Automatic landslide flow direction estimation based on the geometric processing of the bounding box and the geomorphometric analysis of DEMs

Mihai Niculiță

1 Geography Department, Geography and Geology Faculty, Alexandru Ioan Cuza University of Iași, Carol I, 20A, 705505, Iași, Romania

Correspondence to: Mihai Niculiță (mihai.niculita@uaic.ro)

Abstract. Landslides are geomorphologic and hydrologic phenomena, which produce a certain morphology, influenced by the slide/flow of the material downslope. Usually the distance of the movement of the material is greater than the width of the displaced material (especially for flows, but also the majority of the slides enter in the same category), the resulting landslides having the length bigger than the width. In some specific environments (monoclinic regions, with cuesta landforms) or some types of landslides (translational slides, bank failures, complex landslides), the distance of the movement is smaller than the width of the displaced mass, the resulting landslides having the length smaller than the width. When dealing with landslide inventories which contain both the classes of landslides presented above, the analysis of the length and width of landslides will be biased. To correct this aspect, I present an algorithm which uses both the geometry of the landslide polygon minimum oriented bounding box and a DEM of the landslide topography for identifying the long versus wide landslides, in a landslide inventory covering 131.1 km² in Moldavian Plateau, Eastern Romania, and containing 1327 landslides (518 long and 809 wide). In a first step, the difference in elevation of the length and width of the minimum oriented bounding box is used to separate long from wide landslides (long landslides having the greatest elevation difference along the length of the bounding box). In a second step, the long landslides are checked if their length is greater than the slope length (estimated with a flow routing algorithm), in which case the landslide is labelled as wide. By using this approach the AUROC value for the classification of the two cases is 87.8%. An intensive review of the wrong classified cases is made, and discussions are included about the prospects of improving with further steps the approach, to reduce the number of misclassified cases.

1 Introduction

Landslides are a natural phenomenon which have an important impact on human society and environment (Geertsema et al., 2009; Kjekstad and Highland, 2009; Petley, 2012). The topographic imprint of these geomorphologic processes is given by the slide/flow of the material downslope, and it is highly characteristic (Pike, 1988; Berti et al., 2013; Rossi et al., 2013; Marchesi et al., 2014), especially on high resolution DEMs (Schulz, 2004; McKean and Roering, 2004; Van den Eeckhaut
et al., 2005; Glenn et al., 2006; Schulz, 2007; Kasai et al., 2009; Jaboyedoff et al., 2012; Berti et al., 2013). In the last decades, big improvements were made in landslide delineation (Cararra and Merenda, 1976; Keaton and DeGraff, 1996; Ardizzone et al., 2002; Ardizzone et al., 2007; Guzzetti et al., 2012) and landslide inventories analysis (Malamud et al., 2004; Guzzetti et al., 2012; Van den Eeckhaut et al., 2012). Nowadays, a landslide inventory is usually a collection of polygon shapes which represent the boundary of a single event or of a complex landslide, generated by multiple events, but with a clear morphologic signature. Landslides are delineated manually (Van den Eeckhaut and Hervás, 2012), but in several environments where certain geospatial data is available (high resolution imagery, LiDAR) semi-automatic procedures were created (Van den Eeckhaut et al., 2009; Martha et al., 2010; Mondini et al., 2011; Martha et al., 2012; Van den Eeckhaut et al., 2012; Lyon et al., 2013).

The sliding/flowing of the material downslope leaves a scarp and generate a displaced mass, which usually is longer than the width of the displaced mass. This is the case for the flow types (Cruden and Varnes, 1996), but apply also to rotational and translational slides, falls, topples and spreads. I will call this type of landslides, long landslides. Long landslides appear either as single events on hillslopes or as complex landslides (Fig. 1).

While not so frequent for flows, falls and rotational slides, cases when the length of the displacement does not exceed the width of the displaced mass, are relatively common for translational slides and spreads. These type of landslides appear in almost every landslide inventory, in different proportions (such cases are visible in landslide inventory snapshots from Galli et al., 2008, Hattanji et al., 2009, Guzzetti et al., 2012, Van den Eeckhaut et al., 2012, Mondini et al., 2013). I will call this type of landslides, wide landslides.

