INVESTIGATING THE TRANSFERABILITY OF STATISTICAL DISPOSITION MODELS FOR SLOPE-TYPE DEBRIS FLOWS

Tobias Heckmann and Michael Becht

With 5 figures and 4 tables
Received 9 December 2008 · Accepted 20 February 2009

Summary: Statistical disposition models are used both in geomorphology and in natural hazards research to spatially predict the occurrence of mass movements, other geomorphic processes or landforms. This is achieved by establishing some statistical relationship between an inventory of mapped occurrences and relevant geofactor maps. Statistical approaches rely, among others, on the assumption that future events will take place under the same conditions as they did in the past. In the present paper, the Certainty Factor (one of the favourability functions) is used to quantify the association of debris flow initiation sites with the geofactors landcover, slope and CIT index in three alpine study areas. The transferability of the models, i.e. their applicability to other study areas, is investigated using a comparison of model results (CF parameters) from different study areas and an extensive cross-validation. It is shown that, in similar study areas, the model parameters are generally very similar, indicating that the geofactors represent well enough the relevant conditions and processes that lead to debris flow initiation. The cross-validation is generally successful, independent of the similarity of the area for which the model has been established and the target area. Therefore, it can be concluded that statistical disposition models such as the CF model are, in principle, transferable between study areas. Nevertheless, some critical issues such as data quality have to be kept in mind when trying to apply a model to another study area.

Zusammenfassung: Die räumliche Verteilung von Massenbewegungen, anderen Naturgefahrenprozessen und auch Oberflächenformen kann durch statistische Dispositionsmodelle vorhergesagt werden. Diese Modelle stellen einen statistischen Zusammenhang zwischen dem kartierten Vorkommen der Phänomene und einer Reihe von Geofaktoren her, die für das Auftreten der Phänomene von Relevanz sind. Ein solcher Ansatz basiert vor allem auf der Grundannahme, dass zukünftige Ereignisse unter denselben Bedingungen auftreten wie die aus der Vergangenheit bekannten. Die vorliegende Studie nutzt den Certainty Factor (ein statistischer Ansatz aus der Familie der favourability functions) zur räumlichen Modellierung von Muranrissen in drei alpinen Einzugsgebieten. Modelle, welche die Geofaktoren Landnutzung, Hangneigung und CIT-Index betrachten, werden auf ihre Übertragbarkeit zwischen den verschiedenen Untersuchungsgebieten überprüft. In Gebieten, die einander sehr ähnlich sind, ergeben sich sehr ähnliche Parameter der CF-Modelle, was auf eine allgemeine Eignung der drei gewählten Geofaktoren hindeutet. Die Kreuzvalidierung der Modelle in den Untersuchungsgebieten (bzw. Teilen davon) zeigt empirisch, dass unabhängig von der Ähnlichkeit zweier Untersuchungsgebiete generell von einer Übertragbarkeit des statistischen Modells auszugehen ist. Die Vergleichbarkeit und Qualität der in den Gebieten vorhandenen Datenbasis muss jedoch kritisch hinterfragt werden, bevor ein Transfer von Modellen auf andere Untersuchungsgebiete erfolgt.

Keywords: Spatial modelling, mass movements, debris flows, certainty factor, transferability

1 Introduction

Slope-type debris flows represent an important geomorphic process concerning the transfer of sediments through a geosystem (Haas et al. 2004; Becht et al. 2005), and they may pose a serious geomorphic hazard where buildings or infrastructure are threatened. They initiate on steep talus slopes when high-intensity rainfall leads to positive pore water pressures and to failure (Zimmermann et al. 1997; Rieger 1999; Hagg and Becht 2000). Contrary to torrent-type debris flows (debris torrents sensu Slaymaker 1988 which initiate in channels), this process is rather transport-limited than material-limited.

The occurrence of debris flows can be mapped on large scale (hillslope to catchment scale) geomorphological maps (e.g. Glade 2005; Seijmonsbergen and deGraaf 2006). On smaller scales, e.g. on the regional scale, a complete field survey is very time-consuming and therefore can be costly. Moreover, some studies suggest that the traditional field survey may introduce a high degree of subjectivity (Carrara et al. 1992; Ardizzone et al.

2002; VAN ASSELEN and SEIJMONSBERGEN 2006; the latter present a procedure for semi-automated geomorphological mapping). The same has been argued in the discussion of a methodology termed "predictive geomorphological mapping" (LUOTO and HJORT 2005). Along the lines of susceptibility models for natural hazards (e.g. GUZZETTI et al. 1999), the same or similar methods are used here to delineate the spatial distribution of the occurrence of geomorphic processes or of landforms/landscape characteristics indicative of their activity. LUOTO and HJORT (2005) for example developed a model to "predict" the occurrence of patterned ground on the basis of geodata derived from digital elevation models and remotely sensed maps.

In the present study, a similar approach is chosen to delineate the de-facto and potential initiation sites of slope-type debris flows (not the total area affected).

Methods for modelling the spatial occurrence of mass movements can be classified in rule-based, statistical and physically-based approaches (see e.g. HECKMANN and BECHT 2006). A model that includes the relevant physical processes leading to failure and debris flow initiation and that yields a valid prediction of such sites in a given area should, conceptionally, be transferable to any area in the world. This assumption is based on the notion that the mechanism of the process is based on universal principles. Physically based models, however, imply that all necessary parameters have to be known with both their magnitude and their spatio-temporal distribution. This prerequisite is problematic and therefore one of the main reasons why statistical models have been so frequently applied in modelling disposition or susceptibility. With this class of models, statistical relationships, e.g. the parameters of regression or discriminant functions, are sought between the spatial occurrence of the process and the relevant geofactors on these areas. The key assumption of this approach is that future events will occur under the same circumstances as they have previously. While statistical models often yield valid and useful results, e.g. for catchment scale hazard susceptibility analyses (see e.g. CHUNG and FABBRI 2003), their applicability to different areas (i.e. their transferability) is questionable. One of the reasons for doubting the transferability is that any geofactor, e.g. a derivative of a digital elevation model, only represents a proxy for the activity of an underlying physical processes – which may be influenced by the geofactor in an ambiguous way and only to a limited (generally unknown and possibly variable) degree.

The present study is an empirical investigation of the transferability of a statistical method for modelling the spatial distribution of slope-type debris flow initiation sites by comparing and cross-validating model results in three different alpine study areas. After a comparative description of their properties (section 2), the modelling method and the steps of the scientific approach are introduced (section 3). The results (section 4) are discussed in section 5, and some conclusions are drawn.

2 Study areas

Three study areas have been chosen to analyse the transferability of a statistical disposition model for slope-type debris flows. They are located in different parts of the Alps and are different with respect to their lithology, geomorphology and climate. Figure 1 gives an overview of the location of the study areas, table 1 compares them with respect to the most important parameters.

Table 1: Selected properties of the study areas. In the GS study area, models have been calculated for the three subareas MB, WBA, PFO as well as for combinations of any two of them. The disposition map in figure 2 results from a model calculated using a combination of all GS subareas.

