

## Topography-driven satellite imagery analysis for landslide mapping

M. Alvioli , A. C. Mondini, F. Fiorucci, M. Cardinali and I. Marchesini 

Consiglio Nazionale delle Ricerche, Istituto di Ricerca per la Protezione Idrogeologica, Perugia, Italy

### ABSTRACT

We describe a semi-automatic procedure for the classification of satellite imagery into landslide or no landslide categories, aimed at preparing event landslide inventory maps. The two-steps procedure requires knowledge of the occurrence of a landslide event, availability of a pre- and post- event pseudo-stereo pair and a digital elevation model. The first step consists in the evaluation of a discriminant function, applied to a combination of well-known change detection indices tuned on landslide spectral response. The second step is devoted to discriminant function classification, aimed at distinguishing the only landslide class, through an improvement of the usual 'thresholding' method. We devised a multi-threshold classification, in which thresholding is applied separately in small subsets of the scene. We show that using slope units as topographic-aware subsets produces best classification performance when compared to the ground truth of a landslide inventory prepared by visual interpretation. The method proved to be superior to the use of a single threshold and to any multi-threshold procedure based on topography-blind subdivisions of the scene, especially in the validation stage. We argue that the improved classification performance and limited training requirements represent a step forward towards an automatic, real-time landslide mapping from satellite imagery.

### ARTICLE HISTORY

Received 4 November 2017  
Accepted 17 March 2018

### KEYWORDS

Landslides; landslide event; landslide mapping; satellite imagery; discriminant function; slope units

## 1. Introduction

Landslides represent a serious hazard in many areas of the world, and particularly in tropical regions, where storms trigger every year thousands of them. The most effective source of information to document the landslide event extension and magnitude in a region is a landslide inventory map. Different types of landslide inventories exist, and they are the key input to derive landslide hazard and risk maps. Knowledge of the extent of landslide events is fundamental for risk management, preparedness and recovery actions. Landslides also represent one of the drivers of landscape evolution in time, whose study requires monitoring with fast and cost-efficient tools. Despite their importance, landslide inventory maps cover a limited extension of the landslide-prone areas across the global landmass (Guzzetti et al. 2012), and the completeness, accuracy and relevance of many existing inventories for landslide hazard studies are difficult to establish (Marchesini et al. 2014).

Landslide inventory maps are best prepared by visual interpretation of stereoscopic aerial images (Fiorucci et al. 2011). In the last two decades, the images captured by high-resolution (HR) and very high-resolution (VHR) optical satellites are becoming a viable replacement of aerial photographs, encouraging research efforts in the direction of developing semi-automatic and automatic

**CONTACT** M. Alvioli  [massimiliano.alvioli@irpi.cnr.it](mailto:massimiliano.alvioli@irpi.cnr.it)

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

classification algorithms to distinguish different land covers, including vegetation, urban areas, water bodies and landslides.

Guzzetti et al. (2012) compiled a review of the advantages and limitations of producing different kinds of landslide inventory maps using remote sensing data, as compared to conventional methods based on visual interpretation of stereoscopic images. They concluded that a combination of satellite, aerial and terrestrial remote sensing data represents the optimal solution for landslide detection and mapping, facilitating the definition and systematic application of standards and increasing the quality of derivative products of landslide maps.

Casagli et al. (2017) recently reviewed a few different technological options available for landslide mapping from both terrestrial and spaceborne remote sensing. They concluded that spaceborne optical and synthetic aperture radar (SAR) data have proved effective tools for post-disaster damage assessment, landslide detection and rapid mapping, landslide activity and updating of shallow rapid- and slow-moving landslides. Casagli et al. (2017) also stated that unmanned airborne vehicles (UAVs) provide ultra-high-resolution data and can be used at a slope-scale in selected test sites; the combined use of ground-based interferometric SAR, terrestrial laser scanner (TLS) and infrared thermography (IRT) ground-based methods was applied for the surveying, monitoring and characterization of different kinds of slope instabilities.

Additional works not included in the mentioned reviews investigated further steps forward to reach full automation of the mapping process. Moosavi et al. (2014) successfully applied the Taguchi method to find optimized parameters in object-oriented or support vector machine classification schemes. Mondini et al. (2017) proposed an automatic method to systematically produce inventories over selected catchments using synthetic generated (Monte Carlo) training samples. Yu and Chen (2017) used saliency enhancement of potential landslide signatures in Landsat 8 imagery and selective search for large-scale detection.

A few other recent attempts exist exploiting SAR data as well, using classic measures of phase changes but considering new missions (Sentinel-1) data (Barra et al. 2016), using measures of changes of amplitude spatial auto-correlation (Mondini 2017), contouring connection methods applied to LIDAR (Gaidzik et al. 2017), and change detection in aerial photographs and Marcov random fields (Li et al. 2016) for landslide detection and mapping in a systematic way. Plank et al. (2016) combined pre-event HR optical imagery and VHR PolSAR data to mitigate the systematic lack of pre-event polarimetric SAR data.

Automatic and semi-automatic landslide mapping requires image classification methods, including supervised and unsupervised clustering (Borghuis et al. 2007; Martha et al. 2011; Stumpf and Kerle 2011; Keyport et al. 2018), and index thresholding (Rosin and Hervás 2005). Supervised classification calls for a manual training process which can result tough and time-consuming. Reducing the overall effort required to prepare an event landslide inventory map (eLIM) and the time needed to complete the mapping procedure while increasing its level of automation, and key issues to obtain a reliable estimate of the extent and magnitude of a landslide event and, in turn, to quickly prepare response measures.

Event landslides usually show spectral fingerprints ascribable to a generic bare soil class (Mondini and Chang 2014). In this work, we focus on a supervised classification method which assigns individual pixels to user-defined classes. Existing Bayesian-based maximum likelihood (ML) approaches typically assign each pixel to a land cover class according to some decision rules applied to discriminant functions (Richards and Jia 2006) prepared for each land cover class defined in the area. The simplest decision rule is represented by thresholding, the procedure of defining a proper single numerical value among the values of an image (threshold) and assigning the pixels with values above (or below) the threshold to a particular class (Cheng et al. 2004).

The thresholding procedure, applied to the whole image, necessarily implies a compromise among different spectral responses of the same land cover in different geometric conditions, dictated by the combination of satellite point of view, sun position and slope orientation and inclination. We expect that using multiple thresholds, within many sub-areas, allows to overcome this limitation,

provided that overall geometric conditions are homogeneous within individual sub-areas. A topography-driven partition of the study area into small subsets is represented by slope units (SUs) (Carrara 1993; Guzzetti et al. 1999). SUs are particularly suited in the present context, since they encompass areas with similar slope-facing direction (aspect).

In this work, we generated SU using the automatic delineation software of Alvioli et al. (2016). Delineation of SUs can be performed emphasizing a particular morphometric quantity, thus it is not unique. We adopted a specific landslide mapping performance index (Carrara 1993; Fiorucci et al. 2018) to fine-tune the SU delineation, and compared the optimization procedure with the metric, used in the original work of Alvioli et al. (2016), which provides optimal segmentation of the aspect map into different spatial domains. Optimizing SU parameters with respect to aspect segmentation has solid grounds for the purpose of this work, since pixels located in regions homogeneously facing the same direction likely provide consistent spectral response in satellite imagery. As in Alvioli et al. (2016), we worked within the open source software GRASS GIS<sup>1</sup> (Neteler and Mitasova 2007) for all of the analyses presented in this work, if not otherwise specified.

The proposed method was tested in an area of about 1000 m<sup>2</sup> in Myanmar, where torrential rainfall triggered extensive landslides in 2015, which made the news due to the occurrence of the massive Tonzang landslide and the large number of fatalities (Brakenridge et al. 2017). Results of our semi-automatic mapping were calibrated and validated against a landslide inventory map prepared through photo-interpretation by expert geomorphologists.

The paper is organized as follows. Section 2 describes a test application of our procedure, including a description of the study area, of the available data and preparation of an eLIM by expert photo-interpretation. Section 3 describes in detail the method devised in this work for semi-automatic landslide mapping, including the evaluation of a suitable change detection discriminant function, slope units delineation, calibration of the classification procedure and metrics used to calibrate and validate the method. Many technicalities are not included in Section 3, and discussed in the Appendix sections for the interested reader and to allow full reproducibility of the procedure. Results are reported in Section 4 and extensively discussed in Section 5. Eventually, conclusions are drawn in Section 6.

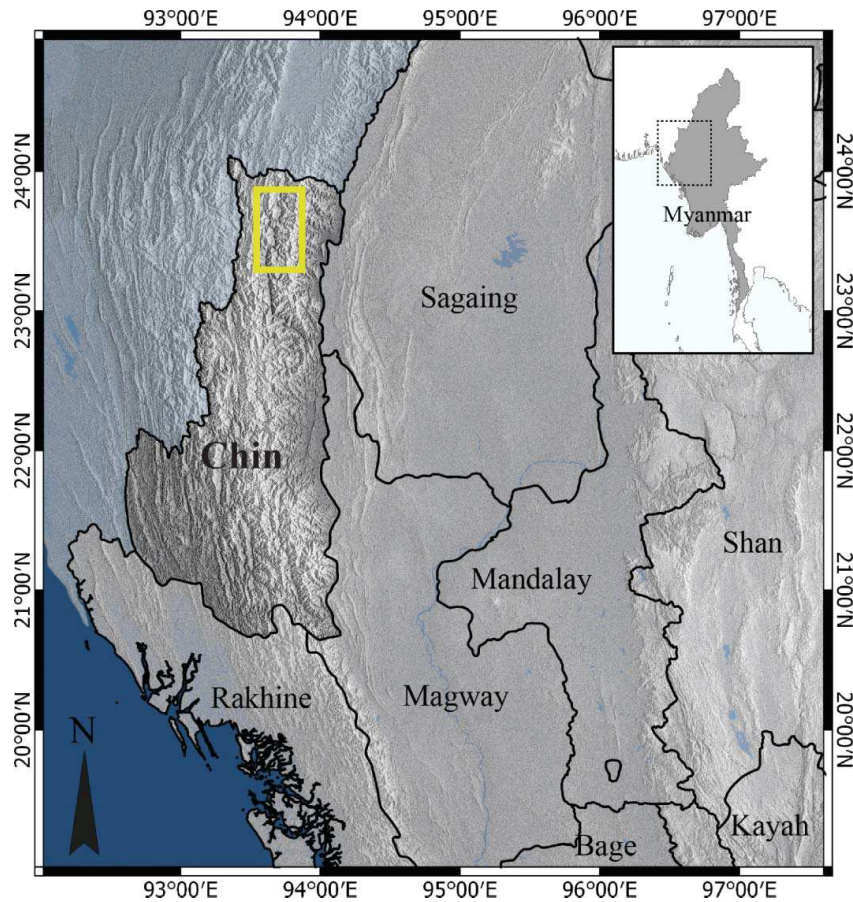
## 2. Test case

### 2.1. Study area

The study area corresponds to 1000 km<sup>2</sup> in the Chin State (western Myanmar) and its location is shown in Figure 1. Myanmar is exposed to a range of natural hazards, including floods, cyclones, earthquakes, tsunamis, and landslides. Natural hazards in Myanmar are accompanied by high economic costs and social consequences. The annual expected losses linked to natural hazards are approximately US\$184.8 million, equivalent to 0.9% of the country's 2008 gross domestic product.

