 1988). The `r.watershed` module can use single flow direction (SFD) or multiple flow direction (MFD) strategies. In the SFD strategy, water is routed to the single neighboring cell with the lowest elevation, and in the MFD strategy water is distributed to all the cells lower in elevation, proportionally to the terrain slope in each direction. The `r.slopeunits` software adopts the MFD strategy to distribute water to the neighboring cells.



**Figure 2.** Flowchart for the `r.slopeunits` software. (a) Input data and parameters. AP, (input parameter name: `plainsmap`) map showing plain areas to be excluded from the processing; DEM, (`demmap`) digital elevation model;  $t$ , (`thresh`) initial FA threshold area;  $a$ , (`areamin`) minimum area;  $c$ , (`cvmin`) circular variance;  $r$ , (`rf`) reduction factor; `maxarea`, maximum SU area; `cleansize`, size of candidate SU to be removed; see text for detailed explanation. (b) `r.watershed` processing in GRASS GIS (Neteler and Mitasova, 2007). (c) Tests if the produced HBs are in a plain area. (d) The average area of each new  $HB_{child}$  in each  $HB_{parent}$  is checked against  $a$ . (e)  $HB_{child}$  is individually checked against  $a$  and  $c$  requirements. (f) The process proceeds to iteration  $i + 1$  for each and every  $HB_{child}$  that still does not meet the requirements, with an updated  $t_{i+1} = t_i - t_i/r$  FA threshold. (g) Small polygons are removed from the candidate SU set.

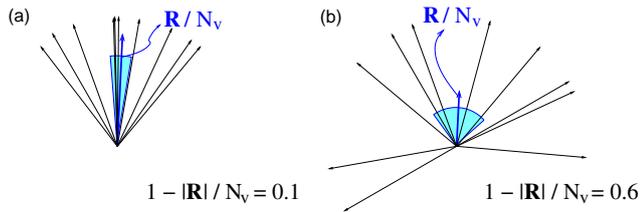
The `r.slopeunits` software requires a DEM, `demmap`, accepts an optional layer showing alluvial plains (APs), `plainsmap`, and the following user-defined numerical parameters (Fig. 2a): (i) the flow accumulation area (FA) threshold, `thresh`; (ii) the minimum surface area for the SU, `areamin` (in square meters); (iii) the minimum

circular variance (Nichols, 2009) of terrain aspect within a slope unit, `cvmin`; (iv) a reduction factor, `rf`; (v) the maximum surface area for the SU, `maxarea` (in square meters, optional); and (vi) a threshold value for the cleaning procedures, `cleansize` (in square meters, optional; see Sect. 3.2). For simplicity, in the following we refer to the numerical values of `thresh`, `areamin`, `cvmin`, and `rf` as  $t$ ,  $a$ ,  $c$ , and  $r$ , respectively.

The software adopts an iterative approach to partition a landscape into SUs. In the first iteration, `r.watershed` uses the threshold  $t$  for controlling the partitioning into HBs. The parameter  $t$  has to be smaller than the maximum surface accumulation area (FA) for the study area to allow for the delineation of at least two HBs that are large enough to be further subdivided. Grid cells with  $FA > t$  are recognized as drainage lines (i.e., streams), and used to delineate the river network by `r.watershed`, grouping all grid cells in the DEM that drain into a given stream segment. These cells collectively represent the catchment drained by the stream segment, which can be further subdivided into left and right HBs (Fig. 2b). A low value of the contributing (FA) area  $t$  results in a dense hydrological network draining a large number of small HBs, and a large value of  $t$  results in a reduced number of streams draining a smaller number of relatively large HBs.

At each iteration, where a GIS layer showing APs is available, `r.slopeunits` identifies the grid cells in the APs and excludes them from the analysis (Fig. 2c). When the SUs are exploited for the analysis of the processes causing slope instability, the exclusion is justified by the empirical observation that landslides do not occur in plain areas. The value of the `areamin` parameter  $a$  defines the smallest possible (planimetric) area for a SU. The circular variance is defined as  $1 - |\mathbf{R}|/N_v$ , in the range between 0 and 1, where  $N_v$  is the number of grid cells in each HB and  $|\mathbf{R}|$  is the magnitude of the vector  $\mathbf{R}$  that results from the sum of all unit vectors describing the orientation of each grid cell. As an example, Fig. 3a shows that a group of unit vectors dispersed  $23^\circ$  apart, on average, is characterized by a circular variance of 0.1, and Fig. 3b shows that a group of unit vectors dispersed  $62^\circ$  apart, on average, is characterized by a circular variance of 0.6. Thus, a value of 0.1 of the circular variance represents grid cells all facing nearly in the same direction, whereas a value of 0.6 represents more dispersed grid cells. In the algorithm, the circular variance is controlled by the value of the `cvmin` parameter  $c$ , affecting the homogeneity of terrain aspect in the HBs. Small values of  $c$  result in more uniform HBs, and large values of  $c$  in less uniform HBs, in terms of terrain aspect.

At each iteration, `r.watershed` splits each existing parent half basin,  $HB_{parent}$  (Fig. 2d) into nested child half basins,  $HB_{child}$ . The average area of the  $HB_{child}$  defined for each  $HB_{parent}$  is checked against  $a$ . When the average area is smaller than  $a$ , the subdivision is rejected and the  $HB_{parent}$  is selected as a candidate SU. When the average area is larger than  $a$ , the procedure keeps the  $HB_{child}$  for the next step



**Figure 3.** Graphical representation of the circular variance of terrain aspect,  $1 - |\mathbf{R}|/N_v$ . Two groups of unit vectors are shown. The unit vectors represent the local direction of terrain aspect for each grid cell, resulting in circular variance (a) of 0.1 and (b) of 0.6. Unit vectors are perpendicular to the local topography, represented by the grid cell.

of the analysis. The rejection procedure prevents very small HBs from being selected as a candidate SU.

The next step of the procedure consists in comparing the size and circular variance of each  $\text{HB}_{\text{child}}$  with the user-defined values  $a$  and  $c$ . Where the circular variance is smaller than  $c$ , or the size is smaller than  $a$ , the  $\text{HB}_{\text{child}}$  is taken as a candidate SU (Fig. 2e). If the user defines the optional input parameter `maxarea`, the control on the circular variance is not performed for the  $\text{HB}_{\text{child}}$  having a size larger than `maxarea`. This imposes a constraint on the maximum size of the candidate SU (not shown in Fig. 2e). Each and every remaining  $\text{HB}_{\text{child}}$  is fed to the next iteration of `r.watershed`, which is initialized using a smaller value of  $t$ . At the  $i$ th iteration, the value of  $t$  depends on the value of the reduction factor  $r$ , according to  $t_{i+1} = t_i - t_i/r$  (Fig. 2f). The decrease of  $t_i$  is faster for small values of  $r \geq 2$ . We checked empirically that values of  $r > 10$  lead to visually better results, at least in our study area. This is due to the fact that a slow decrease of  $t_i$  allows for a better (finer) control on the subdivision of the parent half basin,  $\text{HB}_{\text{parent}}$ . On the other hand, a fast decrease obtained with  $r = 2$  prevents the algorithm from checking if intermediate values of  $t_i$  produce a candidate SU. Thus, a slow decrease in  $t_i$  results in a larger number of iterations, which produce better results at the expenses of a longer computing time required to complete the iterations.

At each iteration, each  $\text{HB}_{\text{child}}$  is processed again by `r.watershed` as  $\text{HB}_{\text{parent}}$ . The iterative procedure ends when the entire study area is subdivided into candidate SUs that match the user requirements, in terms of minimum area ( $a$ ) and circular variance ( $c$ ). In the resulting map, the size and shape of the candidate SUs can be determined by the constraint of minimum surface area, or by the constraint of the minimum circular variance of the terrain aspect. The final terrain subdivision contains candidate SUs whose minimum size approximates  $a$ .

The final SU partitioning is obtained after an additional (cleaning) step intended to identify and process candidate SUs exhibiting unrealistic or unacceptable size or shape (Fig. 2g). This is discussed in the next section.

### 3.2 Slope units cleaning procedure

The `r.slopeunits` software may produce locally unrealistic candidate SUs which are too small, too large, or oddly shaped. As an example, in large open valleys where terrain is flat or multiple channels join in a small area, unrealistically small subdivisions can be produced. Another example is represented by unrealistically elongated or large candidate SUs found along regular and planar slopes. These terrain subdivisions, although legitimate from the algorithm hydrological perspective, are problematic for practical applications and should be revised and removed eventually. Very small candidate SUs consisting of a few grid cells are often the result of artifacts (errors) in the DEM. These candidate SUs should also be removed. To remove candidate SUs with unrealistic or unwanted sizes (Fig. 2g), we have implemented three distinct software tools based on three different methods. The three methods require the user to set the `cleansize` parameter that imposes a strict constraint on the minimum possible size (planimetric area) for a slope unit.