Study Area	Area [km²]	Elevation [m asl]	Average slope [°]	Precipitation [mm/a]	Geology
LWG	16.7	715–1985	28.4	1600–2000	limestone, marls
RT	17.3	1052–2744	41.3	1600–2000	limestone
GS	103	1218–2835	29.3	ca. 850–1000	crystalline rocks (gneiss, mica schists)
-GS_MB	2.2	1349–2490	33.2		
-GS_WBA	4.5	1620–2835	32.6		
-GS_PFO	1.9	2050–2715	30.7		

The mostly subalpine (715–1985 m a.s.l.) **Lahnenwiesgraben** basin (abbrev. LWG; 16.7 km²) is located in the Ammer mountain range in the Northern Calcareous Alps (Bavaria, Germany). The area consists of Triassic sedimentary rocks ("Hauptdolomit" dolostone, "Plattenkalk" limestone, "Koessen" marls and some Jurassic limestones and marls). The V-shaped valley was glaciated both locally and by ice entering the valley from the Loisach valley during the last ice age. Slope-type debris flows occur mostly in the upper parts of the basin where they initiate from talus slopes at the base of steep rock faces (mostly Hauptdolomit lithology) and in glacial cirques.

Some 10 km to the south, the **Reintal** basin (RT; 17.3 km²) has a more alpine character (1052–2744 m a.s.l.) and consists mainly of a very pure and homogenous "Wettersteinkalk" limestone. The Reintal val-

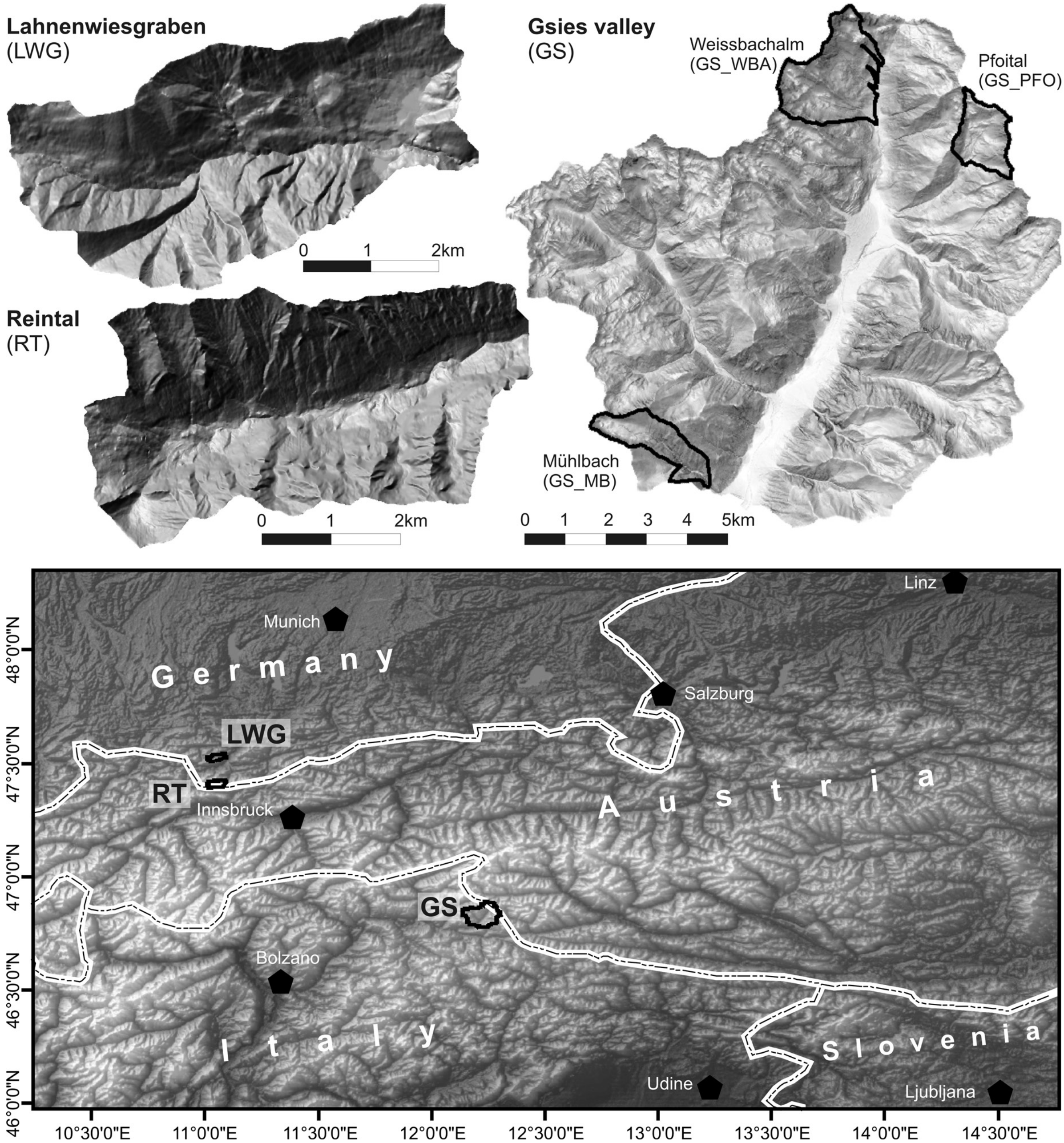

Figure 1: Overview of the study areas

ley is a glacial trough (local glaciation) with hanging valleys/cirques in the upper regions. Debris flows are frequent on the large talus cones and sheets (c.f. Becht et al. 2005).

The alpine (1218–2835 m a.s.l) basin **Gsieser Tal** (Valle di Casies; GS; 103 km²) is located in the Central Alps of South Tyrolia/Italy. Its lithology consists mostly of metamorphic rocks (gneisses and schists) and some quaternary sediments, including talus, moraines and alluvial sediments. Within the GS basin, different valley cross profiles can be found (glacial troughs and V-shaped valleys). Debris flows have been mapped in three small subareas (MB=Mühlbach: 2.2 km², WBA=Weissbachalm: 4.5 km² and PFO=Pfoital: 1.9 km²). They are mainly located on steep scree and soil-covered slopes derived from gneiss and mica schist rocks (Hertel 2007).

3 Material and methods

3.1 Methods

3.1.1 The Certainty Factor

The certainty factor (CF) is a bivariate statistical method belonging to the family of favourability functions (Chung and Fabbri 1993; Pistocchi et al. 2002). Originally developed for the diagnosis of blood diseases under uncertainty, the method has been used in the geosciences for the spatial modelling of mineral resources (Chen 2003) and various mass movements (shallow landslides: Binaghi et al. 1998; earth flows: Keller et al. 2005; snow avalanches: Heckmann et al. 2005; slope-type debris flows: Wichmann et al. in press, Wichmann 2006, Hertel 2007). In these studies, the method is used on a raster-cell basis.

There are two reasons for choosing this method. Firstly, it has a conceptual advantage over other statistical methods as e.g. logistic regression or discriminant analysis. These approaches require both mapped initiation sites and "non-initiation sites" (e.g. as values 1 or 0 of the dependent or group variable, respectively) in order to estimate the parameters of the regression (or discriminant) function. While the former are readily provided as initiation zones are mapped in the field, the identification of "non-initiation sites" implies some uncertainty, as it is questionable to presume that a process definitely cannot occur where it hasn't (yet) been documented to occur. This source of uncertainty or even error is avoided here as the CF approach only requires initiation sites. In a comparative case study on landslides, the CF method provided the best predictive power compared with a discriminant analysis and a heuristic method (Remondo et al. 2003). The second reason for choosing the CF method is the possibility to interpret the results for each geofactor value separately (see below). This avoids the problem of calculating disposition values for combinations of geofactors (e.g. "unique condition subareas") that can be unrealistic for combinations present on only very few (or even single) cells (c.f. Heckmann and Becht 2006). Besides, the calculations are computationally simple and can be done completely within a GIS.