According to the geological map produced and compiled by the Department of Geological Survey and Mineral Exploration of Myanmar, and obtained from the OnegeologyGlobal<sup>2</sup> web portal, a succession of sub-vertical layers of volcanoclastic materials (metasedimentary rocks) and sandstones of the Indo-Burman Ranges and the Central Myanmar Basin (Allen et al. 2008; Licht et al. 2013) crop out in the study area. Climate in Myanmar is tropical with three seasons: a monsoon/rainy season (May–October), a cool season (November–February), and a hot season (March–April). Rainfall during the monsoon season totals more than 500 cm/year in upper Myanmar and over 250 cm/year in lower Myanmar and Yangon, while Central Myanmar and Mandalay both receive about 76 cm/year.

During summer 2015, Chin State was affected by a major torrential rainfall event which triggered thousands of landslides. Torrential rain started on 16 July 2015, saturating the soil. On 30 July, cyclone Komen caused landslides in Bangladesh, due to strong winds and additional torrential rainfall in Chin and Rakhine States and Sagaing, Magway, and Bago Regions. In July and August 2015,



**Figure 1.** Geographical location of the Chin State, Western Myanmar. The yellow box shows the bounding box our study area, covering N23.874°–23.431° and E93.7°–93.948° (EPSG:4032).

widespread floods and landslides affecting 12 out of 14 states in Myanmar caused 132 fatalities and left 1,676,086 displaced people (Mondini 2017).

## 2.2. Available data and pre-processing

We obtained two pre- and two post-event RapidEye satellite images with 5 m resolution, in the framework of the Commons ESA project.<sup>3</sup> Images were available with a 3A processing level, aligned to an UTM/WGS84 map projection, which includes imagery orthorectification with a rigorous camera model and ground control points. The Global Reference 2.0 ground control data set used for the orthorectification allows for the production of orthorectified imagery with positional accuracy under 10 m root mean square error (RMSE) on a global scale.<sup>4</sup> We verified the quality of the relative co-registration among pre- and post-event images using 20 homologous points and we measured a RMSE of about 7 meters. We corrected for atmospheric effects using the Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH<sup>®</sup>) model available in ENVI<sup>®</sup>, which returns atmospherically corrected reflectance images.<sup>5</sup>

We used a portion of DEM with 1 arc second resolution from the ASTER<sup>6</sup> initiative, for slope units delineation and landslide inventory map preparation. Both operations need only be performed once in a given study area. Using a high-resolution DEM for slope unit delineation does not necessarily produce better results, especially from slope unit-derived quantities, as investigated by

Schlögel et al. (2018). All the remaining analyses presented in this paper were performed, and output map were produced, at 5 m resolution.

### 2.3. Event landslide inventory map

Two of us (F. Fiorucci and M. Cardinali) prepared an eLIM by visual interpretation of satellite images. The set of images consisted of three derivative types: (i) a pre-event and post-event false colour composite (NIR, R and G) image, (ii) a Normalized Difference Vegetation Index (NDVI) pre-event and post-event image, (iii) a pre-event and post-event pseudo-stereo pair obtained combining the ASTER DEM with the optical images (Chen and Rau 1993). The three-dimensional pseudo-stereo pair was visualized using dedicated hardware (Planar stereo mirror screen<sup>7</sup>) and software (Image Stereo Analyst). Comparison of pre- and post-event images allowed identifying landslides triggered by the considered event, mapped as vector polygons. The analysis of the distribution of patterns and tones, supported by morphological information provided by the pseudo-stereo pair, allowed identifying different landslide types, including flows, shallow and deep-seated slides. Multiple activations were identified within a few of the larger landslide bodies; all of them were related to the same event and therefore included in the landslide inventory, eventually used as ground truth for this work.

## 3. Method

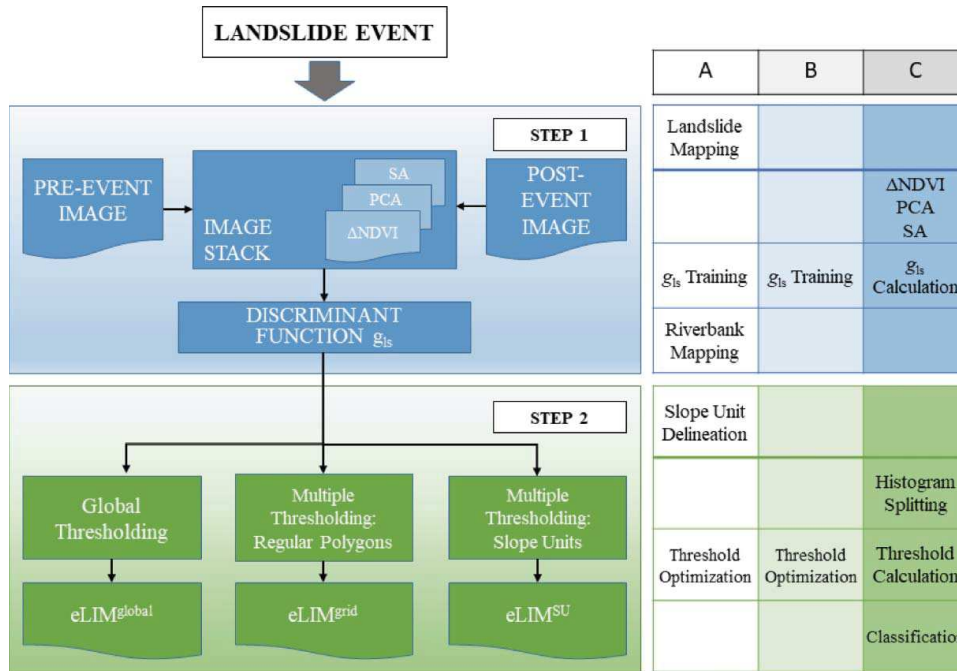
Our method to semi-automatically map landslides or, equivalently, to identify the pixels belonging to the landslide land cover class, relies on the concept of a change detection (CD) function. Here, we denoted the CD function as  $g_{ls}$  (where 'ls' stands for 'landslides'), obtained with a simplified ML classifier in a Bayesian approach. A common approach to binary classification is to identify a threshold value  $T$ : pixels in the  $g_{ls}$  map with values larger than  $T$  are classified as landslides, and no landslides otherwise. We propose to perform the classification procedure within many sub-areas, with a multi-threshold approach.

Figure 2 summarizes the method developed in this work, consisting in two main steps.

In the first step, we define the function  $g_{ls}$  whose values represent the ML distance of each pixel from the landslide class, providing a pixel-by-pixel measure of the presence of new landslides. The  $g_{ls}$  function is obtained measuring changes occurred between a pre- and a post-event image. We measured changes in the satellite images using three different metrics: changes of NDVI (Tucker 1979; Lee 2005), spectral angle (SA) (Sohn and Rebelló 2002; Richards and Jia 2006; Mondini et al. 2011b) and principal component analysis (PCA) (Richards and Jia 2006). The three metrics were combined in a single image stack for the analysis. We stress that the resulting discriminant function defined in this work does not represent the probability of landslide presence, in a mathematical sense.

In the second step, a map is generated by evaluating the  $g_{ls}$  function in each pixel of the study area. Then, the  $g_{ls}$  map pixels are classified as 'landslides' or 'no landslides', either by: (i) thresholding the  $g_{ls}$  values, i.e. selecting as landslides the pixels with  $g_{ls}$  values larger than a single threshold value over the whole study area; (ii) thresholding square and rectangular subsets of the  $g_{ls}$  map, using multiple threshold values; (iii) replacing regular subsets of (ii) with irregular SU polygons, thus introducing local geomorphological information.

The innovative feature, in the first step of Figure 2, is represented by the fact that we only aim at defining one land cover, the landslide class, and thus we only need to train the procedure in one class, while typical approaches to image classification aim at identifying many classes, and focussing on a single class allows easier calibration of the CD function. The calibration area was selected in only one (big) landslide, for a total of 421 pixels (about 10,000 m<sup>2</sup> out of about 1000 km<sup>2</sup>) in the stack of changes. The core innovation of the procedure, in the second step of Figure 2, is that we applied  $g_{ls}$  thresholding in a large number of subsets of the study area, singled out either with and



**Figure 2.** A flow chart of the algorithm proposed in this work. The algorithm can be applied upon knowledge of the occurrence of a landslide event. Step 1 describes the discriminant function calculation, while Step 2 represents the three different classification possibilities by index thresholding considered in this work, resulting in three different eLIMs (cf. Tables 1 and 2). The table on the right describes the level of automation of the individual operations involved in each of the two steps. Column A lists one-time, site-dependent operations; column B, operations that can be optionally performed again in a new study area; column C, fully automatic operations.

without a topographic information. Existing thresholding approaches use a single threshold, necessarily reducing accuracy, while SU provide local topography information and allows to find local custom thresholds.

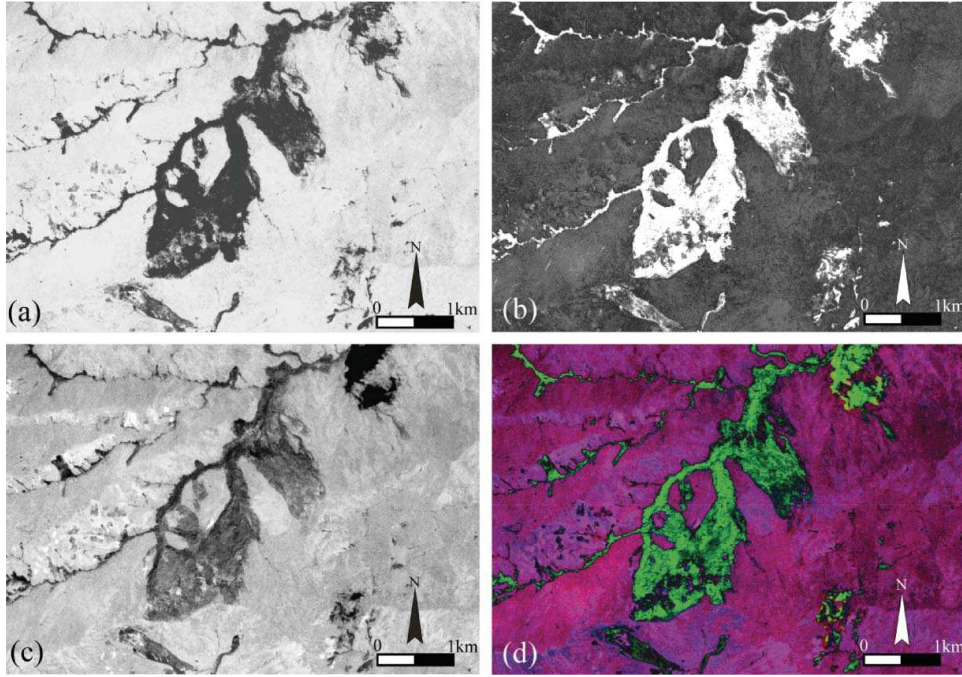
Note that, throughout the paper, we refer to training as the production of the  $g_{ls}$  function based on a certain number of pixels known to contain new landslides; calibration, instead, is the selection of the best binary landslide/no-landslide classification thresholds to be associated to the different subsets of the calibration area. Validation is the application of the overall procedure to a different, larger part of the study area.