The first method simply removes all candidate SUs smaller than `cleansize`. The adjacent candidate SUs are enlarged to fill the area left by the removed units, using the `r.grow` GRASS GIS module. The second method, in addition to removing all candidate SUs smaller than `cleansize`, also removes odd-shaped polygons from the set of the candidate SUs. Enabled via the `-m` flag in combination with `cleansize`, the second method removes markedly elongated candidate SUs with a width smaller than two grid cells. The third method merges small candidate SUs with neighboring ones based on the average terrain aspect. Enabled via the `-n` flag in combination with `cleansize`, the third method calculates the average terrain aspect of all the grid cells in each small candidate SU. The average aspects of the neighboring SUs are compared, and the two adjacent units exhibiting the smallest difference in average terrain aspect are merged, provided that the SU that is removed shares a significant part of its boundary with the neighboring unit. The final result is a terrain subdivision in which the vast majority of the SU has an area larger than `cleansize`. A drawback of the third method is that it is significantly more computer intensive than the other two methods, requiring longer processing times.

## 4 Segmentation metric

To evaluate the terrain partitioning into SUs, we use a simple metric originally proposed for the evaluation of the quality of a segmentation result (Espindola et al., 2006). In digital image processing, segmentation consists in the process of partitioning an image into sets of pixels, such that pixels within the same set share certain common characteristics. Here, we consider the terrain aspect grid map as an image to be segmented, and we assume that the segmentation metric

proposed by Espindola et al. (2006) is appropriate to evaluate the terrain partition into SUs. We base the assumption on the observation that the metric makes a straightforward evaluation of the internal homogeneity of the SUs using the local aspect variance  $V$ , and the external heterogeneity of the SUs using the autocorrelation index,  $I$ . The two quantities are defined as follows:

$$V = \frac{\sum_n s_n c_n}{\sum_n s_n}, \quad (1)$$

and

$$I = \frac{N \sum_{n,l} w_{nl} (\alpha_n - \bar{\alpha}) (\alpha_l - \bar{\alpha})}{\left( \sum_n (\alpha_n - \bar{\alpha})^2 \right) \left( \sum_{n,l} w_{nl} \right)}, \quad (2)$$

where  $n$  labels all the  $N$  SUs in a given partition,  $s_n$  is the surface area of the  $n$ th SU,  $c_n$  is the circular variance of the aspect in the  $n$ th SU,  $\alpha_n$  is the average aspect of the  $n$ th SU,  $\bar{\alpha}$  is the average aspect of the entire terrain aspect map, and  $w_{nl}$  is an indicator for spatial proximity, equal to unity if SU  $n$  and  $l$  are adjacent, zero otherwise. The local variance  $V$ , defined in Eq. (1), assigns more importance to large SUs, avoiding numerical instabilities produced by small SUs. The autocorrelation index  $I$  of Eq. (2) has minima for partitions that exhibit well-defined boundaries between different SUs.

The optimal selection of the input parameters is the one that combines small  $V$  and small  $I$ . This is quantified by the following objective function (Espindola et al., 2006):

$$F(V, I) = \frac{V_{\max} - V}{V_{\max} - V_{\min}} + \frac{I_{\max} - I}{I_{\max} - I_{\min}}, \quad (3)$$

where  $V_{\min(\max)}$  and  $I_{\min(\max)}$  are the minimum (maximum) values of the quantities in Eqs. (1) and (2) as a function of the input parameters. To calculate  $I$ , we rewrite Eq. (2) to make it consistent with the terrain aspect map. The aspect map contains values in degrees, and the average values and products cannot be taken straightforwardly, and the following definitions have to be considered. The average values of the angles are a vectorial sum of unit vectors, so that

$$\bar{\alpha} = \text{Arctan} \left( \frac{\sum_j \sin \alpha_j}{\sum_j \cos \alpha_j} \right), \quad (4)$$

where the index  $j$  runs over all the grid cells in the terrain aspect map. A similar definition holds for  $\alpha_i$ , the average aspect inside the  $i$ th slope unit, if the sum in Eq. (4) is limited to the grid cells belonging to the  $i$ th SU. The difference  $(\alpha_i - \bar{\alpha})$  should also be intended vectorially, as follows:

$$\alpha_i - \bar{\alpha} = \text{Arctan} \left( \frac{\sin \alpha_i - \sin \bar{\alpha}}{\cos \alpha_i - \cos \bar{\alpha}} \right). \quad (5)$$

Lastly, the product at the numerator of Eq. (2) is taken as the scalar product of the vectors  $(\alpha_i - \bar{\alpha})$  and  $(\alpha_j - \bar{\alpha})$ , as follows:

$$(\alpha_i - \bar{\alpha}) \cdot (\alpha_j - \bar{\alpha}) = \cos \theta_i \cos \theta_j + \sin \theta_i \sin \theta_j, \quad (6)$$

where

$$\theta_i = \text{Arctan} \left( \frac{\sin \alpha_i - \sin \bar{\alpha}}{\cos \alpha_i - \cos \bar{\alpha}} \right), \quad (7)$$

$$\theta_j = \text{Arctan} \left( \frac{\sin \alpha_j - \sin \bar{\alpha}}{\cos \alpha_j - \cos \bar{\alpha}} \right). \quad (8)$$

Care must be taken in expressing angles and arcs consistently in degrees or radians.

The segmentation metric  $F(V, I)$  is a measure of the degree of fulfillment of the SU internal homogeneity requirement, and it is related only to the SU geometrical and morphological delineation. The metric can be used to define the optimal (best) partition of the territory in terms of aspect segmentation by maximization of  $F(V, I) = F(V(a, c), I(a, c))$  as a function of the  $a$  and  $c$  user-defined modeling parameters.

## 5 Landslide susceptibility modeling

Landslide susceptibility (LS) is the likelihood of landslide occurrence in an area, given the local terrain conditions, including topography, morphology, hydrology, lithology, and land use (Brabb, 1984; Guzzetti et al., 1999; Guzzetti, 2006). Various types of LS modeling approaches are available in the literature. The approaches differ – among other things – on the type of the TU used to partition the landscape and to ascertain LS (Guzzetti et al., 1999; Huabin et al., 2005; Guzzetti, 2006; Pardeshi et al., 2013). Interestingly, instead of trying to define where landslides are not expected, (e.g., Marchesini et al., 2014), over the last decades a lot of efforts have been spent on the prediction of the spatial probability of the slope failures. Among the many available types of TUs (Guzzetti, 2006), SUs have proved to be effective terrain subdivisions for LS modeling.

To model LS, we considered slow- to very slow-moving shallow slides, deep-seated slides, and earth flows, and we excluded rapid to fast-moving landslides, including debris flows and rock falls. The `r.slopeunits` software performs the delineation of the SUs, and does not perform the LS modeling. For the latter, we exploit specific modeling software (Rossi et al., 2010; Rossi and Reichenbach, 2016).

In this work, we prepare LS models adopting a single multivariate statistical classification model. For the purpose, we use a logistic regression model (LRM) to quantify the relationship between dependent (landslide presence/absence) and independent (geo-environmental) variables. We use the presence/absence of landslides in each SU as the grouping (i.e., dependent) variable. Adopting a consolidated approach in our study area (Carrara et al., 1991, 1995; Guzzetti et al., 1999), SUs with 2% or more of their area occupied by landslides are considered unstable (having landslides), and SUs with less than 2% of the area occupied by landslides are considered stable (free of landslides). The 2% threshold value depends on the accuracy of a typical landslide inventory map

(Carrara et al., 1991, 1995; Guzzetti et al., 1999; Santangelo et al., 2015). Users of the `r.slopeunits` software may select a different value for the landslide presence/absence threshold in the LRM (or any other classification model, where considered more appropriate, in different study areas, and with different input data. Numerical (i.e., terrain elevation, slope, curvature, and other variables derived from the DEM), and categorical (i.e., lithology and land use) variables are used as explanatory (i.e., independent) variables in the LS modeling. Each SU is characterized by (i) statistics calculated for all the numerical variables, and (ii) percentages of all classes for the categorical variables.

The LS evaluation is repeated many times using different SU terrain subdivisions obtained changing the `r.slopeunits` ( $a, c$ ) modeling parameters, and cleaned from small areas using the first (and simplest) method described in Sect. 3.2. We evaluate the performance skills of the different LS models calculating the  $AUC_{ROC}$  (Rossi et al., 2010). ROC curves show the performance of a binary classifier system when different discrimination probability thresholds are chosen (Fawcett, 2006). A ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) for the different thresholds. The area under the ROC curve,  $AUC_{ROC}$ , ranging from 0 to 1, is used to measure the performance of a model classifier that, in our case, corresponds to the LS model.

