

MICHAEL MÄRKER (*,**), VOLKER HOCHSCHILD (***), VIT MACA (****)
& VÍT VILÍMEK (****)

STOCHASTIC ASSESSMENT OF LANDSLIDES AND DEBRIS FLOWS IN THE JEMMA BASIN, BLUE NILE, CENTRAL ETHIOPIA

ABSTRACT: MÄRKER M., HOCHSCHILD V., MACA V. & VILÍMEK V.,
*Stochastic assessment of landslides and debris flows in the Jemma basin,
Blue Nile, Central Ethiopia* (IT ISSN 0391-9838, 2016)

In this paper we evaluate a stochastic method to assess the spatial distribution of landslide and debris flow processes in the Jemma basin, Central Ethiopia. The Jemma basin is draining the highlands (max. 3.676 m a.s.l.) northeast of Addis Ababa towards the Blue Nile. The basin is characterized by a deeply incised stratigraphy made up of volcanic deposits like flood basalts and tephra. Hence, gravitational mass movements as well as water driven erosion processes occur, documented by the respective forms. We mapped these features using Google Earth images, aerial photo interpretation and fieldwork. The information about the spatial distribution of landslide and debris flow forms was taken as dependent variable in the stochastic modelling approach. Moreover, we performed a detailed terrain analysis to derive the independent variables. We applied two different stochastic modelling approaches based on i) Boosted Regression Trees (BRT) and ii) on an Maximum Entropy Method (MEM) to predict the potential spatial distribution of landslides and debris flows in the Jemma basin. The models are statistically evaluated using the training data and a set of performance parameters such as the area under the receiver operating characteristic curve (AUC). Variable importance and response curves provide further insight into controlling factors of landslide and debris flow distribution. The study shows that

both processes can be perfectly identified and distinguished. The spatial distribution of the predicted process susceptibilities generally follows topographic constraints. Model performance parameters show better results for BRT, that outperforms MEM. However, MEM results are quite robust and hence are used for the spatial prediction of process susceptibilities.

KEY WORDS: Landslides, Debris flows, Boosted Regression Trees (BRT), Maximum Entropy Method (MEM), Spatially explicit prediction, Ethiopia, Blue Nile, Jemma Basin.

INTRODUCTION

The Ethiopian highland is severely affected by land degradation due to soil erosion and mass wasting processes. Generally, Ethiopia show the highest degradation rates in Eastern Africa (de Mûelenaere & *alii*, 2014; Gessesse & *alii*, 2014; Lanckriet & *alii*, 2014, Adugna & *alii*, 2015) and particularly the highland, is facing new strategies to combat desertification (Mekonnen & *alii*, 2015). Especially after long dry periods intensive rainfall events wash away fertile topsoils and trigger debris flows and landslides, particularly if no prevention methods are applied. Hence, agricultural production is at risk and thus, food security is a mayor issue for local population. In order to prevent degradation processes in terms of mass movements, susceptible areas have to be identified and prevention measures established. In the rural areas initiatives have been conducted to fight mass movements in small catchments or on smaller local units. However, to assess the incidence of different mass wasting processes on catchment scale more sophisticated techniques must be applied. In order to derive information on the probability distribution of these processes statistical methods may yield valuable results.

In the recent past several authors categorized the approaches utilized in susceptibility assessments as analytic, heuristic, deterministic and stochastic (e.g. Guzzetti & *alii*, 1999; Brenning, 2005, Lombardo & *alii*, 2014). In terms

(*) Dipartimento Scienze della Terra e dell'Ambiente, Università degli studi di Pavia, Via Ferrata 1, 27100 Pavia, Italia.

(**) Heidelberg Academy of Sciences and Humanities, Germany.

(***) Department of Geography, Tübingen University, Germany.

(****) Department of Physical Geography and Geoecology, Charles University Prague, Czech Republic.

Corresponding author: Michael Märker, michael.maerker@unipv.it

The authors would like to thank the German Research Foundation (DFG-Project HOCHSCHILD 2012) and the Czech Grant Agency (Project No. P209/12/J068) for financing the investigation and the field work campaign. Moreover, we would like to thank the Heidelberg Academy of Sciences and Humanities for travel funding and the Department of Geography at University of Tübingen, Germany for hosting the research activities and providing laboratory and computer facilities. Finally we would like to thank also the Marie Curie EU-IRSES project entitled FLUMEN (Project No. PIRSES-GA-2012-318969) for support and assistance.

of stochastic approaches, several statistic and/or data mining techniques were applied, having certain advantages and disadvantages (e.g. Bai & *alii*, 2011; Märker & *alii*, 2011; Akgün 2012; Vorpahl & *alii*, 2012). Commonly, the methods exploit a set of independent variables, which play the role of predictors, to produce a spatial probability distribution of the dependent variable also called target variable. Multi-variate statistical analysis compares the relationship among multiple variables simultaneously. In case of mass movements such as debris flow and landslide processes the relative influence among all thematic layers of driving factors is assessed. One of the multi-variate statistical approaches is the *discriminant analysis method*. It shows the relative impact of every variable to slope stability in the form of a discriminant function. The higher the coefficient, the higher is the probability that a landslides and/or debris flows may occur.

The knowledge about the spatial distribution of the susceptibility to certain mass movement process is a prerequisite for the assessment of these processes by physically based models, scenario analyses and studies of systems sensitivity (e.g. Märker & *alii*, 2008, Märker & *alii*, 2011). We define the susceptibility as a certain landscape potential or disposition to develop a specific mass movement process. Having information on the distribution of the debris flow and landslide processes in a characteristic test area (e.g. mapped from aerial photographs or Google Earth) we can express the susceptibility to certain processes as a combination and specificity of particular driving forces.

