

Automatic landslide length and width estimation based on the geometric processing of the bounding box and the geomorphometric analysis of DEMs

Mihai Niculiță¹

5 ¹Geography 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 morphology is 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 slides), the resulting landslides having a greater length than width. In some specific geomorphologic environments (monoclinic regions, with cuesta landforms type) or the case of some types of landslides (translational slides, bank failures, complex landslides), for the majority of the landslides the distance of the movement of the displaced material can be smaller than its width, thus the landslides having a smaller length than width. When working with landslides inventories containing both types of landslides presented above, the analysis of the length and width of the landslides computed with usual GIS techniques (like bounding boxes) can be flawed. To overcome this flaw, 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. I tested the proposed algorithm for a landslide inventory which covers 131.1 km² in Moldavian Plateau, Eastern Romania. This inventory contains 1327 landslides, from which 518 were manually classified as long and 809 as 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 length of flow downslope (estimated with a flow routing algorithm), in which case the landslide is classified as wide. By using this approach, the AUC value for the classification of the long versus wide landslides is 87.8%. An intensive review of the misclassified cases and the challenges of the proposed algorithm is made, and discussions are included about the prospects of improving with further steps the approach, to reduce the number of misclassifications.

1 Introduction

Landslides are a natural phenomenon which has 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; Marchesini 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 (Carrara and Merenda, 1976; Brunsden, 1993; 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). A landslide inventory is usually a collection of polygon shapes which represent the boundary of a single event or a complex landslide, generated by multiple events (Van Westen, 1993; Guzzetti et al., 2012). Landslides are delineated manually, nowadays digitally in a GIS environment (Carrara et al., 1991; Van Westen, 1993; Van den Eeckhaut and Hervás, 2012), but in several environments (forested areas with frequent debris or earth flows, triggered by earthquakes or typhoons) where certain geospatial data is available (high-resolution imagery, LiDAR) semi-automatic procedures were used (Petley et al., 2002; Barlow et al., 2003; Nichols and Wong, 2004; Martin and Franklin, 2005; Barlow et al., 2006; Borghuis et al., 2007; Van den Eeckhaut et al., 2009; Martha et al., 2010; Mondini et al., 2011; Martha et al., 2012; Van den Eeckhaut et al., 2011; Lyons et al., 2014). Since these automatic methods can delineate a big number of landslides, an automatic method for length and width estimation can be helpful for analyzing the resulted inventories.

The sliding/flowing of the material downslope produces a scarp and a displaced mass, which are usually longer than wide (Fig. 1d). While generalized for flows (Fig. 1d) (Cruden and Varnes, 1996), this mechanism is also quite frequent for rotational (Fig. 1a) and translational slides, falls, topples and spreads (Cruden and Varnes, 1996). I will call this type of landslides, long landslides. Long landslides appear either as single events or as complex landslides (Fig. 1a, d).

While not characteristic 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 (Fig. 1c, e). These type of landslides appear in almost every landslide inventory, in different proportions (such cases are visible in landslide inventory snapshots from Dewitte and Demoulin, 2005; Schulz, 2007; Galli et al., 2008; Baldo et al., 2009; Hattarji 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 (Fig. 1b, c, e).

Wide landslide most commonly appears along gully and river banks (Fig. 1b, 1e), but also on hillslopes (Fig. 1c) especially in regions with monoclinic structure along the scarps of cuesta landforms. Such a situation is found in the Moldavian Plateau (Niculiță, 2015; Mărgărint and Niculiță, 2016) where the monoclinic structure favored cuesta escarpments which extend in width on long distances without being fragmented (Fig. 2, 3).

In both scenarios of landslide delineation (manual or automatic), the landslide length and width can be computed automatically in a GIS environment, using the dimensions of the minimum bounding box of the polygons representing the landslides (Taylor and Malamud, 2012). While for long landslides, the length of the bounding box is a reasonable approximation of the length of the landslide, for wide landslides, the length is actually the width of the bounding box.

To overcome the above-mentioned flaw, which can bias the analysis of landslide length and width (Taylor and Malamud, 2012; Taylor et al., 2015; Niculiță, 2015; Mărgărint and Niculiță, 2016), an automatic algorithm for the estimation of landslide

flow direction is proposed. The algorithm uses the processing of the oriented bounding box, and the elevation and slope length computed from the DEM, for deriving the length along the landslide slide/flow direction and the associated width. The method is validated on a landslide inventory where the long and wide character was assessed during landslide mapping.

5 2 Materials

For testing the proposed method I have used a landslide inventory (Fig. 2, 3), 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. 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. The long and wide type was assessed through expert opinion, during landslide delineation, also using 3D views to better assess the slide/flow direction. Landslide typology in the study area is dominated by translational and rotational slides, with very few flows. A five meters resolution LiDAR DEM was available for the study area.

3 Methods

The proposed method was implemented using R software application (R Core Team, 2015) and several users contributed packages, as a script, which is available at <http://www.geomorphologyonline.com/LandslideFlowDirection.R> (Fig. 4). Oriented bounding boxes are in fact minimum-area enclosing rectangles for convex polygons (Fig. 5). Besides the rectangular minimum bounding box (whose sides are parallel to 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 Shapira, 1975). The rotating calipers algorithm (Toussaint, 1983) is used to create these bounding boxes in the majority of GIS implementations. Its implementation in the R *shotGroups* package (Wollschläger, 2016) was used in the present approach, 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. 4). 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. 5) of the minimum oriented bounding box in the counter-clockwise order (CP1, CP2, CP3, CP4), which allows the creation of the rectangle with the help of the *sp* package functions. While the majority of the resulting rectangles have sides CP1-CP2 and CP3-CP4 (Fig. 4) as the long sides (length) and sides CP1-CP4 and CP2-CP3 as short sides (width), there are some cases where the situation is vice versa. Since the algorithm needs a consistent approach, the coordinates of the four corner points of the minimum oriented bounding box,

were introduced in a data table, and using the length of the sides, as a rule, all the corner points were assigned to correspond to the above notation (Fig. 5 left).

The coordinates of the corner points were used, to compute the midpoints of the sides (MP1 to MP4), 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 (Fig. 5 left). Using these four midpoints, the algorithm defines 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. 5). Using the elevations of the four corners (MP1 to MP4), taken from the DEM, the square root of the elevation difference between MP1 to MP3 and MP2 to MP4 raised to the power of two (for obtaining only positive values) is computed. These values are used to classify the landslides as long or wide according to the following rules:

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

At this step, some wide landslides are classified as long because they are located on short hillslopes but which span on a certain relative elevation along the valley (as the sides of small valleys incised on cuesta escarpments – Fig. 6 c). In this morphologic setting the difference in elevation along the width of the landslides can be the same as the difference in elevation along the length. This is characteristic for the landslides developed on the short (under 500/800 m in length) and laterally extended (2 to 6 km) hillslopes of cuesta escarpments of the Moldavian Plateau (Mărgărint and Niculiță, 2016), and along the banks of gullies which dissect these hillslopes (Fig. 2, 3).

20 For these cases, a further step of processing is required. This step involves the cropping of the DEM for every landslide polygon, the hydrological pre-processing to enforce drainage flow routing and the computation of the slope length (Fig. 6). The slope length estimation requires the use of RSAGA (Brenning and Bangs, 2016) R package, which links SAGA GIS (Conrad et al., 2015). The DEM is preprocessed with the functions Sink Drainage Route Detection and Sink Removal from the Terrain Analysis – Preprocessing module library. The hydrological enforcement is 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) are computed (Fig. 6). The maximum slope length for every landslide polygon is computed using the function Grid Statistics for Polygons from Shapes – Grid Tools tool library.

Using the following rules, only the long landslides class is reclassified:

- 30 - 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.

The confusion matrix and the associated statistics from Table 1 were computed in R stat.