Wide landslides appear especially along gully and river banks, but can also appear on hillslopes, especially in regions with monoclinal structure and cuesta landforms, along the scarps (as is the case of the study area – Fig. 1, 2).

In both scenarios of landslide delineation (manual or automatic), the landslide length and width is usually computed automatically in a GIS environment, using the dimensions of the minimum bounding box. While for long landslides, the length of the bounding box is the correct length, for wide landslides, the correct length is actually the width of the bounding box.

### 2 Materials

For testing the proposed method I have used a landslide inventory (Fig. 2), created and described by Niculiță and Mărgărint (2015), for a small area (131.1 km²) from the Moldavian Plateau (Eastern Romania), where single event and complex landslides were delineated. This inventory contains 1327 polygons, representing landslides, with area ranging from 94 m² to 800 00 m². Wide landslides appear both along channels, gullies, old landslide scarps, hillslopes and landslide displaced masses. The delineation of polygon landslides was made using aerial imagery, high resolution LiDAR DEM and field validation. For every landslide the long or wide type was recorded in the attribute table, during inventory creation. Landslide typology in the study area is dominated by translational and rotational slides, with very few flows. A five meters LiDAR DEM was used for the midpoint elevation extraction and for the slope length estimation.
3 Methods

The automatic algorithm for the estimation of landslide flow direction is based on the use of the geometric bounding box, which is an oriented bounding box, along the geometric length, its geometric processing, and the elevation and slope length computed from the DEM.

The proposed method was implemented using R software application (R Core Team) and several user contributed packages, as a script, which is available at http://www.geomorphologyonline.com/LandslideFlowDirection.R (Fig. 3). Oriented bounding boxes (as are they called in the GIS literature) are in fact minimum-area enclosing rectangles for convex polygons. Beside the rectangular minimum bounding box (whose sides are parallel with the coordinate axes), the oriented minimum bounding box can be defined as the bounding box which has the sides along the orientation of the polygon (Freeman and Shapiro, 1975). The rotating calipers algorithm (Toussaint, 1983) is used to create these bounding boxes, its implementation in the R shotGroups package (Wollschläger, 2016) being used, through the getMinBBox function. This function uses the coordinates of the polygon vertices, stored in an object of the sp package, called SpatialPolygonsDataFrame (Bivand et al., 2013), which is created from a shapefile containing the landslide polygons and a unique ordered sequence of autoincrement IDs, by import with the rgdal package (Fig. 3). Further, sp package functions are used to manipulate the bounding boxes.

The getMinBBox function creates the x, y coordinates of the corners (CP1 to CP4 – Fig. 4) of the minimum oriented bounding box in the counter-clockwise order, which allows the creation of the rectangle with the help of the sp package functions. While the majority of the resulting rectangles have sides 12 and 34 (Fig. 4) as the long sides (length) and sides 14 and 23 as short sides (width), there are some cases where the situation is vice versa. Because the algorithms needs a consistent approach, the coordinates of the four corner points of the minimum oriented bounding box, where introduced in a polygon data table, and using the length as rule, all the corner points were assigned to correspond to the above notation (Fig. 4). Using the coordinates of the corner points, midpoints of the sides (MP1 to MP4) were computed using the Midpoint Formula, in such a way that MP1 and MP3 are on the long sides, and MP2 and MP4 are on the short sides of the minimum oriented bounding box. Using this four midpoints I defined the length of the minimum oriented bounding box (line connecting MP2 and MP4) and the width of the minimum oriented bounding box (line connecting MP1 and MP3) (Fig. 4). Altitudes of MP1 to MP4 were taken from the DEM, the square root of the altitude difference between MP1 to Mp3 and MP2 to Mp4 raised at power of two being computed, and using the following rules, the landslides were classified in a first step as long or wide:

- The landslides which have the biggest altitude difference along the length of the minimum oriented bounding box (line connecting MP2 and MP4) are classified as long;
- The landslides which have the biggest altitude difference along the width of the minimum oriented bounding box (line connecting MP1 and MP3) are classified as wide.