Following this hypothesis we assessed the debris flow and shallow landslides in the Jemma basin. Therefore, we

utilized two stochastic approaches taking into account the relevant environmental information to determine the susceptibility to specific mass wasting processes.

STUDY AREA

The study area is located in the Amhara region 100 km Northeast of Addis Ababa, Ethiopia (fig. 1). Basically, it comprises the basin area of the Jemma River with a drainage area of 15.247 km². The elevations difference is ranging from 770-4.265 m a.s.l. The surface topography can generally be divided into three zones: gorges of Jemma River and tributaries (770-2.500 m a.s.l.); flat plain situated on basalt and ignimbrite plateaus (2.500-3.000 m a.s.l.); mountain peaks and ridges (3.000-4.265 m a.s.l.). The main valley is steeply incised into Tertiary volcanic rocks and Mesozoic sediments (sandstone, mudstone, limestone, gypsum) and at the very bottom into Paleozoic sandstone. The climate depends mainly on the elevation: the tropical zone cover only a limited area at the bottom of deepest gorges while most of valley slopes have a subtropical character. Flat plains have temperate climate and only the highest mountains and ridges are slightly cooler.

The average yearly rainfall in the study area varies between 800 and 1.300 mm a⁻¹. A clear pattern of high inter-annual variability is also evident. The regime includes a dry season in April with the main wet season between June and September with the peak in July. A thin vegetation cover characterizes the area. Main landuse is rain fed agriculture with maize, wheat, sorghum etc. as well as pasture. Close to the villages Eucalyptus plantation

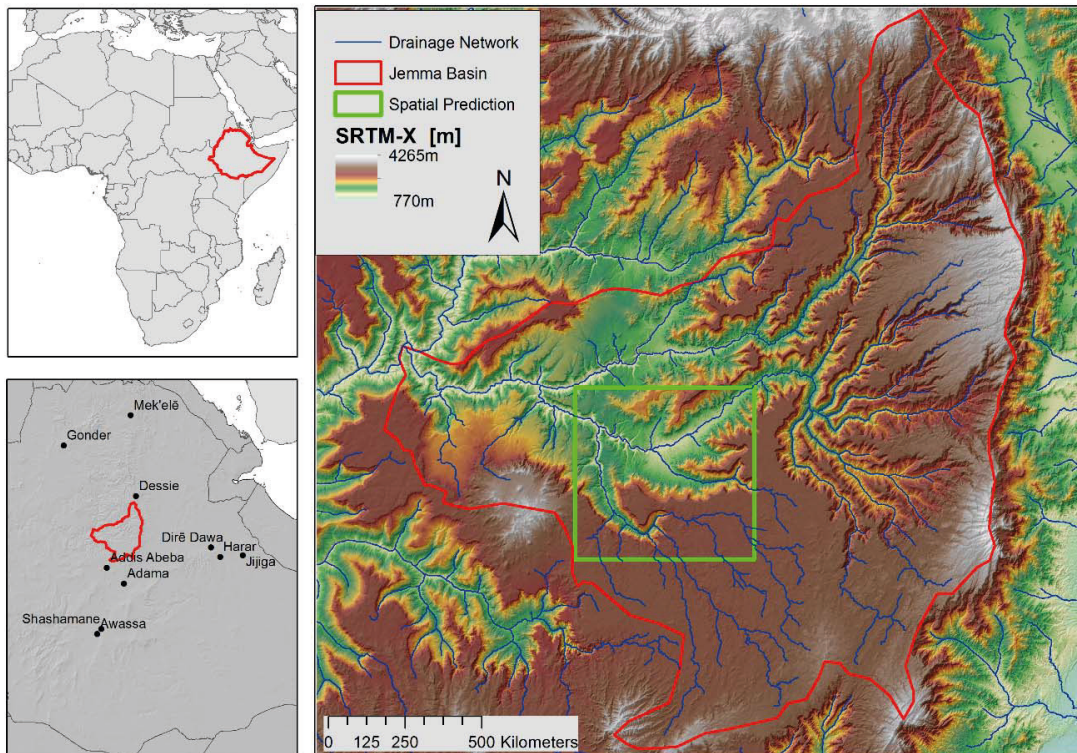


FIG. 1 - Study area of the Jemma basin.

are present. Soils are mainly consisting of Andosols on the plateau areas and Regosols and Lithosols on the steeper slopes according to the Soil Atlas of Africa (Jones & alii, 2013). Zvelebil & alii (2010) carried out a first mapping and zoning of vulnerability to mass wasting processes based on geophysical, geological and hydrogeological information

MATERIALS AND METHODS

Data set

The dataset applied in the modelling consists of 12 parameters: i) topographic wetness index (*TWI*; Beven & Kirkby, 1979), ii) stream power index (Moore & alii, 1991), iii) transport capacity index (Moore & alii, 1997), iv) slope after Horn (1981), v) aspect, vi-viii) curvature (plan, profile, combined), ix) relief elevation above thalweg (Olaya & Conrad, 2008), x) upward flow length after Tarboton (1997), (xi) relative slope position after Böhner & Conrad (2008) and xii) valley depth following Conrad (2007). The topographic indices were calculated from a digital elevation model (DEM) with 30 m cell size based on the Shuttle Radar Topographic Mission (SRTM). In February 2000 a series of Synthetic Aperture Radar (SAR) images was taken by a Space shuttle with the aim to build a high resolution topographic database of the entire world. Here we used the one with 30 m resolution (1-arcsec) that was acquired in the X-band. The major advantage of radar techniques and in particular of the resulting SRTM-X DEM is that microwaves are capable to penetrate clouds and to some degree also the vegetation cover. Even though the SRTM-X DEM offers no global coverage, it covers the study area completely. Primary and secondary topographic attributes as cited above were then calculated from the DEM with Saga GIS based on Wilson & Gallant (2000) and Hengl & Reuter (2009).