4 Results and Discussions

The results of the proposed algorithm need to be validated against the reality identified in the landslide inventory, to prove its efficiency. The long and wide type was assessed during landslide delineation. Beside field validation, the use of 3D views of LiDAR shading and oriented bounding box (similar to the 3D views from Fig. 6 and 7), allowed correct identification of the direction of slide/flow and the comparison of length and width. Since the long versus wide type was correctly assessed, I call the length and width derived from the oriented bounding box along the slide/flow direction, the correct length and width (Fig. 5, 8). The width and length computed using the oriented bounding box without taking in consideration the slide/flow direction are called bounding box length and width (Fig. 5, 8). These two variables will be biased if the wide landslide type appears in a landslide inventory. The length and width obtained using the proposed algorithms are called estimated length and width because the algorithm fails to classify correctly certain types of wide landslides (false negatives – type II error) and certain types of long landslides (false positive – type I error) (Table 1). I argument that although some landslides are misclassified, from a pure morphometric point of view, the impact is minimized by the algorithm. This can be seen in Figure 8. All four plots from this figure show that estimated length and width are closer to correct length and width than bounding box length and width. These plots show the scatter (Fig. 8a, b) of the difference between the correct landslide dimensions and estimated/bounding box dimensions and the frequency (Fig. 8c, d) of landslide dimensions. From the scatter plots it appears that bounding box dimensions are bigger than the correct dimensions since length is switched with width, the biggest impact being registered for elongated/elliptic landslides (flows). In the case of round landslides (rotational and translational slides), switching the length with width has a smaller impact. The histograms show better the impact of changing the length with the width and that the results of the proposed algorithm resemble with fidelity the shape of the real dimensions distributions. The use of the dimensions of the bounding box has instead a huge impact not essentially on the shape but the magnitude of the distribution (since the wide type of landslides is dominant, 61% – Table 1).

While the interpretation of the differences of the dimensions are instructive and argument the quality of the proposed algorithm in producing results which can be used in geomorphometric analyses, the assessment of the accuracy to predict the correct class is made using the confusion matrix (Stehman, 1997; Powers, 2011) and the ROC graphs (Fawcett, 2006; Powers, 2011). The confusion matrix is presented in Table 1, and its graphic synthesis is presented in the folded plots from Figure 9.

In the first step of the proposed method, from 1327 landslides, 518 (39%) long and 809 wide (61%), 642 landslides (48.4%) are classified as long, of which 478 (36%) are indeed long (true positives - TP). The other 40 (3%) long landslides are misclassified as wide (false negative – FN – type II error). Although wide, 164 (12.4%) landslides are misclassified as long (false positives – FP – type I error). The other 645 (48.6%) wide landslides are correctly classified (true negative - TN).

The wide landslides wrongly classified as long in the first step (FP – type I error), are mainly developed along the banks of incised rivers and gullies, landslide scarps, landslide toes, and cuesta scarps, in situations where there a greater difference in elevation between the flanks than between the scarp and the toe (Fig. 5c). The long landslides misclassified as wide (FN – type II error) 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. 5a). This diagonal position generates a difference of elevation along the width which is greater than the difference along the length.

Overall, the first step has an accuracy of 0.85 (Table 1). The receiver operating characteristic (ROC) curve is used to visually assess the performance of a classification (Fawcett, 2006). The curve plots the true positive rate (benefits, sensitivity) versus the false positive rates (costs, specificity) of the classification (Fawcett, 2006; Powers, 2011). The diagonal line is equivalent with randomly guessing the outcome of the classification (Fawcett, 2006; Powers, 2011). In this first step of the classification, the TP and FP rates are higher, showing that the long type is better classified. Overall, the area under the ROC curve (AUC - Fig. 10), shows that 84.3% landslides are correctly classified. The 95% confidence interval was computed using 2000 stratified bootstrap replicates in pROC sp package (Robin et al., 2011). This interval is in the proximity of the ROC curve, the range of the AUC being of ± 1.93 , which shows the stability of the curve.

In the second step, the best performance was acquired by the multiple flow algorithm for computing slope length. This algorithm computes flow distance which is more similar to hillslope length, because flow is modeled 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 classify as wide and only 59 as long. The Multiple Flow Direction algorithm classify 837 as wide and 478 landslides as long.

In the first step of the proposed method, 685 landslides (51.6%) are classified as long, of which 425 (32%) are indeed long (true positives - TP). The other 93 (7%) long landslides are misclassified as wide (false negative – FN – type II error). Although wide, 65 (4.9%) landslides are misclassified as long (false positives – FP – type I error). The other 744 (56.1%) wide landslides are correctly classified (true negative - TN).

The majority of the wide landslides classified as long in the second step (FP – type I error) 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. 5d). The long landslides classified as wide (FN – type II error) 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. 5b). These landslides have gullies or mounds, which interrupts the diffusion of the flow, and the flow algorithms do not route the flow from the crown to the toe of the landslide.

Overall, the second step has an accuracy of 0.88 (Table 1). All the statistics associated with the confusion matrix (Table 1) show that in the second step the classification performs slightly better. In this second step, the ROC curve shows that the TP rate is lower, and FP rate is higher, meaning that the wide type is better classified. Overall, 87.8% landslides are correctly classified (AUC value – Fig. 10). The confidence interval is lower, ± 1.85 , showing an increase in the stability of the classification.

Although the proposed algorithm performs well, any inconsistency between the shape of the earth portrayed by the DEM and the delineated landslide morphology will give misclassification cases. This scenario is to be expected from inventories where 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 1 elevation 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:

- 5 - 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. 5c), but we believe that in general, this situation will not introduce errors since we have not met any such
10 cases.

Inconsistencies of the type described above are to be expected in a substantial 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 also be used for checking landslide inventories for geomorphological topologic inconsistencies.

15 The main challenges 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 failure to separate long from wide landslides because of their diagonal position (mainly related to the false negative cases, wide landslides classified as long).

20 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
25 to the computation of the bounding box. Probably, using a 3D approach in computing the bounding box, from a 3D vector of the landslide boundary (vertex of the polygon with elevation altitude from the DEM) could resolve this, but research is needed. For resolving the flow length estimation, responsible for the type II error, better methods for slope length estimation are needed.

5 Conclusions

The proposed method can discern between long and wide landslides in 87.8% of cases. The misclassified cases have a specific
30 geometry and geomorphologic topology, and the algorithm fails to identify them in the right class. The majority of these landslides are round, rhombic or hexagonal, and oriented diagonally to the downslope direction of the hillslope. The slide/flow direction of these landslides is diagonal to the bounding box. Although the algorithms fail to recognize them as long or wide,

the errors are minimal when their lengths and widths are used, because the difference between them is not big. Of course, if the direction of flow is needed (as aspect value), the algorithm fails to identify the right direction. The minority of misclassified 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 slope length estimation does not exceed the oriented bounding box length, a situation which requires further investigations on the possibilities of using other methods to estimate slope length. Overall, the presented approach managed to identify the majority of the long versus wide landslides from a landslide inventory and represents a starting point in deriving algorithms for automatic geomorphometric analysis of landslides.

Acknowledgments

The author gratefully acknowledges partial support from 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).

15

References

- Ardizzone F., Cardinali, M., Carrara, A., Guzzetti, F. and Reichenbach, P.: Impact of mapping errors on the reliability of landslide hazard maps, *Nat. Hazards Earth Syst. Sci.*, 2, 3-14, doi:10.5194/nhess-2-3-2002, 2002.
- Ardizzone F., Cardinali, M., Galli, M., Guzzetti, F. and Reichenbach, P.: Identification and mapping of recent rainfall-induced landslides using elevation data collected by airborne Lidar, *Nat. Hazards Earth Syst. Sci.*, 7, 637–650, doi:10.5194/nhess-7-637-2007, 2007.
- Baldo, M., Bilocchi, C., Chiocchini, U., Giordan, D., Lollino, G.: LIDAR monitoring of mass wasting processes: The Radicofani landslide, Province of Siena, Central Italy, *Geomorphology*, 105, 193-201, doi:10.1016/j.geomorph.2008.09.015, 2009.
- Barlow, J., Martin, Y., Franklin, S.: Detecting translational landslide scars using segmentation of Landsat ETM+ and DEM data in the northern Cascade Mountains, British Columbia, *Canadian Journal of Remote Sensing*, 29, 510–517, 2003.
- Barlow, J., Franklin, S. and Martin, Y.: High spatial resolution satellite imagery, DEM derivatives, and image segmentation for the detection of mass wasting processes, *Photogramm. Eng. Remote. Sens.*, 72, 687-692, 2006.
- Berti, M., Corsini, A. and Daehne, A.: Comparative analysis of surface roughness algorithms for the identification of active landslides, *Geomorphology*, 182, 1-18, doi:10.1016/j.geomorph.2012.10.022, 2013.