The basis of the CF method is a comparison of prior probability p(R) and the conditional probabilities p(R|G) calculated for each category G of each influencing geofactor (eq. 1). The resulting number is standardised to [-1.1]:

$$CF^{+}=\begin{cases}\dfrac{p(\mathrm{R|G})-p(\mathrm{R})}{p(\mathrm{R|G})\times(1-p(\mathrm{R}))} & \text{if } p(\mathrm{R|G})\geqslant p(\mathrm{R})\\[2ex] \dfrac{p(\mathrm{R|G})-p(\mathrm{R})}{p(\mathrm{R})\times(1-p(\mathrm{R|G}))} & \text{if } p(\mathrm{R|G})< p(\mathrm{R})\end{cases} \qquad \text{(eq. 1)}$$

The prior probability p(R) is calculated by dividing the number of mapped initiation cells by the cell count of the study area – in this way, p(R) can be interpreted as "debris flow density". The conditional probability p(R|G) is calculated by dividing the number of mapped initiation cells within the respective geofactor category G (e.g. the land cover class "unvegetated") by the number of cells of this category.

The resulting CF^{+} parameter assigns a certainty or degree of favourability [-1;1] to the association of a hypothesis („debris flows can initiate here") with some evidence or information ("geofactor category XY is present here"). A CF^{+} value of >0 indicates that the evidence provides some support for the hypothesis (CF^{+}=1 means strong support), values <0 indicate that the evidence somehow opposes the hypothesis (CF+=-1 means strong opposition), and CF^{+}=0 means that the evidence neither increases nor decreases the certainty concerning the association of the evidence with the hypothesis. There is a monotonous, non-linear relation between the CF^{+} and the ratio of conditional and prior probability (i.e. a higher ratio leads to a higher CF^{+}).

The CF values for each geofactor category are listed in a table, where the relative importance of geofactor categories can be interpreted. Additionally, a CF^- is calculated using the prior probability p(R) and the conditional probability p(R|G') which is the probability that initiation takes place when the respective geofactor category is absent. By subtraction of CF^- from CF^+, the so-called contrast CC [-2;2] is calculated which will be used in this study as a measure of disposition.

$$CF_{1\wedge 2}=\begin{cases} CF_1+CF_2-CF_1\times CF_2 & \text{if } CF_1, CF_2 \geqslant 0 \\ \dfrac{CF_1+CF_2}{1-\min(|CF_1|,|CF_2|)} & \text{if } CF_1 \times CF_2<0 \\ CF_1+CF_2+CF_1\times CF_2 & \text{if } CF_1, CF_2<0 \end{cases} \quad \text{(eq. 2)}$$

Using eq. 2, the CF values (both CF+ and CF-) are combined for each raster cell of the study area, according to the geofactor configuration on this cell (CF1 and CF2, e.g. slope classes and land cover). When the CF factors of more than two geofactor maps are to be combined, this is done by stepwise combination of ever two geofactors (the order is equal). From the combined CF^+ and CF^- values, the combined contrast CC is calculated. The combined CF^+, CF^- and CC are displayed as raster maps indicating the spatial distribution of disposition.

For further details and a discussion of the method, the reader is referred to Binaghi et al. (1998), Chung and Fabbri (1993) and Heckmann and Becht (2006).

3.1.2 Approach of the investigation

According to the working hypothesis, statistical models correctly capture the decisive geofactors for debris flow initiation if the results (in this case the CF or CC parameters for the single geofactor categories) are similar in similar study areas. If different geofactors were responsible for debris flow activity in any two study areas, the CF parameters are expected to differ even if the two study areas under comparison are very similar.

In order to empirically test these considerations, CF models are calculated using the mapped debris flow initiation sites in the following (sub-)areas (c.f. Fig. 1):

- Lahnenwiesgraben (LWG)
- Reintal (RT)
- Mühlbach (MB), Weissbachalm (WBA) and Pfoital (PFO), all of which are subareas of the GS study area. Additionally, any two of them were combined (MBWBA, MBPFO and WBAPFO) in order to increase the number of different study areas.

Both categorical and continuous geodata are classified in all study areas using the same classification scheme (i.e. the same categories and class limits) in order to ensure comparability. In a first step, the CF tables of any pair of two models are compared (section 4.3). For the respective CC values, the mean absolute difference and the root mean square (RMS) difference are calculated as distance measures.

The CF tables of the single models are then applied to any of the remaining areas in order to test their "predictive" power (cross validation; see section 4.4). For example, the CF table generated for LWG is applied to the MB basin, and the CC map is produced. Using the validation procedure for spatial prediction models suggested e.g. by Chung and Fabbri (2003), a prediction curve is constructed which shows how many % of the mapped debris flow initiation sites in MB are covered by the p% percentile of the CC map generated by the LWG model (c.f. Fig. 3). The validation is successful if a large percentage of initiation sites are located within the 90- or 80% percentile of the CC map, i.e. if the leftmost part of the curve is very steep. If the validation curve is close to the main diagonal of the diagram, the prediction is as good as random, and the validation can be considered a failure. By calculating the area under the curve (AUC) of the prediction curve, a single value can be given in order to quantify the validation. The AUC is in the interval [0.5 ; 1] with values close to 1 indicating a high predictive power (i.e. a good validation) and values close to 0.5 indicating that the model is not capable of predicting debris flows initiation sites better than random.

In order to analyse both CF/CC comparisons and cross-validations with respect to the (dis-)similarity of any two study areas, a measure for similarity has to be established. The percentage similarity index proposed by Whitaker (1952 fide Lydy et al. 2000; see the latter paper for a compilation of several indices) is commonly used, amongst others, for comparing the species assemblage of biotopes. In this study, it is used to compare two (sub-)areas with respect to their geofactor composition. The index is simply based on two frequency tables (one for each study area) where, for each combination of geofactor categories (e.g. land cover=1, slope=2, lnCIT=6), the lowest of the two relative frequencies is summed up. This results in a similarity index between 0 (if none of the categories present in one area is present in the other – this indicates total dissimilarity) and 100 (all categories are present in both study areas, and their relative frequencies are equal indicating identity of the two study areas). The similarity index for each pair of (sub-)areas is shown in table 2. The similarity indices for pairs of (sub-)areas that are completely independent from each other (i.e. that do neither contain, nor are contained by, nor share a part with another subarea) are set in boldface. It can be seen that LWG and the PFO subarea of GS are the most dissimilar, and the PFO and WBA subareas of GS are the most similar independent study areas. Note that only the geofactors used for the CF analysis (slope, CIT index and land cover, see section 3.2) are contained in the similarity analysis; the similarity index thus does not indicate "overall" similarity of the study areas.

3.2 Data

Preconditions for debris flows involve availability of mobile sediments, steepness of slopes and availability of water. A statistical model therefore has to include geodata that represent these preconditions. As previous works have shown (Wichmann et al. in press; Wichmann 2006; Hertel 2007), the location of potential slope-type debris flows can be effectively modelled using only three geofactors: Slope, CIT index and land cover. Therefore, considering the focus of this study, no exploration of the suitability of other or additional data layers is performed.

The slope map is derived from a digital elevation model using the polynomial approximation method suggested by Zevenbergen and Thorne (1987). The CIT index (c.f. Montgomery and Foufoula-Georgiou 1993) is a sediment transport capacity index used for modelling channel initiation and is calculated by multiplying the specific catchment area (catchment area per unit contour length) with the squared tangent of slope. Raster cells with high CIT index are characterized by both large upslope catchment areas (representing the availability of water) and steep slope. The CIT index is used as a proxy for the erosional power of surface runoff which is considered to cause the initiation of debris flows by progressive erosion of small channel headcuts. Due to the very large range of CIT values, the logarithm of CIT (lnCIT) is used in this study. The range of lnCIT in all study areas is approximately between -15 and 15 while ca. 95% of the values are between -1.1 and 6.6 (mean $+/- 2\sigma$).