### 3.1. Discriminant function definition and single threshold classification

In our approach,  $\Delta$ NDVI, SA and PCA are composed into a single stack, which constitutes our measure of changes.

NDVI is a well-known image differencing index useful to identify vegetated areas and their conditions (Tucker 1979). NDVI can assist in landslide identification, in particular for shallow landslides when they occur in vegetated areas (Lee 2005). Large negative values of  $\Delta$ NDVI are typically found in areas where forest is the predominant class and may signal loss of vegetation, possibly caused by new landslides (Mondini et al. 2011b).

SA measures the generalized angle between two spectral signatures representing two different surface covers (Sohn and Rebello 2002) or, when measured in the same pixel at two different times ‘post’ and ‘pre’, its temporal evolution. Spectral angles different from zero measure changes that may be not directly related to loss of vegetation (Mondini et al. 2011b).



**Figure 3.** Zoom of the three change detection metrics used in this work in a sample subset of the study area, calculated from the pre- and post-event RapidEye images. (a)  $\Delta\text{NDVI}$ ; (b) spectral angle, SA; (c) principal component analysis, PCA; (d) the RGB composite image with R =  $\Delta\text{NDVI}$ , G = SA and B = PCA. (d) is the stack of changes, which we used to obtain the numerical values of  $\mu_{\beta=ls}$  and  $\Sigma_{\beta=ls}$  needed to calibrate the  $g_{ls}$  CD function.

PCA is a linear transformation of a number of potentially correlated variables into a number of uncorrelated variables in a different orthogonal system (Richards and Jia 2006) with the axes oriented along the directions of the largest possible variances.

When ML is applied, it assigns the class membership of each pixel in the stack of changes as follows (Richards and Jia 2006):

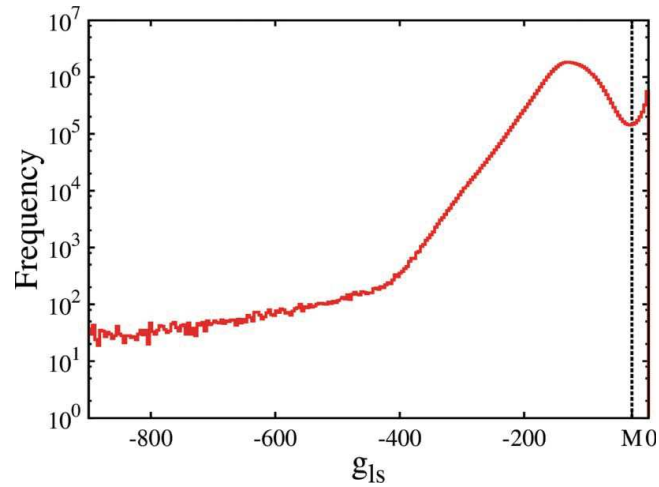
$$\{\vec{x} \in \omega_{\beta} \mid g_{\beta}(\vec{x}) > g_{\gamma}(\vec{x}), \forall \gamma \neq \beta\}, \quad (1)$$

where  $g_{\beta}$  is the discriminant function associated to the  $\omega_{\beta}$  land cover class changes,  $\beta$  and  $\gamma$  label all the land cover classes, and  $\vec{x}$  the value of the changes measured in the pixel position. The function  $g_{\beta}(\vec{x})$  is defined as follows:

$$g_{\beta}(\vec{x}) = \ln(p(\omega_{\beta})) - \frac{1}{2} \ln|\Sigma_{\beta}| - \frac{1}{2} (\vec{x} - \vec{\mu}_{\beta})^t \Sigma_{\beta}^{-1} (\vec{x} - \vec{\mu}_{\beta}), \quad (2)$$

where  $\mu_{\beta}$  and  $\Sigma_{\beta}$  are, respectively, the multidimensional mean and covariance matrix of a multinormal Gaussian distribution estimated modelling the statistical behaviour of samples of pixels selected in  $\beta$  land cover changes present in the scene. In this work, (a)  $\vec{\mu}_{\beta=ls}$  and  $\Sigma_{\beta=ls}$  (where 'ls' stands for 'landslides', as in the  $g_{ls}$  definition) were obtained selecting samples areas representative of landslides occurred during the event, and (b) only the discriminant function related to landslides was estimated ( $g_{\beta=ls}$ ).

The three change detection metrics are illustrated in Figure 3. The Figure shows the three individual quantities (Figure 3(a),  $\Delta\text{NDVI}$ ; Figure 3(b), SA and Figure 3(c), PCA) and the combined stack of changes. One can see that  $\Delta\text{NDVI}$  and SA are partially correlated, though they still contain



**Figure 4.** Histogram of the  $g_{ls}$  function values over the whole study area. The vertical line situated at  $M = -26.725$  represents the divide between the two existing modes. The mode located right from the divide is due to pixels with  $g_{ls}$  values close to zero, i.e. with spectral behaviour very similar to pixels known to be within the landslides selected for the training procedure.

different information. PCA instead is rather different, mainly due to the PCA being obtained by global quantities, while the others are calculated locally. In conclusion, the three quantities provide different information as, within the same landslide body, the combination of them (Figure 3(d)) does not show a uniform pattern.

The training stage considered a total of 421 landslide pixels in the stack of changes. To check that the selected pixels were multinormal Gaussian distributed (as requested by the ML function), we performed a Mardia's test, which showed a degree of confidence on the null hypothesis (normality) higher than 90%.

Figure 4 shows a histogram of the values for the CD discriminant function  $g_{ls}$  introduced in this work (Equation (2)). The histogram of  $g_{ls}$  values in Figure 4 reveals a rather broad distribution of negative values, ranging from about  $-1200$  (the figure does not show the whole range) to  $0$ , but mostly bounded in the  $(-400, 0)$  region. Landslides correspond, by construction, to values close to  $0$ , i.e. small ML distance from landslide response, meaning that those pixels have spectral properties close to the landslide pixels selected for training  $g_{ls}$ .

A distinctive feature of the histogram in Figure 4 is a bi-modal behaviour, characterized by a small peak around  $g_{ls} = 0$ , overwhelmed by a broad peak centred at  $g_{ls} = -150$  and containing the vast majority of pixel values in the  $g_{ls}$  map with spectral properties dissimilar from the landslide ones. The two peaks (modes) are separated by a well-defined local minimum, occurring at some  $g_{ls}$  value denoted in the following as  $M$ . The first approximation to a binary classification of the  $g_{ls}$  values is to flag as 'landslide' the pixels with  $M < g_{ls} < 0$ , and to flag as 'no-landslide' the remaining pixels.

It is straightforward that a sharp cut on the  $g_{ls}$  values in correspondence to  $M$  (Figure 4) introduces false negatives, i.e. pixels that are incorrectly flagged as free of landslides resulting from cutting the left tail of the landslide-related peak in the distribution, as well as false positives, i.e. pixels that are incorrectly flagged as landslides resulting from the right tail of the broad peak of the distribution. There is no straightforward way to overcome the misclassification from the sole analysis of the distribution of Figure 4.

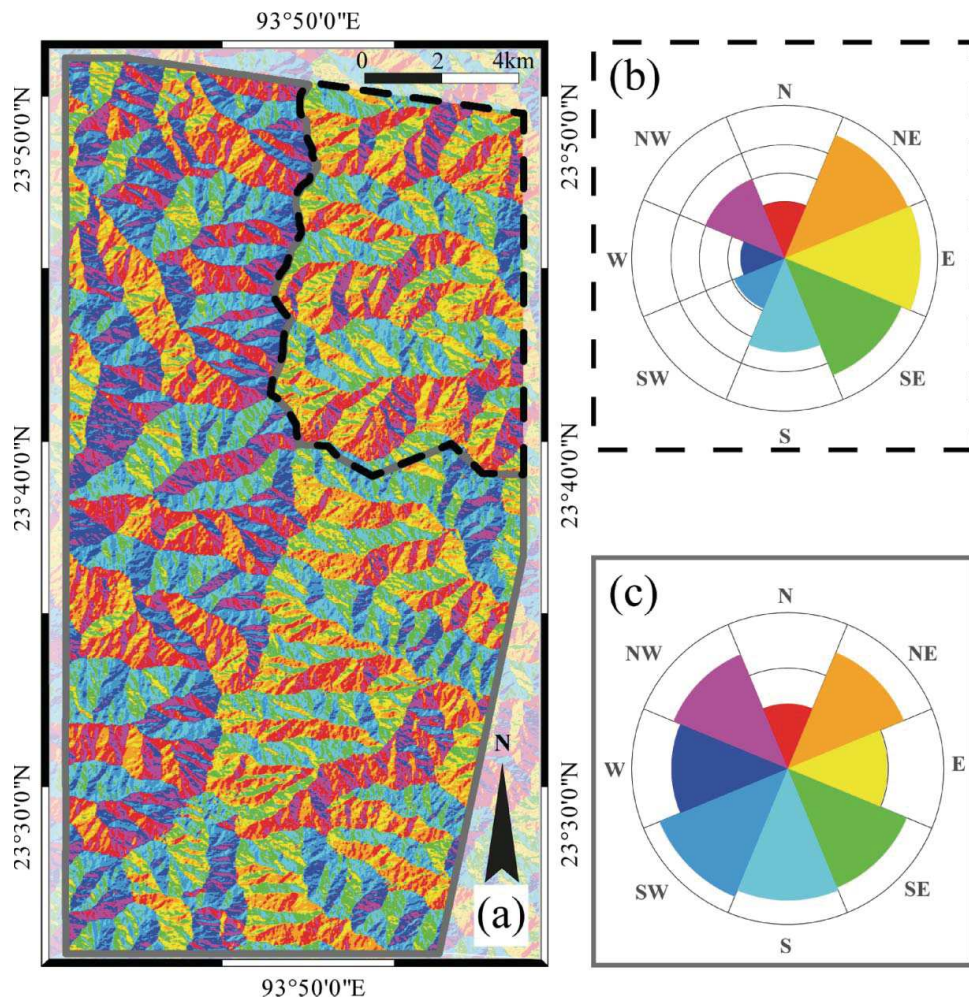
### 3.2. Automatic mode detection within regular and topography-driven subsets

A strategy to minimize misclassification imposed by the use of a single threshold consists in splitting the study area into a variable number of polygonal subsets, with variable size and shape, and repeat, within each subset, the  $g_{ls}$  mode analysis operated globally on the histogram of Figure 4, described



in Section 3.1. In each polygon, we can investigate the histogram of  $g_{ls}$  values and, in principle, single out a custom threshold for binary classification. We considered different subsets, namely a large number of (i) regular (square or rectangular) polygons resulting from a grid with variable number of rows and columns, (ii) SU polygons with variable size and shape.