Hereafter, we quantify the  $AUC_{ROC}$  metric with the  $R(a, c)$  function, for each value of the  $a$  (surface area) and  $c$  (circular variance) input parameters of the `r.slopeunits` software. We stress that we use the  $AUC_{ROC}$  metric to evaluate the fitting performance of the LS model, i.e., to evaluate the ability of the LS model to fit the same landslide set used to construct the LS model (the landslide training set), and not an independent landslide validation set (Guzzetti et al., 2006; Rossi et al., 2010). This is done purposely, because the purpose of the procedure is not to evaluate the performance of the LS classification but to help determine an optimal terrain subdivision for LS modeling, and thus before any LS model is available for proper validation. When an optimal SU subdivision is obtained (see Sect. 6), and a corresponding LS model is prepared, the prediction skills of the model can be evaluated using independent landslide information (where this is available) (Guzzetti et al., 2006; Rossi et al., 2010).

In addition to the  $AUC_{ROC}$ , which is a direct measure of the fitting/prediction performance of a binary classifier, the performance of the LRM model can be analyzed in terms of how the different input variables (both numerical and categorical) contribute to the final result (Budimir et al., 2015). In the model, a  $p$  value can be associated to each variable, and used to establish the significance of the variable in the LRM. The fraction of significant variables used by the LRM can be used to qualitatively understand the behavior of the classification model as a function of the `r.slopeunits` software input parameters or, in turn, as a function of the average size of the SUs.

## 6 Optimization of SU partitioning for LS zonation

Once, for all the SU sets computed using different modeling parameters (i.e., different  $a$  (surface area) and  $c$  (circular variance) values), (i) the SU terrain aspect segmentation metric  $F(a, c)$  – that assesses how well the requirements of internal homogeneity/external heterogeneity are fulfilled – has been established (2a in Fig. 1 and Sect. 4; Espindola et al., 2006); and (ii) the performance of the individual LS models are estimated using the  $AUC_{ROC}$  metric  $R(a, c)$  – that assesses the calibration skills of the LRM used for LS modeling – has also been established (2b in Fig. 1 and Sect. 5; Rossi et al., 2010), a proper objective function  $S$  that combines the aspect segmentation metric ( $F(a, c)$ ) and the  $AUC_{ROC}$  calibration metric ( $R(a, c)$ ) is established. Maximization of  $S(a, c)$  as a function of the  $a$  (the minimum surface area of the slope unit) and  $c$  (the slope unit circular variance) modeling parameters allows to single out the optimal SU terrain subdivision for LS modeling in the given study area.

The reason for proposing a combination of the aspect segmentation and the  $AUC_{ROC}$  metrics in the search for an optimal terrain subdivision for LS modeling is the following. The single segmentation metric  $F(a, c)$  is a measure of the degree of fulfillment of inter-unit homogeneity and intra-unit heterogeneity for a SU delineation obtained with given ( $a, c$ ) values. As such,  $F(a, c)$  is related only to the geometrical delineation of the SUs, and does not consider the subsequent application of a LS model, the presence/absence of landslides, or any quantity other than terrain aspect.

To combine the two functions  $R(a, c)$  and  $F(a, c)$  into a single objective function, we normalized them to  $[0, 1]$  as follows:

$$R_o(a, c) = \frac{R(a, c) - R_{\min}(a, c)}{R_{\max}(a, c) - R_{\min}(a, c)} \quad (9)$$

$$F_o(a, c) = \frac{F(a, c) - F_{\min}(a, c)}{F_{\max}(a, c) - F_{\min}(a, c)}. \quad (10)$$

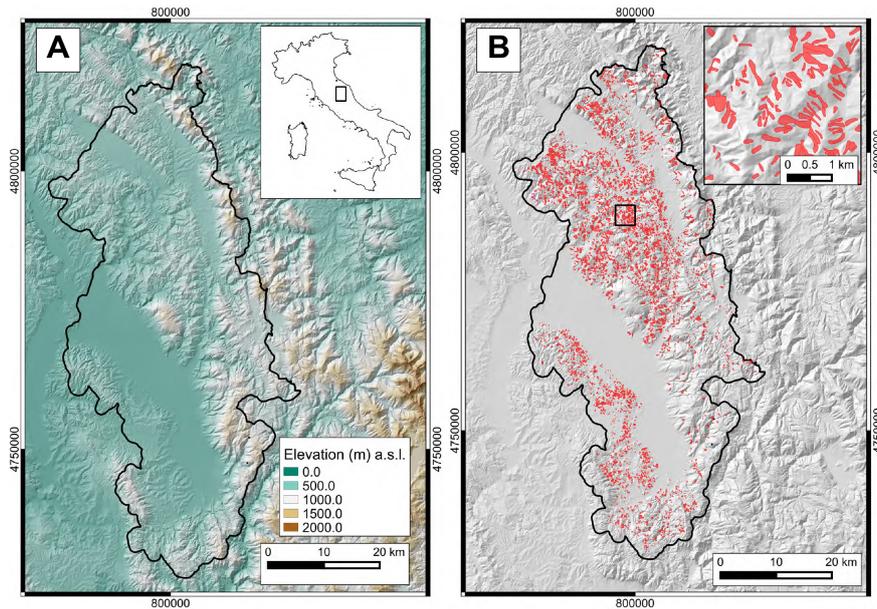
The functions  $R_o(a, c)$  and  $F_o(a, c)$  are then multiplied to obtain the final objective function  $S(a, c)$ :

$$S(a, c) = R_o(a, c)F_o(a, c), \quad (11)$$

which embodies information on the quality of the terrain aspect map segmentation and on the performance of the LS model in a consistent, objective, and reproducible way. The function  $S(a, c)$  assumes values in the range  $[0, 1]$ : the larger the value the better the SU partitioning in terms of (i) SU internal homogeneity and external heterogeneity, and (ii) suitability of the subdivision for LS zonation in our study area.

## 7 Test area

We tested our proposed modeling framework for the delineation of SUs, and for the selection of the optimal modeling parameters for LS assessment, in a portion of the upper



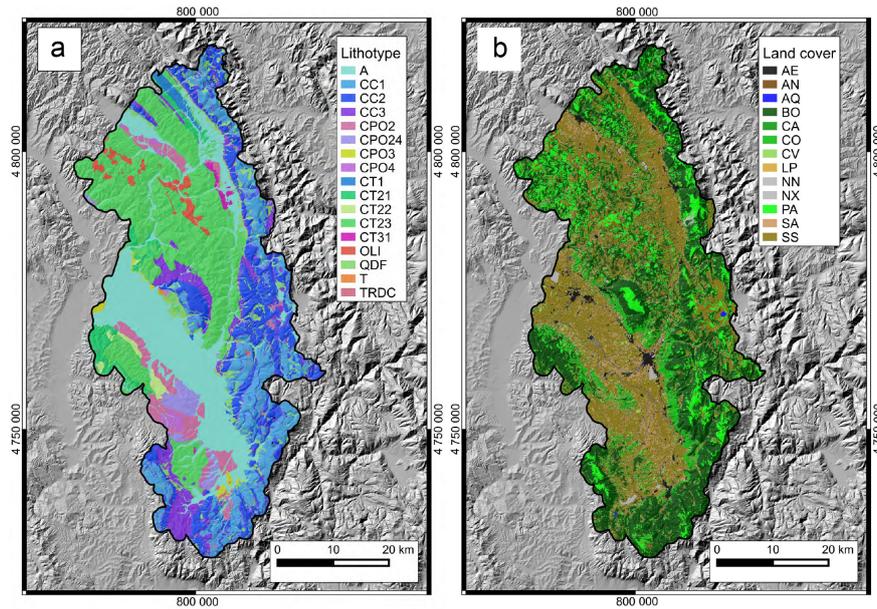
**Figure 4.** (a) Shaded relief image of the study area located on the upper Tiber River basin, Umbria, central Italy. Shades of green to brown show increasing elevation. Inset shows location of the study area in Italy. (b) Landslide inventory map (Cardinali et al., 2001). Inset shows the detail of the landslide mapping. Landslides (shown in red) were used to prepare the landslide susceptibility zonations shown in Fig. 10. The maps are in the UTM zone 32, datum ED50 (EPSG:23032) reference system.

Tiber River basin, central Italy (Fig. 4a). In the 2000 km<sup>2</sup> area, elevation ranges from 175 to 1571 m, and terrain gradient from almost zero along the river plains, to more than 70° in the mountains and the steepest hills. Four lithological complexes, or groups of rock units (Cardinali et al., 2001), crop out in the area, including: (i) sedimentary rocks pertaining to the Umbria–Marche sequence, Lias to lower Miocene in age; (ii) rocks pertaining to the Umbria turbidites sequence, Miocene in age; (iii) continental, post-orogenic deposits, Pliocene to Pleistocene in age; and (iv) alluvial deposits, recent in age. Each lithological complex comprises different sedimentary rock types varying in strength from hard to weak and soft rocks. Hard rocks are massive limestone, cherty limestone, sandstone, travertine, and conglomerate. Weak rocks are marl, rock-shale, sand, silty clay, and stiff over-consolidated clay. Soft rocks are clay, silty clay, and shale. Rocks are mostly layered and locally structurally complex. Soils in the area reflect the lithological types, and range in thickness from less than 20 cm where limestone and sandstone crop out along steep slopes, to more than 1.5 m in karst depressions and in large open valleys.