- Bivand, R.S., Pebesma, E. and Gomez-Rubio, V.: Applied spatial data analysis with R, UseR! Series, Springer, 2013.
- Borghuis, A.M., Chang, K. and Lee, H.Y.: Comparison between automated and manual mapping of typhoon triggered landslides from SPOT-5 imagery, *Int. J. Rem. Sens.*, 28, 1843-1856, doi:10.1080/01431160600935638, 2007.
- Brenning, A. and Bangs, D.: RSAGA: SAGA geoprocessing and terrain analysis in R, <https://cran.r-project.org/web/packages/RSAGA/index.html>, 2016.
- 5 Brunnsden, D.: Mass movement; the research frontier and beyond: a geomorphological approach, *Geomorphology*, 7, 85-128, doi:10.1016/0169-555X(93)90013-R, 1993.
- Carrara, A. and Merenda, L.: Landslide inventory in northern Calabria, southern Italy, *Geol. Soc. Am. Bull.*, 87, 1153-1162, doi: 10.1130/0016-7606(1976)87<1153:LIINCS>2.0.CO;2, 1976.
- 10 Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., Reichenbach, P.: GIS techniques and statistical models in evaluating landslide hazard, *Earth. Surf. Proc. Landf.*, 16, 427-445, doi: 10.1002/esp.3290160505, 1991.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Whichmann, V. and Böhner, J.: System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, *Geosci. Model Dev.*, 8, 1991-2007, doi:10.5194/gmd-8-1991-2015, 2015.
- 15 Cruden D.M. and Varnes, D.J.: Landslide types and processes, in: *Landslides: Investigation and Mitigation*, edited by: Turner, A. K. and Schuster, R. L., Washington, D.C., National Research Council, Transportation Research Board Special Report 247, 36-75, 1996.
- Dewitte, O. and Demoulin, A.: Morphometry and kinematics of landslides inferred from precise DTMs in West Belgium, *Nat. Hazards Earth Syst. Sci.*, 5, 259-265, doi:10.5194/nhess-5-259-2005, 2005.
- 20 Fawcett, T.: An Introduction to ROC analysis, *Patt. Recog. Lett.*, 27, 861 – 874, doi:10.1016/j.patrec.2005.10.010, 2006.
- Freeman, H. and Shapira, R.: Determining the minimum-area encasing rectangle for an arbitrary closed curve, *Commun. ACM*, 18, 409-413, doi:10.1145/360881.360919, 1975.
- Freeman, G.T.: Calculating catchment areas with divergent flow based on a regular grid, *Comput. Geosci.*, 17, 413-422, doi:10.1016/0098-3004(91)90048-I, 1991.
- 25 Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F. and Reichenbach, P.: Comparing landslide inventory maps, *Geomorphology*, 94, 268-289, doi:10.1016/j.geomorph.2006.09.023, 2008.
- Geertsema, M., Highland, L. and Vaugeouis, L.: Environmental impact of landslides, in: *Landslides – Disaster Risk Reduction*, edited by: Kyoji, S., Canuti, P., Springer, 589-607, doi: 10.1007/978-3-540-69970-5_31, 2009.
- Glenn, N.F., Streutker, D.R., Chadwick, D.J., Thackray, G.D. and Dorsch, S.J.: Analysis of LiDAR-derived topographic information for characterizing and differentiating landslide morphology and activity, *Geomorphology*, 73, 131-148, doi:10.1016/j.geomorph.2005.07.006, 2006.
- 30 Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K.-T.: Landslide inventory maps: New tools for an old problem, *Earth-Sci. Rev.*, 112(1-2), 42-66, doi:10.1016/j.earscirev.2012.02.001, 2012.

- Hattanji, T. and Moriwaki H.: Morphometric analysis of relic landslides using detailed landslide distribution maps: implications for forecasting travel distance of future landslides, *Geomorphology*, 103, 447-454, doi:10.1016/j.geomorph.2008.07.009, 2009.
- Jaboyedoff, M., Oppikofer, T., Abellán, A., Derron, M.-H., Loye, A., Metzger, R. and Pedrazzini, A.: Use of LIDAR in landslide investigations: a review, *Nat. Hazards*, 61, 5-28, doi:10.1007/s11069-010-9634-2, 2012.
- Keaton, J.R. and DeGraff, J.V.: Surface observation and geologic mapping, in: *Landslides: Investigation and Mitigation*, edited by: Turner, A. K. and Schuster, R. L., Washington, D.C., National Research Council, Transportation Research Board Special Report 247, 178-230, 1996.
- Kjekstad, O. and Highland, L.: Economic and social impacts of landslides, in: *Landslides – Disaster Risk Reduction*, edited by: Kyoji, S., Canuti, P., Springer, 573-587, doi:10.1007/978-3-540-69970-5_30, 2009.
- Lyons, N.J., Mitasova, H. and Wegmann, K.W.: Improving mass-wasting inventories by incorporating debris flow topographic signatures, *Landslides*, 11, 385-397, doi:10.1007/s10346-013-0398-0, 2014.
- Mărgărint, M.C., and Niculiță, M.: Landslide type and pattern in Moldavian Plateau, NE Romania, in: *Landform Dynamics and Evolution in Romania*, edited by: Rădoane, M. and Stroe-Vespremeanu, A., Springer, 271-304, doi: 10.1007/978-3-319-32589-7_12, 2016.
- Malamud, B.D., Turcotte, D.L., Guzzetti, F. and Reichenbach, P.: Landslide inventories and their statistical properties, *Earth Surf. Proc. Land.*, 29 (6), 687-711, doi: 10.1002/esp.1064, 2004.
- 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.
- Martha, T. R., Kerle, N., Jetten, V., Van Westen, C. J. and Vinod Kumar, K.: Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods, *Geomorphology*, 116, 24-36, doi:10.1016/j.geomorph.2009.10.004, 2010.
- Martha, T. R., Kerle, N., Van Westen, C. J., Jetten, V. and Vinod Kumar, K.: Object-oriented analysis of multi-temporal panchromatic images for creation of historical landslide inventories, *ISPRS J. Photogramm.*, 67, 105-119, doi:10.1016/j.isprsjprs.2011.11.004, 2012.
- Martin, Y. and Franklin, S.: Classification of soil- and bedrock-dominated landslides in British Columbia using segmentation of satellite imagery and DEM data, *Int. J. Remote Sens.*, 26, 1505–1510, 2005.
- McKean, J and Roering J.: Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry, *Geomorphology*, 57, 331-351, doi:10.1016/S0169-555X(03)00164-8, 2004.
- Mondini, A. C., Guzzetti, F., Reichenbach, P., Rossi, M. and Ardizzone, F.: Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images, *Remote Sens. Environ.*, 115(7), 1742-1757, doi:10.1016/j.rse.2011.03.006, 2011.