SUs are morphological terrain units, bounded by drainage and divide lines (Carrara 1993; Guzzetti et al. 1999) delineated in such a way that terrain homogeneity is maximized within the units, and inhomogeneity is maximized across neighbouring units. We obtained SUs for our study area using the *r.slopeunits* specialized software (Alvioli et al. 2016). The software is adaptive, in that SUs are delineated with varying sizes and shapes in different regions of the study area. Optimized SU can be obtained by selecting values of the software's input parameters that maximize fitness of the output SU set for a particular purpose. We investigated the use of two metrics for optimization of the software's parameters: (i) an aspect segmentation metric first introduced by Espindola et al. (2006) for the segmentation of generic digital images and adapted by Alvioli et al. (2016) to work with the aspect circular variance; (ii) the error index  $E_l$  first introduced by Carrara (1993), and described later on in Section 3.4. To illustrate the aspect distribution across the study area, Figure 5 (a) shows an aspect map of the study area, along with rose diagrams of the distribution of aspect



**Figure 5.** (a) Aspect map of the study area, with the subdivision into calibration (black dashed contour) and validation (grey contour) areas. (b) Rose diagram of the aspect distribution within the calibration area. (c) As in (b), for the validation area.

values within the calibration (Figure 5(b)) and the validation (Figure 5(c)) areas. Figure 6 shows the final SU subdivision of the study area, obtained optimizing the error index  $E_f$ . This point will be further discussed in the following.

We stress here that SUs must be obtained once and for all, making the mapping procedure readily applicable to any future landslide event with comparable computational cost (running time) with respect to applying a single threshold to the global histogram. The multi-threshold calculation of the final map on the whole study area requires about two hours on a 64-core machine, thanks to the possibility of developing scripts in GRASS GIS, which can be efficiently run in parallel.

We calculated histograms of the values of the discriminant function  $g_{ls}$  (Equation (2)), within either regular (square or rectangular) polygons obtained from a grid and within irregular SU polygons. Then, histograms were processed using the software of Delon et al. (2007). The software contains an automatic, non-parametric algorithm for one-dimensional histogram segmentation without a-priori assumptions about the number or shape of the histogram modes. The method tests the simplest multimodal law that fits the data coupled with a test which presents the advantage of being simultaneously local and global over the histogram range. The histogram is first split into monotone chunks, then the algorithm makes use of the so-called meaningful rejection for a decreasing hypothesis, leading to a piece-wise multimodal hypothesis. The automatic histogram-splitting algorithm of Delon et al. (2007) was used to determine, for each polygon, the number of modes (number of maxima)  $N_{mod}$  in the corresponding  $g_{ls}$  distribution and the numerical values of the separations (positions of minima) between modes,  $g_{s}^{(i)}$ ,  $i = 2, \dots, N_{mod}$ , if any.

The information about the number of modes and the positions of minima between the peaks, alone, is actually not enough to define individual thresholds for each polygon. The possible situations are qualitatively illustrated in Figure 7, for  $N_{mod} = 1, 2, 3$ . The figure shows that for each given number of modes (each row in the Figure), the histogram may (type HA histograms) or may not (type HB histograms) present a mode peaked about  $g_{ls} = 0$ , which likely corresponds to landslide presence, by  $g_{ls}$  construction. As a matter of fact, knowing the number and positions of the separations only enable us to know if the histogram belongs to the first, second or third row, but does not allow to know if they are of kind HA or HB in Figure 7, so that further analysis is needed.

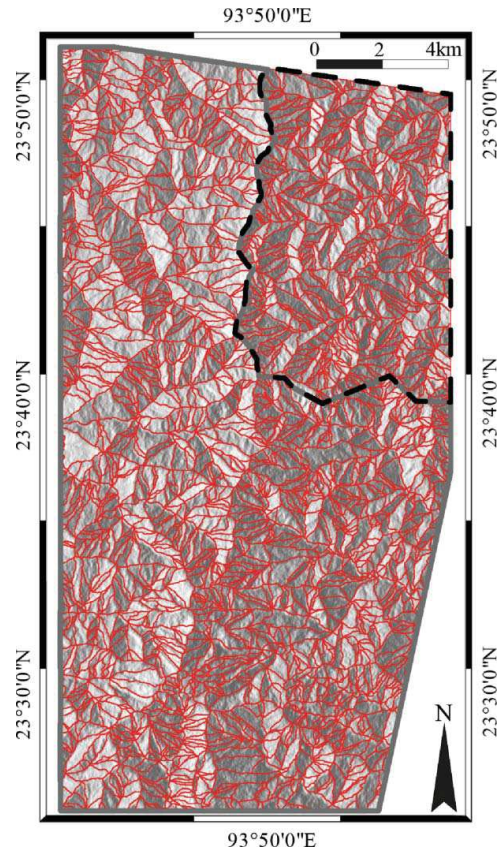
### 3.3. Calibration of the classification thresholds

In our study area, histograms built for both regular or SU subsets of the  $g_{ls}$  map were found to present one, two or three distinct modes. Our goal is to find custom thresholds  $T$  for each polygon or, more realistically, for each different kind of the corresponding histograms on the basis of the different situations sketched in Figure 7. We adopted a different strategy for  $N_{mod} = 1$  (first row-like histograms in Figure 7),  $N_{mod} = 2$  (second row-like) or  $N_{mod} = 3$  (third row-like).

For the purpose, we further characterized the histograms considering the average values of mode separations. We denoted as  $m^{(2)}$  the average value of separations in the  $N_{mod} = 2$  cases, and with  $m^{(3)_a}$  ( $m^{(3)_b}$ ), the average value of the leftmost (rightmost) separations in the  $N_{mod} = 3$  cases. We established if  $N_{mod} = 2$  cases are more likely to fall in the  $HA_2$  or  $HB_2$  classes by analyzing  $m^{(2)}$  in relation to  $m^{(3)_a}$  and  $m^{(3)_b}$ . This is easily done by assuming Gaussian distributions of the  $g_{ls}$  values of the leftmost (labelled with  $a$  in the following) and rightmost (labelled with  $b$ ) separations in the set of  $N_{mod} = 3$  cases, and proceeding as illustrated in Appendix 1. We end up classifying  $N_{mod} = 2$  cases either as left-like (type  $HA_2$ ) or right-like (type  $HB_2$ ).

In a subset area (calibration area shown in Figure 5(a)), we collected the following information to characterize the different polygons:

- (i) the number of intervals  $N_{mod}$  found by the automatic histogram-splitting algorithm polygon-specific quantity;
- (ii) the values of the separations  $g_{s}^{(i)}$ ,  $i = 2, \dots, N_{mod}$  between the different peaks in the associated histogram (polygon-specific);



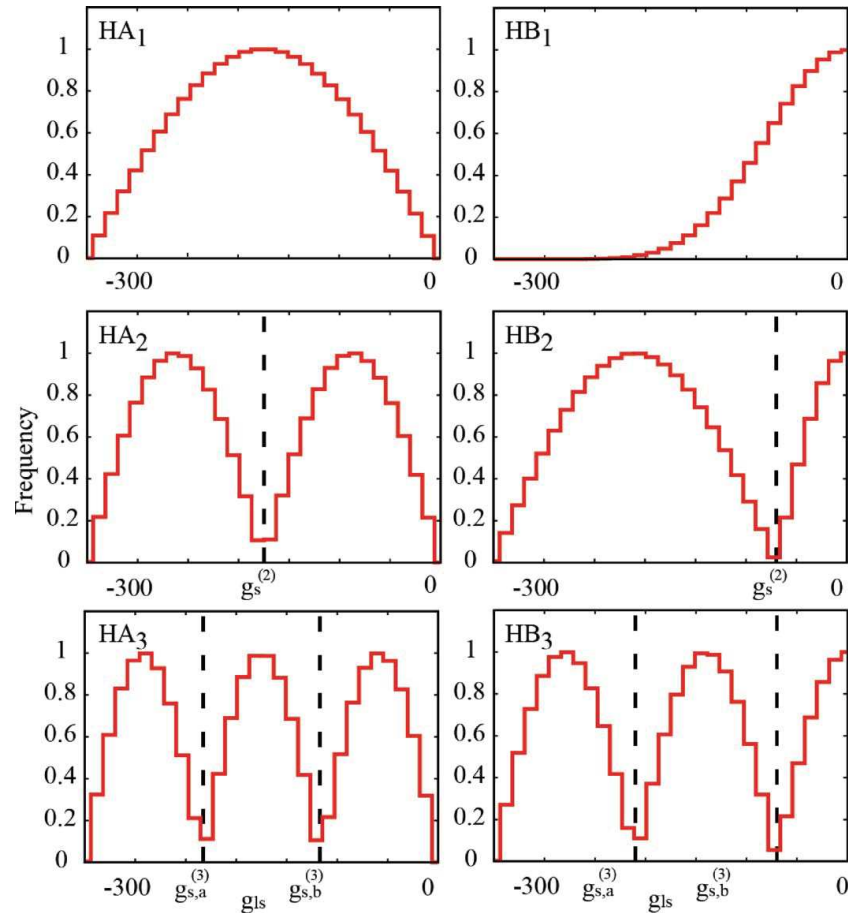
**Figure 6.** The optimal slope units partition of the study area, corresponding to the values  $(a_E, c_E) = (200,000 \text{ m}^2, 0.15)$  of the *areamin* and *cvmin* input parameters of the software *r.slopeunits* introduced by Alvioli et al. (2016). The background image is a shaded relief of the study area.

- (iii) the average values  $m^{(2)}$ ,  $m^{(3)}_a$  and  $m^{(3)}_b$  of the separations in each  $N_{mod}$  class (not polygon-specific);
- (iv) the value  $M$  of the separation of the two peaks in the global  $g_{ls}$  histogram (obviously not polygon-specific);
- (v) distinction of  $N_{mod} = 2$  cases in left-like or right-like (not polygon-specific).

We stress here that the object of the calibration procedure is the method we devised to obtain the best result, i.e. how to use the information contained in (i)–(v). Specific numerical values of the quantities (i)–(iv), instead, are directly calculated from the histograms, they depend on the study area, and they are different between the calibration and validation area (also shown in Figure 5(a)).

Different  $T$  values within the different classes of polygons should be introduced in such a way that misclassification is minimized polygon-wise instead of on the global map. Following a trial-and-error procedure, we found the best results for the  $g_{ls}$  classification threshold  $T$  for different polygon classes, as compared to the ground truth of expert mapping within the calibration area. The details of the procedure are rather cumbersome, and are described in Appendix 2 for the interested reader and to allow full reproducibility of the procedure.

The best results were found adopting polygon-specific thresholds in the cases with  $N_{mod} = 3$  (one threshold value for each polygon  $HA_3$  and  $HB_3$  in Figure 7), and common thresholds for the following classes of polygons: (a) one threshold value for all the  $N_{mod} = 1$  polygons ( $HA_1$  and  $HB_1$  in Figure 7); (b) one threshold value for all the  $N_{mod} = 2$ , left-like polygons ( $HA_2$  in Figure 7); (c) one



**Figure 7.** Qualitative sketches of different histogram shapes with single mode (first row, HA<sub>1</sub> and HB<sub>1</sub>), two modes (second row, HA<sub>2</sub> and HB<sub>2</sub>) and three modes (third row, HA<sub>3</sub> and HB<sub>3</sub>). Landslides correspond to  $g_{ls}$  values close to zero (cf. Figure 4) and are likely present in histograms shapes in the right column. Vertical black dashes depict the separations between different modes, to be calculated by the automatic histogram-splitting algorithm of Delon et al. (2007). The positions of the modes are for illustration purposes and are not the actual ones found during the  $g_{ls}$  image processing.

threshold value for all the  $N_{mod} = 2$ , right-like polygons (HB<sub>2</sub> in Figure 7). The actual numerical values of the thresholds are given by the polygon-specific value  $g_{ls}^{(s)}_{b,3}$ , in the  $N_{mod} = 3$  cases, and can be calculated from the quantities listed in (i)–(v) for the  $N_{mod} = 1, 2$  cases as specified in Appendix 2.