The inspection of the results showed that some wide landslides are classified as long, because they are located on short hillslopes but which span on a certain relative altitude (sides of small valleys incised on cuesta scarps).
For these cases a further step of processing was required. This step involved the cropping of the DEM for every landslide polygons, the hydrological pre-processing to inforce drainage flow routing and the computation of the slope length (Fig. 4). This step required the use of R package RSAGA (Brenning and Bangs, 2016) which links SAGA GIS (Conrad et al., 2015). The DEM was preprocessed with the functions Sink Drainage Route Detection and Sink Removal from the Terrain Analysis – Preprocessing module library. The hydrological enforcement was done using the Deepen Drainage Routes method and the flow path length was computed using the function Flow Path Length from the Terrain Analysis – Hydrology module library. Both flow routing methods implemented in this tool: Deterministic 8 (O’Callaghan and Mark, 1984) and Multiple Flow Direction (Freeman, 1991; Quinn et al., 1991) were computed (Fig. 5). The maximum slope length for every landslide polygon was computed using the function Grid Statistics for Polygons from Shapes – Grid Tools tool library.

Using the following rules, only the long landslides class was reclassified:
- The landslides which have the length greater than the slope length are reclassified to wide;
- The landslides which have the length smaller than the slope length remain classified as long.

3 Results and Discussions

In the first step of the proposed method, 642 landslides are classified as long, from which 40 (3%) wrongly (false positive - FP) and 685 as wide, from which 165 (12.4%) wrongly (false negative) (Fig. 6). In the second step the best performance was acquired by the multiple flow algorithm for computing slope length. This algorithm give flow distance which are more similar with the hillslope length, because models the flow as a diffusion process, while the Deterministic 8 algorithm creates a pattern of interrupted flows, although the DEM was preprocessed hydrologically (Fig. 5). Using D8, in the second step 1268 landslides are classified as wide and only 59 as long. The Multiple Flow Direction algorithm classify 490 landslides as long, and 837 as wide. Overall, 87.8% landslides are correctly classified (AUROC value – Fig. 7), either as long or wide. From the 837 wide landslides, 66 (4.97%) landslides are classified as long although they are wide, and from 490 long landslides, 93 (7%) are classified as wide although they are long.

The wide landslides wrongly classified as long in the first step, are mainly developed along the banks of incised rivers and gullies, landslide scarps, landslide toes and cuesta scarps, in situations where there a bigger difference in altitude between the flanks than between the scarp and the toe (Fig. 8c). The long landslides wrongly classified as wide are almost round, rhombic, hexagonal or elongated landslides which, are diagonal to the minimum enclosed bounding box and the downslope direction of the hillslope, which in general is gentle sloped (Fig. 8a). This diagonal position generates a difference of altitude along the width which is greater than the difference along the length.

The majority of the wide landslides classified as long in the second step are almost round, rhombic or hexagonal, and diagonal to the downslope direction, and only a few are developed along the gully banks or scarps (Fig. 8d). The long landslides classified as wide are related to areas where the slope length computed by the flow algorithm fails to model the length of the hillslope, and in step two their length is greater than the slope length (Fig. 8b). These landslides have gullies
or mounds, which interrupts the diffusion of the flow, and the flow algorithms does not routes the flow from the crown to the
toe of the landslide.

Any inconsistency between the shape of the earth portrayed by the DEM and the delineated landslide morphology will give
wrong classification cases. This scenario is to be expected from inventories were the landslides are delineated from high
resolution aerial/satellite imagery, but the DEM has not the same resolution, a frequent scenario in the case of automatically
delineated landslides. Normally if all the data has the same resolution, the proposed algorithm can provide good results. We
also tested the SRTM 1elevation data with the proposed method for the landslide inventory, and the AUROC values decreased
to 86.6%.

The quality of the landslide inventory is also of great importance, since any geomorphologic topologic inconsistency can
introduce wrong identification cases:
- Amalgamation of landslides by crossing channels;
- Shifting of landslides and touching channels;
- Landslides spread across several hillslopes.

In specific topographic situations, the midpoints will not be on the edge of the landslide, and not even the same hillslope as
the landslide (Fig. 8c), but we believe that in general this situation will not introduce errors.

