Assessment of Landslides Activity in Maily-Say Valley, Kyrgyz Tien Shan

Romy Schlögel, Anika Braun, Almaz Torgoev, Tomas M. Fernandez-Steeger, and Hans-Balder Havenith

Abstract

There is a strong possibility that environmental change (whether climate or land use) will be manifest as changes in the size-frequency distribution of landslides in Maily-Say Valley, Kyrgyzstan. The evolution of the landslide activity over the past 50 years has been analysed on the basis of five landslide inventories of 1962, 1984, 1996, 2002 and 2007. Their size-frequency analyses show that both the number and size of unstable slopes are increasing from 1962 (162 objects) to 2007 (208 objects) and the power-law exponent is decreasing over time. This might indicate that there is an evolution in size probably corresponding to an increase of landslide-related hazards. Remote sensing and spatial analysis through the image subtraction method based on NDVI allowed an accurate detection of new sliding activation in that area. Another aim is to evaluate if and how Data Mining, with its wide range of tools, may support automatic landslide recognition.

Keywords

Landslide activity • Frequency density function • Data Mining • Kyrgyzstan

Introduction

Kyrgyz Tien Shan is prone to different natural hazards due to high seismicity, active tectonics, glacier retreat, and locally intense rainfall (Torgoev et al. 2002). Due to the high landslide activity along the rim of the Fergana Basin in the southern part of the country (Roessner et al. 2002), disasters causing social and economic losses occur almost each year. Especially, the Maily-Say Valley can be considered as a geological and environmental hazard hotspot within the Tien Shan Mountains.

In order to detect those landslides, which might be dangerous for population, two methods are processed. Normalized Differenced Vegetation Index (NDVI), including specific spectral signature of new landslides, were calculated for two images (taken at different times) and subtracted. The second process was to evaluate how Data Mining, may support automatic landslide recognition (e.g. from remote sensing data), for susceptibility mapping or knowledge discovery regarding the causes of landslides. Data Mining is the application of specific algorithms, under acceptable computational efficiency limitations, for extracting patterns from data (Fayyad et al. 1996). Here, the Data Mining approach allows the prediction of landslides using classification methods (Fernandez-Steeger et al. 2002).

R. Schlögel (✉)
Department of Geology, Georisks and Environment, University of Liège, B20 Sart Tilman, 4000 Liège, Belgium

School and Observatory of Earth Sciences, University of Strasbourg, 5 rue Descartes, 67084 Strasbourg Cedex, France
e-mail: schlogel@unistra.fr

A. Braun • T.M. Fernandez-Steeger
Department of Engineering Geology and Hydrogeology, RWTH Aachen University, Lochnerstraße 4-20, 52064 Aachen, Germany

A. Torgoev • H.-B. Havenith
Department of Geology, Georisks and Environment, University of Liège, B20 Sart Tilman, 4000 Liège, Belgium

C. Margottini et al. (eds.), *Landslide Science and Practice*, Vol. 1,
DOI 10.1007/978-3-642-31325-7_14, © Springer-Verlag Berlin Heidelberg 2013

