

# LANDSLIDES IDENTIFICATION USING AIRBORNE LASER SCANNING DATA DERIVED TOPOGRAPHIC TERRAIN ATTRIBUTES AND SUPPORT VECTOR MACHINE CLASSIFICATION

Kamila Pawłuszek, Andrzej Borkowski

Institute of Geodesy and Geoinformation, Wrocław University of Environmental and Life Sciences

[kamila.pawluszek@igig.up.wroc.pl](mailto:kamila.pawluszek@igig.up.wroc.pl)

[andrzej.borkowski@igig.up.wroc.pl](mailto:andrzej.borkowski@igig.up.wroc.pl)

Commission VIII, WG VIII/1

**KEY WORDS:** Airborne laser scanning, support vector machine, landslide mapping, Principal Component Analysis

## ABSTRACT:

Since the availability of high-resolution Airborne Laser Scanning (ALS) data, substantial progress in geomorphological research, especially in landslide analysis, has been carried out. First and second order derivatives of Digital Terrain Model (DTM) have become a popular and powerful tool in landslide inventory mapping. Nevertheless, an automatic landslide mapping based on sophisticated classifiers including Support Vector Machine (SVM), Artificial Neural Network or Random Forests is often computationally time consuming. The objective of this research is to deeply explore topographic information provided by ALS data and overcome computational time limitation. For this reason, an extended set of topographic features and the Principal Component Analysis (PCA) were used to reduce redundant information. The proposed novel approach was tested on a susceptible area affected by more than 50 landslides located on Rożnów Lake in Carpathian Mountains, Poland. The initial seven PCA components with 90% of the total variability in the original topographic attributes were used for SVM classification. Comparing results with landslide inventory map, the average user's accuracy (UA), producer's accuracy (PA), and overall accuracy (OA) were calculated for two models according to the classification results. Thereby, for the PCA-feature-reduced model UA, PA, and OA were found to be 72%, 76%, and 72%, respectively. Similarly, UA, PA, and OA in the non-reduced original topographic model, was 74%, 77% and 74%, respectively. Using the initial seven PCA components instead of the twenty original topographic attributes does not significantly change identification accuracy but reduce computational time.

## 1. INTRODUCTION

Landslides are natural hazard causing significant damages to the environment in many countries. Landslides can be fatal and can also destroy or damage natural landforms. Moreover, landslides have disruptive impact on man-made structures such as buildings, agricultural and forestal lands and on water in rivers and streams (Akgun and Erkan, 2016; Schuster and Fleming, 1986). Because of the increasing socio-awareness of landslide impact on the environment, efficient landslide assessment is required (Aleotti and Chowdhury, 1999). Therefore, many countries created or are creating their own national or regional landslide databases (LDBs). This is a fundamental source for quantitative zoning of landslide susceptible areas (Van Den Eeckhaut and Hervás, 2012).

Currently, landslide mapping methods involve field inventories, which are time consuming. Alternative landslide mapping methods applied photogrammetric approaches and analysis of Digital Elevation Model (DEM). Here airborne laser scanning (ALS) achieved big popularity by providing high resolution topographic information overwhelming over traditional surveying techniques (Tarolli, 2014). Nowadays, applying ALS data in landslide mapping seems to be a standard tool. Nevertheless, many studies beside the ALS data use expert knowledge, experience and very often familiarity with the study sites. Unfortunately, very few researchers made attempt to automate the process of landslide mapping using computer-aided methods (Van Den Eeckhaut et al., 2012). In automatic approaches, DEM and first and second-order derivatives of DEM such as: slope, aspect, curvature, topographic roughness among