The dependent response variable was coded in two classes namely shallow landslides and debris flows showing different levels of geomorphic process dynamics. The two classes were mapped based on Google Earth (GE) using a CNES/Astrium SPOT 6 images with 1.5m ground resolution. Class zero was attributed to landslides and class two to debris flows. We mapped an area within the Jemma basin covering a section from the so called Portuguese bridge near the locality Fiche and the village of Deneba ca. 35 km upwards the Jemma basin from the Portuguese bridge. The training data set was extracted using a randomised sampling scheme covering 15 percent of randomly drawn raster cells for each class ($N = 1097$). These samples were extracted from the Jemma River basin where geo-referenced information on the present debris flows and landslides was available.

Statistical modelling

Methods such as artificial neural networks (Kohonen, 1984), logistic regression (Hosmer and Lemeshow 2000) or classification and regression trees (CART, i.e. recursive binary partitioning into regions that are increasingly

homogeneous with respect to the class variable; Breiman & alii, 1984; Breiman, 2001 De'ath and Fabricius, 2000, De'ath, 2007) have been applied in a wide range of geomorphologic and pedologic studies in the recent past (Moore & alii, 1993; Gessler & alii, 1995; Paruelo and Tomasel, 1997; Mertens & alii, 2002; Brenning, 2005; Grimm & alii, 2008, Vorpahl & alii, 2012). These methods allow for deriving transparent algorithms that can be assessed by applying a range of performance criteria. In this study we tested two different explorative statistical models: i) a classification regression tree approach and ii) a mechanical statistics method. The application of two models with the same dataset guarantees the achievement of consistent and robust modeling results.

The first method is based on stochastic gradient boosting (Salford Systems implementation: TN, cf. Friedman, 1999, also called Boosted Regression Trees (BRT), Elith & alii, 2008). Gradient boosting constructs additive regression models by sequentially fitting a simple parameterized function to current 'pseudo'-residuals by least squares at each iteration. The pseudo residuals are the gradient of the loss function being minimized, with respect to the model values at each training data point, evaluated at the current step (Friedman, 1999). Practically, the method derives several hundreds to thousands of small trees (in our case fixed to a maximum of six nodes). Each tree is devoted to contributing a small portion of the overall model whereas the final model prediction is constructed by adding up each of the individual tree contributions. The advantage of the methodology is that it is not sensitive to data errors in the input variables.

The second modelling approach originates in Bayesian statistics (Jaynes, 1957) and is called Maximum Entropy Method (MEM). Basically, MEM estimates a distribution function of the predictors by finding a distribution of maximum entropy for the single predictors that is closest to uniform (Vorpahl & alii, 2012). Furthermore, the expected value of each predictor under the estimated distribution has to match its empirical average (Phillips & alii, 2006). The advantage of MEM is, that it can handle presence only data, thus it does not need classified variables or a binary presence-absence dataset.

For this study, we run the model with the following parameters: The entire dataset ($N = 1097$) used for the modelling was separated into a training fraction and a test fraction ($N_{\text{train}} = 932$, 0.85 of N ; and $N_{\text{test}} = 165$, 0.15 of N). The separation was performed by randomly selecting a sample of cases from the entire data set. In both approaches, the maximum number of trees to use was 2000. The TN model uses a regression model with the Huber-M loss function.

Model evaluation

The BRT and MEM models were applied to predict the landslides (class 0) and debris flows (class II) of the entire basin in a spatially explicit way. The models were applied to the entire data set containing all explanatory variables for the entire basin. In the last step, the resulting tabulated data – or to be more precisely, the predicted probabilities of each pixel to fall into one of the two classes – were post-

processed to produce susceptibility maps for each erosion process class using a GIS. In order to evaluate the models' predictive performance besides classification matrices, we calculated the receiver operating characteristics (ROC) curves for each class and both models for training and test data. In an ROC curve the true positive rate (sensitivity) is plotted over the false positive rate (1-specificity) for all possible cut-off points (Swets, 1988). Each point on the ROC plot represents a sensitivity/specificity pair corresponding to a particular decision threshold. A perfect discrimination between positives and negatives has an ROC plot that passes through the upper left corner (100% sensitivity, 100% specificity), so that the area under the ROC curve, *AUC*, is 1 (cf. Reineking & Schröder, 2006). Therefore, the closer the ROC plot to the upper left corner, the higher the overall accuracy of the test. According to Hosmer & Lemeshow (2000), *AUC* values exceeding 0.7/0.8/0.9 indicate acceptable/excellent/outstanding predictions.

For improving model interpretation, we calculated predictor variable importance ranking all predictors due to their contribution to the final models by calculating the improvements of all splits associated with a given predictor across all trees in the model. Moreover, we analysed partial dependency plots as univariate response curves by calculating non-parametric loess regressions from predicted values for each continuous predictor variable as a basis for interpreting the effect of predictors. Finally, we illustrated the spatial patterns of prediction results using erosion susceptibility maps.