As the calculation of the CIT index is based on slope, these two geofactors are not statistically independent which formally violates an underlying assumption of the statistical concept (see e.g. Agterberg and Cheng 2002). In spite of the correlation between the two geofactors (highly significant exponential correlation with $r^2=0.43$), both of them remain included in the model calculations as they represent two different mechanisms (slope as a factor governing the instability of loose sediments, CIT as a proxy for fluvial erosion potential) for debris flow initiation.

The next three subsections give some information on the data used in this study.

Table 2: Pivot table showing the similarity of any two (sub-)areas as indicated by the percentage similarity index [0;100]

(Sub-)Area	LWG	RT	GSIES	GS_MB	GS_WBA	GS_PFO	GS_MBWBA	GS_MBPFO	GS_WBAPFO
LWG	100								
RT	**36**	100							
GSIES	**17**	**26**	100						
GS_MB	**21**	**18**	63	100					
GS_WBA	**13**	**26**	87	**51**	100				
GS_PFO	**8**	**24**	78	**46**	**79**	100			
GS_MBWBA	**18**	**25**	94	67	84	**71**	100		
GS_MBPFO	**19**	**23**	85	75	**72**	71	87	100	
GS_WBAPFO	**12**	**26**	87	**51**	94	85	83	73	100

3.2.1 Digital elevation models

For the northern alpine study areas (Lahnenwiesgraben, Reintal), a DEM with a resolution of 5x5 m has been interpolated from official elevation data (Bavarian topographical survey, Munich: VM-DLZ-LB0628) using the TOPOGRID algorithm implemented in the ArcGIS spatial analyst (ESRI).

The province of Bolzano provides a LiDAR DEM with a resolution of 2.5x2.5 m (http://www.provinz.bz.it/umweltagentur/geobrowserpro/). In order to make the DEM data as comparable as possible, the Gsies valley DEM was resampled to 5x5 m resolution.

3.2.2 Land cover

Land cover maps were produced manually on the basis of aerial photos (with field validation) in the Lahnenwiesgraben and Reintal basins (Wichmann 2006). Five land cover classes (unvegetated surfaces or pioneer vegetation, grass, krummholz, shrubs, forest) of the original seven were derived for this study.

The 19 classes of the Bolzano province "Realnutzungskarte" (land use map; http://www.provinz.bz.it/umweltagentur/geobrowserpro/) of the Gsies valley were reclassified in order to match the land cover classes in the two other study areas as closely as possible. An additional sixth class ("other") was introduced to include land use types not present in the northern study areas (agricultural land, settlements, buildings etc.). Due to the scale of the original vector map, small unvegetated areas, above all channels and their tributary slopes, are not resolved inside larger forest areas. As several debris flows initiated on those slopes, the drainage system of the GS basin was generated from the DEM, and the cells belonging to the drainage system were set to "unvegetated" in order to improve the land cover map with respect to potential initiation sites.

3.2.3 Debris flows initiation sites

Debris flows initiation sites (both for calibration and validation of the model) have been mapped in the field in all study areas, supported by aerial photos. The resulting field map was digitised and rasterised; a single initiation site is mostly represented by one single raster cell. In this context, it is important to mention that initiation sites/zones rather than whole debris flows areas (including initiation, transit and deposition zones) are used. Conceptionally and geomorphologically, this approach is undoubtedly more meaningful than mapping the total process area, though it has been reported that it does not automatically lead to better results (c.f. Magliulo et al. 2008).

In the two northern study areas, a complete survey was conducted (Wichmann 2006), while in the Gsies valley, detailed field mapping was done only in the three small subareas mentioned above (Hertel 2007, see also Fig. 2). From the orthophotos provided by the Bolzano province (http://www.provinz.bz.it/umweltagentur/geobrowserpro/), additional initiation sites could be mapped in order to increase the sample size. In total, the inventory comprises 73 initiation sites in the GS subareas, 211 in the LWG area and 75 in the RT area.

4 Results

4.1 CF Model results

In this section, the most relevant geofactors in all study areas are discussed based on their CC values calculated by CF analysis (Tab. 3), and a disposition map of the Gsies Valley based on all mapped initiation sites is given (Fig. 2) as an example.

The CF tables calculated for each (sub-)area can be interpreted with respect to geofactor categories that influence the initiation. CF [-1;1] and CC [-2;2] values substantially greater than zero indicate positive influence, CF and CC values substantially less than zero indicate negative influence. Values of $|CF| < 0.2$ and $|CC| < 0.4$ are arbitrarily chosen to identify classes lacking interpretable influence. Table 3 contains positive CC values only (as only geofactors associated with initiation are discussed here), and those larger than the 0.4 threshold are set in boldface.

If those geofactors are most important that have high CC values in the majority of study areas, the raster cells most prone to debris flow initiation are generally unvegetated and steep (35–50°) with a medium to large CIT index.

Looking at the CC values of a particular geofactor in a single area, the change of CC values is generally consistent. In the GS area, for example, the CC values increase from 0.2 to 0.8 for the categories 3–6 and decrease again to 0.4 for the larger lnCIT classes. In some instances, however, CC values change inconsistently: The lnCIT records for the LWG basin, for example, show an increase of CC towards catego-