other derivatives are very beneficial in the landslide recognition. Very promising results are obtained based on surface roughness investigation. McKean and Roering (2004) applied local variability of unit vector orientations of slope and aspect. Their results indicate that contrast analysis of surface roughness can be used to identify bedrock landslides, outline their spatial extent and even examine landslide internal kinematics. Furthermore, impressive results were obtained by Booth et al. (2009) using spectral analysis and high-resolution topographic data. They compared the results with independent landslide inventory maps and correctly classified an average of 82% of the terrain in five study areas in Washington and Portland Hills, Oregon. Chen et al. (2014) used DEM-derived features and random forests algorithm for semi-automatic landslide mapping. They achieved overall accuracy of 78.24%. Another approach using object-oriented image analysis (OOA) was proposed by Van Den Eeckhaut et al. (2012). The results obtained show that OOA using DEM-derivatives allows them to recognize more than 90% of the main scarps and 70% of the landslide bodies. Recently, another approach was presented by Leshchinsky et al. (2015). They proposed a novel algorithm for the automatic and consistent landslide deposits mapping. The authors applied the Contour Connection Method (CCM) and a high agreement with manual delineated landslides deposit was achieved.

Applying many DEM-derivatives and sophisticated classifiers including Support Vector Machine (SVM), Artificial Neural Network or Random Forests are often computationally time consuming. Moreover, high resolution of the DEM data and their derivatives for large areas are often big data sets, which require powerful processing platforms to handle them.

Therefore, the objective of this research is to deeply explore topographic information provided by the ALS data and to decrease the computational time in semi-automatic and computer-aided landslide mapping. For this reason, an extended set of topographic features and the Principal Component Analysis (PCA) were used to reduce redundant information and to accelerate the classification process.

## 2. DEM AND ITS DERIVATIVES

The ALS data was obtained in the framework of the ISOK project (Pawłuszek et al., 2014). Point cloud with resolution of 4-6 points/m<sup>2</sup> and overall accuracy less than 15 cm was used. According to Pawłuszek et al. (2014) the height component accuracy of the ISOK data does not exceed 23 cm for forested areas. Based on the DEM, 19 DEM-derivative layers were prepared. Because of the large number of the DEM-derivatives, only multiple shaded relief derivative is presented in figure 1. Main information, calculation patterns and references for other derivatives are collected in table 1.

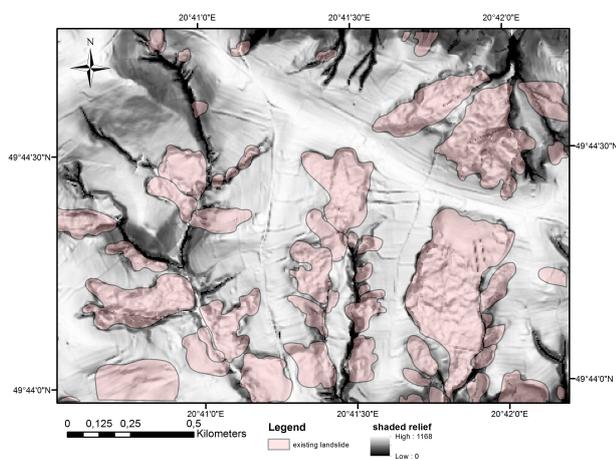


Figure 1: Multiple shaded relief and the existing landslides (pink areas)

## 3. GENERAL SETTINGS OF THE STUDY AREA

The study area is located in the central part of the Outer West Carpathians in Poland and it approximately covers an area of 2.8 km<sup>2</sup> (figure 2). The geographical location of this area is 49° 44'N to 49° 45'N latitude and 20° 40'E to 20° 43'E longitude. The altitude of the study area ranges from 267.48m to 477.77m. Maximum slope angle in the study area is 58.63°. From the geological point of view, the study area is situated within the Cieżkowice Foothills, close to the bank of Rożnów Lake (Starkel, 1972). According to the hydrological data, precipitation occurs frequently in the form of rain and snow throughout the winter. The annual mean precipitation of this area over the period of 1981–2010 is 800 mm (Woźniak et al., 2013). The main reason of the landslide occurrence within the study area is sedimentary rocks and rainfalls. Moreover, landslide activity is mostly associated to the abundant rainfall, fluctuation of water level in the Rożnów Lake and the flysch type of rocks (Borkowski et al., 2011). According to figure 2 three diverse land uses can be observed: forestial, agricultural and urban areas. As can be seen, landslides mainly occur in forested areas and cropland. Therefore, it is worth to emphasize usefulness of ALS in landslide mapping within study area, where traditional field inventories are challenging.