To model LS, we use a digital representation of the terrain elevation, an inventory of known landslides, and relevant geo-environmental information. We use a DEM with a ground resolution of 25 m × 25 m obtained through linear interpolation of elevation data along contour lines shown on 1 : 25 000 topographic base maps (Cardinali et al., 2001) (Fig. 4a). The landslide inventory (Fig. 4b) was obtained from the visual interpretation of multiple sets of aerial pho-

tographs flown between 1954 and 1977, aided by field surveys, review of historical and bibliographical data, and geological, geomorphological, and other available landslide maps (Cardinali et al., 2001; Galli et al., 2008; Guzzetti et al., 2008; Alvioli et al., 2014). To prepare the LS model, we considered only the shallow slides, the deep-seated slides, and the earth flows. These landslides are (i) slow- to very slow-moving failures, and (ii) they typically remain in the slope (i.e., the slope unit) where they occur. For the deep-seated landslides and the earth flows, the landslide source (depletion) area and the deposit were considered together. This is a standard approach in modeling LS for these types of landslides. We excluded from the LS modeling all the rapid to fast-moving landslides, including debris flows and rock falls that may travel outside the slope (i.e., the slope unit) where they form.

We obtained lithological information (Fig. 5a) from available geological maps, at 1 : 10 000 scale (Regione Umbria, 1995–2015), prepared in the framework of the Italian national geological mapping project CARG and in other regional geological mapping projects. The original maps, in vector format, were edited to eliminate and reclassify polygons coded as landslide deposits or debris flow deposits, and to reclassify the original 186 geologic formations and 20 cover types in five lithological complexes or lithological domains (Table 1). Definition of the five lithological complexes was made on the basis of the characteristics of the rock types (carbonate, terrigenous, volcanic, post-orogenic sediments) and the degree of competence or composition of the differ-



**Figure 5.** (a) Lithological map of upper Tiber River basin, Umbria, central Italy (Regione Umbria, 1995–2015). Details of the lithological types are given in Table 1. (b) Land use map of the study area. Legend: AE: urban area, AN: bare soil, AQ: water, BO: forest, CA: orchard, CO: olive grove, CV: vineyard, LP: poplar trees, NN/NX: unclassified, PA: meadow/pasture, SA: treed seminative, SS: simple seminative. The maps are in the UTM zone 32, datum ED50 (EPSG:23032) reference system.

ent geological units (massive, laminated sandstone/pelitic-rock ratio). Applying these criteria, the original 206 geological units were grouped in 17 lithological units. We obtained information on land cover from a map, at 1 : 10 000 scale (Regione Umbria, 1995–2015), prepared through the visual interpretation of color aerial photography at 1 : 13 000 scale, acquired in 1977 (Fig. 5b). The map contains 13 land cover classes, of which the most common are forest (42 %), arable land (31 %), meadow and pasture (10 %), built up areas (10 %), and vineyards, live trees, and orchards (6 %).

The study area (Fig. 4) corresponds to an alert zone used by the Italian National Department for Civil Protection to issue landslide (and flood) regional warnings. The boundary of the alert zone is partly administrative, and does not correspond locally to drainage and divides lines. As a result, our SU partitioning intersects locally the boundary of the alert zone. For convenience, for our analysis we considered only SUs that fall entirely within the alert zone. As a drawback, the extent of the study area varies slightly depending on the combination of the selected *a* and *c* parameters. We maintain that this has a negligible effect on the final modeling results.

## 8 Results

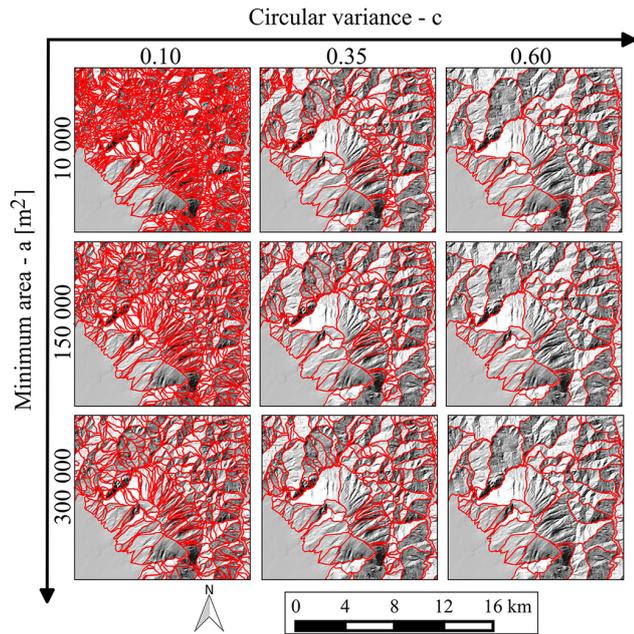
### 8.1 Slope units delineation

In the study area, we ran the `r.slopeunits` software for 99 different combinations of user-defined input parameters, resulting in 99 different terrain subdivisions. In the itera-

**Table 1.** Lithological codes used in Fig. 5. SD is sandstone, P is pelitic rock, CO is conglomerate, S is sand, G is clay (Regione Umbria, 1995–2015).

Complex		Code	
Carbonate	Massive layers	CC1	
	Thin layers	CC2	
	Marl, calcareous marl	CC3	
Terrigenous	Massive sandstone	CT1	
	Variable SD/P fraction	Thick layers	CT21
		Medium layers	CT22
		Thin layers	CT23
	Marly, SD/P << 1	Stratified	CT31
Post-orogenic	Conglomerate, gravel, pebble	CPO2	
	Sand	CPO3	
	Silt and clay	CPO4	
	Variable CO/S/G fraction	CPO24	
Others	Olistostrome	OLI	
Quaternary deposit	Alluvial deposit	A	
	Debris cone	QDF	
	Red soil in karstic landscape	TRDC	
	Travertine	T	

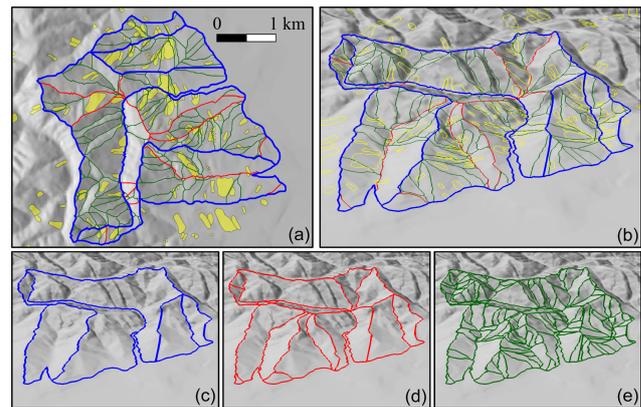
tive procedure (Fig. 2), the initial FA threshold (*t*) area and the reduction factor (*r*) control the numerical convergence, and do not have an explicit geomorphological meaning. On the other hand, the minimum area (*a*) and the circular variance (*c*) determine the size and control the aspect of the SUs. For the analysis, we selected a large value for the FA area ( $t = 5 \times 10^6 \text{ m}^2$ ) keeping it constant for all the different



**Figure 6.** A group of 9 (out of 99) SU terrain subdivisions for a portion of the study area. The SUs were obtained changing the  $a$  and  $c$  parameters used by the `r.slopeunit` software. The maps are in the UTM zone 32, datum ED50 (EPSG:23032) reference system.

model runs. The large value was chosen to obtain large initial HBs that could be further subdivided into smaller SUs by the iterative procedure (see Sect. 4). The value is also consistent with (i.e., substantially larger than) (i) the size of the average SU used to partition the same area in a different LS modeling effort (Cardinali et al., 2001, 2002), and (ii) the average area of the considered landslides (shallow slides, the deep-seated slides, and the earth flows) in the study area (Guzzetti et al., 2008). Selection of the reduction factor ( $r = 10$ ) was heuristic and motivated by the fact that this figure provided stable results compared to those obtained using larger values. The two most relevant parameters, the minimum area  $a$  and the circular variance  $c$ , were selected from broad ranges:  $a = 10\,000, 25\,000, 50\,000, 75\,000, 100\,000, 125\,000, 150\,000, 200\,000, \text{ and } 300\,000 \text{ m}^2$ , and  $0.1 < c < 0.6$ , at evenly spaced values with an increment of 0.05. For all the `r.slopeunits` model runs, we used the same value for `cleansize` =  $20\,000 \text{ m}^2$ , to remove candidate SUs with area  $< 20\,000 \text{ m}^2$  using the first (and simplest) of the three methods described in Sect. 3.2.