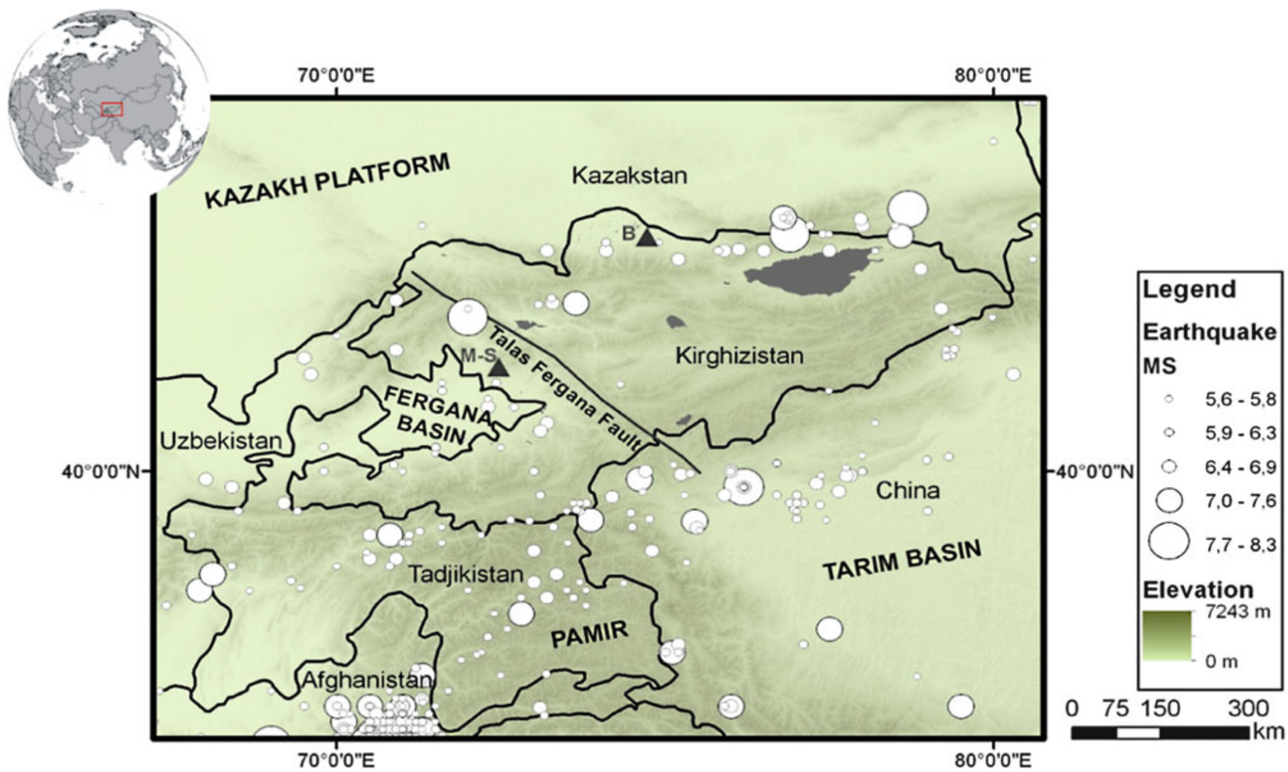

Fig. 1 Topographic map of the Kyrgyz Tien Shan and neighbouring regions. The location of Maily-Say (MS) and the capital city (Bishkek) is indicated by a *black triangle*. The map also shows all historical and instrumental earthquake events with Ms > 5.5 (Schlögel et al. 2011)

Study Area

The Maily-Say Valley is situated in the north of the Fergana Basin, about 25 km from the border with Uzbekistan. The study area is located in a transitional zone bounded in the northeast by the Talas-Fergana Fault (Fig. 1) and in the western foothills of the seismically active Tien Shan mountain belt. Bedrock geology is constituted of Cenozoic and Mesozoic sedimentary rocks (Vandenhove et al. 2003; Havenith et al. 2006). The relief is rough with heights culminating at an altitude ranging from 700 to 4,000 m.

Over the last decades, the town of Maily-Say (estimated population 10,000) has experienced a series of environmental disasters related to earthquakes, landslides and groundwater pollution (in particular by radioactive waste). According to Abdrakhmatov et al. (2003), the seismic hazard of the region is moderate to high. The last strong earthquake (Ms 6.2) hit the region on May 15, 1992, and was located at about 30 km south-southeast of the town of Maily-Say. Evidence for high landslide activity and slope instability is observed throughout the area. Today, more than 200 landslides are located in the vicinity of the town, 80 of which are unstable and have the potential to move under adverse conditions.

The largest landslide in the region, located about 7 km east of the town of Maily-Say, is the Kochkor-Ata landslide, was activated as a whole on April 13, 1994, forming an almost 4 km long earth flow with a total volume of about 10 Mm^3 (Roessner et al. 2002).

Data and Pre-processing

Multi-temporal, remotely sensed images were collected to detect landslides and to determine their evolution. Aerial photographs of the years 1962, 1984 and 1996 (panchromatic, approximate 1:21,000 scale, roughly 2 m ground-resolution) for the Maily-Say Valley were studied in detail and compared with panchromatic and multi-spectral 10.3 by 11.1 km Quickbird images of 2002 and 2007. The multi-spectral images provide four spectral bands, (blue, green, red, near-infrared) with a spatial resolution of 2.44 m; while panchromatic Quickbird images have a spatial resolution of 0.61 m.