RESULTS AND DISCUSSION

We applied the BRT and MEM models training them with the 85% of data selected randomly from the entire dataset ($N_{tot}=1097$). The remaining test sample fraction of 15% ($N_{test}=165$) was used to test model performance. The latter are shown as ROC integral or *AUC* for the specificity/1-sensitivity plots shown in figure 2 and 3. As illustrated in fig. 2 the BRT approach shows very high accuracy for the training data with ROC integrals of 99%. The test data set also reveals an outstanding model performance of 97% for both processes. Generally, the good performance of the BRT method is also reported by other studies like Vorpahl & alii (2011), Märker & alii (2012) or Brenning (2005).

The MEM approach yield also respectable model performances as illustrated in fig 3. The *AUC* values for the train data according to Hosmer & Lemeshow (2000) are acceptable having values of 0.74 for debris flows and 0.7 for landslides. Moreover, also the train data set show nearly acceptable values with *AUC* values of 0.69 for landslides and 0.67 for debris flows.

The variable importance yield insights into the most relevant topographic indices driving landslide and debris flow distribution. We excluded the aspect from the stochastic modelling procedure to avoid a bias due to the fact that mainly slides and debris flows on the north and south facing slopes have been mapped. Moreover, other authors like Ayalew & Yamagishi (2003, 2004) or Asfaw (2010) stated that aspect do play a minor role in the spatial distribution of landslides.

Table 1 shows the combined variable importance for the BRT model for both processes. The indices valley depth, vertical distance to river network, and catchment network base level play the dominant role triggering both processes. These parameters indicate a strong stratigraphic/ lithologic influence since the study area is characterized by prevailing horizontally stratified lithological layer and hence, the processes seem to be correlated to specific geological formations and/ or a certain connectivity to the drainage network. Catchment area and slope are the parameters mainly involved also in physically based models like SINMAP (Pack & alii, 2005).

For the MEM approach we present the variable importance separately since there are distinct differences between both processes. Table 2 shows the variable importance for debris flows whereas Tab. 3 the variable importance for landslides. For both models slope is the most important variable but thereafter we find particular differences. The debris flow distribution are more related to valley depth and topographic wetness index, hence indicating pre-existing linear incisions as well as surface runoff accumulation areas. Whereas landslides are related to catchment area and channel network distance representing a stratigraphic relation or undercutting processes creating instabilities at the base of the slopes. As already stated above slope and catchment area are also used in physically based landslide models as a proxy for soil or substrate weight and potential shear stress.

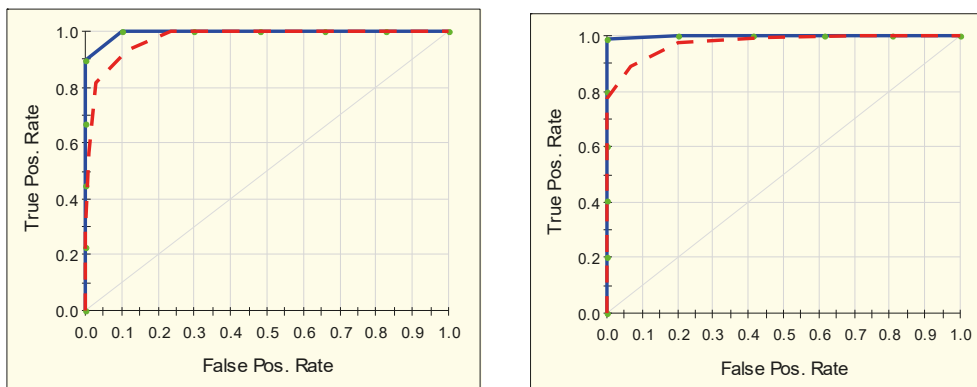


FIG. 2 - Receiver Operator Curve (ROC) diagrams. Left: landslides (class 2); right debris flows (class0). Blue line training data red line test data. *AUC* landslides train 0.99 (test 0.975); *AUC* debris flow train 0.99 (test 0.975).

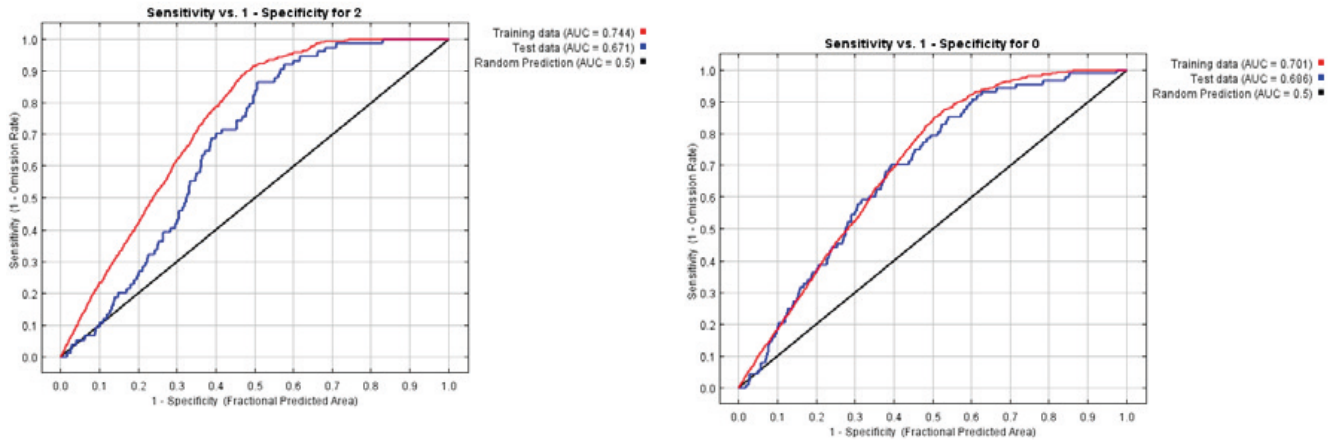


Fig. 3 - Receiver Operator Curve (ROC) diagrams. Left: landslides; right debris flows. Blue line test data red line training data. AUC landslides train 0.7 (test 0.69); AUC debris flow train 0.74 (test 0.67).

