Mountain geocoeosystems. GIS modelling of rockfall and protection
Dorren, L.K.A.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Abstract
This chapter investigates whether a GIS-based distributed model developed for rockfall assessment at slope scale, which uses input data with a support of 2.5 m × 2.5 m, could be used for rockfall assessment at the regional scale, using input data with a support of 25 m × 25 m and with poorer quality. It was anticipated that in the latter case the model error would increase. The hypothesis was that poorer data quality is the main cause of a larger model error rather than the fact that the model simulates a similar process at a larger support. Three simulation schemes were applied to the same model and the outcomes were validated with field data. The first simulation scheme used input data with 2.5 m × 2.5 m support and aggregated the output to 25 m × 25 m support. The second simulation scheme used the same input data as in the first simulation scheme, but these data were aggregated to a support of 25 m × 25 m before running the model. The third simulation scheme used input data with poorer quality directly obtained at a support of 25 m × 25 m. The results show that simulating the maximum rockfall runout zones with a distributed model using input data with 25 m × 25 m support is realistic and feasible. This also accounts for data with poorer quality as the third simulation scheme produced only a slightly larger mean squared error than the first scheme, but it surprisingly produced a lower error than the second scheme. Therefore the cause of the large error produced by scheme 2 is investigated. It is concluded that the combination of a high quality digital elevation model and the loss of important spatial structure in the tree distribution map due to aggregation mainly caused this.

Geomorphological research focuses on understanding processes, patterns and landforms in geocosystems. Therefore, this type of research contributes to a sound basis needed for identifying, assessing and perhaps solving or preventing environmental problems such as soil erosion, flooding and mass wasting. Increasingly popular tools within geomorphological research are distributed models (Willgoose et al., 1991; Montgomery and Dietrich, 1994; Tucker et al., 2001), since these models provide excellent frameworks for conceptualisation of developed theories and for simulating current or future processes and patterns in geocosystems. Nowadays, increased computing capacity has strongly increased opportunities to apply such models.

Models are simplified representations of reality and therefore model outcomes always deviate from ‘the truth’ to some extent. In other words, model outcomes contain errors or uncertainties. These are firstly caused by the introduced simplifications of the real world in the model itself. Secondly, model input data are hardly ever exactly known and therefore likely contain errors (Burrough and McDonnell, 1998). An important factor that determines the degree of model simplification is the availability of input data, which is often determined by the feasibility of measuring the model inputs with sufficient detail. Clearly, it is more difficult to obtain detailed input data for a large catchment than for a small monitoring plot. Therefore, models and their input and output data are mostly less detailed when moving up from a smaller to a larger spatial scale (Heuvelink, 1998). Uncertainties or errors of model outcomes are thus related to spatial scale, which is mainly prescribed by the support of the model input data. Here, support is defined as the largest area treated as homogenous in a sense that an average value of the property of interest within the area is known and not the variation within (Bierkens et al., 2000). Errors of model outcomes do not necessarily increase at larger spatial scales. For example, the averaging-out effect caused by aggregating data to large supports could decrease uncertainties about the average value in a large raster cell compared to a small cell (van Rompaey et al., 1999). On the other hand, if the model input data is obtained at a large support, errors of model outcomes might increase due to the loss of terrain variability. The input data for a model should be of sufficient detail to capture the spatial variations that are essential for the process or pattern being modelled (Goodchild, 2001). In most cases a simpler or at least another model structure is required for larger areas, since processes controlling changes in spatial patterns are different at different spatial scales.
(Bergkamp, 1995, 1996; Kirkby et al., 1996). Only in some cases the same model may be used for both a small and a large spatial scale.

This chapter focuses on a GIS-based distributed model developed for rockfall assessment at a slope scale, which was designed to use input data with a support of 2.5 m × 2.5 m. The main objective was to analyse whether such a model may also be used for an accurate assessment at a regional scale, using input data obtained at a support of 25 m × 25 m with poorer quality due to lesser detail and reduced sampling or mapping effort. It was anticipated that the model that uses data with a support of 25 m × 25 m produces a larger error than the model that uses data with a support of 2.5 m × 2.5 m. The hypothesis was that poorer data quality is the main cause of a larger model error rather than the fact that the model simulates a similar process at a larger support. To investigate this, the distributed rockfall model using three different modelling schemes was applied. The main question of this study was whether rockfall simulation is realistic and/or feasible when using input data with 25 m × 25 m support.