DEM-derivatives	Information and references
DEM	[ArcGIS™] Van Westen et al. (2008)
slope	[Spatial Analyst in ArcGIS™] Van Westen et al. (2008)
standard deviation of shaded relief	moving standard deviation filter using 3 x 3 pixel kernel [Spatial Analyst in ArcGIS™]
openness	difference between original DEM and $DEM_{ki}$ , where $DEM_{ki}$ is interpolated DEM with 9 x 9 moving average kernel Van Den Eeckhaut et al. (2012) [Raster Calculator in ArcGIS™]
topographic roughness	GIS Geomorphometry & Gradient Metrics toolbox by Evans et al. (2015)
contour density	20cm contour density per circle with radius of 3m [Python in ArcGIS™]
area solar radiation (ASR)	ASR represents the solar energy for a given pixel and specific date [Spatial analyst toolbox in ArcGIS™]
morphological gradient	represents difference between the dilation and the erosion of the DEM-image [Python in ArcGIS™]
topographic position index (TPI)	Land Facet Corridor Designer by Jenness et al. (2013)
skewness	represents the asymmetry of the probability distribution, moving skewness index filter wit 3x3 pixel kernel [Python in ArcGIS™]
curvature	[Spatial analyst toolbox in ArcGIS™] (Van Den Eeckhaut et al., 2012)
integrated moisture index (IMI)	[GIS Geomorphometry & Gradient Metrics toolbox by Evans et al. (2015)]
stream power index (SPI)	describes potential flow erosion at the specific location of the surface. [Spatial analyst Tools and Raster Calculator in ArcGIS™] (Akgun et al., 2008)
compound topographic index (CTI)	[GIS Geomorphometry & Gradient Metrics toolbox by Evans et al. (2015)]
semivariogram	moving semivariogram index filter using 9x9 pixel kernel [Python in ArcGIS™]
slope/aspect transformation	[GIS Geomorphometry & Gradient Metrics toolbox by Evans et al. (2015)]
difference between DEM and polynomial surface fitted into DEM	[Raster calculator and interpolation function in ArcGIS™]
deviation from trend	[GIS Geomorphometry & Gradient Metrics toolbox by Evans et al. (2015)]
standard deviation of slope	moving standard deviation filter using 3 x 3 pixel kernel [ArcGIS™]
multiple shaded relief	[Spatial analyst toolbox in ArcGIS™] (Eeckhaut et al., 2007) figure 2

Table 1: Main information, calculation patterns and references about DEM-derivative layers

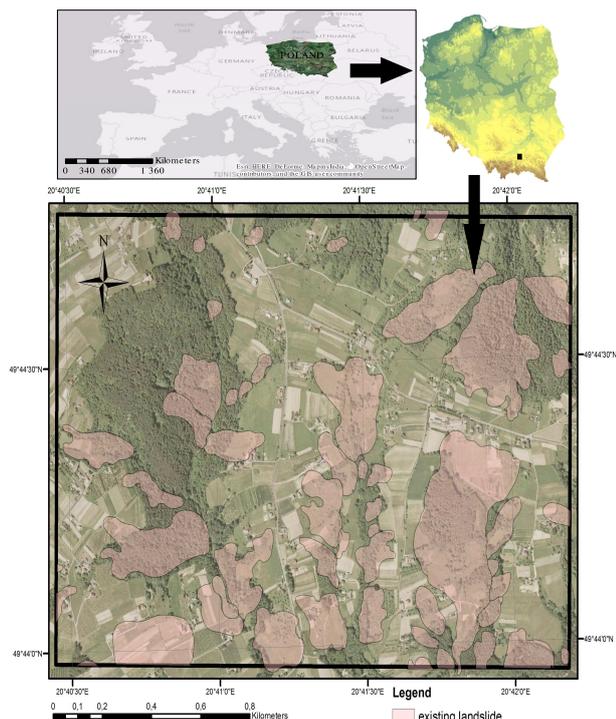


Figure 2: Location of the study area with the existing landslides and orthoimage

#### 4. METHODOLOGY

The methodology is presented in figure 3. Based on the DEM, 19 first or second order DEM-derivatives were calculated. The analysis were performed for two data sets. The first data set includes all layers derived from the DEM. The second data set consists of only seven layers derived from the PCA. For the training samples, 24.5% of the study area was used. Within the training data set 62.6% were the landslide areas. The training samples were selected randomly. Afterwards, based on the same training data set, the SVM classification was performed for the second data set.