TABLE 1 - Variable importance of BRT model.

VARIABLE	SCORE
VALLEY_DEPTH	32.8
VERTICAL_DISTANCE	31.0
SLOPE	30.2
CHANNEL_NETWORK BASE LEVEL	26.5
ANALYTICAL HILLSHADING	25.9
CATCHMENT_AREA	24.4
RELATIVE_SLOPE	21.4
CROSS_SECTION	18.2
SRTM30_JEMMA	17.7
LONGITUDINAL CURVATURE	17.6
TOPOGRAPHIC WETNESSINDEX	17.2
LS_FACTOR	14.1
CONVERGENCE INDEX	13.6

TABLE 2 - Variable Importance for debris flows.

VARIABLE	% CONTRIBUTION
SLOPE	29.0
VALLEY DEPTH	24.6
TOPOGRAPHIC WETNESS INDEX	10.5
CROSS-SECTIONAL CURVATURE	10.4
VERTICAL DISTANCE TO CHANNEL NETWORK	8.9
CATCHMENT AREA	7.0
LONGITUDINAL CURVATURE	3.5
CHANNEL NETWORK BASE LEVEL	3.4
RELATIVE SLOPE POSITION	1.7
CONVERGENCE INDEX	0.6
LS FACTOR	0.4

TABLE 3 - Variable importance for landslides.

VARIABLE	% CONTRIBUTION
SLOPE	32.8
CATCHMENT AREA	20.4
CHANNEL NETWORK BASE LEVEL	17.8
VERTICAL DISTANCE TO CHANNEL NETWORK	16.2
VALLEY DEPTH	9.3
RELATIVE SLOPE POSITION	0.9
LS FACTOR	0.7
CONVERGENCE INDEX	0.6
CROSS-SECTIONAL CURVATURE	0.5
TOPOGRAPHIC WETNESS INDEX	0.5
LONGITUDINAL CURVATURE	0.2

The response curves shown in fig. 4 represent a Maxent model created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables. Fig. 4 illustrates the response curves of the three most important parameters of the MEM model runs for debris flows and landslides. Landslides seem to be related to higher slope degrees than debris flows. Medium valley depth and medium to high TWI values depict debris flows. Instead, catchment areas up to 18 ha (2.000 pixel * 900 m²) and distances to channel network between 1.5 and 2 km seem to characterize high landslide susceptibilities.

As shown by Lombardo & alii (2015) and Vorpahl & alii (2012) BRT normally outperforms the MEM approach. However, it was reported that if the model is used for spatial prediction MEM is performing quite robustly. Hence, we show the spatial distribution of the susceptibility to landslides and debris flows using the MEM approach.

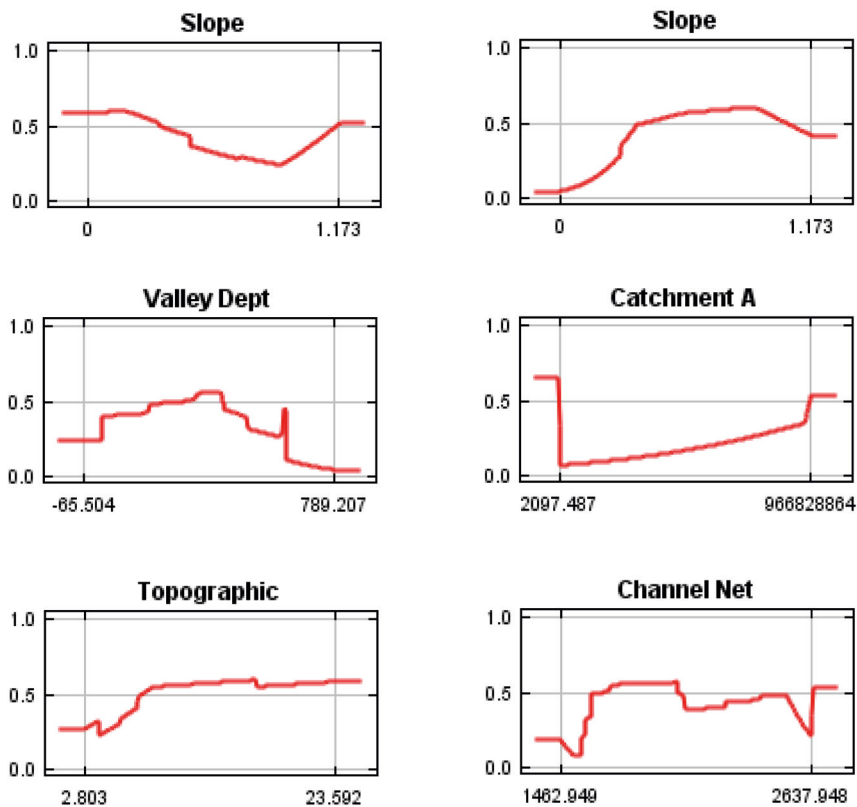


FIG. 4 - Response curves of the three most important variables for debris flow (left) and landslides (right) using MEM.