### 7.2 Method and model overview

Three simulation schemes were defined to analyse the effect of input data with different aggregation levels on the accuracy of the used model. These schemes are shown in Figure 7.1.

![Figure 7.1](image)

Figure 7.1. Three simulation schemes tested in this study. For explanation see text.

In simulation scheme 1, input data with a support of 2.5 m × 2.5 m was used. The output was aggregated to a support of 25 m × 25 m by averaging the values of the aggregated output...
obtained at 2.5 m × 2.5 m support. Finally, the output data was compared with validation data with a same support of 25 m × 25 m. In simulation scheme 2 the same input data as in simulation scheme 1 was used, however, these data were aggregated to a support of 25 m × 25 m before running the model. In simulation scheme 3 the support of the used input data was identical to scheme 2, but the input data for simulation scheme 3 was obtained at a support of 25 m × 25 m from other data sources than the input data from simulation scheme 1 and 2 (details will be explained in section 7.3). Since the data used in simulation scheme 3 were obtained at the regional scale and thus less detailed than the data used in simulation scheme 1 and 2, it may be stated that the quality of this data was considerably poorer.

Within the three simulation schemes the same rockfall simulation model was used, which simulates a falling rock by calculating the kinetic energy balance during sequences of motion though the air and collisions on the slope surface or against trees. Start locations for the rockfall simulations were derived from field mapped rockfall source areas. From each start location one single falling rock was simulated. The simulation was repeated one hundred times according to the Monte Carlo method (Lewis and Orav, 1989). This was done for all the applied simulation schemes.

Standard algorithms for a uniform accelerated parabolic movement through the air calculated the motion through the air. Algorithms modified from Pfeiffer and Bowen (1989) calculated the energy balance before and after collisions with the slope surface and tree stems. The modifications were such that the factor compensating for the effect of the rockfall velocity on the elasticity of the collision was left out, since the empirical constants required for calculating this factor were not available for the study site. The algorithms for bouncing and motion through the air were combined with a procedure that calculated the fall direction on the basis of a digital elevation model (DEM) as described in chapter 5. A flow diagram of the model is given in Figure 5.8.

The model inputs that are affected by a change of support, which were also the model inputs focused at in this study, are the DEM, the rasters containing values for the tangential and normal coefficient of restitution and the tree distribution raster, which includes both the number of trees per raster cell and the range of tree stem diameters. The DEM determined the mean slope gradient and therefore the acceleration and deceleration of a falling block. Furthermore the DEM determined the fall direction. The coefficients of restitution determined the amount of energy lost during a bounce, where the tangential coefficient of restitution determines energy loss parallel to the slope surface (due to surface roughness or vegetation).
The normal coefficient of restitution determines energy loss perpendicular to the slope surface (due to elasticity of the material covering the slope surface). The tree distribution determined the probability of a falling rock hitting a tree.

The outcomes produced by the three simulation schemes were all compared with a validation dataset on the basis of the mean error (ME) and the mean squared error (MSE) following equations 5.9 and 5.10.

7.3 Test site and available data

A forested rockfall slope in the most western part of the Austrian Alps, located at 47°00' latitude and 10°01' longitude, was chosen as a test site for this study. The test site may be divided in two areas. Firstly, the rockfall source area, which is a steep cliff face dissected by large denudation niches and secondly, an accumulation area, which is a large post-glacial developed talus cone mainly consisting of rockfall scree, but also some debris flow material. The slope length of the talus cone is approximately 900 meters. An overview of the site is shown in Figure 7.2.

Two DEMs were available for this test site, both arranged as rasters with a support of respectively 25 m × 25 m (LRDEM) and 2.5 m × 2.5 m (HRDEM). The LRDEM was created by interpolation of photogrammetric height measurements at a ground distance of 50 m, enhanced and supplemented with prominent terrain structures. The given maximum error in this DEM was 20 m (BEV, 2002).