Figure 6 shows 9 of the 99 results of the terrain subdivisions obtained using different combinations of the  $a$  and  $c$  modeling parameters. The map in the upper left (lower right) corner shows the finest (coarsest) SU partitioning, determined using small (large) values of  $a$  and  $c$ . The map in the center was obtained using intermediate values for the two user-defined modeling parameters. For a limited portion of the study area, Fig. 7 shows different partitioning



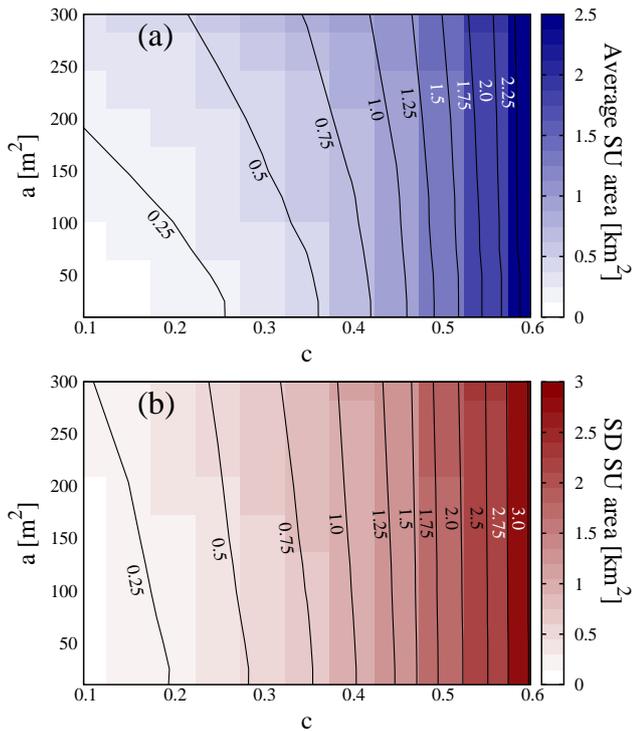
**Figure 7.** Example of subdivisions into SUs for a portion of the study area. Legend: blue, red, and green lines show boundaries of SUs of increasing density and corresponding decreasing average size. Yellow areas are landslides. The five maps show the same area in plan view (a) and in perspective view (b, c, d, e).

results. In particular, Fig. 7a and b show the overlay of the three different partitions with the landslide inventory map (Fig. 4), using a two- and three-dimensional visualization, respectively, and a shaded image relief of the terrain, whereas Fig. 7c, d, and e show separately the same partitions using a three-dimensional representation. Visual inspection of Fig. 7a, b, and c reveals that the coarser subdivision ( $c = 0.60$  and  $a = 0.3 \text{ km}^2$ ), shown in blue, defines large SUs characterized by heterogeneous orientation (aspect) values. On the other hand, the finest SU subdivision (shown in green in Fig. 7a, b and e) obtained using  $c = 0.10$  and  $a = 0.01 \text{ km}^2$ , is too small to completely include many of the landslides (shown in yellow). The combination that uses intermediate values  $c = 0.35$  and  $a = 0.15 \text{ km}^2$  resulted in the SU subdivision shown in red in Fig. 7a, b, and d.

Figure 8 shows the effect of different combinations of  $a$  and  $c$  on the average SU size. The average area of the SU increases significantly with  $c$  (less homogeneous, more irregular slope), and it is less sensitive to the increment of  $a$  (Fig. 8a). The effect of  $c$  becomes predominant in Fig. 8b. The standard deviation of the area of the SU varies significantly with  $c$ , highlighting that larger values of  $c$  increase the average and the variability of the SU size (Fig. 8a and b, respectively).

## 8.2 Segmentation metric

For each of the 99 terrain subdivisions obtained using the procedure described above, we calculated the segmentation objective function value given by Eq. (3). The segmentation metric  $F(a, c)$  (Fig. 9) is a measure of the performance of our SU delineation algorithm, and an assessment of how well the requirements of internal homogeneity/external heterogeneity are fulfilled by the procedure, as a function of  $a$  and  $c$  (2a in Fig. 1, Sect. 4). Where the SUs are relatively small, their de-



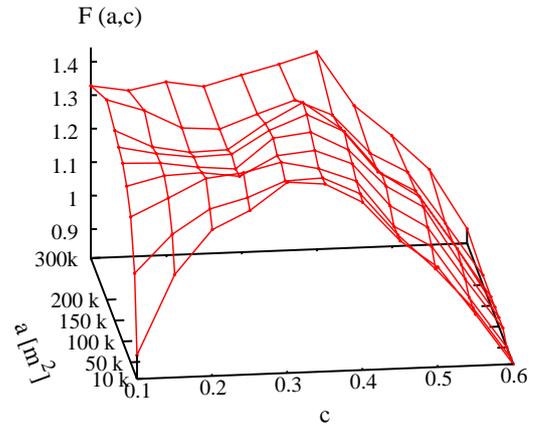
**Figure 8.** (a) Average and (b) standard deviation of the size of the SUs, for different combinations of  $a$  and  $c$  parameters. The minimum size of the SUs is fixed by the  $a$  parameter, and the maximum size of the SUs is independent of  $a$ .

gree of internal homogeneity is large, but the requested heterogeneity between adjacent units is not completely fulfilled. Where the SUs are large, the requested internal homogeneity is not fulfilled entirely, because in each SU the aspect variability is large. The analysis of the  $F(a, c)$  values in Fig. 9 suggests that SU subdivisions obtained using  $c$  smaller than about 0.2 and  $a$  smaller than about 50 000 m<sup>2</sup>, or  $c$  larger than about 0.5, should not be considered in the analysis because they are too small or too large to satisfy the aspect variability requirement.

### 8.3 Landslide susceptibility modeling

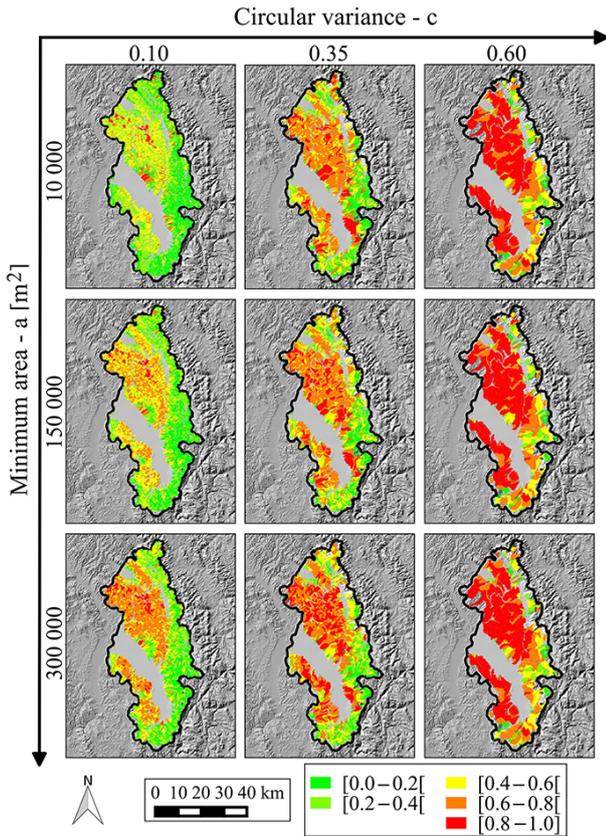
For each of the 99 SU delineations, we prepared a different LS zonation using a LRM (Sect. 6) adopting the modeling scheme described by Rossi et al. (2010); Rossi and Reichenbach (2016). Figure 10 shows the results obtained for 9 (out of 99) combinations of the  $a$  and  $c$  parameters. For each of the 99 susceptibility assessments, we evaluated the fitting (calibration) performance of the models computing AUC<sub>ROC</sub> (Rossi et al., 2010), and Fig. 11 shows the obtained AUC<sub>ROC</sub> as a function of the  $a$  and  $c$  parameters (Sect. 5).

The larger values of AUC<sub>ROC</sub> were obtained for LS zonations based on SU partitions resulting from large values of  $c$  and  $a$ . Such combinations may locally result in SUs with extremely large internal heterogeneity. Signatures of the het-



**Figure 9.** Segmentation objective function values (Eq. 3) calculated for 99 SU partitions obtained using different combinations of the  $a$  and  $c$  parameters.