Quickbird imagery was pre-processed with the ENVI software before being classified (unsupervised) through the image subtraction in order to make radiometrically and geometrically comparable 2002 and 2007 scenes. Factors affecting geometry can be related to the sensor system, due to variations in platform altitude, in the sensor view angle or caused by the rotation of the earth (Gupta 2003). The images were orthorectified on the basis of a 20 m resolution Satellite Pour l'Observation de la Terre (SPOT) digital elevation model (DEM) and georeferenced to UTM zone 43 north. Panchromatic and multi-spectral images were co-registered separately to allow superimposition of multi-spectral images. In addition, the images were normalized using a relative radiometric correction process. Gains and offsets were applied to create a normalized image of the 2007 scene while the 2002 image is taken as reference as it shows the larger contrasts (Jensen 1996).

Methods

Landslide Inventories

Landslide inventories are the simplest form of landslide susceptibility mapping (Guzzetti et al. 1999). Two new landslide inventories were created by remote sensing analysis from the Quickbird images for the Maily-Say Valley, verified by field observations and used in specific techniques of landslide detection described further. Using Geographical Information System (GIS) tools, a total of 189 and 208 landslides were mapped from 2002 and 2007 images, respectively.

The initial inventories were created by local scientists through field surveys and air photo interpretation, and corrected by comparison with recent data to make them accurate and comparable (Brardinoni et al. 2003). Previous landslide inventories (already used by Havenith et al. 2006) were improved to produce new ones for the years of 1962 (162 landslides), 1984 (206 landslides) and 1996 (222 landslides). All five inventories contain only landslides that could be identified from remote imagery of a certain year and, therefore, are likely to have been (re-)activated in recent times.

Evolution of Landslides

The landscape in the Maily-Say Valley is continuously changing owing to the high landslide activity. The total area affected by landslides increased from 1.0 % (~1.2 km^2) in 1962 to 3.3 % in 1984, 4.5 % in 1996, 4.3 % in 2002 and 5.6 % (~6.5 km^2) in 2007 compared to the entire investigated area along the Maily-Say Valley. The slight decrease of 0.2 % from 1996 to 2002 is likely to be related to the changed dataset used for mapping (aerial photographs and existing inventories in 1996 and before, Quickbird imagery for 2002 and 2007). The mean landslide size increased from 15,170 m^2 in 1962 to 31,000 m^2 in 2007. Landslide size is quite variable; it ranges from 335 m^2 for the smallest detected landslide to 348,425 m^2 for the largest one in 2007. The evolution of the formation of the largest mass movement is marked by the formation of landslide Tektonik in 1992. Due to this, the maximum landslide size increased from 110,000 m^2 in 1984 to 325,000 m^2 in 1996.

NDVI Subtraction

Methodology

The NDVI, initially developed for classification of vegetation type and health, is here applied to detect fresh translational mass movements (Chang and Liu 2004) such as debris flows and earth flows which are quite common in Maily-Say valley. It is calculated on the basis of the red and near-infrared bands which depend strongly on the presence of vegetation, according to the following equation:

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{R})/(\mathrm{NIR} + \mathrm{R})$$

This processing allows for improved differentiation between spectrally-different surfaces, compared to the use of one band only. Further, band ratioing greatly mitigates shadow effects that sometimes make visual interpretation difficult. The values of this index, also an indicator of biomass, are in the range between -1 and 1; negative values indicate bare land and positive values indicate a greater level of photosynthetic activity due to the presence of vegetation (often included between 0.2 and 0.8).

Image differencing, or subtraction, is based on a pair of co-registered raw or transformed images of the same area collected at different times. The process simply subtracts pixel values on a pixel-by-pixel basis to generate a third image composed of numerical differences between the pairs of pixels (Ridd and Liu 1998).

The process leading to a new image, using the available Quickbird satellite images of these two periods, can be quantified by the following the equation (here applied to the b-band of the images):

$$\mathrm{D}x^{b}{}_{ij} = x^{b}{}_{ij}(t_{2007}) - x^{b}{}_{ij}(t_{2002}) \tag{1}$$

where $x_{ij}{}^{b}$ (t_{2007}) and $x_{ij}{}^{b}$ (t_{2002}) are the digital numbers (DNs) of a pixel (i,j) of b-band for the 2007 and 2002 images, respectively. In the new image ($\mathrm{D}x_{ij}{}^{b}$), positive or negative values denote the region whose radiation value has increased or decreased, respectively, between 2002 and 2007; a 0-value corresponds to no change in the region.