The HRDEM was derived from a TIN (Triangular Irregular Network), which was created from contour lines with an equidistance of 5 m. The contour lines were derived from a combined dataset consisting of slope transects measured in the field, a detailed geomorphological field map (1:2000) and existing contour lines with an equidistance of 20 m. The maximum error of the HRDEM was 6 m. The main difference between the HRDEM and the LRDEM was that important terrain structures, which determine the variation in the slope gradient of the test site, were well represented in the HRDEM, whereas in the LRDEM such structures were not present at all (see Fig. 7.3). Since the study area was fully covered with forest, a tachymeter or photogrammetric height measurements on the basis of stereographic aerial photographs for creating the HRDEM were not used.
Both from the HRDEM and the LRDEM mean slope gradients maps were derived using the method described by Zevenbergen and Thorne (1987). In addition, fall directions were calculated using a modified multiple-flow algorithm (e.g. Quine et al., 1991; Wolock and McGabe, 1995; Tarboton, 1997). To randomise the fall direction from a central cell for each simulation run, this method calculates a fall direction raster repeatedly for each simulation run by sampling the fall direction for each cell randomly from a probability distribution, which was determined by the steepness of the mean slope gradients between the central cell and all its downslope neighbouring cells. In this procedure, the probability of a rock falling to a downslope cell increases with the gradient between the two cells.

Furthermore, the tangential coefficient of restitution ($r_t$), the normal coefficient of restitution ($r_n$) and the tree distribution were both represented as a raster with 25 m × 25 m support (hereafter respectively LR$r_t$, LR$r_n$ and LR$\text{tree}$) and as a raster with 2.5 m × 2.5 m support (hereafter HR$r_t$, HR$r_n$ and HR$\text{tree}$). LR$r_t$ and LR$r_n$ were created by combining literature data and a land cover map, which was obtained by classifying a September 1998 Landsat TM image (see section 6.3). LR$\text{tree}$ was also derived from this land cover map in combination with data provided by a regional forest inventory (Maier, 1993). HR$r_t$ and HR$r_n$ were based on a detailed slope surface cover map (scale 1:2000) from which $r_t$ and $r_n$ could be
estimated on the basis of literature data (Pfeiffer and Bowen, 1989; Van Dijke and van Westen, 1990; Kobayashi et al., 1990; Giani, 1992; Azzoni et al., 1995; Chau et al., 1998; Meißl, 1998). HRtree was created using a combination of a tree crown map and forest inventory data. The tree crown map was derived from an object-based classification of high-resolution digital colour-infrared (CIR) orthophotos (0.25 m × 0.25 m) on which each individual tree crown was visible. The forest inventory, which was done with the Winkelzählpisode (Bitterlich, 1948), provided amongst others data on the distribution of tree stem diameters throughout the study site. This inventory method is also known as prism plot sampling or probability proportional to size sampling. Within this method the probability of a tree being selected within a sampling plot is proportional to the basal area of a tree (Shiver and Borders, 1995). For this detailed inventory, a grid covering the study area was defined in which tree stem diameters and the number of trees as well as the number of damages per tree stem were measured. On the basis of these the tree volume per hectare, the number of trees per hectare as well as the number of damages per hectare were calculated for each inventory grid cell.

The HRDEM, HRr, HRn and HRtree provided input data for simulation scheme 1. As described in section 7.2, aggregation of these data provided input data for simulation scheme 2 and LRDEM, LRR, LRRn and LRtree provided input data for simulation scheme 3. The data input for the three simulation schemes is visualised in Figure 7.3. In simulation scheme 2 the input data suffer from a considerable loss of detail compared to the input data used in simulation scheme 1. As expected this is even worse in simulation scheme 3.