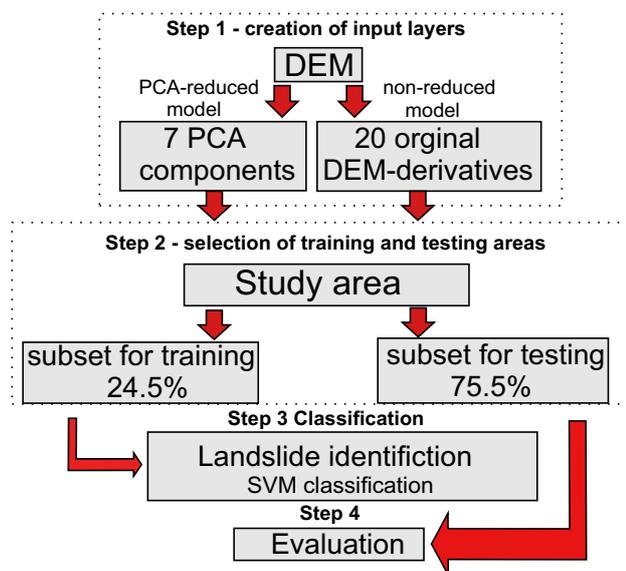


Figure 3: Methodology flowchart

#### 4.1 Principal Component Analysis

PCA is well known method to reduce redundant information between highly correlated variables. It is widely used in hyper-spectral analysis, where bands are highly correlated with each other. The PCA allows reducing the elements necessary to describe large number of inter-correlated variables (Abdi and Williams, 2010). Many classification method from machine learning theory are time-consuming, therefore using full data set of DEM-derivatives for big study areas is ineffective. For this reason the PCA was performed for the DEM and the 19 DEM-derivatives. The normalization of the DEM-layers was required before performing the PCA in order to overcome influence one layer over the others. The initial seven PCA components provide 93.2% of information contained in 20 DEM-layers. Figure 4 shows RGB composition of the three initial PCA components.

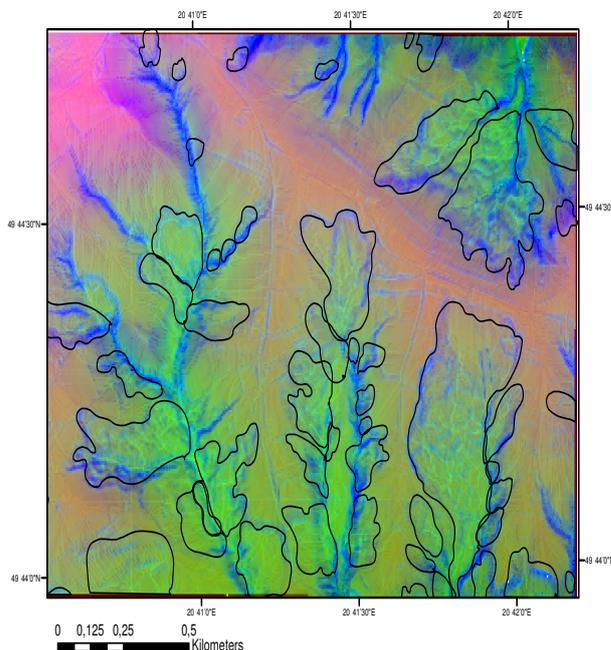


Figure 4: RGB composition of the initial three PCA component with borders (black line) of the existing landslides

#### 4.2 Support Vector Machine

SVM is the useful tool for data classification. It was introduced by Vapnik in 1995 as a new technique from machine learning theory. The main idea of the SVM is to transform data using kernel function into high-dimensional feature space. Discrimination of classes in a new feature space can be separable by means of decision hyperplane (figure 5). Our implementation of SVM classification was made in ENVI using four degree of kernel polynomial function, bias term equal to three and kernel bandwidth  $\gamma = p^{-1}$ , where p is the number of DEM-derivatives (Hsu et al., 2003).