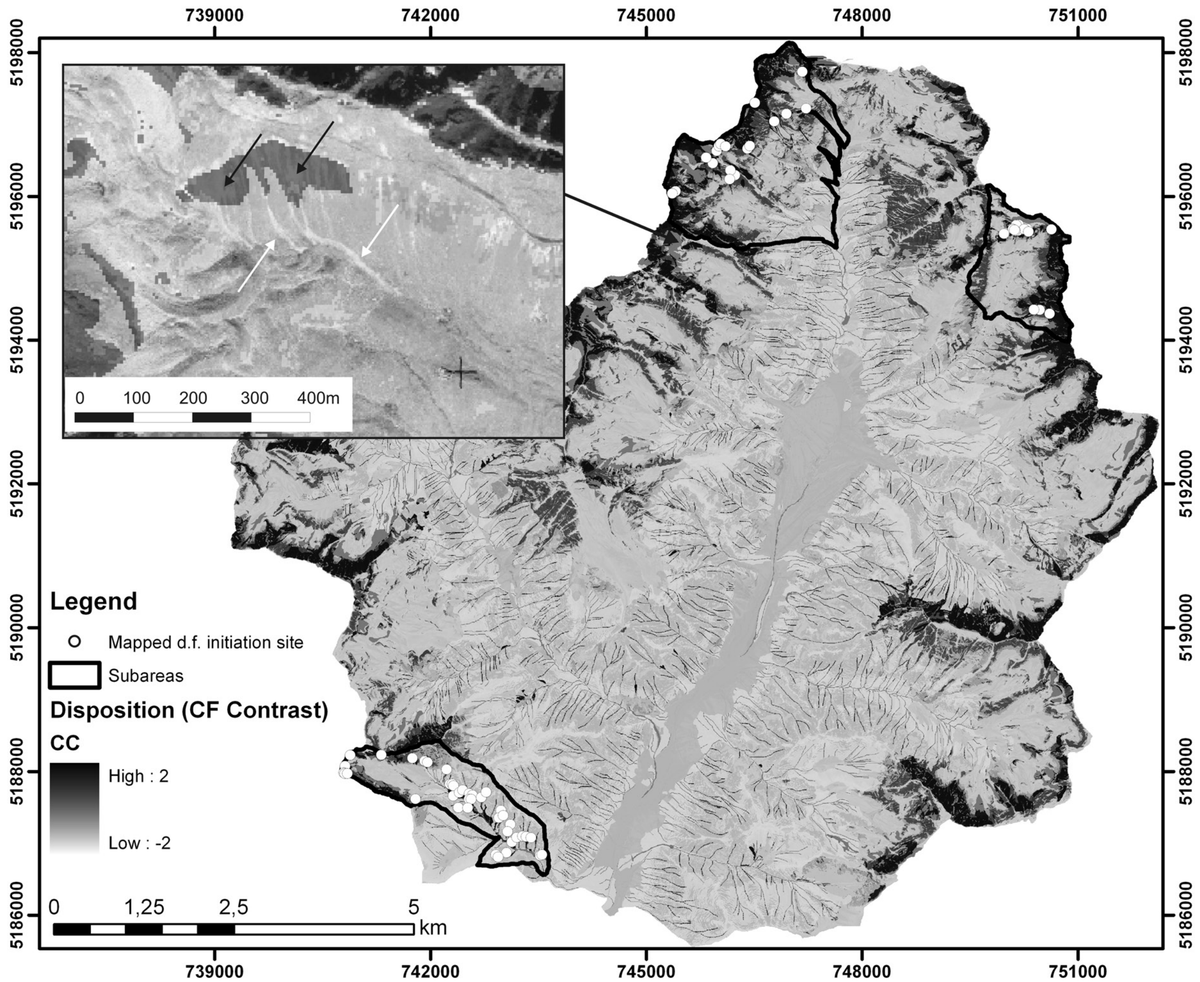

Figure 2: Disposition map of the GSIES area (CC values [-2;2] calculated using the model based on the combined subareas MB, WBA and PFO) and debris flow initiation sites mapped in these subareas. The inset shows a location (outside of the mentioned subareas) where the model correctly "predicts" debris flow initiation (dark grey areas, black arrows; white arrows point to the corresponding runout zones). Background of inset: Orthophoto © Bolzano Province.

ry 5-6, then a decrease towards category 7–8 before increasing again in category 8–9. Another example is the existence of high CC values for the highest slope class, notwithstanding the largely consistent behaviour of the CC values in the lower classes. For a discussion of these observations, the reader is referred to section 5.

Table 3 also shows that some categories are, according to the model results, "important" in only few or even single study areas, and that there are considerable differences between the CC values for each category. These differences are explored in section 4.3 where they are compared to the degree of similarity between the respective study areas.

The CC map (Fig. 2) of the study area GS indicates that the zones most prone to debris flow initiation are located in the uppermost parts of the area and near the channels (the latter is due to lnCIT geofactor and the attribution of the drainage network as being unvegetated). A qualitative validation can be undertaken by comparing the disposition map with additional debris flow initiation sites mapped outside of the subareas in which the field survey was completed. The inset in figure 2 exemplifies that the model indeed predicts debris flow initiation sites (black arrows). Additionally, it is important to mention that the runout areas further downslope (white arrows) do not show high CC values, indicating that the model specifically identifies (potential) initiation sites. This differentiation would have been very difficult (if not even impossible) if the model had been calculated for terrain units instead of raster cells.

Table 3: Comparison of model results (here: CC-values ≥0.4) for each geofactor category and each (sub-)area

Geofactor	Category/Class	LWG	RT	GSIES	GS_MB	GS_WBA	GS_PFO	GS_MBWBA	GS_MBPFO	GS_WBAPFO
Landcover	unvegetated	**1.6**	**1.2**	**1.4**	**1.7**	**1.1**	**0.9**	**1.5**	**1.5**	**1.1**
Landcover	krummholz	**0.4**								
Landcover	shrubs		**0.6**							
lnCIT	2 - 3				0.2			0.0	0.2	
lnCIT	3 - 4	**0.4**		0.2	0.1	0.2	0.3	0.1	0.2	0.2
lnCIT	4 - 5	**0.7**	**0.4**	0.2	0.1	**0.4**	0.3	0.2	0.1	**0.4**
lnCIT	5 - 6	**0.7**		**0.8**	**0.4**	**1.0**	**0.9**	**0.7**	**0.6**	**0.9**
lnCIT	6 - 7	0.3	**0.4**	**0.6**	**0.8**			**0.7**	**0.8**	
lnCIT	7 - 8	0.1	**0.8**	**0.4**	**0.8**			**0.5**	**0.7**	
lnCIT	8 - 9	**0.7**	**1.0**							
lnCIT	9 - 10		**1.1**							
lnCIT	>10		**0.9**							
Slope	15-20°		0.0							
Slope	20-25°		**0.5**							
Slope	25-30°		**0.7**							
Slope	30-35°		**0.7**							
Slope	35-40°	**0.6**	0.3	**0.7**	**0.4**	**0.6**	**1.3**	**0.5**	**0.8**	**1.0**
Slope	40-45°	**0.9**	0.1	**0.9**	**0.9**	**0.9**	**0.5**	**0.9**	**0.8**	**0.8**
Slope	45-50°	**0.9**		**0.6**	**0.5**	**0.9**	0.2	**0.7**	**0.5**	**0.7**
Slope	50-55°	**0.8**			**0.4**			0.2	0.2	
Slope	60-65°	**0.4**								
Slope	65-70°	**0.7**		**1.0**	**1.0**			**1.0**	**1.0**	

4.2 Validation of the model

As the focus of this study is on transferability of models between study areas, the validity of each single model (e.g. the model calculated for LWG) is assessed only by a "goodness of fit" measure ("success rate" according to Chung and Fabbri 2003) as explained in section 3.1.2. Generally, the predictive power of a single model should be validated using different starting zones for model training/calculation and validation, which can be done by splitting the initiation sites dataset randomly, spatially or temporally (c.f. Chung and Fabbri 2003).

The fields shaded in grey in table 4 show that basically all models are successful (indicated by *area under the curve* values between 0.88 and 0.96) in detecting the initiation sites based on which they were calculated. According to the *area under the curve* property of the success rate curve, the least performing models are the one for the GS_WBAPFO subarea (a combination of WBA and PFO subarea within the GS study area) and the GSIES model. Figure 3 shows the success rate curves for the GS model (Fig. 2) and the GS_MB model, which has the best goodness of fit. Even in the GS model, >80% of the mapped initiation zones are correctly identified on only 20% of the study area (the 80% percentile of the disposition map). Chung and Fabbri (2003) who calculated a statistical disposition model for shallow landslides have ca 70% of the slides predicted on 20% of the study area (the success curve identifies 90% of the slides on 20%) and speak of a successful validation. The

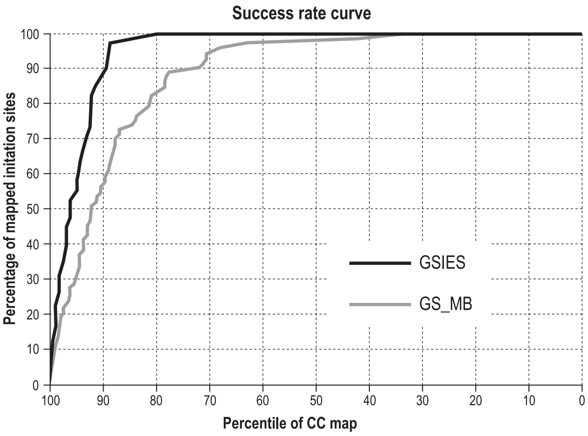

Figure 3: Success rate curve (c.f. CHUNG and FABRI 2003) showing a good (model GS_MB) and a comparatively poor fit (model GSIES). Prediction rate curves are constructed accordingly (but validating the application of a model A to a different area B). Model quality is assessed using the area under the success or prediction curve, respectively.

model with the best fit (GS_MB) "predicts" 90% of the mapped sites on only 10% of the study area (90% percentile) and has an AUC of 0.96. More important than the "goodness of fit" validation within a single study area, however, is the validation of models in different areas (cross validation, section 4.4).