### 3.4. Evaluation of classification performance

To compare the outcome of our binary classification procedure with the expert-mapped landslides, we used the  $E_I$  error index proposed by Carrara (1993) and defined as follows:

$$E_I = \frac{A_U - A_{\cap}}{A_U}, \quad (3)$$

where  $A_U$  is the area of the region where either the automatically classified and the expert-mapped landslides exist (union), while  $A_{\cap}$  is the area of the region where both exist (intersection). We also discuss the results in terms of confusion matrices, in particular false, either positive (FP) or negatives (FN), assignments. Calibration was performed in the area shown in Figure 5(a), and validation in

the remaining part of the area, also shown in Figure 5(a), after the calibration area was excluded, where landslides were mapped by visual interpretation as well.

In order not to misclassify riverbanks, which present spectral behaviour similar to landslides, pixels belonging to such features were removed from the comparison. River features were also mapped by expert photo-interpreters, since this was the most straightforward thing to do in order to show the relevant features of our new method. In principle, riverbanks may change across different landslide events and need be delineated again. At the present stage, this is not included in our procedure, which makes it not fully automatic. Nevertheless, examples exist in the literature of automatic and relatively simple ways of recognizing rivers based, for example, on the correlation with strong changes from the pre- to the post-event images and values of slope below some small threshold. Including such functionalities will make the overall procedure fully automatic, once the one-time operations (training of the change detection function, slope unit delineation and calibration of the classification thresholds) are performed.

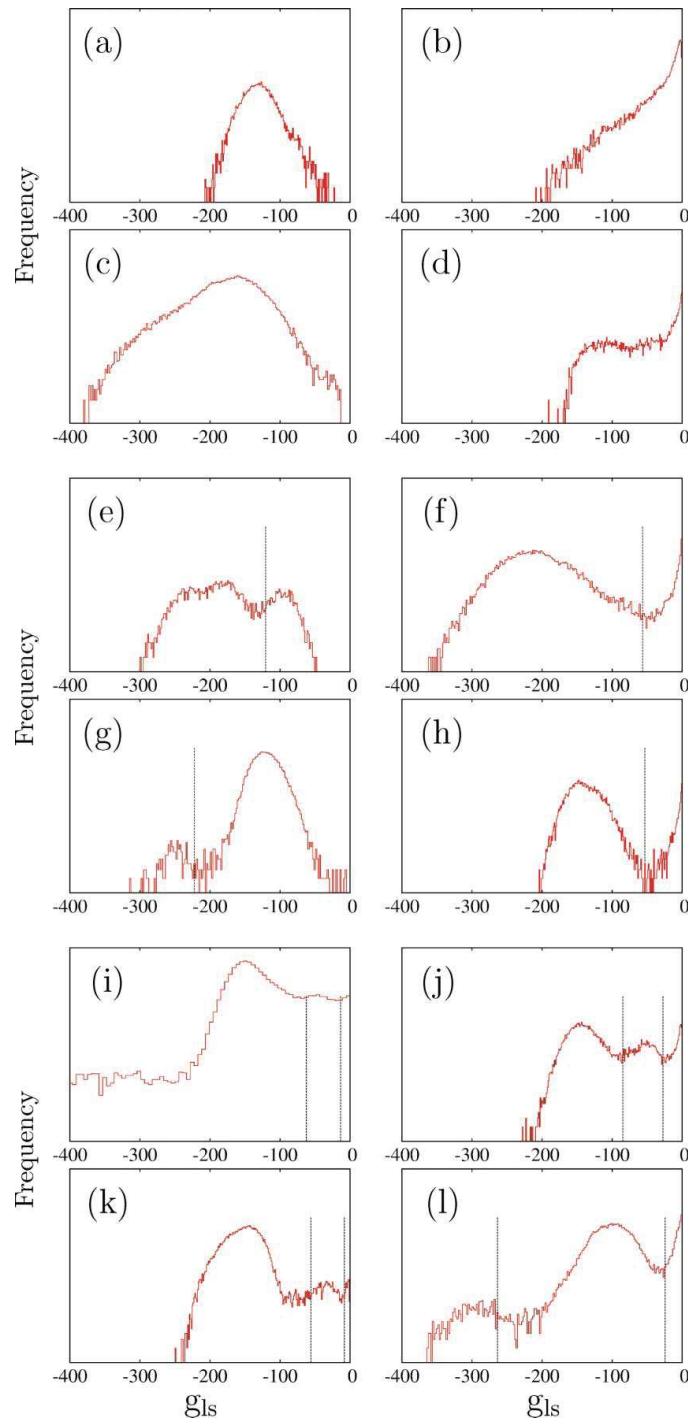
#### 4. Results

We proposed a procedure to go beyond the single-threshold classification of a CD discriminant function. Defining a large number of subsets of the study area, either using topography-blind rectangular regions or using topography-aware slope units, we adopted a multiple-threshold classification following the prescriptions outlined in Section 3.

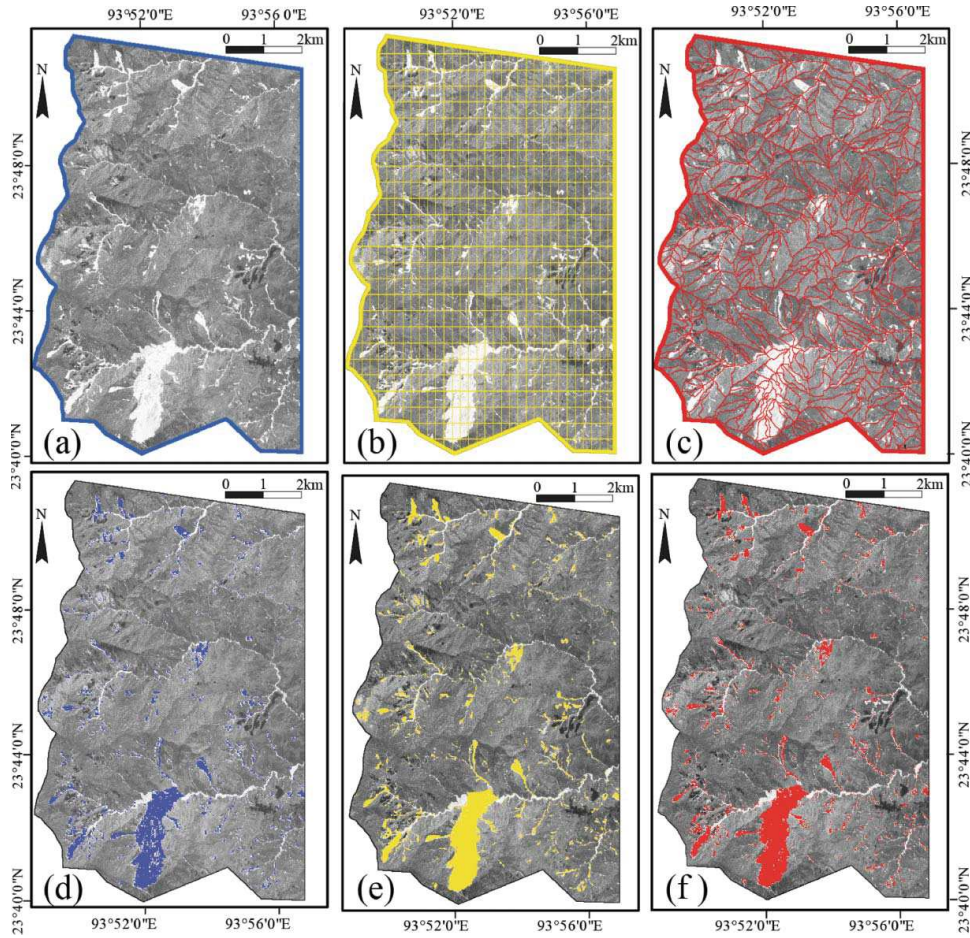
The multi-threshold classification was automated by a histogram-splitting software (Delon et al. 2007). Figure 8 shows a few examples of histogram shapes obtained in a few selected cases, within SU polygons, in the validation area. We selected the examples in order to illustrate the actual shapes of the histograms that we previously sketched in a qualitative way in Figure 7. We acknowledge that, among the several histograms obtained within our procedure (one for each SU polygon; cf. Figure 6), there were many shapes that did not perfectly fit into the shape classification of Figure 7, even if the automatic algorithm of Delon et al. (2007) classified them as belonging to one of those classes. This is the trade-off for using a high level of automation. Moreover, the relative abundance of the histograms of the six types defined in Figure 7 is not at all balanced. In the calibration area, we used an SU set containing 564 polygons, with minimum area 50,550 m<sup>2</sup>, maximum area 2,082,150 m<sup>2</sup>, average area 440,000 m<sup>2</sup> with standard deviation 350,000 m<sup>2</sup>. Out of 564 polygons, 292 were classified as  $N_{mod} = 1$ , 258 as  $N_{mod} = 2$  and 14 as  $N_{mod} = 3$ . In the validation area, we used an SU set containing 1755 polygons, with minimum area 50,400 m<sup>2</sup>, maximum area 3,809,600 m<sup>2</sup>, average area 507,000 m<sup>2</sup> with standard deviation 450,000 m<sup>2</sup>. Out of 1755 polygons, 1132 were classified as  $N_{mod} = 1$ , 604 as  $N_{mod} = 2$  and 19 as  $N_{mod} = 3$ . In both calibration and validation, the  $N_{mod} = 1$  histograms were of type A1 (cf. Figure 7) for the vast majority, and  $N_{mod} = 3$  histograms were a very small number as compared to the total number of polygons. These conditions were addressed by the proposed algorithm, introduced in Section 3.3 and further detailed in Appendix 2. As a matter of fact, we defined the final combination of thresholds through a trial-and-error procedure, so we conclude that the resulting algorithm effectively accounts for the relative abundances of histograms presenting different  $N_{mod}$  values.

We show results of the calibration stage (Section 3.3) for the following approximations: (i) a global threshold  $T = M$  is used (cf. Figure 4), labelled by 'global' ( $T_{global}$ ), resulting in the map shown in Figure 9(d); (ii) the multiple thresholds defined above are used, within polygons from a regular grid, labelled by 'grid' ( $T_{grid}$ ) (Figure 9(e)); (iii) multiple thresholds are used, within polygons from the optimal SU partition, labelled by 'SU' ( $T_{SU}$ ) (Figure 9(f)); (iv) individual thresholds for each SU are used, heuristically optimized in order to obtain the best agreement (minimum value of  $E_I$ ) between classification and expert mapping SU-wise; labelled by 'optimal' ( $T_{opt}$ ).