A suitable threshold value allows us to distinguish the areas of 'real' changes from those marked by the impact of random factors. One interesting application of this method is to apply it to images acquired before and after landslide events in order to identify landslides according to their NDVI value (Lin et al. 2005). In the frame of this study, we analysed where new landsides appeared or old ones were reactivated between 2002 and 2007. A number of 11 classes with 7 negative classes give the most representative results to recognise change in the vegetation canopy due to the disrupted vegetation because of landslide activation. The pre- and post-NDVI images have been created following the equation:

$$\Delta\mathrm{NDVI} = \mathrm{NDVI}_{2007} - \mathrm{NDVI}_{2002} \tag{2}$$

Data Mining Approach

Data Mining uses computational techniques (algorithms) from advanced statistical methods and employs pattern recognition such as neural network techniques. Data Mining also allows discovering previously unknown relationships among the data (pragmatic approach). In contrast to classical modelling attempts or statistical analysis, Data Mining is output-driven. This means that each successful strategy is allowed if it fulfils the previously defined aims and respects the requirements of the applied tools.

Since this method can handle large input datasets, multiple factors such as spectral information (from Quickbird images), recent landslide inventory, and data from the DEM (slope, curvature, topographic roughness index, landform index) were used as input data. The preparation of the data for modelling requires generally various adjustments, which have to be made to the data prior to modelling. Datasets were compiled on a GIS platform and transformed into raster data. The powerful sample command from ArcInfo GRID was used to export the raster in a flat file (like CSV for comma separated values) that could be directly imported in Data Mining tool PASW Modeler. The software then guides the planning in a systematic way to the goal of the study: detecting landslide. It focuses on the process of running data through a series of nodes, referred as a stream. Whereas ANN (Fernández-Steeger et al. 2002; Falaschi et al. 2009) and more recently Bayesians networks (Braun et al. 2011) have successfully been implemented for landslide susceptibility analysis, decision trees (C5.0, CHAID) were developed for landslides recognition in the framework of this study. The models use in advance digitised landslides to train and develop the model in a well-known area. Afterwards the model is tested in another area to evaluate the model skills.

Results

Size-frequency Distribution of Landslides

The frequency-area distribution quantifies the number of landslides per class of surface area. Following the method of Malamud et al. (2004), the size-frequency relationships of these five datasets described earlier were analysed in terms of the frequency density function (f) of the landslide areas (A_L) using the following equation:

$$f(A_L) = \delta \mathrm{N_L} / \delta \mathrm{A_L} \tag{3}$$

where $\delta \mathrm{N_L}$ is the number of landslides with areas between $\mathrm{A_L}$ and $\mathrm{A_L} + \delta \mathrm{A_L}$. $\delta \mathrm{A_L}$ is the width of the respective area class. In a log-log graph (Fig. 2), this function shows a more-or-less linear tail (in log-log graph) which can be fit by a power-law (Turcotte 1997). A rollover (inflection) is observed for smaller events and objects, here represented by the decreasing frequency density for very small landslide areas (Malamud et al. 2004; Havenith et al. 2006). The deviation from the power-law trend on the left part of the graph is partly explained by undercounting of the small landslides. As discussed by Malamud et al. (2004), many factors cause the incompleteness, such as the quality of imagery, landslide age and freshness and the potential removal of geomorphologic and spectral properties, the experience of the scientist involved. Pelletier et al. (1997) attributed the rollover shown by the probability density function of landslide areas, to a transition from a resistance controlled chiefly by friction (for large landslides) to a resistance controlled by cohesion (for small landslides).

Here, only the linear trend of the size-frequency relationships (e.g. for a size greater than the rollover, which is roughly 10,000 m^2) of the five inventories has been considered, which should be complete above that threshold. The fits of these linear tails of a complete relationship by a power-law are characterized by coefficients of determination, R^2, larger than 0.8.

From the size-frequency analysis presented in Fig. 2, it can be seen that the trends slightly but constantly change over time. The related exponent-values continuously decrease from 1.93 in 1962 to 1.73 in 2007. The value obtained for the 2002 dataset, 1.8, is similar to the power-law exponent determined by Havenith et al. (2006) for Maily-Say landslide records obtained from field observations in 2003, which is equal to 1.9. This confirms the observed trend indicating the more frequent formation of large landslides (in terms of objects and not in terms of events or movements) in the region. The increasing frequency of large landslides can be related to the growth of existing landslides or the coalescence of smaller landslides, as already suggested by Havenith et al. (2006). The latter process probably also contributes to the fact that the number of small landslides is decreasing over time.