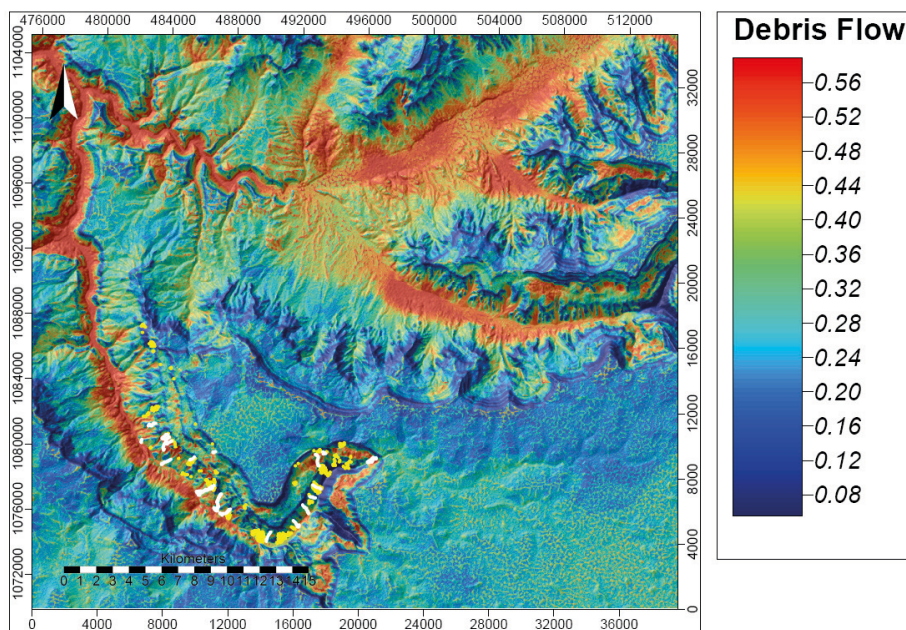


FIG. 5 - Spatial distribution of debris flows using the MEM approach. High susceptibilities in red/yellow, low susceptibilities in green/blue colours. Yellow dots: landslides data set; White dots: debris flow data set.

Fig. 5 illustrates the spatial distribution of the predicted susceptibilities for debris flows whereas Fig 6 highlights the spatial prediction of the landslide susceptibilities. We observe that debris flows (fig. 5) are more related to drainage lines which were already stressed by the variable importance. Areas in yellow indicate high susceptibilities

for debris flows especially in the lower slope areas and larger tributaries with higher runoff concentration potential.

Figure 6 highlights the areas prone to landslides in the yellow orange colours. It can be noticed that high susceptibilities for landslides are concentrating in certain slope positions related to specific stratigraphic layer.

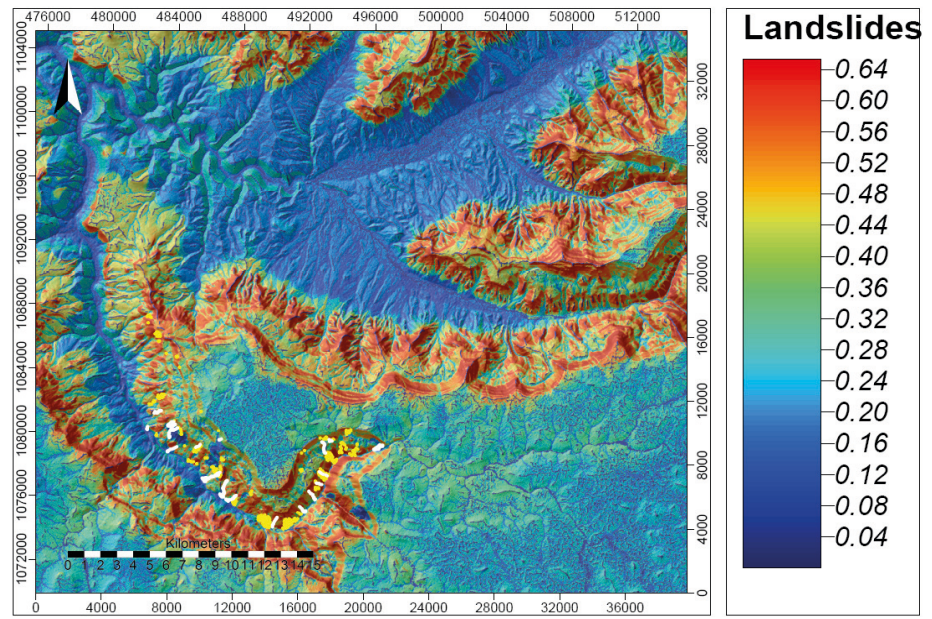


FIG. 6. Spatial distribution of landslides using the BRT model. High susceptibilities in red/yellow, low susceptibilities in green/blue colours. Yellow dots: landslides data set; White dots: debris flow data set.

Actually, two zones with higher landslide potential can be identified: The first one close to the escarpment of the Jemma canyon might be also affected by rockfalls due to very steep slopes and a second area which is further down the slopes and can be related to less steep slopes and maybe substrates with higher clay content that are present in the stratigraphy and of probable lacustrine origin. Since stratigraphy seems to play a role for the landslide potential in further project steps detailed stratigraphic information should be taken into account.