- Mondini, A. C., Marchesini, I., Rossi, M., Chang, K.-T., Pasquariello, G. and Guzzetti, F.: Bayesian framework for mapping and classifying shallow landslides exploiting remote sensing and topographic data, *Geomorphology*, 201, 135-147, doi:10.1016/j.geomorph.2013.06.015, 2013.
- Nichol, J. and Wong, M.S.: Satellite remote sensing for detailed landslide inventories using change detection and image fusion, *Int. J. Remote Sens.*, 29, 1913-1926, doi:10.1080/01431160512331314047, 2005.
- Niculita, M.: Automatic extraction of landslide flow direction using geometric processing and DEMs, in: *Geomorphometry for Geosciences*, edited by Jasiewicz, J., Zwoliński, Z., Mitasova H. and Hengl, T., Bogucki Wydawnictwo Naukowe, Adam Mickiewicz University in Poznań - Institute of Geoecology and Geoinformation, Poznań, Poland, 201-203, <http://geomorphometry.org/system/files/Niculita2015geomorphometry.pdf>, 2015.
- Niculita, M. and Mărgărint, M.C.: Testing landslide susceptibility uncertainty propagation due to the data source of the landslide inventory: satellite imagery versus LIDAR, *Geophysical Research Abstracts*, 17, 10043-2, 2015.
- Petley, D.N.: Global patterns of loss of life from landslides, *Geology*, 40, 927-930, doi:10.1130/G33217.1, 2012.
- Petley, D.N., Crick, W.D.O. and Hart, A.B.: The use of satellite imagery in landslide studies in high mountain areas, *Proceedings of the 23rd Asian Conference on Remote Sensing (ACRS 2002)*, Kathmandu, November 2002.
- Pike, R.J.: The geometric signature: quantifying landslide-terrain types from Digital Elevation Models, *Math. Geol.*, 20(5), 491-511, doi: 10.1007/BF00890333, 1988.
- O'Callaghan, J.F. and Mark, D.M.: The extraction of drainage networks from digital elevation data, *Comp. Vision Graph.*, 28, 323-344, doi:10.1016/S0734-189X(84)80011-0, 1984.
- Powers, D.M.W.: Evaluation: from precision, recall and f-measure to ROC, informedness, markedness & correlation, *J. Mach. Learn. Tech.*, 2, 37-63, 2011.
- Quinn, P.F., Beven, K.J., Chevallier, P. and Planchon, O.: The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models, *Hydrol. Process.*, 5, 59-79, doi:10.1002/hyp.3360050106, 1991.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C. and Muller, M.: pROC: an open-source package for R and S+ to analyze and compare ROC curves, *BMC Bioinform.*, 12, doi:10.1186/1471-2105-12-77, 2011.
- R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>, 2015.
- Rossi, M., Mondini, A.C., Marchesini, I., Santangelo, M., Bucci, F. and Guzzetti, F. (2013), Landslide morphometric signature. 8th IAG/AIG International Conference on Geomorphology Geomorphology and Sustainability Paris, France, August 27-31, 2013.
- Schulz, W.H.: Landslides mapped using LIDAR imagery, Seattle, Washington, U.S. Geol. Surv. Open-File Rep., 2004-1396, U.S. Department of the Interior, U.S. Geol. Surv., 11 pp, 2004.
- Schulz, W.H.: Landslide susceptibility revealed by LIDAR imagery and historical records, Seattle, Washington, *Eng. Geol.*, 89, 67-87, doi:10.1016/j.enggeo.2006.09.019, 2007.

- Stehman, S.V.: Selecting and interpreting measures of thematic classification accuracy, *Remote Sens. Environ.*, 62, 77-89, doi:10.1016/S0034-4257(97)00083-7, 1997.
- Taylor, F.E. and Malamud, B.D.: The statistical distributions of landslide length to width ratios, *Geophysical Research Abstracts*, 14, 826, 2012.
- 5 Taylor, F.E., Malamud, B.D., Witt, A.: What Shape is a Landslide? Statistical Patterns in Landslide Length to Width Ratio, *Geophysical Research Abstracts*, 14, 10191, 2015.
- Toussaint, G. T.: Solving geometric problems with the rotating calipers, in: *Proceedings of the 1983 IEEE MELECON*, Athens, Greece: IEEE Computer Society, 1983.
- Van Den Eeckhaut, M. and Hervás, J.: Landslide inventories in Europe and policy recommendations for their interoperability and harmonization. A JRC contribution to the EU-FP7 SafeLand project, *JRC Sci. and Policy Rep.*, EUR 25666EN, doi:10.2788/75587, 2012.
- 10 Van Den Eeckhaut, M., Kerle, N., Poesen, J. and Hervás, J.: Object oriented identification of forested landslides with derivatives of single pulse LiDAR data, *Geomorphology*, 173-174, 30-42, doi:10.1016/j.geomorph.2012.05.024, 2012.
- Van Den Eeckhaut, M., Reichenbach, P., Guzzetti, F., Rossi, M. and Poesen, J.: Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardennes, Belgium, *Nat. Hazards Earth Syst. Sci.*, 9, 507-521, doi:10.5194/nhess-9-507-2009, 2009.
- 15 Van Den Eeckhaut, M., Poesen, J., Gullentops, F., Vandekerckhove, L. and Hervás, J.: Regional mapping and characterisation of old landslides in hilly regions using LiDAR-based imagery in Southern Flanders, *Geomorphology*, 75, 721-733, doi:10.1016/j.yqres.2011.02.006, 2011.
- 20 Van Den Eeckhaut, M., Poesen, J., Verstraeten, G., Vanacker, V., Moeyersons, J., Nyssen, J. and van Beek, L.P.H.: The effectiveness of hillshade maps and expert knowledge in mapping old deep-seated landslides, *Geomorphology*, 67, 351-363, doi:10.1016/j.geomorph.2004.11.001, 2005.
- Van Westen, C.J.: *Application of Geographic Information Systems to landslide hazard zonation*, ITC Publication no. 15, 1993.
- Wollschläger, D.: Analyzing shape, accuracy, and precision of shooting results with shotGroups, [https://cran.r-](https://cran.r-project.org/web/packages/shotGroups/vignettes/shotGroups.pdf)
- 25 [project.org/web/packages/shotGroups/vignettes/shotGroups.pdf](https://cran.r-project.org/web/packages/shotGroups/vignettes/shotGroups.pdf), 2016.

Table 1 The confusion matrix for the classification performed in the two steps of the algorithm and the associated statistics

Total population		True condition				Prevalence		
		Condition positive (Long) = 518		Condition negative (Wide) = 809				
Predicted condition	1327					0.39		
	Predicted condition positive (Long)	True positive		False positive (Type I error)		Positive prediction value (PPV), Precision	False discovery rate (FDR)	
	Step 1 = 642	Step 1	478	Step 1	164	0.74	0.26	
	Step 2 = 685	Step 2	425	Step 2	65	0.62	0.62	
	Predicted condition negative (Wide)	False negative (Type II error)		True negative		False omission rate (FOR)	Negative predictive value (NPV)	
	Step 1 = 490	Step 1	40	Step 1	645	0.08	1.32	
	Step 2 = 837	Step 2	93	Step 2	744	0.11	0.89	
	Accuracy (ACC)	True positive rate (TPR), Sensitivity, Recall	0.92	False positive rate (FPR), Fall-out	0.2	Positive likelihood ratio (LR+)	4.6	Diagnostic odd ratio (DOR)
	0.85		0.82		0.1		8.2	
	0.88							
	False negative rate (FNR)	0.08	True negative rate (TNR), Specificity (SPC)	0.8	Negative likelihood ratio (LR-)	0.1	46.00	
		0.18		0.9		0.2	41.00	