Inconsistencies of the type described above are to be expected in an important proportion for landslide inventories based on
topographic maps and metric aerial imagery/DEM, their proportion decreasing in landslide inventories based on centimeter
resolution aerial imagery/DEM. The proposed methodology could be used also for checking landslide inventories for
geomorphological topologic inconsistencies.

The main fails of the method are:
- The inability to separate wide from long landslides based on the flow (mainly related to the false positive cases, long
  landslides classified as wide);
- The inability to separate long from wide landslides because of their diagonal position (mainly related to the false
  negative cases, wide landslides classified as long).

Possibilities to resolve the failure to separate between long and wide landslides based on the flow length estimation require a
new approach in slope length computation. I have tried the use in the second step, of the slope length estimated as the slope
compensated along the length side of the bounding box, but this approach gave results worse than the hydrologically slope
length estimation one.

Possibilities to resolve the failure to separate between long and wide landslides because of their diagonal position, are related
to the computation of the bounding box. Probably using a 3D approach some of the caveats of the current approach could be
dealt with. For the majority of diagonal landslides (false negative - 93 cases) the difference between the length and width is
not big (under 10%), in their case the impact in the morphometric analysis being small.
3 Conclusions

The proposed method has the ability to discern between long and wide landslides in 87.8% of cases. There is a proportion of landslides, which because of their geometry and geomorphologic topology is hard to be identified in the correct class. The majority of these landslides are round, rhombic or hexagonal, and oriented diagonal to the downslope direction of the hillslope. The flow direction of these landslides is diagonal to the bounding box. If their length and width is needed, although the algorithms fails to recognize them as long or wide, the errors are minimal, even if the “wrong” length is chosen. Of course, if the direction of flow is needed, the values obtained for them is erroneous. The minority of wrong identified landslides is represented by the elongated landslides, which are either long (similar to flows) or wide (developed along gully banks and landslide scarps). For these landslides the proposed method fails, because of the failure in slope length estimation which require further investigations on the possibilities of using other methods to estimate slope length. Overall, the presented approach manage to identify the majority of the long versus wide landslides from a landslide inventory and represent a starting point in deriving algorithms for automatic geomorphometric analysis of landslides.

Acknowledgments

The author gratefully acknowledge partial support by the European Social Fund in Romania, under the responsibility of the Managing Authority for the Sectoral Operational Program for Human Resources Development 2007-2013 [grant POSDRU/159/1.5/S/133391]. I am grateful to Prut-Bârlad Water Administration who provided me with the LIDAR data. I have used the computational facilities given by the infrastructure provided through the POSCCE-O 2.2.1, SMIS-CSNR 13984-901, No. 257/28.09.2010 Project, CERNESIM (L4).

References


Marchesini, I., Rossi, M., Mondini, A., Santangelo, M. and Bucci, F.: Morphometric signatures of landslides, Third Open Source Geospatial Research and Education Symposium (OGRS), Espoo, Finland, 10-13 June 2014.


Figure 1: Geomorphologic cases of long (L) and wide (W) landslides from the study area (up there is a Google Earth capture, down is a 3D view of the DEM shading, both with the landslide polygons overlayed).
Figure 2: The study area and the landslide inventory (the red boxes from the right map indicated the insets from Fig. 5 and Fig. 8).
Figure 3: Workflow in R.
Figure 4: Geometry of the minimum oriented bounding box.
Figure 5: Flow lengths for a landslides DEM patch (MFD is the flow length computed with the Multiple Flow Direction algorithm and D8 is the flow length computed with the Deterministic 8).
Figure 6: Four folded plots of the confusion matrices for the two steps of the algorithm (TP = true positive, TN = true negative, FP = false positive, FN = false negative).
Figure 7: ROC curves for the two steps of the algorithm and their 95% confidence interval computed with 2000 stratified bootstrap replicates.

AUROC = 84.3%

AUROC = 87.8%
Figure 8: False positive and false negative cases: a) false positive cases (long landslides classified as wide landslides) in step 1, b) false positives in step 2, c) false negative cases (wide landslides classified as long landslides) in step 1, d) false negative cases in step 2.