### 5. RESULT

#### 5.1 Accuracy assessment

Figure 6 and figure 7 present classification results for the PCA-reduced model and the non-reduced model with the landslide inventory map, respectively. Comparing results with

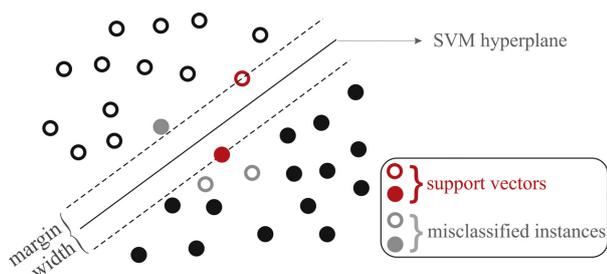


Figure 5: Example of the linear SVM by (San, 2014)

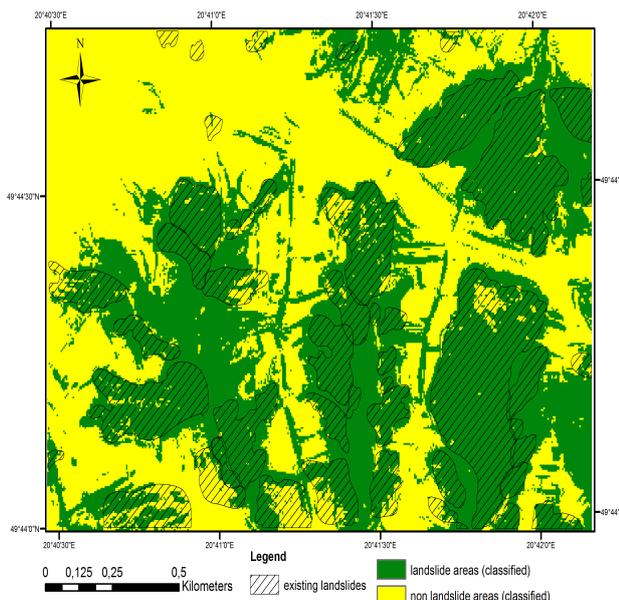


Figure 6: SVM classification results (green areas) for the seven PCA components and the existing landslides (hatched polygons)

the landslide inventory map, the average user's accuracy (UA), the producer's accuracy (PA) and the overall accuracy (OA) were calculated for two models. Thereby, for the PCA-feature-reduced model the UA, the PA and the OA were found to be 72%, 76%, and 72%, respectively. Similarly, UA, PA, and OA in the non-reduced original topographic model, was 74%, 77% and 74%, respectively. Table 2 and table 3 present detailed accuracy assessment as a confusion matrix.

## 5.2 Computational time processing

The SVM classifications were performed using ENVI 32bit, version 5.2. The computations were performed on two Intel (R) Xeon (R) E5649 CPUs 2.53 GHz with 48 GB DDR3 RAM memory. According to the time required for the calculation, user time was measured. The user time is the amount of the CPU time spent in user-mode code within the process. This is only actual CPU time used in executing the process. Other processes and time the process spends blocked do not count towards this figure. According to the results, the SVM classification for the PCA-reduced model and full dataset with the 20-DEM layers took 32 and 65 minutes of the user time, respectively. The computational time of the PCA is not crucial, because using ArcGIS software the computation took 4s.

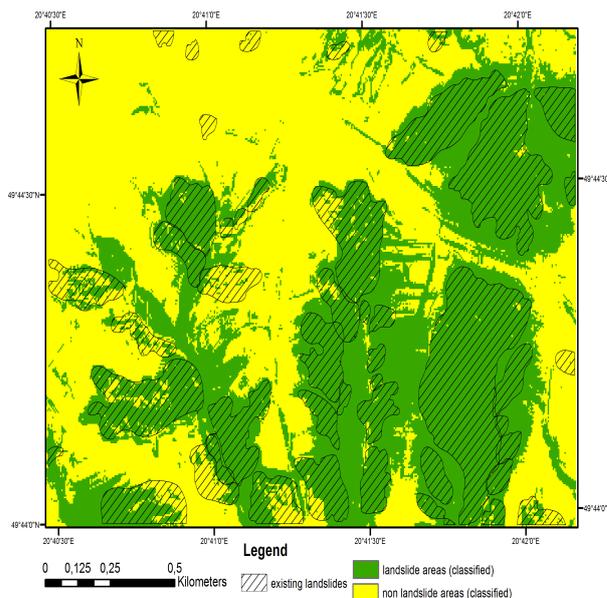


Figure 7: SVM classification results (green areas) for all DEM-derivatives and the existing landslides (hatched polygons)