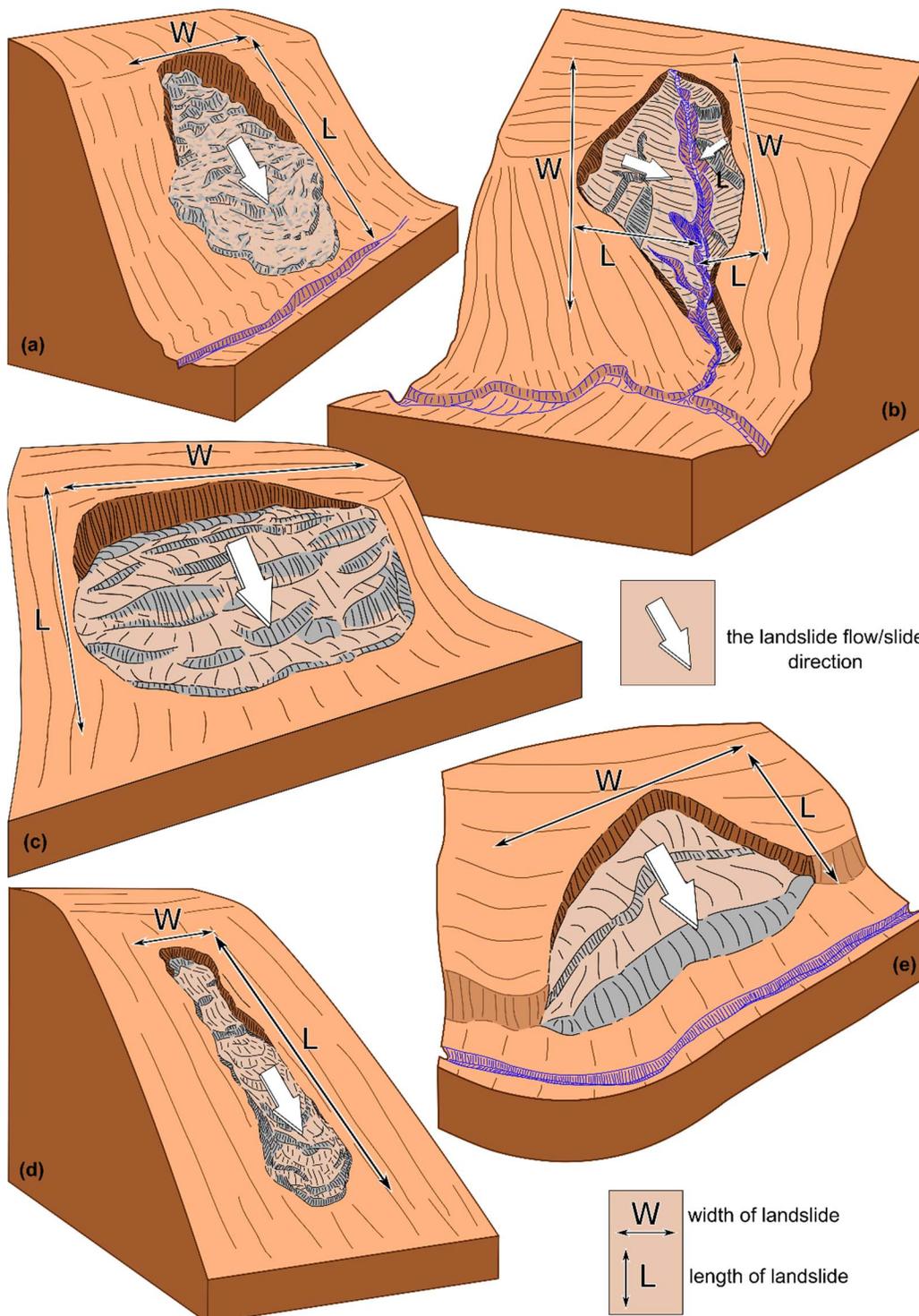


Figure 1: Schematic drawings of long and wide cases of landslides: a) rotational slide – long type, b) gully bank slides – wide type, c) translational slide – wide type, d) flow – long type, e) river bank slide – wide type.

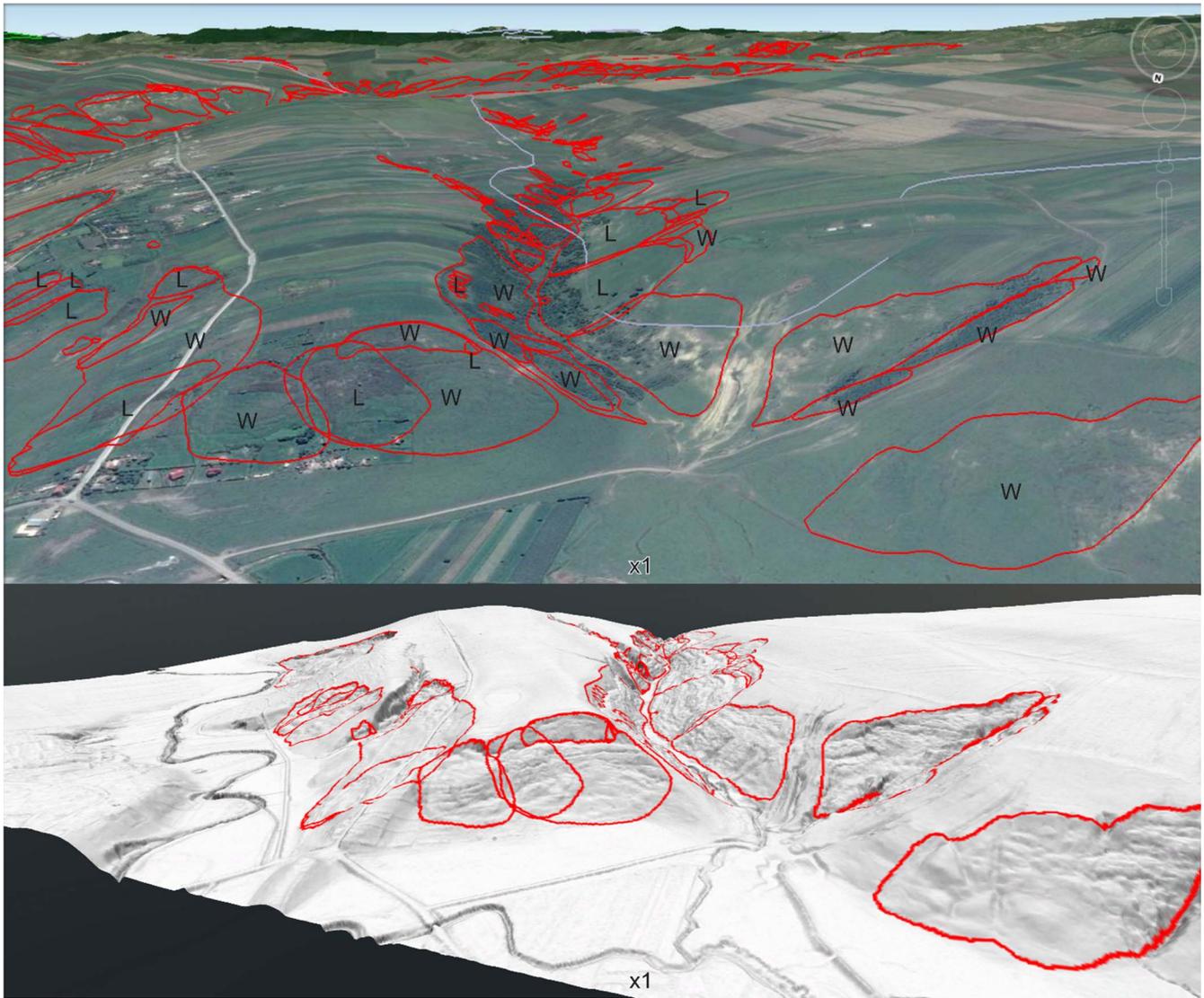


Figure 2: Geomorphologic cases of long (L) and wide (W) type of landslides from the study area (top is a 3D view Google Earth capture; bottom is a 3D view of the LiDAR DEM shading; both views have the landslide polygons in red, overlaid).