NDVI Subtraction

The results of the NDVI calculation for the 2007 Quickbird image (Fig. 3) show that values close to zero represent landslides partly denudated or sites covered by sparse vegetation. The histogram analysis of the NDVI map of the entire zone shows that, the average NDVI is lower in 2007 (NDVI of 0.26) than in e (NDVI of 0.32). This difference may be due in part to the increasing landslide activity and to minor differences in vegetation growth stage. The mass movements of the Valley are generally complex combining deep-seated deformation with surficial erosion. Stable bare soil may be misclassified as landslides.

Figure 3 shows low NDVI where no landslide is located, but some of these false positives can be excluded from the difference map. The NDVI 2007–2002 subtraction maps

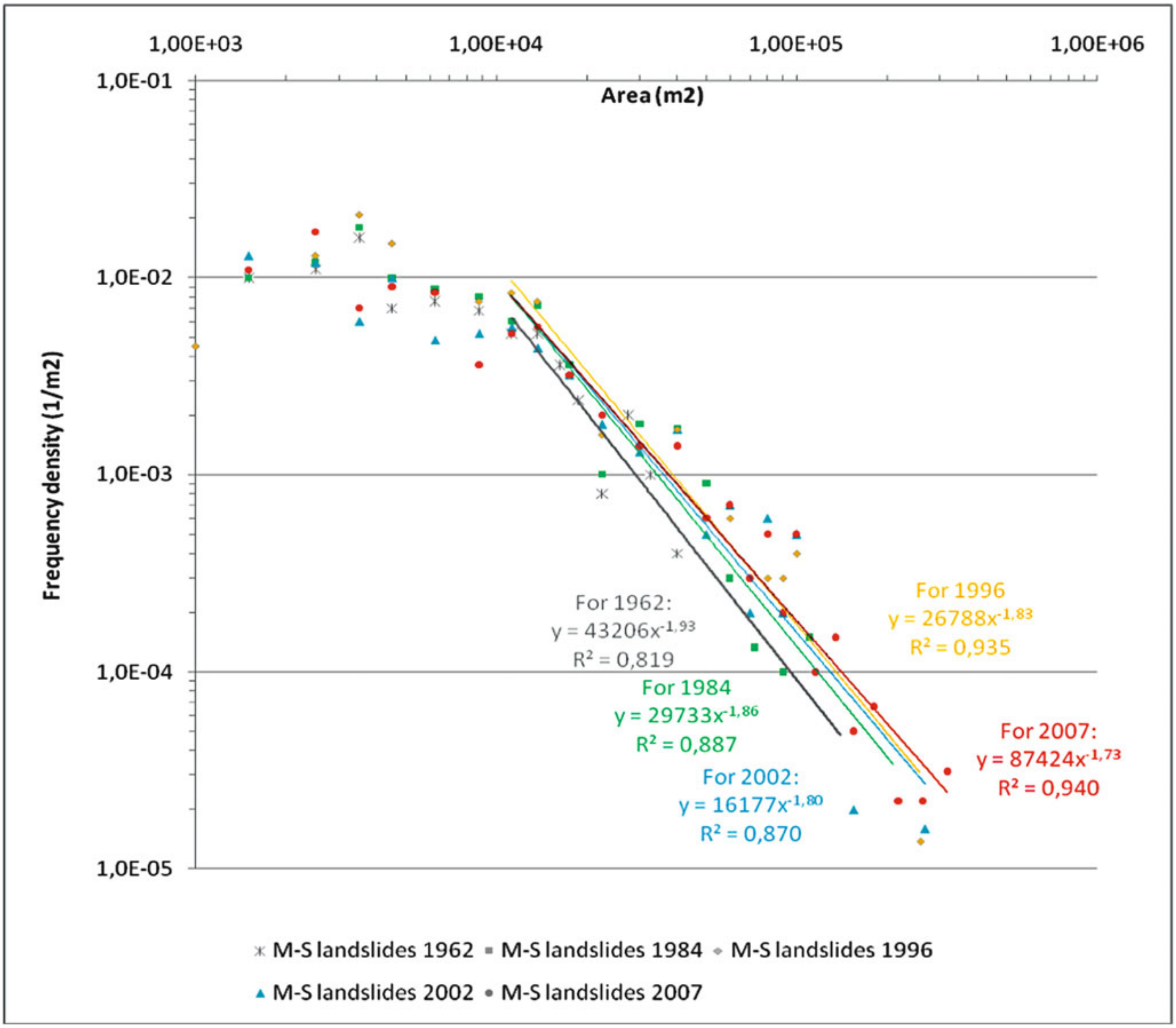

Fig. 2 Frequency density function for landslide areas in the Maily-Say Valley in 1962 (162 objects), 1984 (206 objects), 1996 (222 objects), 2002 (189 objects) and 2007 (208 objects)