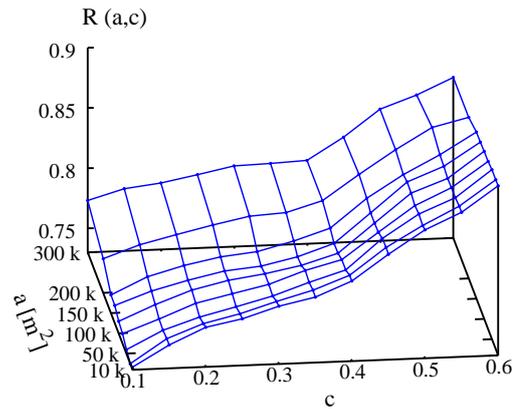
erogeneity within very large slope units can be found both in the segmentation metric and in the LRM model results. In the segmentation metric case, this is very straightforward since the values of  $F(a, c)$ , which is a direct measure of the slope aspect homogeneity within SUs, clearly decrease with increasing values of  $a$  and  $c$ , as shown in Fig. 9. Very large SUs, however, are not only heterogeneous in terms of terrain aspect, but also in terms of the morphometric and thematic variables used as input of the LRM. This is reflected in the number of input variables that significantly contribute to the susceptibility model results as a function of  $a$  and  $c$ . We computed the number of statistically significant variables in each realization of the LRM (i.e., the variables with a  $p$  value < 0.05). Figure 12 shows that the number of significant variables ranges from 5% (of the 50 morphometric and thematic variables) for large SUs, to 35% for small SUs. From this analysis we observe that for very large slope units, only very few variables are effectively used by the LRM. A more detailed analysis reveals that, in our test case, lithological variables significantly control the results, whereas other local settings (e.g., terrain slope) are neglected. The relevant variables are typically the terrigenous sediments and carbonate lithological complexes, where landslides are expected and not expected, respectively. As a result, in the region of large ( $a, c$ ) parameters, even if we obtain high values of AUC<sub>ROC</sub>, the LRM can be replaced by a simple heuristic analysis of the lithological map. The fine details of the remaining input variables are lost and a multivariate statistical approach is of little use. We clarify that to evaluate the model performance, any statistical metric based on the comparison of observed and predicted data (e.g., confusion matrices and derived indexes), would exhibit the same or similar trend as the AUC<sub>ROC</sub> as a function of the SU size.



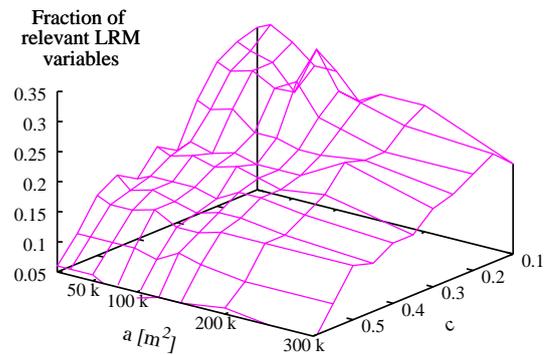
**Figure 10.** The group of 9 (out of 99) LS maps obtained with different SU partitions resulting from different combinations of the  $a$  and  $c$  parameters. The maps are in the UTM zone 32, datum ED50 (EPSG:23032) reference system.

**9 Discussion**

We have run the `r.slopeunits` software with a significant number of combinations (99) of the  $(a, c)$  input parameters, and a corresponding number of realizations of the LS model. Results showed that new `r.slopeunits` software was capable of capturing the morphological variability of the landscape and partitioning the study area into SU subdivisions of different shapes and sizes well suited for LS modeling and zonation. As a matter of fact, depending on the type of landslides, the scale of the available DEM, the morphological variability of the landscape, and the purpose of the zonation, the detail of the terrain subdivision may vary. A detailed terrain partitioning, with many small SUs, is required to capture the complex morphology of badlands, or to model the susceptibility to small and very small landslides (i.e., soil slips). A coarse terrain subdivision is best suited for modeling the susceptibility of very old and very large, deep-seated, complex and compound landslides. Coarse subdivisions can also be used to model the susceptibility to channeled debris flows that travel long distances from the source areas to the depositional areas. Subdivisions of intermediate size may be



**Figure 11.** Values of the  $AUC_{ROC}$  calculated for LS maps estimated using SU partitions derived for different combination of the  $a$  and  $c$  user-defined modeling parameters.

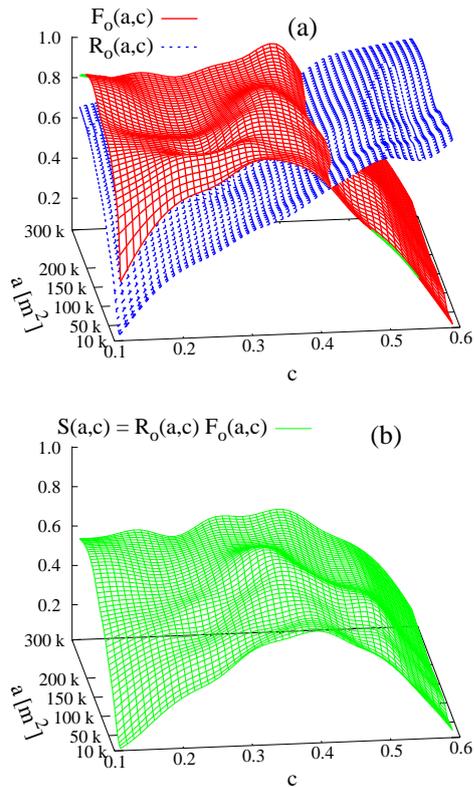


**Figure 12.** Percentage of relevant variables in the LRM used in this work to prepare the LS maps as a function of the  $a$  and  $c$  user-defined modeling parameters. Combination of  $a$  and  $c$  are the same used to prepare Fig. 11. Note that the direction of the axes is reversed with respect to Fig. 11.

required for medium to large slides and earth flows (Carrara et al., 1995). By tuning the set of user-defined model parameters, `r.slopeunits` can prepare SU terrain subdivisions for LS modeling in different geomorphological settings.

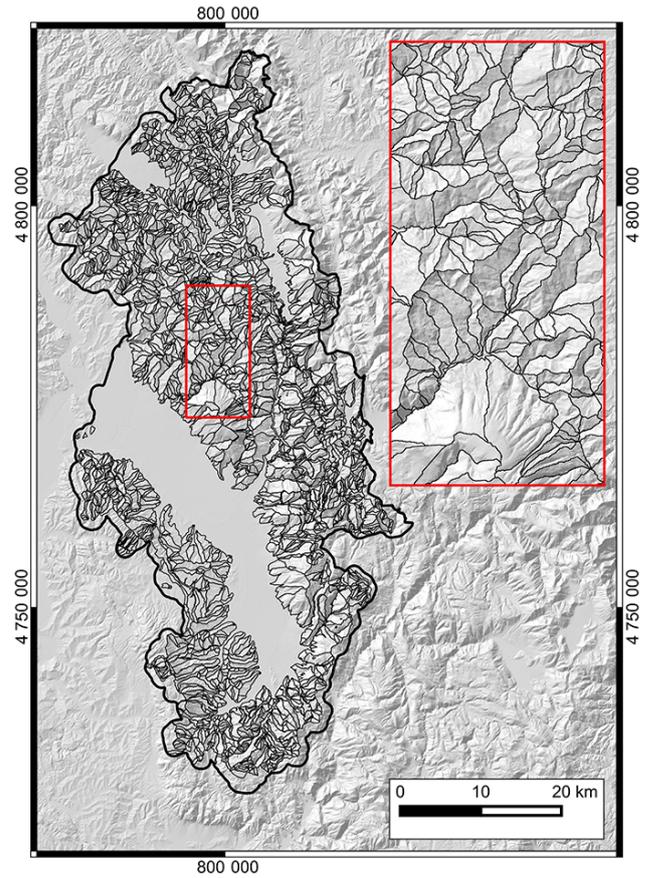
Concerning the LS model, we acknowledge that our selection of the 2% presence/absence threshold may influence the production of the appropriate SU subdivision, and may affect the results of the LS zonation. Examination of different thresholds is not investigated in the present work, because it is not an input parameter of the `r.slopeunits` software and does not change the logic of the approach or the rationale behind our optimization procedure.

We clarify that the subdivisions produced by `r.slopeunits` using different  $(a, c)$  parameters are nested, i.e., the boundaries of a coarse resolution subdivision encompass the boundaries of intermediate and finer subdivisions (see Fig. 7c, d, e). This is a significant operational advantage where landslides of different sizes and types



**Figure 13.** (a)  $R_o(a,c)$  (blue) and  $F_o(a,c)$  (red) functions, defined by Eqs. (9) and (10). (b)  $S_o(a,c)$  function defined by Eq. (11). The functions were interpolated on a denser mesh for improved visual representation.

coexist, posing different threats and requiring multiple and combined susceptibility assessments, each characterized by a different terrain subdivision (Carrara et al., 1995). Optimal values of the  $(a,c)$  parameters have to be determined to obtain the best SU subdivision for a particular goal, in our case, LS modeling. We defined a custom objective function  $S(a,c)$  to determine such optimal values.  $S(a,c)$  is the product of a segmentation quality measure,  $F(a,c)$ , and LS model performance in calibration,  $R(a,c)$ . If only the  $F(a,c)$  metric is used to select a particular set of modeling parameters, the resulting optimal (best) set of SUs has the only meaning of “best partition of the territory in terms of aspect segmentation”. Similarly, the  $R(a,c)$  metric considers solely the classification results of the LS model and not the geometry of the single SU, some of which may be inadequate (e.g., too large, too irregular, too small) for the scope of the terrain zonation. Values of  $F(a,c)$  indicate that there are combinations of the  $c$  and  $a$  parameters that result in SU subdivisions that do not satisfy the user requirements in terms of SU internal homogeneity and external heterogeneity (Fig. 8) (Sect. 8.2). On the other hand, the  $\text{AUC}_{\text{ROC}}$  metric increases with the average size of the SU (Fig. 11) (Sect. 8.3). To select the optimal terrain partitioning for



**Figure 14.** The SU subdivision corresponding to the best parameters selected by our optimization procedure. The values of the parameters are  $a = 150\,000\text{ m}^2$ ,  $a = 0.35$ . The associated LS result is shown in Fig. 10, in the central box, corresponding to the optimal parameters values.