Result (iv) in Table 1 ( $T_{opt}$ ) is given for reference and it is not our final result, due to the optimization being performed in each SU of the calibration subset of the study area. Thus, it is impossible to generalize result (iv) to the validation area: ground truth is available to us in the whole study area,



**Figure 8.** A few histograms of the  $g_{ls}$  values, corresponding to 12 individual slope units. We selected the example histograms to show the actual shapes that were qualitatively sketched in Figure 7. In particular, (a) and (c) correspond to sketch  $HA_1$  in Figure 7; (b) and (d) to  $HB_1$ ; (e) and (g) to  $HA_2$ ; (f) and (h) to  $HB_2$ ; (i) and (k) to  $HA_3$ ; (j) and (l) to  $HB_3$ . We drew black vertical dashed lines at  $g_{ls}$  values corresponding to the histogram mode separations  $g_{ls}^{\theta}$ ,  $i = 2, 3$ , as calculated by the automatic procedure of Delon et al. (2007) used in this work.



**Figure 9.** The different approximations to the binary classification of the discriminant function map. (a) The  $g_i$  map, providing the  $T_{global}$  result in Table 1 with the single-threshold classification (Figure 4); (b) the set of rectangular polygons providing the classification result  $T_{grid}$ ; (c) the set of SU providing the best result,  $T_{SU}$ ; (d)–(f) show the eLIMs corresponding to the ‘global’, ‘grid’, and ‘SU’ approximations, respectively: pixels corresponding to automatically mapped landslides are shown in colour.

but we used it to build the method itself in the sole calibration area and to validate in the remaining part of the study area (cf. Figure 5(a)). We assume that the combination of the best values  $E_I$  in each SU also provides the best theoretical  $E_I$  for the whole map, which we believe to be a very good approximation to the real overall best possible classification using a custom binary classification threshold for each of the different polygons.

The numerical values of the confusion matrix indices and the  $E_I$  index of Equation (3), obtained from the comparison of the expert mapping from orthophotos and automatic mapping with the procedure introduced in this work, are listed in Table 1 for the calibration stage, and in Table 2 for the validation stage.

The calibration stage (Table 1) provides a performance gain  $(E_I^{grid} - E_I^{global})/E_I^{global} = 6.7\%$  when using regular (square) polygons, with respect to single threshold, and  $(E_I^{SU} - E_I^{global})/E_I^{global} = 8.1\%$  when using SUs; the ideal, ‘optimal’ result would provide  $(E_I^{opt} - E_I^{global})/E_I^{global} = 23\%$  gain.

The relative small gain observed in the calibration area when going from grid polygons to SUs increases when looking at validation stage results (Table 2): regular polygons provide a performance gain of 0.4% (with square polygons of the same size as in the calibration stage), while we gained 4.8% by using SUs.

**Table 1.** Comparison of the agreement of the result of our automated procedure with an expert-mapping procedure in the calibration area. TP (TN) = true positive (negative); FP (FN) = false positive (negative);  $E_I$  is the error index of Equation (3) introduced in Carrara (1993) for the comparison of two inventory maps.

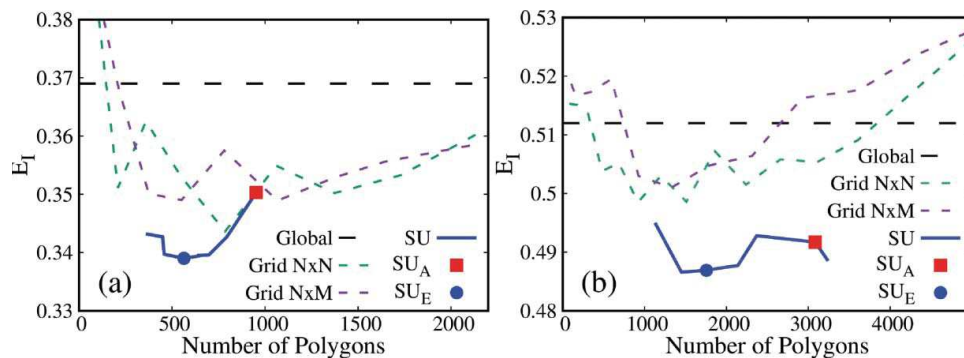
	$T_{global}$	$T_{grid}$	$T_{SU}$	$T_{opt}$
TN	92.55%	92.39%	92.27%	92.30%
FN	1.78%	1.49%	1.39%	0.97%
FP	0.97%	1.13%	1.22%	1.22%
TP	4.70%	4.99%	5.09%	5.51%
$E_I$	0.369	0.344	0.339	0.284

**Table 2.** As in Table 1, but for the validation area. In this case, the results corresponding to the optimal threshold  $T_{opt}$  are not available, by construction.

	$T_{global}$	$T_{grid}$	$T_{SU}$
TN	97.03%	96.82%	96.98%
FN	0.66%	0.71%	0.72%
FP	0.70%	0.91%	0.75%
TP	1.45%	1.56%	1.54%
$E_I$	0.512	0.510	0.487

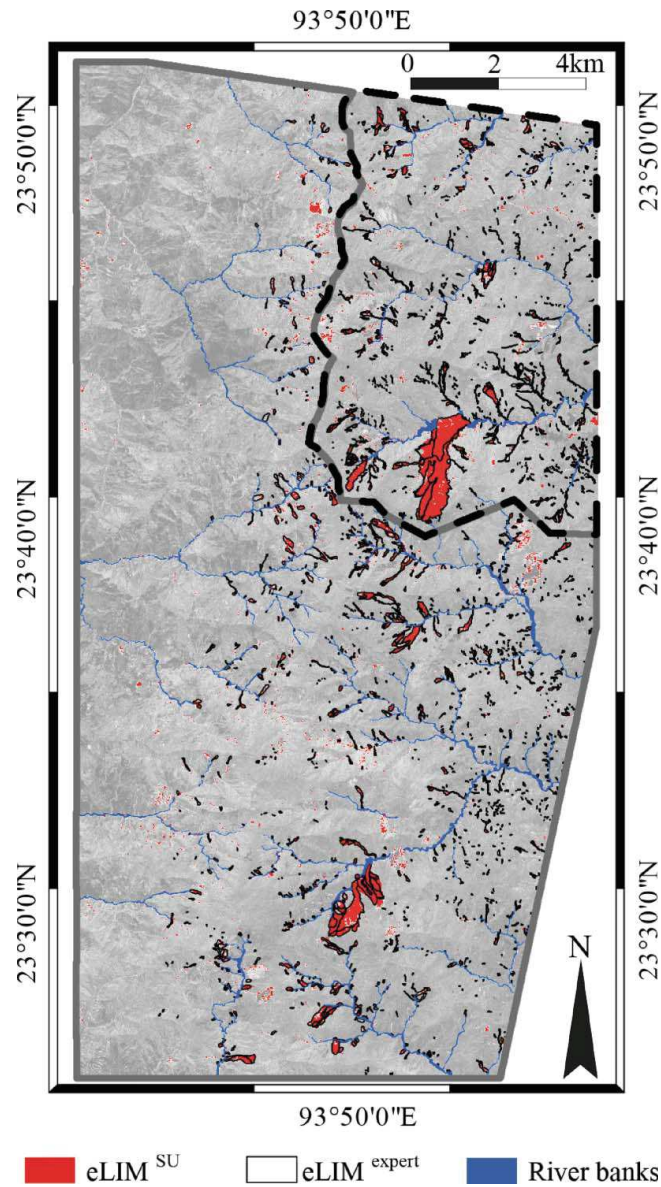
The dependence of the results upon the number of polygons used in the analysis, either square, rectangular or irregular SU polygons, is shown in Figure 10. The figure compares the constant  $T_{global}$  with  $T_{grid}$ , separately for the square and rectangular cases, and with  $T_{SU}$ , both in the calibration (Figure 10(a)) and the validation (Figure 10(b)) areas. For the latter, we highlighted in the figure the two results corresponding to the best SU set with respect to aspect segmentation (Alvioli et al. 2016) and to  $E_I$  performance (cf. Equation (3)). We see that, in the validation case (Figure 10(b)), the SU results are always better than the regular grid results, irrelevant of the number of polygons in the different polygon sets.

Tables 1 and 2 list, along with the values of  $E_I$  in the different approximations investigated in this work, the values of the FP, FN, TP, and TN indices, for completeness. We maintain that  $E_I$  is an easier overall agreement measure, since it is a single index specifically developed for the comparison of different mapping efforts of the same landslides bodies. Fiorucci et al. (2018) recently provided a detailed application of the index for the analysis of remote sensing imagery for landslide mapping.



**Figure 10.** Classification performance index  $E_I$  (Equation (3)) results for the different approximations used in this work to automatically classify landslides versus anything else. NxM represents the results obtained with rectangular polygons grids; MxM, with square polygons; SU, with different SU sets; SU<sub>E</sub> with the optimal SU set, obtained by minimizing  $E_I$  in the calibration area; SU<sub>A</sub>, with the optimized SU set with respect to aspect segmentation (Alvioli et al. 2016) in the whole study area. (a) Calibration area; (b) validation area.





**Figure 11.** Comparisons of eLIMs obtained by photo-interpretation (black polygons) and by the automatic mapping procedure developed in this work (red pixels). Blue polygons represent riverbanks, excluded by the analysis. The map clearly shows many false positives, predicting landslides where they did not actually occur. False negatives, instead, typically fall within detected landslides, with a few missing pixels. The background image is the  $g_{ls}$  discriminant function.

Figure 11 compares the ground truth of eLIM<sup>expert</sup>, mapped by expert photo-interpreters, to the best automatically mapped eLIM<sup>SU</sup>. The map shows both the calibration and validation portions of the study area.

## 5. Discussion

Systematic production of landslide inventories requires a high level of automation. In this work, we addressed the issue of reducing the time to train a supervised classifier, exploiting the possibility of

obtaining an inventory map as a simplified land cover map containing only two classes: 'landslide' and 'no landslide' (i.e. 'everything else'). Such an approach involves: (a) estimating a single discriminant function for the only landslide land cover class; (b) minimal training information to prepare a statistical model for landslide spectral properties; (c) automated binary classification, after proper calibration. All the (a)–(c) features go in the direction of reducing the overall effort to obtain the inventory map.

In this work, we used a maximum likelihood discriminant function, applied to a combination of change detection indices (step 1 in Figure 2). When a discriminant function is calculated, one is left with assigning the pixels either to the 'landslide' or 'no landslide' classes (step 2 in Figure 2). This can be done by a thresholding procedure, which we accomplished in three different ways, showing that a multi-threshold procedure using automatic histogram segmentation in conjunction with topographic-aware subdivision of the study area into many sub-areas provides the best result.