4.3 Comparison of CF parameters for different areas

In each (sub-)area, debris flow initiation sites have been mapped, and CF models have been calculated. Given that the employed geofactors (land cover, slope, CIT index) represent the mechanisms of debris flows initiation well enough, and that these mechanisms are valid in the same way in any study area, the CF parameters (CF+, CF- and CC) of the single geofactor categories are expected to be equal or similar in similar study areas (see section 5). Otherwise, if e.g. the CIT index played a different role in two very similar study areas, the CF and CC parameters of the CIT categories should be different. Some degree of difference between the CF parameters is generally expected because of the different number of mapped initiation sites, their different spatial distribution in any two study areas and because of possible positional errors in the mapping process.

Figure 4 shows the difference between the CF tables of ever two models (quantified by the RMS distance measure for the CC parameters) plotted against the percentage similarity index of the respective study areas. (Sub-)areas that are neither part of nor contain the respective area of comparison are distinguished from those that are somehow dependent on each other (the same distinction was made for tables 2 and 3). It can clearly be seen that the CF parameters of two models tend to be more similar the more similar the two study areas are. The coefficient of determination of the resulting linear correlation is r^2=0.47, the rank correlation coefficient after Spearman is r_S=-0.60 ($p < 0.001$).

4.4 Cross-validation

In this section, the predictive power of each model is evaluated by applying it to another study area and calculating the area under the curve for the prediction rate curve (CHUNG and FABBRI 2003). For the validation, this approach is more suitable as the initiation sites used for model calculation are totally independent of those that are used for validation.

Table 4 contains the validation results (area under the curve [0.5;1]). Cross-validation between independent (sub-)areas is set in boldface. Additionally, the "goodness of fit" validations (model A applied to area A) have a grey background (see section 4.2).

Generally, the results are good (the median is 0.88); high AUC values, however, are not evenly distributed over the table. There is a conspicuous cluster of very low to low values in the "RT" column (i.e. different models do not perform well in the Reintal study area). On the contrary, the RT model applies comparatively well to other study areas ("RT" row). The reason for this fact is discussed in section 5.

In order to find out whether the applicability of a model to another study area (i.e. the transferability)

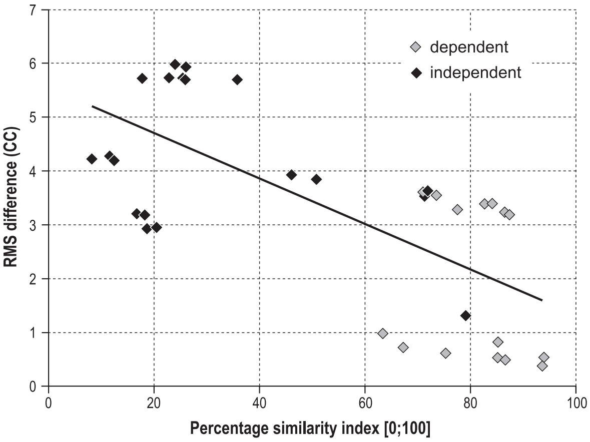

Figure 4: RMS difference between the CC parameters of ever two models plotted against the similarity of the two respective study areas.

Table 4: Cross-validation (area under prediction/success curve [0.5; 1]) for each pair of (sub-areas). Values near 1 indicate successful predictions. Row headings denote the area where a model has been calculated, the area to which a model has been applied is indicated in the column headings. The success rates for the validation of single models (see section 4.2) are shaded. Boldface indicates cross-validation of two areas completely independent of each other.

	LWG	RT	GS_MB	GS_WBA	GS_PFO	GS_MBWBA	GS_MBPFO	GS_WBAPFO	GSIES
GS_MB	**0.92**	**0.73**	0.96	**0.86**	**0.82**	0.91	0.91	**0.83**	0.88
GS_WBA	**0.91**	**0.63**	**0.93**	0.90	**0.87**	0.89	**0.89**	0.87	0.88
GS_PFO	**0.89**	**0.54**	**0.91**	**0.87**	0.91	**0.87**	0.89	0.87	0.88
GS_MBWBA	**0.91**	**0.61**	0.95	0.88	**0.83**	0.91	0.91	0.84	0.89
GS_MBPFO	**0.92**	**0.55**	0.95	**0.87**	0.85	0.91	0.91	0.85	0.89
GS_WBAPFO	**0.90**	**0.63**	**0.93**	0.90	0.89	0.89	0.89	0.88	0.88
RT	**0.85**	0.93	**0.88**	**0.76**	**0.69**	**0.80**	**0.80**	**0.71**	**0.76**
LWG	0.94	**0.60**	**0.93**	**0.86**	**0.82**	**0.88**	**0.88**	**0.82**	**0.86**
GSIES	**0.94**	**0.61**	0.95	0.88	0.85	0.91	0.91	0.86	0.89

depends on the similarity of the two study areas, the validation results (area under the curve) are plotted against the percentage similarity index (Fig. 5). The point signatures show the (sub-)area that different models have been applied to. There is no substantial dependence of transferability on the similarity. There seems to be only a slight tendency of better transferability when the study areas are similar, but it is also clear that the most unsuccessful predictions (application to RT, cross signature) are not associated with the smallest similarity index values.

5 Discussion

Most aspects of the disposition modelling results suggest both validity and transferability of the chosen approach. In this section, the results are discussed also with respect to problems and their implications.

The geofactors identified as important by the interpretation of the CF tables generated by the debris flow models (section 4.1) are largely similar to those reported in other studies. Zimmermann et al. (1997), for example, give a minimum slope angle of 25° for the initiation of slope-type debris flows, Rieger (1999) reports slopes of 20.9–55.9° on initiation sites (mean: 35.1°, standard deviation 5.1°; the minimum slope angle is greater with smaller catchment sizes). In table 3, slopes between 35 and 55° are indicative of debris flow initiation in most study areas.

The association of high CC values with certain geofactor categories or classes is roughly consistent with the theory of debris flow genesis. Debris flow disposition, for example, tends to increase with increasing slope up to a certain maximum, then decreases again. Nevertheless, the problem of inconsistent CC values for the classes of a geofactor (c.f. section 4.1) has to be addressed. This problem can be caused by

empty or infrequent classes containing one or few initiation sites (this leads to an "artificially" high conditional probability and thus to higher CF/CC values), or

- an uneven distribution of initiation sites over the range of geofactor classes (classes with few or zero initiation sites where the latter have either not been mapped or have not occurred for some reason not discernible from the geodata contained in the analysis).

This problem may be aggravated by positional errors during the mapping process, especially regarding the fact that an initiation site is represented by a single raster cell (or at most a few cells).

It is clear that the way continuous data layers are categorised (e.g. choice of bin number and bin width) is very important for the interpretation of the distribution of initiation sites over the range of these geofactor values. This influence could be reduced by using methods of kernel density estimation (see e.g. Cox 2007; Chung 2006). Further discussion of these problems is beyond the scope of this study; it will be addressed in future work.