class	<b>Commission (%)</b>
landslide areas	46.91
non landslide areas	8.87
class	<b>Omission (%)</b>
landslide areas	13.92
non landslide areas	34.70
class	<b>Producer Accuracy (%)</b>
landslide areas	86.08
non landslide areas	65.30
class	<b>User Accuracy (%)</b>
landslide areas	53.09
non landslide areas	91.13
	<b>Overall Accuracy= 71.81%</b> <b>Kappa coefficient=0.44</b>

Table 2: Confusion Matrix of SVM classification using the seven PCA components

class	<b>Commission (%)</b>
landslide areas	43.62
non landslide areas	9.11
	<b>Omission (%)</b>
landslide areas	14.75
non landslide areas	30.94
class	<b>Producer Accuracy (%)</b>
landslide areas	85.25
non landslide areas	69.06
	<b>User Accuracy (%)</b>
landslide areas	56.38
non landslide areas	90.89
	<b>Overall Accuracy= 74.23%</b> <b>Kappa coefficient=0.48</b>

Table 3: Confusion Matrix of SVM classification using full data set

## 6. SUMMARY AND CONCLUSIONS

The objective of this research was to deeply explore topographic information provided by the ALS data and to overcome computational time limitation. The PCA was used to reduce redundant information in an extended set of topographic features. The PCA was used also to decrease computational time in semi-automatic and computer-aided landslide mapping. The proposed novel approach was tested on the susceptible area affected by more than 50 landslides located on Rożnów Lake in the Carpathian Mountains, Poland. Based on accuracy parameters presented in table 2 and table 3, seven initial PCA components with 90% of the total variability in the original topographic attributes were used for the SVM classification.

Comparing results with landslide inventory map, the average user's accuracy (UA), the producer's accuracy (PA) and the overall accuracy (OA) were calculated for two models. Thereby, the PCA-feature-reduced model the UA, the PA, and the OA were found to be 72%, 76%, and 72%, respectively. Similarly, the UA, the PA and the OA in the non-reduced original topographic model, was 74%, 77% and 74%, respectively.

Using the initial seven PCA components instead of all 20 original topographic attributes do not significantly change identification accuracy but reduce the computational time. The tests were performed on the relatively small study area (2.8 km<sup>2</sup>) and the classification for the seven PCA components took 32 minutes while for the non-reduced model 65 minutes.

Authors will continue research in further studies on this topic taking into account an extended test sites. Nevertheless, based on the results achieved in this study, the proposed procedure, which combines the DEM-derivatives and the SVM algorithm, can effectively identify landslide areas in the region of the Carpathian Mountains. Moreover, applying PCA for the DEM-derivatives effectively decrease the computational time in the semi-automatic landslide mapping presented in this study.