The proposed method presents several advantages. In first place, the part strictly related to image preparation is substantially simplified with respect to existing remote sensing land cover classification methods. In fact, considering only one land cover class, we reduce the number of points and time needed to define the landslide class itself in the training stage. In second place, the classification algorithm parameters are specific of the study area but are expected to stay constant across different landslide events. This means that expert landslide mapping, CD image definition and topography-related tasks are required only once, to train the CD function and calibrate the thresholds for SU-based classification, in each newly considered study area. In third place, class assignment is automatic and it does not require a-posteriori identification of the different classes. The method can thus be used on a routine basis, and run whenever the occurrence of a new landslide event is otherwise detected with specialized methods (Martha et al. 2016; Mondini 2017) or simply a new image (or a pair of images) of the area becomes available. We combined different indices to obtain the discriminant function  $g_{ls}$ , to cope with the natural heterogeneity showed by the spectral response of the landslide surface.  $\Delta$ NDVI, PCA and SA proved to be reliable in identifying changes introduced by landslides in vegetated areas (Mondini et al. 2011a, 2011b), since they take into account both the radiometric ( $\Delta$ NDVI and SA) and the geometric (PCA) information contained in the satellite images. The combination of the three indices highlights the presence of landslides, making even easier the selection of a training area; we do not exclude that in other geographical regions different indices may be more effective. The computation and use of additional indices, together or in place of the ones used in this work, would easily be integrated in the method. In general terms, the computation of indices is not time-consuming even in a large scene like the ones offered by typical Sentinel-2 images.

The approximation provided by the global thresholding is actually a good one in our test case (cf. Tables 1 and 2), due to the very definition of  $g_{ls}$  as a high-contrast CD map. The study area is not characterized by sharp illumination differences, which makes the landslide-related peak in the histogram of  $g_{ls}$  values to be rather well separated from the remaining of the distribution. We stress, however, that this might not be the case for other case studies in which, for example, different combinations of latitude, season, and slope aspect in combination with image acquisition angle and sun illumination may determine a much less evident bi-modal behaviour of the global histogram of  $g_{ls}$ . In that eventuality, we expect the multi-threshold procedure to be even more effective than it does in the study area considered in this work.

We stress that the values of the average separations of peaks in the histograms of the  $g_{ls}$  values,  $m^{(2)}$ ,  $m^{(3)}_a$  and  $m^{(3)}_b$  are peculiar of the study area and the peculiar partition into sub-areas, so they are newly calculated in the validation stage. Moreover, unlike SU polygons, the numerical values of such variables depend upon the CD discriminant function and thus have to be re-calculated once a new  $g_{ls}$  is obtained from a new pair of images, and this is a rather fast calculation. Obtaining a new  $g_{ls}$  map is an automatic procedure, since the spectral properties of the same landslide type are believed to stay the same, in the same area, across multiple events.

In order to better illustrate which parts of the proposed procedure are site-dependent or not, and which parts are automatic or supervised, Figure 2 includes a table describing such characteristics of both the  $g_{ls}$  calculation step and of the classification step. We can summarize the Table as follows:

- the  $g_{ls}$  definition (step 1 in Figure 2) requires training; this amounts to single out (visually or by prior knowledge) one landslide, heuristically chosen as representative of the landslides possibly occurring in the area. The training step is a one-time operation, and allows to automatically calculate the  $g_{ls}$  raster map for the whole study area and for any pair of images available at present or in the future. Only deployment of the method in a different study area, or for use with images from a different sensor, requires new training. Running the method with any newly acquired image (or image pair) of the same study area, instead, is fully automatic;
- the topography-driven classification (step 2 in Figure 2) requires: (i) slope unit delineation, (ii) calculation of  $g_{ls}$  histograms in each slope unit and their grouping according to the classes sketched in Figure 7, (iii) definition of the quantities best suited to act as classification thresholds in each class of histograms, (iv) actual classification of the raster map into landslide/no landslide by multiple thresholding. SU delineation and definition of best quantities to be used as thresholds are one-time operations, since the first is only dictated by the DEM and the second is obtained by trial-and-error procedure against the ground truth in a calibration area. Calculation of the actual histograms in each SU and of the actual threshold values, and final classification, are fully automatic. Deployment of the method in a different study area requires SU delineation and, optionally, check if the set of quantities defined here as thresholds is still valid for the new area, if a landslide event can be mapped by expert geomorphologists in a calibration area.

There exist additional reasons for analyzing the distribution of  $g_{ls}$  values within single slope units instead of across the whole map. The spectral response of the terrain is a function of the relative orientations of the local slope, sun azimuth at the time of image acquisition, and the line of sight of the satellite. As a consequence, pixels facing a homogeneous aspect direction, like the pixels within each SU do on average, are likely to produce a similar response and, eventually, similar  $g_{ls}$  values for the same land cover class; in our case of interest, the landslide class. These expectations were borne out by our results, as one can see in Tables 1 and 2 and, in particular, in Figure 10, where the superiority of the SU-based multi-thresholding approach is evident especially in the validation area.

Tables 1 and 2 show that the effects of using a topographic-aware subdivision of the study area into sub-areas is numerically more effective in the validation area than in the calibration area with respect to a topographic-blind subdivision, though the gain with respect to the global thresholding decreases from about 8% in calibration to about 5% in validation. Nevertheless, the performance of SU-based classification with respect to regular polygon-based classification increased from the calibration to the validation areas. In addition, we observe from Figure 10(b) that a miscalibration of SUs would produce a smaller error than a miscalibration of square polygons size: the SU results are always better than the polygon results, in validation. In other words, the SU approach is more effective in a larger area than it is in a smaller one, regardless of small variations in the number and size of individual SU polygons.

Figure 5(a) shows an aspect map of the study area. No clear distinction exists between different regions containing preferred aspect directions. As a matter of fact, the histogram selection algorithm of Delon et al. (2007) could not distinguish different modes in the histograms of SU-based (average values in each SU) aspect values in our study area, neither within the calibration (Figure 5(b)) nor the validation (Figure 5(c)) areas. We explicitly checked that further splitting the study area and performing the analysis presented in this work separately in sub-areas where the SU aspect, loosely speaking, faces two or three preferred directions do not substantially improve the results. Should another study area present a clear-cut distinction between aspect directions in different groups of SU, the method would probably be more effective if the calibration procedure was performed, and classification applied, separately in such sub-areas.

## 6. Conclusions

We presented a novel approach to binary classification of satellite imagery, aimed at supervised landslide mapping, with a high level of automation. Our approach requires knowledge of the occurrence of a landslide event, obtained by external means, availability of a high-resolution pre- and post-event pseudo-stereo pair, and a digital elevation model.

The mapping procedure, outlined in Figure 2, consists of two steps. In the first step, we defined a discriminant function  $g_s$ , a combination of well-known  $\Delta$ NDVI, PCA and SA change detection indices, tuned to highlight differences in landslide spectral response versus anything else. In the second step, we devised a multi-threshold binary classification method for the  $g_s$  map, aimed at distinguishing the only landslide class. The multi-thresholding procedure is applied within a large number of sub-areas, namely slope units (Alvioli et al. 2016), known to be particularly suited for landslide studies.

From the results obtained in this work, illustrated in Tables 1 and 2 and Figures 9–11, we can draw the following conclusions:

- comparison of the results of the semi-automatic mapping procedure with the ground truth of an eLIM prepared by visual interpretation reveals that the topographic-aware subdivision of the territory allows for a better performance both than thresholding applied globally on the study area, and within a topographic-blind subdivision;
- the increased performance of the proposed multi-threshold approach can be obtained with a comparable computational cost of the single-threshold approach, once proper calibration has been performed;
- the overall procedure is fully automatic once the preliminary steps of SU delineation,  $g_s$  training and threshold calibration are performed. In principle, the definition of river is left out from the second step.

As a matter of fact, a severe rainfall event modifies the rivers in a non-negligible way and they have to be drawn for each new event. In our case, this was done at photo-interpretation time, for simplicity. The literature contains several sound automatic ways of mapping riverbanks, for example with pixel-based methods (Mondini and Chang 2014; Mondini et al. 2017) or object-oriented methods (Martha et al. 2010; Stumpf and Kerle 2011). Thus, in principle, the method is fully automatic after riverbank detection is included.

Applicability of the method in other study areas, in which different environmental conditions may exist, remains to be investigated. We expect the stack of changes (Figure 3(d)) to exhibit signatures produced by new landslides occurring in different lithologies to be more similar among themselves than signatures of changes due to other features (or signatures of no-changes), which entitles us to assume the portability of a single landslide class, possibly similar or invariant across different areas. We also expect the relative performance of multi-thresholding with respect to global thresholding to increase in study areas where a sharp difference in orientation of hillslopes, generating a different spectral response between the different orientations, may exist.

We argue that the improved performance and limited training requirements of the classification procedure represent a step forward towards an automatic, real-time landslide mapping from satellite imagery.

### Geolocation information

Our study area covers N23.874°–23.431° and E93.7°–93.948° (EPSG:4032), in Myanmar.

### Notes

1. <https://grass.osgeo.org>.
2. <http://portal.onegeology.org/OnegeologyGlobal/>.

3. <http://www.seom-commonsproject.net/>.
4. [https://resa.blackbridge.com/files/RapidEye\\_Image\\_Positional\\_Accuracy\\_Whitepaper\\_V1.0\\_ENG.pdf](https://resa.blackbridge.com/files/RapidEye_Image_Positional_Accuracy_Whitepaper_V1.0_ENG.pdf).
5. <http://www.harrisgeospatial.com/docs/FLAASH.html>.
6. <https://asterweb.jpl.nasa.gov>.
7. <http://www.planar.com>.

## Acknowledgments

We thank NERC for providing partial support within the LANDSLIP project (<http://www.landslip.org>), grant number NERC/DFID NE/P000681/1. M. Alvioli and F. Fiorucci were also supported by a grant from Dipartimento della Protezione Civile, Italy, 2017.

## Disclosure statement


No potential conflict of interest was reported by the authors.

## Funding

Partial support for this work was provided by the Natural Environment Research Council (NERC) [grant number NERC/DFID NE/P000681/1]; MA and FF were also supported by a grant from Dipartimento della Protezione Civile, Italy, 2017 [grant number DPC 2016] .

## ORCID

M. Alvioli  <http://orcid.org/0000-0003-1543-4349>

I. Marchesini  <http://orcid.org/0000-0002-8342-3134>

## References

- Allen R, Najman Y, Carter A, Barfod D, Bickle M, Chapman H, Garzanti E, Vezzoli G, Ando S, Parrish RR. 2008. Provenance of the tertiary sedimentary rocks of the Indo-Burman Ranges, Burma (Myanmar): Burman arc or Himalayan-derived? *J Geol Soc.* 165(6):1045–1057.
- Alvioli M, Marchesini I, Reichenbach P, Rossi M, Ardizzone F, Fiorucci F, Guzzetti F. 2016. Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling. *Geosci Model Dev.* 9:3975–3991.
- Barra A, Monserrat O, Mazzanti P, Esposito C, Crosetto M, Mugnozza GS. 2016. First insights on the potential of Sentinel-1 for landslides detection. *Geomat Nat Hazards Risk.* 7(6):1874–1883.
- Borghuis AM, Chang K, Lee HY. 2007. Comparison between automated and manual mapping of typhoon triggered landslides from spot 5 imagery. *Int J Remote Sens.* 28(8). 1843–1856.
- Brakenridge G, Syvitski J, Niebuhr E, Overeem I, Higgins S, Kettner A, Prades L. 2017. Design with nature: causation and avoidance of catastrophic flooding, Myanmar. *Earth Sci Rev.* 165(Supplement C):81–109.
- Carrara A. 1993. Uncertainty in evaluating landslide hazard and risk. In: Nemeč J, Nigg JM, Siccardi F, editors. *Prediction and Perception of Natural Hazards: Proceedings Symposium; Oct 22–26, 1990, Perugia, Italy.* Dordrecht: Springer Netherlands. p. 101–109.
- Casagli N, Frodella W, Morelli S, Tofani V, Ciampalini A, Intrieri E, Raspini F, Rossi G, Tanteri L, Lu P. 2017. Spaceborne, UAV and ground-based remote sensing techniques for landslide mapping, monitoring and early warning. *Geoenviron Disasters.* 4(1):9.
- Chen LC, Rau JY. 1993. A unified solution for digital terrain model and orthoimage generation from spot stereopairs. *IEEE Trans Geosci Remote Sens.* 31(6):1243–1252.
- Cheng KS, Wei C, Chang SC. 2004. Locating landslides using multi-temporal satellite images. *Adv Space Res.* 33(3):296–301.
- Delon J, Desolneux A, Lisani JL, Petro AB. 2007. A nonparametric approach for histogram segmentation. *IEEE Trans Image Process.* 16(1):253–261.