show that in general, a strong decrease of NDVI could be attributed to the loss of vegetation due to recent activation of landslides. However, a simple correlation between the decrease of vegetation and the presence of a landslide cannot be established. Each landslide has particular characteristics due to its type of movement (fall, topple, slide, slump, spread or flow) and its type of material (rock, debris, earth material). In addition, the age of a landslide must be considered. With this method, primarily, fresh earth flows can be detected.

By visual analysis, it seems that new activations are represented by the classes 7 to 10, i.e. by the NDVI change values of −0.1 to −0.5. A statistical analysis shows that 88 % of the landslide areas are marked by NDVI change values of 0.0 to −0.5. However, only 27 % of the pixels inside the landslides belong to the NDVI change values of −0.1 to −0.5. This shows that only parts of the landslides can be clearly detected by the method.

Data Mining Using Decision Trees

The CHAID model was largely used to recognize if a pixel corresponds to a landslide area or not in the Maily-Say Valley.

In this study, additionally advanced parameters were taken into account to improve the modelling. This are the band 2 of 2007 and 2002 images, the slope, the topographical roughness index and the aspect. They contribute to about 80 % of the detections of landslide and no landslide using the CHAID algorithm as modelling tool. In addition, other types of models have been created and tested trying to improve the results. With the C5.0 model, the obtained results are quite good and