## References

- Abdi, H. and Williams, L. J., 2010. [Principal component analysis](#). *Wiley Interdisciplinary Reviews: Computational Statistics* 2(4), pp. 433–459.
- Akgun, A. and Erkan, O., 2016. [Landslide susceptibility mapping by geographical information system-based multivariate statistical and deterministic models: in an artificial reservoir area at Northern Turkey](#). *Arabian Journal of Geosciences* 9(2), pp. 1–15.
- Akgun, A., Dag, S. and Bulut, F., 2008. [Landslide susceptibility mapping for a landslide-prone area \(Findikli, NE of Turkey\) by likelihood-frequency ratio and weighted linear combination models](#). *Environmental Geology* 54(6), pp. 1127–1143.
- Aleotti, P. and Chowdhury, R., 1999. [Landslide hazard assessment: summary review and new perspectives](#). *Bulletin of Engineering Geology and the environment* 58(1), pp. 21–44.
- Booth, A. M., Roering, J. J. and Perron, J. T., 2009. [Automated landslide mapping using spectral analysis and high-resolution topographic data: Puget Sound lowlands, Washington, and Portland Hills, Oregon](#). *Geomorphology* 109(3), pp. 132–147.
- Borkowski, A., Perski, Z., Wojciechowski, T., Jóźków, G. and Wojcik, A., 2011. [Landslides mapping in Rożnów Lake vicinity, Poland using airborne laser scanning data](#). *Acta Geodyn. Geomater* 8(3 (163)), pp. 399–412.
- Chen, W., Li, X., Wang, Y., Chen, G. and Liu, S., 2014. [Forested landslide detection using LiDAR data and the random forest algorithm: A case study of the Three Gorges, China](#). *Remote Sensing of Environment* 152, pp. 291–301.
- Eeckhaut, M., Poesen, J., Verstraeten, G., Vanacker, V., Nyssen, J., Moeyersons, J., Van Beek, L. and Vandekerckhove, L., 2007. [Use of LIDAR-derived images for mapping old landslides under forest](#). *Earth Surface Processes and Landforms* 32(5), pp. 754–769.
- Evans, J., Oakleaf, J., Cushman, S. and Theobald, D., 2015. [An ArcGIS Toolbox for surface gradient and geomorphometric modeling, version 2.0-0](#).
- Hsu, C.-W., Chang, C.-C., Lin, C.-J. et al., 2003. [A practical guide to support vector classification](#).
- Jenness, J., Brost, B. and Beier, P., 2013. [Land facet corridor designer: Topographic position index tools](#).
- Leshchinsky, B. A., Olsen, M. J. and Tanyu, B. F., 2015. [Contour Connection Method for automated identification and classification of landslide deposits](#). *Computers & Geosciences* 74, pp. 27–38.
- McKean, J. and Roering, J., 2004. [Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry](#). *Geomorphology* 57(3), pp. 331–351.
- Pawłuszek, K., Ziąja, M. and Borkowski, A., 2014. [Accuracy assessment of the height component of the airborne laser scanning data collected in the ISOK system for the Widawa River Valley](#). *Acta Sci. Pol. Geod. Descr. Terr.* 13(3-4), pp. 27–38.
- San, B. T., 2014. [An evaluation of SVM using polygon-based random sampling in landslide susceptibility mapping: the Candir catchment area \(western Antalya, Turkey\)](#). *International Journal of Applied Earth Observation and Geoinformation* 26, pp. 399–412.
- Schuster, R. L. and Fleming, R. W., 1986. [Economic losses and fatalities due to landslides](#). *Bulletin of the Association of Engineering Geologists* 23(1), pp. 11–28.
- Starkel, L., 1972. An outline of the relief of the polish carpathians and its importance for human management. *Problemy Zagospodarowania Ziemi Górskich* 10, pp. 75–150.
- Tarolli, P., 2014. [High-resolution topography for understanding Earth surface processes: opportunities and challenges](#). *Geomorphology* 216, pp. 295–312.
- Van Den Eeckhaut, M. and Hervás, J., 2012. [State of the art of national landslide databases in Europe and their potential for assessing landslide susceptibility, hazard and risk](#). *Geomorphology* 139, pp. 545–558.
- Van Den Eeckhaut, M., Kerle, N., Poesen, J. and Hervás, J., 2012. [Object-oriented identification of forested landslides with derivatives of single pulse LiDAR data](#). *Geomorphology* 173, pp. 30–42.
- Van Westen, C. J., Castellanos, E. and Kuriakose, S. L., 2008. [Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview](#). *Engineering geology* 102(3), pp. 112–131.
- Vapnik, V., 2013. [The nature of statistical learning theory](#). Springer Science & Business Media.
- Woźniak, A. et al., 2013. [Precipitation in the Polish Carpathian Mountains in 2010 compared to the period 1881–2010](#). *Prace Geograficzne* 2013(Zeszyt 133), pp. 35–48.