- Espindola G, Camara G, Reis I, Bins L, Monteiro A. 2006. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. *Int J Remote Sens.* 14:3035–3040.
- Fiorucci F, Cardinali M, Carlà R, Rossi M, Mondini A, Santurri L, Ardizzone F, Guzzetti F. 2011. Seasonal landslide mapping and estimation of landslide mobilization rates using aerial and satellite images. *Geomorphology.* 129(1–2):59–70.
- Fiorucci F, Giordan D, Santangelo M, Dutto F, Rossi M, Guzzetti F. 2018. Criteria for the optimal selection of remote sensing optical images to map event landslides. *Nat Hazards Earth Syst Sci.* 18(1):405–417.
- Gaidzik K, Ramírez-Herrera MT, Bunn M, Leshchinsky BA, Olsen M, Regmi NR. 2017. Landslide manual and automated inventories, and susceptibility mapping using LIDAR in the forested mountains of Guerrero, Mexico. *Geomat Nat Hazards Risk.* 8(2):1054–1079.
- Guzzetti F, Carrara A, Cardinali M, Reichenbach P. 1999. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology.* 31:181–216.
- Guzzetti F, Mondini AC, Cardinali M, Fiorucci F, Santangelo M, Chang KT. 2012. Landslide inventory maps: new tools for an old problem. *Earth Sci Rev.* 112(1–2):42–66.
- Keyport RN, Oommen T, Martha TR, Sajinkumar K, Gierke JS. 2018. A comparative analysis of pixel- and object-based detection of landslides from very high-resolution images. *Int J Appl Earth Obs Geoinf.* 64(Supplement C): 1–11.
- Lee S. 2005. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int J Remote Sens.* 26(7):1477–1491.
- Li Z, Shi W, Lu P, Yan L, Wang Q, Miao Z. 2016. Landslide mapping from aerial photographs using change detection-based Markov random field. *Remote Sens Environ.* 187(Supplement C):76–90.
- Licht A, France-Lanord C, Reisberg L, Fontaine C, Soe AN, Jaeger JJ. 2013. A palaeo Tibet–Myanmar connection? Reconstructing the Late Eocene drainage system of central Myanmar using a multi-proxy approach. *J Geol Soc.* 170(6):929–939.
- Marchesini I, Ardizzone F, Alvioli M, Rossi M, Guzzetti F. 2014. Non-susceptible landslide areas in Italy and in the Mediterranean region. *Nat Hazard Earth Syst Sci.* 14(8):2215–2231.
- Martha TR, Kamala P, Jose J, Vinod Kumar K, Jai Sankar G. 2016. Identification of new landslides from high resolution satellite data covering a large area using object-based change detection methods. *J Indian Soc Remote Sens.* 44(4):515–524.
- Martha TR, Kerle N, Jetten V, van Westen CJ, Kumar KV. 2010. Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. *Geomorphology.* 116(1): 24–36.
- Martha TR, Kerle N, van Westen CJ, Jetten V, Kumar KV. 2011. Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. *IEEE Trans Geosci Remote Sens.* 49(12):4928–4943.
- Mondini AC. 2017. Measures of spatial autocorrelation changes in multitemporal SAR images for event landslides detection. *Remote Sens.* 9(6):554.
- Mondini AC, Chang KT. 2014. Combining spectral and geoenvironmental information for probabilistic event landslide mapping. *Geomorphology.* 213:183–189.
- Mondini AC, Chang KT, Chiang SH, Schlögel R, Notarnicola C, Saito H. 2017. Automatic mapping of event landslides at basin scale in Taiwan using a Monte Carlo approach and synthetic land cover fingerprints. *Int J Appl Earth Obs Geoinf.* 63(Supplement C):112–121.
- Mondini AC, Guzzetti F, Reichenbach P, Rossi M, Cardinali M, Ardizzone F. 2011a. Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. *Remote Sens Environ.* 115(7):1743–1757.
- Mondini AC, Tsung Chang K, Yuan Yin H. 2011b. Combining multiple change detection indices for mapping landslides triggered by typhoons. *Geomorphology.* 134(3–4):440–451.
- Moosavi V, Talebi A, Shirmohammadi B. 2014. Producing a landslide inventory map using pixel-based and object-oriented approaches optimized by Taguchi method. *Geomorphology.* 204(Supplement C):646–656.
- Neteler M, Mitasova H. 2007. *Open source GIS: a GRASS GIS approach.* New York: Springer.
- Plank S, Twele A, Martinis S. 2016. Landslide mapping in vegetated areas using change detection based on optical and polarimetric SAR data. *Remote Sens.* 8(4):307.
- Richards J, Jia X. 2006. *Remote sensing digital image analysis.* Berlin: Springer-Verlag.
- Rosin PL, Hervás J. 2005. Remote sensing image thresholding methods for determining landslide activity. *Int J Remote Sens.* 26(6):1075–1092.
- Schlögel R, Marchesini I, Alvioli M, Rossi M, Reichenbach P, Malet JP. 2018. Optimizing landslide susceptibility zonation: effects of DEM spatial resolution and slope unit delineation on logistic regression models. *Geomorphology.* 301:10–20.
- Sohn Y, Rebello N. 2002. Supervised and unsupervised spectral angle classifiers. *Photogramm Eng Remote Sens.* 68(12):1271–1280.

- Stumpf A, Kerle N. 2011. Object-oriented mapping of landslides using random forests. *Remote Sens Environ.* 115 (10):2564–2577.
- Tucker CJ. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens Environ.* 8(2):127–150.
- Yu B, Chen F. 2017. A new technique for landslide mapping from a large-scale remote sensed image: a case study of Central Nepal. *Comput Geosci.* 100(Supplement C):115–124.

## Appendices

### Appendix 1. Analysis of polygon-wise discriminant function histograms

The cases  $N_{mod} = 2$ , shown qualitatively in the second row of Figure 7, can be further characterized by their similarity to the cases HA<sub>3</sub> or HB<sub>3</sub> with  $N_{mod} = 3$ . This is easily done if one assumes Gaussian distributions of the  $g_{ls}$  values of the leftmost (and the related quantities being labelled with  $a$ ) and rightmost (labelled with  $b$ ) separations, separately. Associating standard deviation  $d^{(3)}_a$  and  $d^{(3)}_b$  to the distributions centred in  $m^{(3)}_a$  and  $m^{(3)}_b$ , respectively, and denoting with  $G_{m,d}(g)$  the normalized Gaussian distribution with average  $m$  and standard deviation  $d$ , one can calculate the probability of a separation  $g^{(2)}_s$  from an  $N_{mod} = 2$  case to belong to the Gaussian distributions  $G_{m,d}(g)$  at the distance  $\delta = |g^{(2)}_s - m|$  from its central vale  $m$  as follows:

$$P_{m,d}(\delta) = \int_{m-\delta}^{m+\delta} dg G_{m,d}(g). \quad (\text{A1})$$

Calculating the probabilities  $P_{m_a^{(3)},d_a^{(3)}}(\delta)$  and  $P_{m_b^{(3)},d_b^{(3)}}(\delta)$  allows to label the separation  $g^{(2)}_s$  to be either:

- more similar to the HA<sub>3</sub> cases with  $N_{mod} = 3$  of Figure 7, if  $P_{m_a^{(3)},d_a^{(3)}}(\delta) < P_{m_b^{(3)},d_b^{(3)}}(\delta)$ , which we define as *left-like*;
- more similar to the HB<sub>3</sub> cases with  $N_{mod} = 3$  of Figure 7 if  $P_{m_a^{(3)},d_a^{(3)}}(\delta) > P_{m_b^{(3)},d_b^{(3)}}(\delta)$ , which we define as *right-like*.

### Appendix 2. Definition of custom thresholds for polygon-wise classification

We have devised a procedure to obtain different  $T$  values within different classes of polygons (ideally, the six classes identified by the histogram types of Figure 7; in practice, the classes listed below) to reduce misclassification. Following a trial-and-error procedure, and using the quantities  $M$ ,  $m^{(2)}$ ,  $g^{(2)}_s$ ,  $m^{(2)}_a$ ,  $m^{(2)}_b$ ,  $g^{(3)}_{s,a}$  and  $g^{(3)}_{s,b}$  introduced in Section 3, we obtained the following procedure:

- $N_{mod} = 1$ : we set  $T = m^{(2)}$ , except for those cases in which either (i) there is a percentage of cells with  $g_{ls} > M$  smaller than 0.1%, where we set  $T = 0$  (assuming no landslides at all in the polygon), or (ii) there is a percentage of cells with  $g_{ls} < M$  larger than 50%, where we set  $T = -200$  (assuming most of the polygon is a landslide). Cases (i) likely correspond to histograms similar to HA<sub>1</sub> in Figure 7, while cases (ii) to HB<sub>1</sub>.
- $N_{mod} = 2$ : in this case, we further distinguish (*cf.* Appendix 1):
  - left-like: we set  $T = m^{(2)}$ , except for those cases in which there is a percent of cells with  $g_{ls} > M$  smaller than 0.1%, corresponding to histograms similar to HA<sub>2</sub> in Figure 7, and where we set  $T = 0$ . We did not find advantages setting a large negative threshold when there are many cells situated at the left of  $M$ , as it is the case for  $N_{mod} = 1$ , hinting that the ansatz of histograms similar to the sketch HB<sub>2</sub> is not realistic.

- right-like: we set  $T = M$ . We do not find advantages in setting a threshold at zero or at large negative values, which is reasonable for right-like  $N_{mod} = 2$  cases: as a matter of fact, both  $T = 0$  or a large negative  $T$  suggest a histogram similar to the sketch  $HA_1$  of Figure 7, which is not likely to occur when one separation is found, as in the case we are discussing.
- $N_{mod} = 3$ : we set  $T = g_{s, b}^{(3)}$ , the rightmost of the two existing separations. The fact that this turns out the best choice, in this case, hints that the ansatz of histograms similar to the sketch  $A_3$  does not actually show up often.