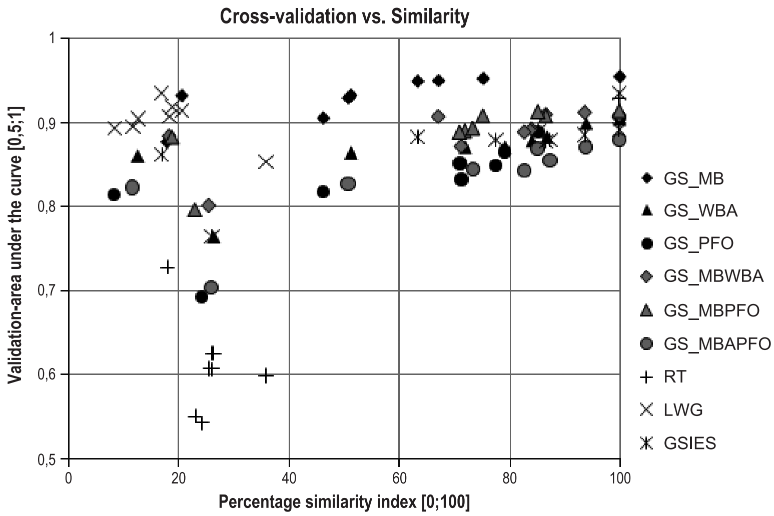

Figure 5: Cross validation (area under prediction curve) plotted against similarity of every pair of (sub-)areas. The point signature indicates the (sub-)area another model has been applied to (columns in Tab. 4)

According to the results of the validation (section 4.2), a CF model based on the geofactors land cover, slope and CIT index is able to identify (or predict) debris flows initiation sites. The fact that only initiation sites are identified or predicted by the raster-based models (see inset of Fig. 2) could be viewed as an advantage over model approaches using different spatial units (e.g. slope units or small subbasins). CARRARA et al. (1992), however, argue that the latter also have advantages over the pixel-based approaches, e.g. that landscape units are geomorphologically more meaningful than singe raster cells analysed without their larger spatial context. In addition, positional errors of mapped events (that are proven to play a major role in model uncertainty) should be better compensated using spatial aggregates. Conceptionally, the spatial unit approaches aim at resolving the "fingerprint" of landslide/debris flow terrain (this is why they mostly include both initiation and runout areas) within the elements of a landscape, while the pixel-based approach used here is focused on delineating the spatial pattern of geofactors responsible for process initiation only. Starting from initiation sites, debris flow pathways and runout can be modelled (see e.g. BECHT et al. 2005; WICHMANN 2006). We suggest that the pixel-based approach might be superior to the aforementioned alternative, but only where geodata of high quality and high resolution are available. In addition, we feel that it makes more sense to calculate pixel-based models with the initiation areas only instead of analysing the whole process area. There are, however, case studies in which the former were not substantially superior to the latter (MAGLIULO et al. 2008); further research should clarify this issue.

In section 4.3, it is observed that CF or CC values calculated by models in different study areas are increasingly similar with increasing study area similarity. This means that in similar (sub-)areas, the ratio of prior probability (i.e. the "debris flow density" in the area) and conditional probability (i.e. the "debris flow density" on areas of a specific geofactor category) is more or less equal (c.f. section 3.1.1). In the extreme case of two study areas with identical geofactor configuration, the conditions should lead to the same density of (i.e. the same prior probability for) initiation sites. A constant ratio of conditional and prior probabilities (i.e. constant CF values) is then only preserved if the conditional probabilities for each geofactor are equal for the two areas. Both statements (concerning prior and conditional probability) can only be true if the geofactors influence the initiation of debris flows in the same way in each study area. The empirical results (section 4.3) confirm this assumption regarding the importance of the chosen geofactors and their influence on debris flow initiation. This constitutes an important conceptional precondition for transferability.

The transferability of the models has empirically been tested by applying each model to every other study area and by evaluating its predictive power. As this cross-validation has shown (section 4.4), the applicability of a model to another study area is largely independent of the degree of similarity between the two areas. Only in the area RT, the predictive power of models calculated in other study areas proved to be significantly lower (with the RT model being successfully validated both in the RT area itself and in other areas). Looking at the properties of both the

Reintal basin and of the geodata, this fact can be explained. The „unvegetated" category of the land cover datasets does not distinguish talus and bedrock. In the glacial trough of the Reintal valley, many slopes are made up of very steep rock walls where debris flows do not occur. In the other study areas, the vast majority of unvegetated, steep slopes are covered with talus, other types of debris or soils. The poor transferability of different models to the RT area is therefore neither a matter of similarity nor does it challenge the general transferability of statistical models. It is in fact the result of a conceptual problem in disposition models: "to what degree do available geofactor maps represent the underlying physical properties and processes relevant for debris flow initiation?" To overcome the particular problem described here, land cover maps should make the distinction between bare rock and loose sediments in order to improve both the validity of the model itself and its transferability to other study areas. However, even if this distinction was made, the resolution of the raster maps would continue to be a limiting factor in this case, as debris flows may also initiate in small debris-filled channels within rock slopes that are not resolved by the mapping process or the raster size.

6 Conclusions and outlook

The results have shown empirically that statistical disposition models are basically transferable to both similar and different study areas. It is clear, however, that expert knowledge is still required (see also Thiery et al. 2007); for example it has to be kept in mind (a) that data quality that can be very different for different study areas, (b) that various assumptions have to be considered when applying statistical approaches (Guzzetti et al. 2005), (c) that a good and comprehensive inventory of initiation sites is required and (d) that the geofactors used must bear reference to the physical processes governing process activity. The similarity of model parameters in similar study areas (section 4.3) can be seen as evidence for the validity of this requirement in the present case study. Generally, susceptibility models have been shown to be valid on both regional and larger scales (Thiery et al. 2007).

Future work will check the suitability of models combined from many study areas, as a larger sample size is expected to improve the accuracy of the prediction models (c.f. Hjort and Marmion 2008). Furthermore, more study areas and different statistical approaches will be tested for transferability using the approach described here, i.e. the parameters of regression or discriminant functions, for example, will be compared for different areas.

Acknowledgements

The present paper presents a re-analysis of previously published and unpublished data that would not have been possible without the cooperation of several colleagues. The authors would like to thank Volker Wichmann who mapped debris flows in the LWG and RT study areas for the SEDAG project (SEDiment cascades in Alpine Geosystems, c.f. Becht et al. 2005), and Christian Hertel, a former diploma student who did the fieldwork in South Tyrolia (Gsieser Tal). The SEDAG studies were funded by a DFG grant to Michael Becht, which is gratefully acknowledged. The comments of an anonymous referee were very helpful in restructuring and clarifying the manuscript.

References

Agterberg, F. P. and Cheng, Q. (2002): Conditional independence test for the weights-of-evidence modeling. In: Natural Resources Research 11 (4), 249–255. Doi: 10.1023/A:1021193827501

Ardizzone, F.; Cardinali, M.; Carrara, A.; Guzzetti, F. and Reichenbach, P. (2002): Impact of mapping errors on the reliability of landlide hazard maps. In: Natural Hazards and Earth System Sciences 2, 3–14.

Becht M.; Haas, F.; Heckmann, T. and Wichmann, V. (2005): Investigating sediment cascades using field measurements and spatial modelling. In: IAHS Publication 291, 206–213.

Binaghi, E.; Luzi, L.; Madella, P.; Pergalani, F. and Rampini, A. (1998): Slope instability zonation: a comparison between certainty factor and fuzzy Dempster-Shafer approaches. In: Natural Hazards 17, 77–97. Doi: 10.1023/A:1008001724538

Carrara, A.; Cardinali, M. and Guzzetti, F. (1992): Uncertainty in assessing landslide hazard and risk. In: ITC journal 1992 (2), 172–183.