LS zonation in our study area, we exploit the objective function  $S(a,c)$ , which simultaneously quantifies (Sect. 6): (i) the SU internal homogeneity and external heterogeneity (Fig. 13a), and (ii) the (fitting) performance of the LS model (Fig. 13a). Maximization of  $S(a,c)$  (Fig. 13b) provides the best combination of the  $(a,c)$  modeling parameters for a terrain subdivision optimal for LS modeling in our study area.

In addition to the  $\text{AUC}_{\text{ROC}}$ , we have analyzed the performance of the LRM model by studying the fraction of significant variables used by the LRM, to qualitatively understand the behavior of the classification model as a function of the `r.slopeunits` software input parameters or, in turn, as a function of the average size of the SU. The LRM is expected to use the input data less efficiently when the average SU size grows, resulting in a smaller number of significant input variables, which is indeed what we observed. This is due to the LRM inability to discriminate between input variables when the SUs are too large, since each unit usually contains all the possible values of the variables. Using

less data makes it easier for the LRM model to produce a high  $AUC_{ROC}$  result, which does not necessarily correspond to the optimal SU set. The complication is removed using the  $S(a, c) = F_o(a, c)R_o(a, c)$  function, whose  $F(a, c)$  component prevents unrealistic SU sets (both large and small) to have a high overall score.

The function  $S(a, c)$  (Fig. 13b), calculated in our test case for different combinations of the  $a$  and  $c$  modeling parameters, has a maximum value at  $a = 150\,000\text{ m}^2$  and  $c = 0.35$ . The set of SUs that corresponds to the optimal combination of the modeling parameters can be singled out as our optimal (best) result. The LS results corresponding to the optimal combination were already shown in the central box of Fig. 10, and the optimal SU map is presented in Fig. 14.

## 10 Conclusions

Despite the clear advantages of SUs over competing mapping units for LS modeling (Guzzetti, 2006), inspection of the literature reveals that only a small proportion (8%) of the LS zonations prepared in the last three decades worldwide was performed using SUs (Malamud et al., 2014). The limited use of SUs for LS modeling and zonation is due to (among other factors) the unavailability of readily available, easy-to-use software for the accurate and automatic delineation of SUs, and to the intrinsic difficulty in selecting a priori the appropriate size of the SUs for proper terrain partitioning in a given area.

To contribute to filling this gap, we developed new software for the automatic delineation of SUs in large and complex geographical areas based on terrain elevation data (i.e., a DEM) and a small number of user-defined parameters. We further proposed and tested a procedure for the optimal selection of the user parameters in a  $2000\text{ km}^2$  area in Umbria, central Italy.

We expect that the `r.slopeunits` software will be used to prepare terrain subdivisions in different morphological settings, contributing to the preparation of reliable and robust

LS models and associated zonations. We acknowledge that further work is required to investigate the optimization of SU partitions for different statistically based tools used in the literature for LS modeling and zonation (e.g., discriminant analysis, neural network). Guzzetti et al. (2012) have argued that lack of standards hampers landslide studies. This is also the case for the production of landslide susceptibility models and associated maps. We expect that systematic use of the modeling framework proposed in this work (Fig. 1, Sect. 2) and of the `r.slopeunits` software for the objective selection of the user-defined modeling parameters, will contribute to the production of more reliable landslide susceptibility models. It will also facilitate the meaningful comparison of landslide susceptibility models produced, e.g., in the same area using different modeling tools, or in different and distant areas using the same or different modeling tools.

Finally, we argue that the proposed modeling framework and the `r.slopeunits` software are general and not site- or process-specific, and can be used to prepare terrain subdivisions for scopes different from landslide susceptibility mapping, including, e.g., definition of rainfall thresholds for possible landslide initiation, distributed hydrological modelling, statistically based inundation mapping, and the detection and mapping of landslides and other instability processes from satellite imagery.

## 11 Code availability

The code `r.slopeunits` is a free software under the GNU General Public License (v2 and higher). Details about the use and redistribution of the software can be found in the file <https://grass.osgeo.org/home/copyright/> that comes with GRASS GIS. The software and a short user manual can be downloaded at <http://geomorphology.irpi.cnr.it/tools/slope-units> (Geomorphology Research Group, 2016).

## Appendix A: Notation

In Table A1, we list the main variables and the acronyms used in the text.

**Table A1.** Main variables and acronyms used in the text.

Variable	Explanation	First introduced	Dimensions
$a$	minimum surface area for the SUs	Sect. 3.1	$m^2$
$c$	minimum circular variance	Sect. 3.1	–
$r$	reduction factor	Sect. 3.1	–
$t$	flow accumulation threshold	Sect. 3.1	$m^2$
$I$	Autocorrelation index	Eq. (1)	–
$V$	Local aspect variance	Eq. (2)	–
$F(V, I)$	Aspect segmentation metric, also $F(a, c)$	Eq. (3)	–
$R(a, c)$	AUC <sub>ROC</sub> metric for LRM calibration	Sect. 3.1	–
$S(a, c)$	Combined segmentation & AUC <sub>ROC</sub> metric	Eq. (11)	–

Acronym	Explanation
AP	Alluvial plain
AUC	Area under the curve
DEM	Digital elevation model
FA	Flow accumulation
FPR	False positive rate
GIS	Geographical information system
HB	Half basin (left and right portion of a slope unit)
LRM	Logistic regression model
LS	Landslide susceptibility
MFD	Multiple flow direction
ROC	Receiver operating characteristic
SFD	Single flow direction
SU	Slope unit, a morphological terrain unit bounded by drainage and divide lines
TPR	True positive rate
TU	Terrain unit, a subdivision of the terrain

The Supplement related to this article is available online at doi:10.5194/gmd-9-3975-2016-supplement.

*Acknowledgements.* This work was supported by a grant of the Italian National Department of Civil Protection, and by a grant of the Regione dell'Umbria under contract POR-FESR (Repertorio Contratti no. 861, 22/3/2012). M. Alvioli was supported by a grant of the Regione Umbria, under contract POR-FESR Umbria 2007–2013, asse ii, attività a1, azione 5, and by a grant of the DPC. F. Fiorucci was supported by a grant of the Regione Umbria, under contract POR-FESR 861, 2012. We thank A. C. Mondini (CNR IRPI) for useful discussions about segmentation algorithms.