CONCLUSIONS

In this study we conducted a stochastic analysis using a Boosted Regression Tree method (BRT) and a Maximum Entropy Method (MEM) to analyse the main driving factors characterized by topographic indices to assess landslide and debris flow occurrence in the Jemma basin in Central Ethiopia. Therefore, we mapped the debris flow and landslide forms in a small tributary of the Jemma catchment showing the typical characteristics of the entire basin. The mapping of forms was performed using GE high resolution images. The resulting dataset was divided into train and test data in order to evaluate the performance of the chosen BRT and MEM approaches. The analysis showed a general excellent model performance with very high AUC values for BRT and acceptable values for the MEM approach for both, debris flows and landslides. Variable importance reveal that valley depth, channel network base level and slope play a major role in terms of variable importance, whereas especially MEM yield particular parameters to distinguish between landslides and debris flows. Slope and channel base level show high correlations and hence seem to trigger landslides and debris flows. Finally, the spatial

distribution of landslide and debris flow susceptibilities demonstrate a specific spatial pattern that is clearly related to stratigraphic units and the drainage system. Only certain slopes and stratigraphic units are able to build up specific slope morphologies and drainage patterns allowing the accumulation of material that in turn is subject to these mass movement processes.

REFERENCES

- ADUGNA A., ABEGAZ A. & CERDÀ A. (2015) - *Soil erosion assessment and control in Northeast Wollega, Ethiopia*. Solid Earth Discussions, 7, 3511–3540.
- AKGÜN A. & TÜRK N. (2011) - *Mapping erosion susceptibility by a multivariate statistical method: a case study from the Ayvalık region, NW Turkey*. Computational Geosciences, 37, 1515–1524.
- ASFAW Y. E. (2010) - *Landslide Assessment in Blue Nile Gorge, Central Ethiopia*. Master Thesis, University Gent, 82 pp.
- AYALEW L. & YAMAGISHI H. (2003) - *Slope failures in the Blue Nile basin, as seen from landscape evolution perspective*. Geomorphology, 136, 1–22
- AYALEW L. & YAMAGISHI H. (2004) - *Slope movements in the Blue Nile basin, as seen from landscape evolution perspective*. Geomorphology, 57, 97–116.
- BAI S.B., WANG J., LÜ G.N., ZHOU P.G., HOU S.S. & XU S.N. (2010) - *GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China*. Geomorphology, 115, 23–31.
- BEVEN, K.J. & KIRKBY, M.J. (1979) - *A physically-based variable contributing area model of basin hydrology*. Hydrological Science Bulletin, 24, 43–69.
- BÖHNER J. & CONRAD O. (2008) - *SAGA-GIS Module Library Documentation (v2.1.3). Relative slope position*. http://www.saga-gis.org/saga_module_doc/2.1.3/ta_morphometry_12.html.