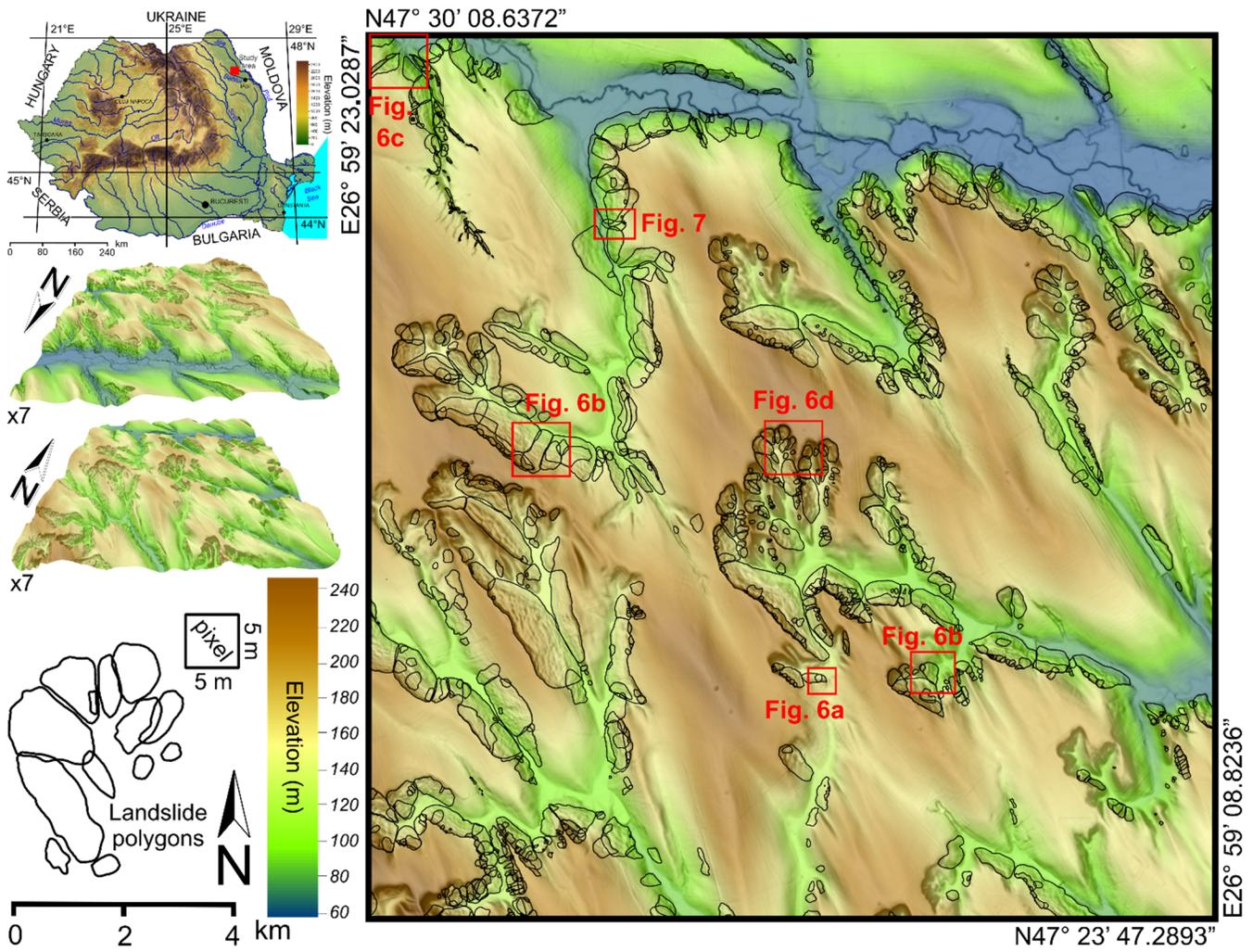


Figure 3: The study area and the landslide inventory overlaid on shaded LiDAR DEM (upper-left is an inset showing the position of the study area in Romania; under the inset is displayed a set of 3D views of the study area landforms from north and south; the red boxes from the right map indicate the insets from Fig. 6 and Fig. 7).

5

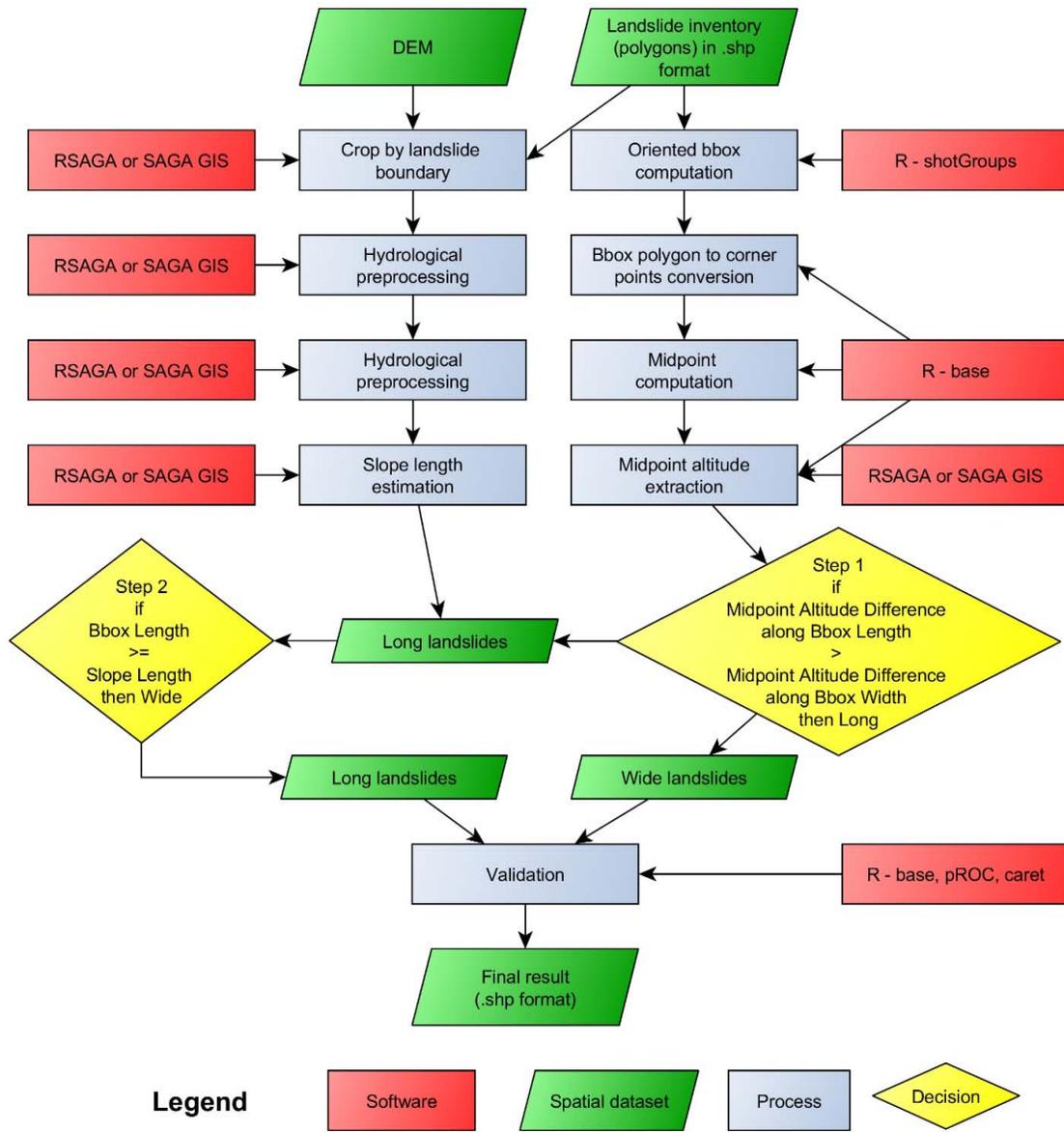


Figure 4: Workflow of the proposed algorithm implemented as an R script.