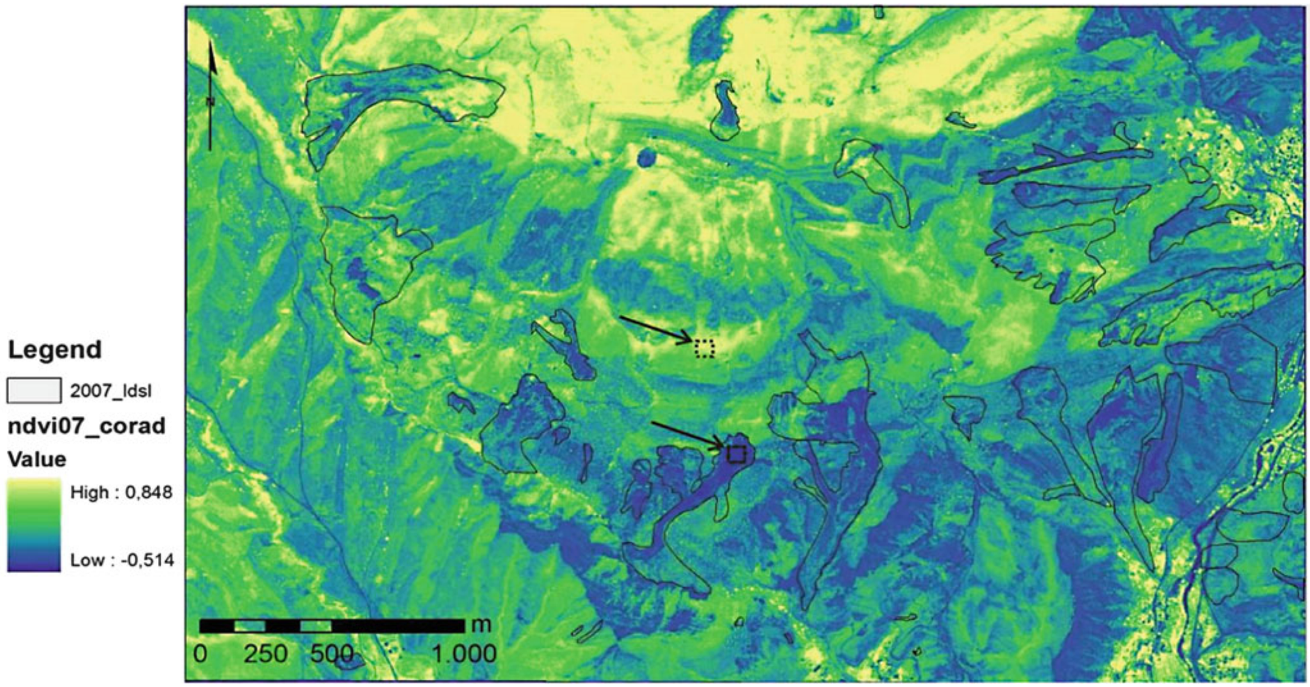

Fig. 3 NDVI of the radiometrically corrected 2007 Quickbird image of the Bedre Coupola area where *bluish* colours represent low NDVI and *greenish* to *yellow* colours represent high NDVI, and the 2007 landslide inventory for the zone (see also an example of denudation processes with low NDVI outlined by a *dashed ellipse*)

probably more relevant to detect landslides but it is often easier for the models to detect places where there are no landslides.

Following a partition at random of 30 % for testing and 70 % for training, concentrated on landslides, the model is only 69 % relevant. In inverting this partition (training area becoming testing area), the model accuracy reveals a more performing model with 82 %. Unfortunately, only 75 % of the pixels are well classified in the training model compared to 79.5 % obtained if we consider the test model. But the latter allows us to detect landslides in a defined environment which is the most interesting in detection.

Discussion and Conclusion

All statistics of the evolution of landslides indicate that the hazard is likely to increase in the Maily-Say Valley. Size-frequency analyses applied to landslide datasets of 1962, 1984, 1996, 2002 and 2007 show that the exponent-values of the power-law fits to the linear tails continuously decrease from 1.93 in 1962 to 1.73 in 2007. This reveals a trend towards the formation of increasingly larger landslides on the Maily-Say slopes.

The size-frequency behaviour of the landslides in Maily-Say can be compared with some observations made by other researchers concerning landslide distributions. All exponent-values obtained for the five inventories are well below those calculated by several other researchers, e.g., by Malamud et al. (2004) studying landslide inventories of landslide types and distributions as well (the values are between 2.11 and 2.48). However, the exponent values of the five inventories presented here compare well with those calculated by Chen (2009) for the landslides triggered by the 1999 Chi-Chi earthquake (value of 1.8). He noted that the exponent value increased for post-seismic landslides triggered by Typhoons in 2001 and 2004 (values of 2 to 2.1). This change of the exponent values of about 0.2–0.3 considered as significant by Chen (2009) is similar to the one observed for the Maily-Say landslides.

Here, we relate the continuous decrease of the exponent value of the landslide size-frequency distribution to the growth and coalescence of existing landslides which are relatively more frequent than the formation of new small slope instabilities. In this regard, it can be assumed that the changing size-frequency distribution of the landslide bodies in Maily-Say is related to the increasing density of slope instabilities. We think that a certain landslide density threshold has been surpassed, beyond which landslide occurrences become interdependent and are not only related to changing environmental conditions and external triggers. Furthermore, power-law exponent-values decreasing over time could indicate a trend towards a cataclysmic situation – this means that in near future larger parts of slopes could fail catastrophically. However, to be able to confirm this, we will also quantitatively assess changes in the landslide sizes and continuously monitor landslide reactivations in hazardous places. Therefore, automatic detection methods are needed.

As many new landslides are translational and constituted by clay and loess deposits, a change detection method, based on image subtraction to detect (re)activations was developed. Differencing of NDVI values determined from the multispectral Quickbird images ($NDVI_{2007} - NDVI_{2002}$) allowed us to outline zones with removed vegetation due to active slope failures considering inherent errors of commission in mapping. This analysis allows us to demonstrate that the multi-temporal differencing method is quite useful to detect both (re)activations of landslides and stabilizing slopes (marked by re-vegetation). However, it is not well adapted to map landslides or to create an inventory of them. For that purpose, a uni-temporal method is more appropriate such as Data Mining, which is able to combine many different parameters influencing slope stability and/or affected by the presence of landslides (e.g. spectral bands, curvature, slope, roughness). The data mining approach is also promising for the task of predicting landslide susceptibility but it requires more time and mathematical skills.

Acknowledgments These studies have been supported by the NATO Science for Peace Project 'Prevention of landslide dam disasters in the Tien Shan, LADATSHA'. We would like to thank reviewers of the 2nd World Landslide Forum for their helpful comments.