Chen, Y.-L. (2003): Indicator pattern combination for mineral resource potential mapping with the general C-F model. In: Mathematical Geology 35 (3), 301–321. Doi: 10.1023/A:1023870231452

Chung, C.-J. F. (2006): Using likelihood ratio functions for modeling the conditional probability of occurrence of future landslides for risk assessment. In: Computers and Geosciences 32, 1052–1068. Doi: 10.1016/j.cageo.2006.02.003

Chung, C.-J. F. and Fabbri, A. (1993): The representation of geosciences information for data integration. In: Nonrenewable Resources 2, 122–139.

– (2003): Validation of spatial prediction models for landslide hazard mapping. In: Natural Hazards 30, 451–472. Doi: 10.1023/B:NHAZ.0000007172.62651.2b

Cox, N. J. (2007): Kernel estimation as a basic tool for geomorphological data analysis. In: Earth surface processes and landforms 32, 1902–1912. Doi: 10.1002/esp.1518

Glade, T. (2005): Linking debris flow hazard assessments with geomorphology. In: Geomorphology 66, 189–213. Doi: 10.1016/j.geomorph.2004.09.023

Guzzetti, F.; Carrara, A.; Cardinali, M. and Reichenbach, P. (1999): Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. In: Geomorphology 31, 181–216. Doi: 10.1016/S0169-555X(99)00078-1

Guzzetti, F.; Reichenbach, P.; Cardinali, P.; Galli, M. and Ardizzone, F. (2005): Probabilistic landslide hazard assessment at the basin scale. In: Geomorphology 72, 272–299. Doi: 10.1016/j.geomorph.2005.06.002

Haas, F.; Heckmann, T.; Wichmann, V. and Becht, M. (2004): Change of fluvial sediment transport rates after a high magnitude debris flow event in a drainage basin in the Northern Limestone Alps, Germany. In: IAHS Publication 288, 37–43.

Hagg W. and Becht, M. (2000): Einflüsse von Niederschlag und Substrat auf die Auslösung von Hangmuren in Beispielgebieten der Ostalpen. In: Zeitschrift für Geomorphologie N.F. Suppl. 123, 79–92

Heckmann, T. and Becht, M. (2006): Statistical disposition modelling of mass movements. In: Böhner, J.; McCloy, K. R. and Strobl, J. (eds): SAGA – analysis and modelling applications. Göttinger Geographische Abhandlungen 115. Göttingen, 61–73.

Heckmann, T.; Wichmann, V. and Becht, M. (2005): Sediment transport by avalanches in the Bavarian Alps revisited – a perspective on modeling. In: Zeitschrift für Geomorphologie N. F. Suppl. 138, 11–25.

Hertel, C. (2007): Modellierung von Hangmuren im Gsieser Tal/Südtirol mit einem GIS. Unpubl. Diploma Thesis. Eichstätt.

Hjort, J. and Marmion, M. (2008): Effects of sample size on the accuracy of geomorphological models. In: Geomorphology 102, 341–350. Doi: 10.1016/j.geomorph.2008.04.006

Keller, D.; Heckmann, T. and Moser, M. (2005): Felduntersuchungen und GIS-Analysen zu flachgründigen alpinen Schuttströmen bei Garmisch-Partenkirchen. In: Moser, M. (ed.) Veröff. von der 15. Tagung für Ingenieurgeologie, 6.-9. April 2005. Erlangen, 189–194.

Luoto, M. and Hjort, J. (2005): Evaluation of current statistical approaches for predictive geomorphological mapping. In: Geomorphology 67, 299–315. Doi: 10.1016/j.geomorph.2004.10.006

Lydy, M. J.; Crawford, C. G. and Frey, J. W. (2000): A comparison of selected diversity, similarity, and biotic indices for detecting changes in benthic-invertebrate community structure and stream quality. In: Archives of environmental Contamination and Toxicology 39, 469–479. Doi: 10.1007/s002440010129 PMid:11031307

Magliulo, P.; di Lisio, A.; Russo, F. and Zelano, A. (2008): Geomorphology and landslide susceptibility assessment using GIS and bivariate statistics: a case study in southern Italy. In: Natural Hazards 47, 411–435. Doi: 10.1007/s11069-008-9230-x

Montgomery, D. R. and Foufoula-Georgiou, E. (1993): Channel network source representation using digital elevation models. In: Water Resources Research 29, 3925–3934. Doi: 10.1029/93WR02463

Pistocchi, A.; Luzi, L. and Napolitano, P. (2002): The use of predictive modeling techniques for optimal exploitation of spatial databases: a case study in landslide hazard mapping with expert system-like methods. In: Environmental Geology 41, 765–775. Doi: 10.1007/s002540100440

Remondo, J.; Gonzales, A.; de Teran, J.; Cendrero, A.; Fabbri, A. and Chung, C.-J. F. (2003): Validation of landslide susceptibility maps: examples and applications from a case study in Northern Spain. In: Natural Hazards 30, 437–449. Doi: 10.1023/B:NHAZ.0000007201.80743.fc

Rieger, D. (1999): Bewertung der naturräumlichen Rahmenbedingungen für die Entstehung von Hangmuren. Möglichkeiten zur Modellierung des Murpotentials. Münchener Geographische Abhandlungen A51. München.

Seijmonsbergen, A. C. and de Graaf, L. W. S. (2006): Geomorphological mapping and geophysical profiling for the evaluation of natural hazards in an alpine catchment. In: Natural Hazards and Earth System Sciences 6, 185–193.

Slaymaker, O. (1988): The distinctive attributes of debris torrents. In: Hydrological Sciences Journal 33, 567–573.

Thiery, Y.; Malet, J.-P.; Sterlacchini, S.; Puissant, A. and Maquaire, O. (2007): Landslide susceptibility assessment by bivariate methods at large scales: application to a complex mountainous environment. In: Geomorphology 92, 38–59. Doi: 10.1016/j.geomorph.2007.02.020

van Asselen, S. and Seijmonsbergen, A. C. (2006): Expert-driven semi-automated geomorphological mapping for a mountainous area using a laser DTM. In: Geomorphology 78, 309–320. Doi: 10.1016/j.geomorph.2006.01.037

Wichmann, V. (2006): Modellierung geomorphologischer Prozesse in einem alpinen Einzugsgebiet. Abgrenzung und Klassifizierung der Wirkungsräume von Sturzprozessen und Muren mit einem GIS. Eichstätter Geographische Arbeiten 15.

Wichmann, V.; Heckmann, T.; Haas, F. and Becht, M. (in press): A new modelling approach to delineate the spatial extent of alpine sediment cascades. In: Geomorphology.

Zevenbergen, L. and Thorne, C. (1987): Quantitative analysis of land surface topography. In: Earth Surface Processes and Landforms 12, 47–56. Doi: 10.1002/esp.3290120107

Zimmermann, M.; Mani, P.; Gamma, P.; Gsteiger, P.; Heiniger, O. and Hunziker, G. (1997): Murgefahr und Klimaänderung – in GIS-basierter Ansatz. Zürich.

Authors

Dr. Tobias Heckmann
Catholic University Eichstätt-Ingolstadt
Physical Geography
Ostenstraße 18
85072 Eichstätt
Tobias.Heckmann@ku-eichstaett.de

Prof. Dr. Michael Becht
Catholic University Eichstätt-Ingolstadt
Physical Geography
Ostenstraße 18
85072 Eichstätt
michael.becht@ku-eichstaett.de