- BREIMAN L., FRIEDMAN J.H., OLSHEN R.A. & STONE C.J. (1984) - *Classification and Regression Trees*. Chapman and Hall, Boca Raton, 358 pp.
- BREIMAN, L., (2001) - Random Forests. *Machine Learning*, 45, 5–32.
- BRENNING, A., (2005) - *Spatial prediction models for landslide hazards: review, comparison and evaluation*. *Natural Hazards and Earth System Sciences*, 5, 853–862.
- CONRAD O. (2007) - *SAGA - Entwurf, Funktionsumfang und Anwendung eines Systems für Automatisierte Geowissenschaftliche Analysen*. Dissertation, University of Göttingen.
- DE'ATH G. & FABRICIUS K. E., (2000) - *Classification and regression trees: a powerful yet simple technique for ecological data analysis*. *Ecology*, 81, 3178–3192.
- DE'ATH G. (2007) - *Boosted trees for ecological modeling and prediction*. *Ecology*, 88, 243–251.
- DE MÜELENAERE S., FRANKL A., HAILE M., POESEN J., DECKERS J., MUNRO N., VEREVERBEKE S. & NYSSSEN J. (2015) - *Historical landscape photographs for calibration of landsat land use/cover in the Northern Ethiopian highlands*. *Land Degradation & Development*, 25, 319–335.
- ELITH J., LEATHWICK J.R. & HASTIE T. (2008) - *A working guide to boosted regression trees*. *Journal of Animal Ecology*, 77, 802–813.
- FRIEDMAN J.H. (1999) - *Stochastic gradient boosting*. Technical Report. Department of Statistics, Stanford University, USA. <http://www.salford-systems.com/treenet.html>.
- GESSESE B., BEWKET W., & BRÄUNING A. (2014) - *Model-based characterization and monitoring of runoff and soil erosion in response to land use/land cover changes in the Modjo watershed, Ethiopia*. *Land Degradation and Development*, 26, 711–724.
- GESSLER P.E., MOORE I.D., MCKENZIE N.J. & RYAN, P.J. (1995) - *Soil-landscape modelling and spatial prediction of soil attributes*. *International Journal of Geographical Information Systems*, 9, 421–432.
- GRIMM R., BEHRENS T., MÄRKER M. & ELSENBEER H. (2008) - *Soil organic carbon concentrations and stocks on Barro Colorado Island – digital soil mapping using Random Forest analysis*. *Geoderma*, 146, 102–113.
- GUZZETTI F., CARRARRA 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, pp. 181–216.
- HENGL T. & REUTER H.I. (200) - *Geomorphometry. Concepts, Software, Applications*. *Development in Soil Science* 33, Elsevier, Amsterdam, Oxford. 765 pp.
- HORN B.K.P. (1981) - *Hills shading and the reflectance map*. *Proceedings of the IEEE* 69(1).
- HOSMER D.W. & LEMESHOW S. (2000) - *Applied Logistic Regression*. (2nd ed.), Wiley, New York. 392 pp.
- JAYNES E.T. (1957) - *Information theory and statistical mechanics*. *Physical Review*, 106, 620–630.
- JONES A., BREUNING-MADSEN H., BROSSARD M., DAMPHA A., DECKERS J., DEWITTE O., GALLALI T., HALLETT S., JONES R., KILASARA M., LE ROUX P., MICHELI E., MONTANARELLA L., SPAARGAREN O., THIOMBIANO L., VAN RANST E., YEMEFACK M. & ZOUGMORÉ R. (2013) - *Soil Atlas of Africa*. European Commission, Publications Office of the European Union, Luxembourg, 176 pp.
- KOHONEN T. (1984) - *Self-organization and Associative Memory*. Springer, Berlin.
- LANCKRIET S., DERUDDER B., NAUDTS J., BAUER H., DECKERS J., HAILE M. & NYSSSEN J. (2014) - *A political ecology perspective of land degradation in the north Ethiopian highlands*. *Land. Degrad. Dev.*, 26, 521–530.
- LOMBARDO L., CAMA M., MÄRKER M. & ROTIGLIANO E. (2014) - *A test of transferability for landslides susceptibility models under extreme climatic events: application to the Messina 2009 disaster*. *Natural Hazards*, 74 (3)1951–1989.
- MÄRKER M., ANGELI L., BOTTAI L., COSTANTINI R., FERRARI L., INNOCENTI L. & SICILIANO G. (2008) - *Assessment of land degradation susceptibility by scenario analysis. A case study in southern Tuscany, Italy*. *Geomorphology*, 3, 120–129.
- MÄRKER M., PELACANI S. & SCHRÖDER B. (2011) - *A functional entity approach to predict soil erosion processes in a small Plio-Pleistocene Mediterranean catchment in Northern Chianti, Italy*. *Geomorphology*, 125, 530–540.
- MEKONNEN M., KEESSTRA S. D., BAARTMAN J. E., RITSEMA C. J., & MELESSE A. M. (2015) - *Evaluating sediment storage dams: structural off-site sediment trapping measures in northwest Ethiopia*. *Cuadernos de Investigación Geográfica*, 41, 7–22.
- MERTENS M., NESTLER I. & HUWE B. (2002) - *GIS-based regionalization of soil profiles with Classification and Regression Trees (CART)*. *Journal of Plant Nutrition and Soil Science*, 165, 39–44.
- MOORE I.D., GARSON R.B. & LADSON A.R. (1991) - *Digital terrain modelling: a review of hydrological, geomorphological and biological applications*. *Hydrological Processes*, 5, 3–30.
- MOORE I.D., GESSLER P.E., NIELSEN G.A. & PETERSON G.A. (1993): *Soil attribute prediction using terrain analysis*. *Soil Science Society of America Journal*, 57, 443–452.
- MOORE I.D., GARSON R.B. & LADSON A.R. (1997) - *Digital terrain modelling: a review of hydrological, geomorphological and biological applications*. *Soil Science Society of America Journal*, 57, 443–452.
- OLAYA V. & CONRAD O. (2008) - *Geomorphometry in SAGA*. In: Hengl, T., Reuter, H.I. (Eds.), “*Geomorphometry: Concepts, Software, Applications*”. Elsevier, Amsterdam, 293–308.
- PACK R., TARBOTON D., GOODWIN C. & PRASAD A. (2005) - *A Stability Index Approach to Terrain Stability Hazard Mapping User's Manual*; CEE Faculty Publications: Logan, UT, USA. Available online: <http://hydrology.usu.edu/sinmap2/> accessed on 8 August 2015.
- PARUELO J.M. & TOMASEL F. (1997) - *Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models*. *Ecological Modelling*, 98, 173–186.
- PHILLIPS S.J., DUDÍK M. & SCHAPIRE R.E. (2006) - *Maximum entropy modeling of species geographic distributions*. *Ecological Modelling*; 190, 231–259.
- REINEKING B., SCHRÖDER B. (2006) - *Constrain to perform: regularization of habitat models*. *Ecological Modeling*, 193, 675–690.
- SWETS J.A. (1988) - *Measuring the accuracy of diagnostic systems*. *Science*, 240, 1285–1293.
- TARBOTON D.G. (1997) - *A new method for the determination of flow directions and contributing areas in grid Digital Elevation Models*. *Water Resources Research*, 33(2), 309–319.
- VORPAHL P., ELSENBEER H., MÄRKER M. & B. SCHROEDER (2012) - *How can statistical models help to determine driving factors of landslides?* *Ecological Modelling*, 239, 27–39
- WILSON J.P., GALLANT J.C. (2000) - *Digital terrain analysis*. In: Wilson, J.P., Gallant, J.C., 2000. *Terrain Analysis. Principles and Applications*. Wiley, New York.
- ZVELEBIL J., ŠÍMA J. & VILÍMEK V. (2010) - *Geo-risk management for developing countries – vulnerability to mass wasting in the Jemma River Basin, Ethiopia*. *Landslides*, 7(1), 99–103.

(Ms. Received 30 July 2015; accepted 1 May 2016)