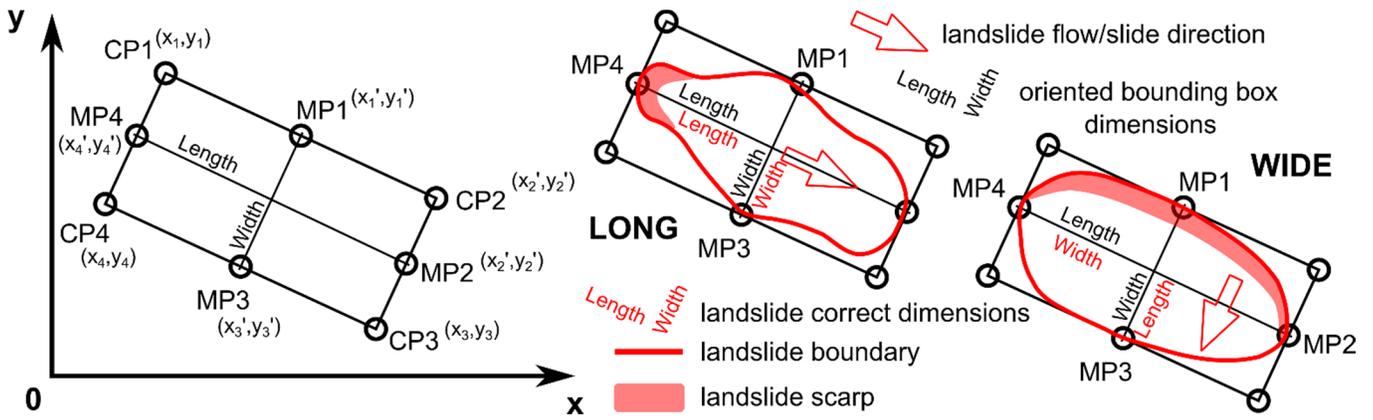


Figure 5: Geometry of the minimum oriented bounding box (left is the geometry of the oriented bounding box; center is the geometry of a long landslide bounding box; right is the geometry of a wide landslide box geometry).

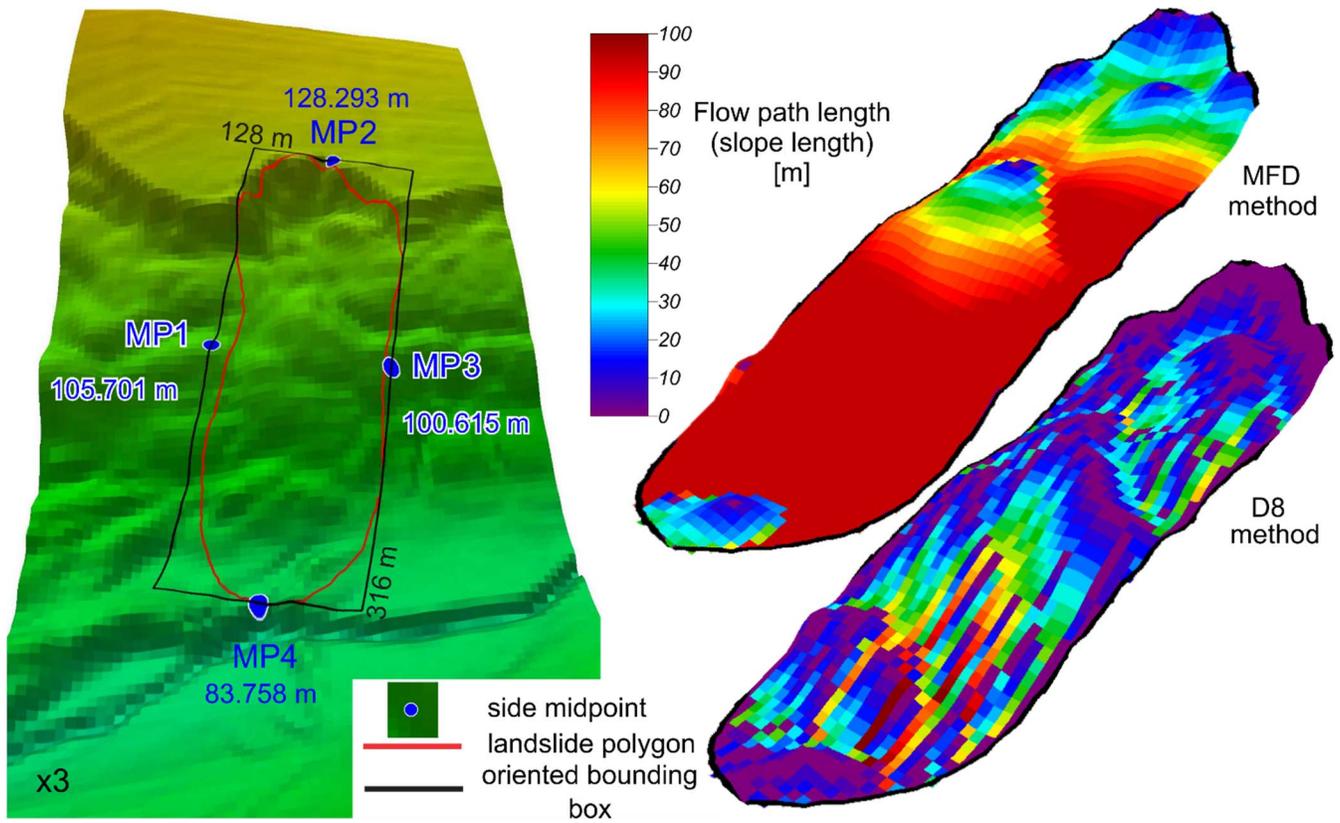


Figure 7: 3D view of a long landslide (left) and the flow lengths for it (right) (MFD is the flow length computed with the Multiple Flow Direction algorithm and D8 is the flow length computed with the Deterministic 8).