References

Abdrakhmatov K, Havenith H-B, Delvaux D, Jongmans D, Trefois P (2003) Probabilistic PGA and arias intensity maps of Kyrgyzstan (Central Asia). J Seismol 7:203–220

Brardinoni F, Slaymaker O, Hassan MA (2003) Landslide inventory in rugged forested watershed: a comparison between air photo and field survey data. Geomorphology 54:179–196

Braun A, Fernandez-Steeger TM, Havenith H-B, Torgoev A, Schlögel R (2011) Analysing the landslide susceptibility in Maily-Say, Kyrgyzstan, with statistical methods. In: Proceedings of the 2nd INQUA-IGCP-567 international workshop on active tectonics, earthquake geology, archaeology and engineering, Corinth

Chang KT, Liu JK (2004) Landslide features interpreted by neural network method using a high-resolution satellite image and digital topographic data. Int Arch Photogramm Remote Sens Sp Inf Sci 35:574–579

Chen CY (2009) Sedimentary impacts from landslides in the Tachia River Basin, Taiwan. Geomorphology 105:355–365

Falaschi F, Giacomelli F, Federici PR, Pucinelli A, D'Amato Avanzi G, Pochini A, Ribolini A (2009) Logistic regression versus artificial neural networks: landslides susceptibility evaluation in a sample area of the Serchio River valley, Italy. Nat Hazards 50:551–569

Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. AI Mag 17(3):37–54

Fernández-Steeger TM, Rohn J, Czurda K (2002) Identification of landslide areas with neural networks for hazard analysis. Landslides. In: Proceedings of the first European conference on landslides, Prague, Czech Republic; Rotterdam, Balkema, pp 163–168

Gupta RP (2003) Remote sensing geology, 2nd edn. Springer, Berlin, p 655p

Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation; a review of current techniques and their application in a multi-scale study, central Italy. Geomorphology 31:181–216

Jensen JR (1996) Introductory digital image processing: a remote sensing perspective, 2nd edn. Prentice Hall, Upper Saddle River, p 316

Havenith H-B, Torgoev I, Meleshko A, Alioshin Y, Torgoev A, Danneels G (2006) Landslides in the Maily-Say valley, Kyrgyzstan: hazards and impacts. Landslides 3:137–147

Lin WT, Chou WC, Lin CY, Huang PH, Tsai JS (2005) Vegetation recovery monitoring and assessment at landslides caused by earthquake in Central Taiwan. For Ecol Manage 210:55–66

Malamud BD, Turcotte DL, Guzzetti F, Reichenbach P (2004) Landslide inventories and their statistical properties. Earth Surf Proc Landf 29:687–711

Pelletier JD, Malamud BD, Blodgett T, Turcotte DL (1997) Scale-invariance of soil moisture variability and its implications for the frequency-size distribution of landslides. Eng Geol 48:255–268

Ridd MK, Liu J (1998) A comparison of four algorithms for change detection in an urban environment. Remote Sens Environ 63:95–100

Roessner S, Wetzel HU, Kaufmann H, Sarnagoev A (2002) Satellite remote sensing for regional assessment of landslide hazard in Kyrgyzstan (Central Asia). Zweites forum katastrophenvorsorge. In: Tetzlaff G, Trautmann T, Radktke KS (eds) Deutsches komitee für Katastrophenvorsorge e.V. (DKKV). Boon. pp 433–441

Schlögel R, Torgoev I, De Marneffe C, Havenith HB (2011) Evidence of a changing size-frequency distribution of landslides in the Kyrgyz Tien Shan, Central Asia. Earth Surf Proc Landf 36(12), pp 1658–1669

Torgoev IA, Alioshin YG, Havenith H-B (2002) Impact of uranium mining and processing on the environment of mountainous areas of Kyrgyzstan. In: Merkel BJ, Planer-Friedrich B, Wolkersdorfer C (eds) Uranium in the aquatic environment. Springer, Berlin/Heidelberg/New York, pp 93–98

Turcotte DL (1997) Fractals and chaos in geology and geophysics, 2nd edn. Cambridge University Press, Cambridge, p 398p

Vandenhove H, Quarch H, Clerc J, Lejeune J, Sweeck L, Sillen X, Mallants D, Zeevaert T (2003) Remediation of uranium mining and milling tailing in Maily-Say district of Kyrgyzstan. Final report of EC-TACIS Project N°SCRE1/N°38, Vandenhove H Q H, Mol (Belgium), 614p