Edited by: L. Gross

Reviewed by: two anonymous referees

## References

- Alvioli, M., Guzzetti, F., and Rossi, M.: Scaling properties of rainfall-induced landslides predicted by a physically based model, *Geomorphology*, 213, 38–47, doi:10.1016/j.geomorph.2013.12.039, 2014.
- Aplin, P. and Smith, G.: Advances in object-based image classification, in: *Remote Sensing and Spatial Information Sciences XXXVII*, Vol. B7 of The International Archives of the Photogrammetry, 725–728, Beijing, China, 2008.
- Brabb, E. E.: Innovative approaches to landslide hazard mapping, in: *Proceedings 4th International Symposium on Landslides (Toronto)*, Vol. 1, 307–324, Canadian geotechnical Society, Toronto, 1984.
- Budimir, M., Atkinson, P., and Lewis, H.: A systematic review of landslide probability mapping using logistic regression, *Landslides*, 12, 419–436, doi:10.1007/s10346-014-0550-5, 2015.
- Cardinali, M., Antonini, G., Reichenbach, P., and Fausto, G.: Photo-geological and landslide inventory map for the Upper Tiber River basin – CNR GNDCI, publication no. 2116, scale 1:100,000, available at: <http://geomorphology.irpi.cnr.it/publications/repository/public/maps/UTR-data.jpg/> (last access: 3 November 2016), 2001.
- Cardinali, M., Carrara, A., Guzzetti, F., and Reichenbach, P.: Landslide hazard map for the Upper Tiber River basin – CNR GNDCI, no. 2634, Scale 1:1,200,000, available at: <http://geomorphology.irpi.cnr.it/publications/repository/public/maps/UTR-hazard.jpg/> (last access: 3 November 2016), 2002.
- Carrara, A.: Drainage and divide networks derived from high-fidelity digital terrain models, in: *Quantitative Analysis of Mineral and Energy Resources*, edited by: Chung, C. F., Fabbri, A. G., and Sinding-Larsen, R., Vol. 223 of *Mathematical and Physical Sciences*, 581–597, D. Reidel Publishing Company, 1988.
- Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., and Reichenbach, P.: GIS techniques and statistical models in evaluating landslide hazard, *Earth Surf. Proc. Land.*, 16, 427–445, doi:10.1002/esp.3290160505, 1991.
- Carrara, A., Cardinali, M., Guzzetti, F., and Reichenbach, P.: GIS Technology in Mapping Landslide Hazard, in: *Geographical Information Systems in Assessing Natural Hazards*, edited by: Carrara, A. and Guzzetti, F., Vol. 5 of *Advances in Natural and Technological Hazards Research*, 135–175, Springer Netherlands, doi:10.1007/978-94-015-8404-3\_8, 1995.
- Dragut, L. and Blaschke, T.: Automated classification of landform elements using object-based image analysis, *Geomorphology*, 81, 330–344, doi:10.1016/j.geomorph.2006.04.013, 2006.
- Dragut, L., Tiede, D., and Levick, S.: ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data, *Int. J. Geogr. Inf. Sci.*, 24, 859–871, 2010.
- Dragut, L., Csillik, O., Eisank, C., and Tiede, D.: Automated parameterisation for multi-scale image segmentation on multiple layers, *ISPRS J. Photogramm. Remote Sens.*, 88, 119–127, doi:10.1016/j.isprsjprs.2013.11.018, 2014.
- Erener, A. and Düzgün, S.: Landslide susceptibility assessment: what are the effects of mapping units and mapping method?, *Environ. Earth Sci.*, 66, 859–877, doi:10.1007/s12665-011-1297-0, 2012.
- Espindola, G., Camara, G., Reis, I., Bins, L., and Monteiro, A.: Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation, *Int. J. Remote Sens.*, 14, 3035–3040, doi:10.1080/01431160600617194, 2006.
- Fall, M., Azzam, R., and Noubactep, C.: A multi-method approach to study the stability of natural slopes and landslide susceptibility mapping, *Eng. Geol.*, 82, 241–263, doi:10.1016/j.enggeo.2005.11.007, 2006.
- Fawcett, T.: An introduction to ROC analysis, *Pattern Recogn. Lett.*, 27, 861–874, doi:10.1016/j.patrec.2005.10.010, 2006.
- Flanders, D., Hall-Beyer, M., and Pereverzoff, J.: Preliminary evaluation of eCognition object-oriented software for cut block delineation and feature extraction, *Can. J. Remote Sens.*, 29, 441–452, doi:10.5589/m03-006, 2003.
- Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., and Reichenbach, P.: Comparing landslide inventory maps, *Geomorphology*, 94, 268–289, 2008.
- Geomorphology Research Group: Slope Units Delineation (*r.slopeunits*), available at: <http://geomorphology.irpi.cnr.it/tools/slope-units>, last access: 3 November 2016.
- Guzzetti, F.: Landslide Hazard and Risk Assessment, PhD Thesis, available at: <http://hss.ulb.uni-bonn.de/2006/0817/0817.htm> (last access: 3 November 2016), 2006.
- Guzzetti, F., Carrara, A., Cardinali, M., and Reichenbach, P.: Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy, *Geomorphology*, 31, 181–216, doi:10.1016/S0169-555X(99)00078-1, 1999.
- Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., and Ardizzone, F.: Probabilistic landslide hazard assessment at the basin scale, *Geomorphology*, 72, 272–299, 2005.
- Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., and Galli, M.: Estimating the quality of landslide susceptibility models, *Geomorphology*, 81, 166–184, doi:10.1016/j.geomorph.2006.04.007, 2006.
- Guzzetti, F., Ardizzone, F., Cardinali, M., Galli, M., Reichenbach, P., and Rossi, M.: Distribution of landslides in the Upper Tiber River basin, central Italy, *Geomorphology*, 96, 105–122, 2008.
- Guzzetti, F., Mondini, A., Cardinali, M., Fiorucci, F., Santangelo, M., and Chang, K.-T.: Landslide inventory maps: New tools for an old problem, *Earth-Sci. Rev.*, 112, 42–66, doi:10.1016/j.earscirev.2012.02.001, 2012.

- Huabin, W., Gangjun, L., Weiya, X., and Gonghui, W.: GIS-based landslide hazard assessment: an overview, *Prog. Phys. Geogr.*, 29, 548–567, doi:10.1191/0309133305pp462ra, 2005.
- Komac, M.: A landslide susceptibility model using the Analytical Hierarchy Process method and multivariate statistics in perialpine Slovenia, *Geomorphology*, 74, 17–28, doi:10.1016/j.geomorph.2005.07.005, 2006.
- Komac, M.: Regional landslide susceptibility model using the Monte Carlo approach – the case of Slovenia, *Geol. Q.*, 56, 41–54, 2012.
- Li, Y., Chen, G., Tang, C., Zhou, G., and Zheng, L.: Rainfall and earthquake-induced landslide susceptibility assessment using GIS and Artificial Neural Network, *Nat. Hazards Earth Syst. Sci.*, 12, 2719–2729, doi:10.5194/nhess-12-2719-2012, 2012.
- Malamud, B., Reichenbach, P., Rossi, M., and Mihir, M.: Report on standards for landslide susceptibility modelling and terrain zonations, LAMPRE FP7 Project deliverables, available at: <http://www.lampre-project.eu> (last access: 3 November 2016), 2014.
- Marchesini, I., Ardizzone, F., Alvioli, M., Rossi, M., and Guzzetti, F.: Non-susceptible landslide areas in Italy and in the Mediterranean region, *Nat. Hazards Earth Syst. Sci.*, 14, 2215–2231, doi:10.5194/nhess-14-2215-2014, 2014.
- Mashimbye, Z., de Clerq, W., and Van Niekerk, A.: An evaluation of digital elevation models (DEMs) for delineating land components, *Geoderma*, 213, 312–319, doi:10.1016/j.geoderma.2013.08.023, 2014.
- Metz, M., Mitasova, H., and Harmon, R. S.: Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search, *Hydrol. Earth Syst. Sci.*, 15, 667–678, doi:10.5194/hess-15-667-2011, 2011.
- Neteler, M. and Mitasova, H.: *Open Source GIS: A GRASS GIS Approach*, Springer, New York, 2007.
- Nichols, G.: *Sedimentology and stratigraphy*, John Wiley & Sons, Chichester, West Sussex, UK, 2009.
- Pardeshi, S. D., Autade, S. E., and Pardeshi, S. S.: *Landslide hazard assessment: recent trends and techniques*, SpringerPlus, 2, 253, doi:10.1186/2193-1801-2-523, 2013.
- Regione Umbria: *Carte Tematiche*, available at: <http://www.umbriageo.regione.umbria.it/pagine/carte-tematiche/ambnat/amnat2.9.htm> (last access: 3 November 2016), 1995–2015.
- Rossi, M. and Reichenbach, P.: LAND-SE: a software for statistically based landslide susceptibility zonation, version 1.0, *Geosci. Model Dev.*, 9, 3533–3543, doi:10.5194/gmd-9-3533-2016, 2016.
- Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A., and Peruccacci, S.: Optimal landslide susceptibility zonation based on multiple forecasts, *Geomorphology*, 114, 129–142, doi:10.1016/j.geomorph.2009.06.020, 2010.
- Saito, H., Nakayama, D., and Matsuyama, H.: Preliminary study on mountain slope partitioning addressing the hierarchy of slope unit using DEMs with different spatial resolution, in: *Geomorphometry 2011*, 143–146, Redlands, USA, 2011.
- Santangelo, M., Marchesini, I., Bucci, F., Cardinali, M., Fiorucci, F., and Guzzetti, F.: An approach to reduce mapping errors in the production of landslide inventory maps, *Nat. Hazards Earth Syst. Sci.*, 15, 2111–2126, doi:10.5194/nhess-15-2111-2015, 2015.
- Schaetzl, R., Enander, E., Luehmann, M., Lusch, D., Fish, C., Bigsby, M., Steigmeyer, M., Guasco, J., Forgacs, C., and Pollyea, A.: Mapping the physiography of Michigan with GIS, *Phys. Geogr.*, 34, 2–39, 2013.
- Sharma, R. and Mehta, B.: Macro-zonation of landslide susceptibility in Garamaura-Swarghat-Gambhar section of national highway 21, Bilaspur District, Himachal Pradesh (India), *Nat. Hazards*, 60, 671–688, doi:10.4236/ijg.2012.31005, 2012.
- Zhao, M., Li, F., and Tang, G.: Optimal Scale Selection for DEM Based Slope Segmentation in the Loess Plateau, *Int. J. Geosci.*, 3, 37–43, doi:10.4236/ijg.2012.31005, 2012.