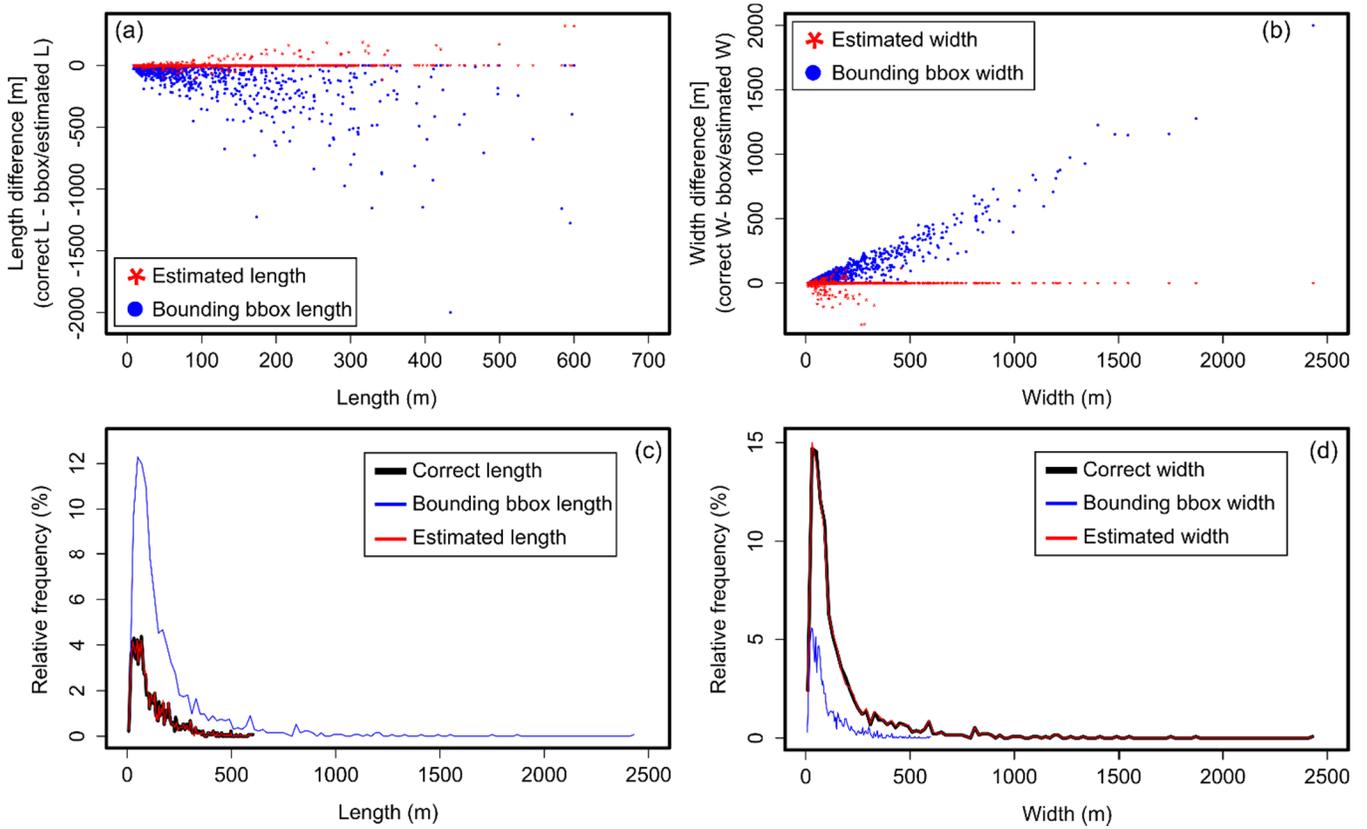


Figure 8: Plots of the difference between the real length/width, length/width of the oriented bounding box and length/width estimated with the proposed algorithm: a) scatter plot of the difference between correct length and the bbox/estimated length, b) scatter plot of the difference between correct width and the bbox/estimated width, c) histograms of the three types of lengths, d) histograms of the three types of widths.

5

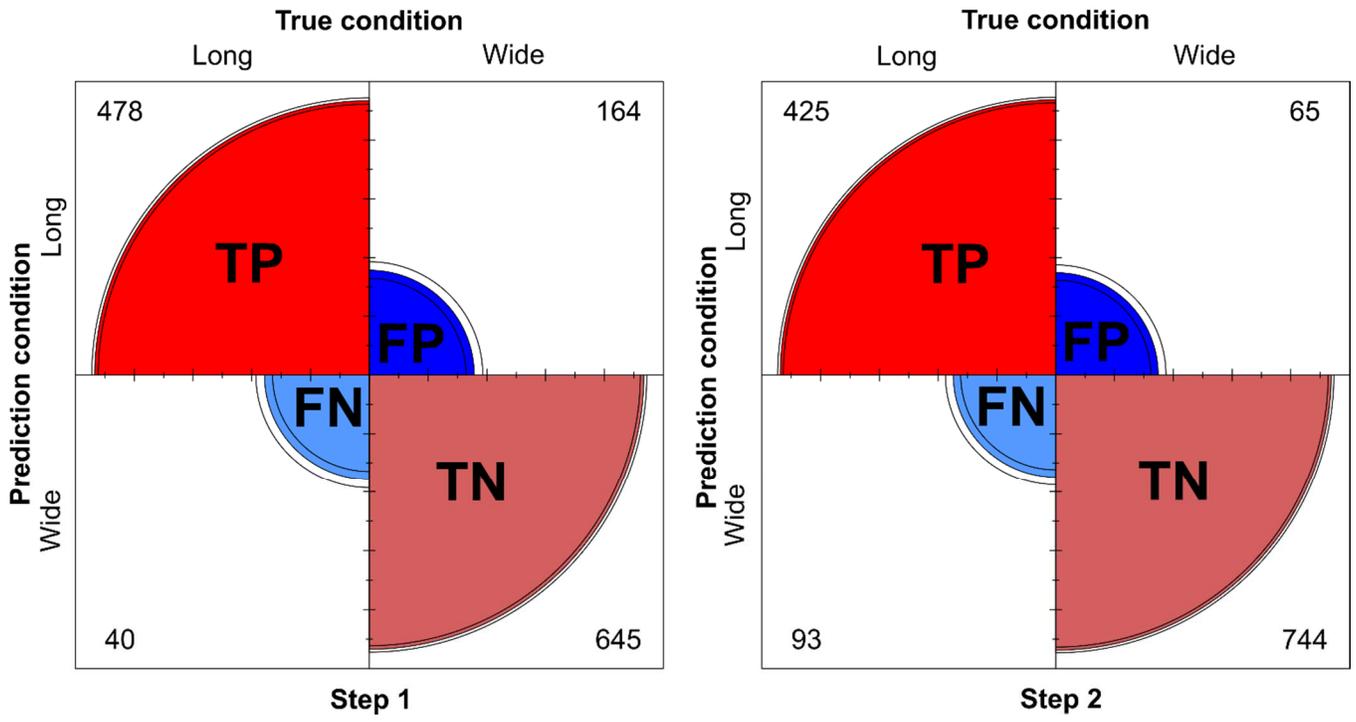


Figure 9: Four folded plots of the confusion matrices for the two steps of the algorithm (TP = true positive, TN = true negative, FP = false positive, FN = false negative).

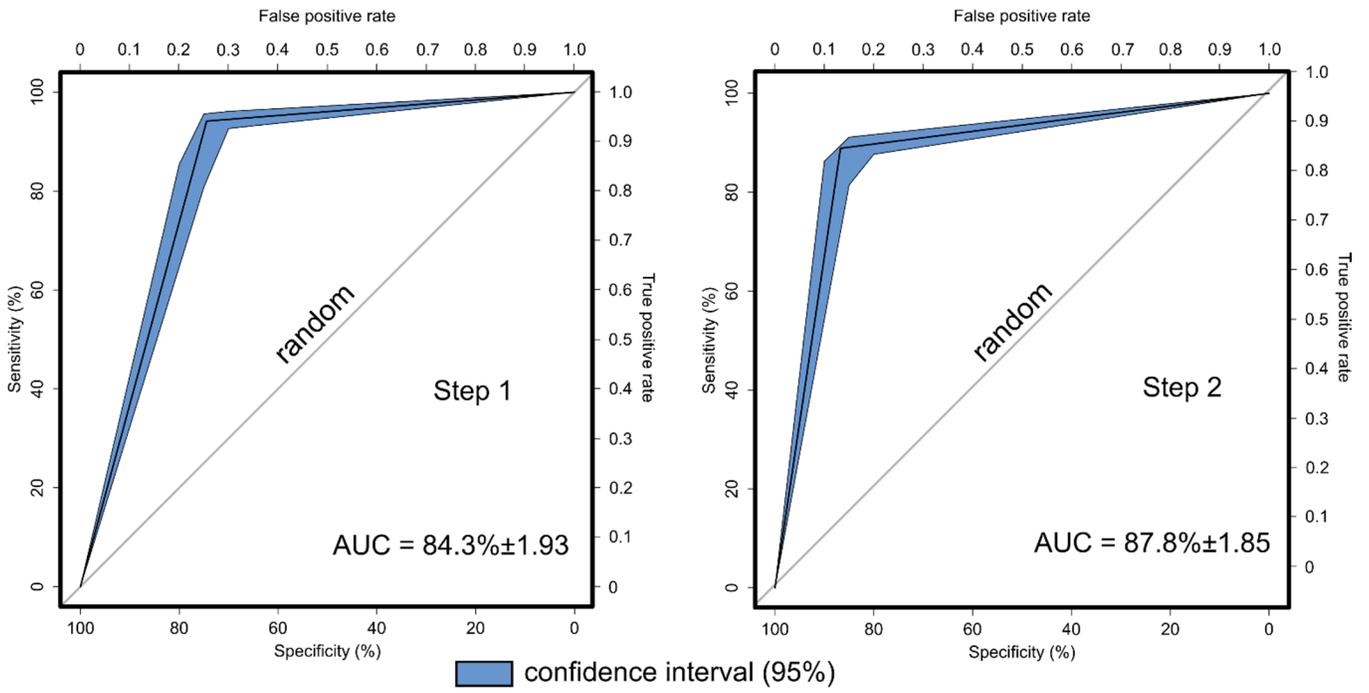


Figure 10: ROC curves for the two steps of the algorithm, their 95% confidence interval and the AUC values.

5