Validation data were extracted from the detailed forest inventory data. In the upper part of the accumulation area 18 squares of 25 m × 25 m were randomly selected. The available forest inventory data in the lower areas of the accumulation area are discarded because in these areas rockfall activity originating from source areas outside the area used for this modelling study also occurred. The relatively small number of validation squares (n=18) only provides a small data set. Therefore results have to be interpreted carefully on the one hand, but on the other hand the data for each individual validation square are derived from a large number of detailed measurements on trees. For these 18 squares the number of scars per tree volume unit were compared with the number of impacts as simulated by the model (the calculation of the validation data is described in detail in section 5.5.5).
Figure 7.3. Mean slope gradient, tangential coefficient of restitution and number of trees per cell used in the simulation schemes 1, 2 and 3.
7.4 Results

Figure 7.4 shows the raster maps representing the relative amount of impacts within each raster as produced by the three simulation schemes. The three raster maps show a similar pattern in the number of impacts per cell. Most impacts occur in the upper right part of the raster in the transition zone between the source area and the accumulation area. Generally the number of impacts decreases towards the lower left, which is the accumulation area or rockfall runout zone. A notable difference between the three raster maps in Figure 7.4 is the location with the maximum number of impacts. For simulation scheme 1 this is located in the lower parts of the rockfall source areas, but for simulation schemes 2 and 3 this is located on the upper part of the accumulation area, although for simulation scheme 3 the location of the maximum number of impacts is more evenly distributed over the accumulation area.

Comparison of the simulated impacts with the number of observed scars per tree volume unit provided the results shown in Figure 7.5. The histograms in this figure show that simulation scheme 2 produced the largest errors and simulation scheme 1 produced the smallest errors. In general, all errors are quite high. The scatter plots in Figure 7.5 show the difference between each predicted and observed value for the three simulation schemes in detail. The plots show that the degree of scatter between observed and predicted values is generally smaller for simulation scheme 1 than for schemes 2 and 3. This is especially caused by a better estimation for the cells with a high number of scars per tree volume unit. Nevertheless, simulation
scheme 1 also produced some considerable mismatches for cells with low observed values. The latter considerably affects the mean squared error produced by simulation scheme 1 (MSE1) as shown in Table 7.1 (MSE1 = 24.9).

Table 7.1. Mean error (ME), mean squared error (MSE) and correlation coefficient (r) of the output of the three simulation schemes.

<table>
<thead>
<tr>
<th>Simulation scheme</th>
<th>ME</th>
<th>MSE</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>24.9</td>
<td>0.47</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>44.4</td>
<td>0.08</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>31.3</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Simulation scheme 2 produced the largest error (MSE3 = 44.4) and simulation scheme 3 produced an intermediate error (MSE2 = 31.3). To some extent this is also indicated by the correlation coefficient (Table 7.1). Here it should be noted that the correlation coefficient represents the degree of any existing correlation between the observed and predicted data. This coefficient is not necessarily indicative for the strength of the 1:1 relationship between observed and predicted values. The correlation coefficient is given here because it might indicate a systematic error in the model in case both the correlation coefficient and the MSE are high.

Figure 7.5. Histograms of the errors and scatterplots with observed versus predicted number of scars per tree volume unit, as produced by simulation schemes 1, 2 and 3.
7.5 Additional tests and discussion

7.5.1 ‘Intermediate’ simulation schemes

It was not expected that simulation scheme 3 would produce a smaller MSE than scheme 2. Rather, it was anticipated that scheme 3 would perform the worst, because it uses input data with the poorest quality. It is therefore appropriate to investigate why simulation scheme 2 produced larger errors, although the input data quality was higher. The differences between simulation schemes 2 and 3 are the values for four model inputs (tree density, normal coefficient of restitution, tangential coefficient of restitution and the DEM), which all have been changed simultaneously. Therefore, ‘intermediate’ simulation schemes were analysed to assess which input data, related to one of the four parameters, caused the lower MSE as produced by simulation scheme 3. Table 7.2 shows the data used in the ‘intermediate’ simulation schemes and it shows the produced mean squared errors.

Table 7.2. The MSE of model output produced with the original schemes 2 and 3 and with the ‘intermediate’ simulation schemes 4 to 7.

<table>
<thead>
<tr>
<th>Simulation scheme</th>
<th>Used data</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>HRtree*, HRr*, HRr*, HRDEM*</td>
<td>44.4</td>
</tr>
<tr>
<td>3</td>
<td>LRtree, LRr_n, LRr_t, LRDEM</td>
<td>31.3</td>
</tr>
<tr>
<td>4</td>
<td>LRtree, LRr_n, LRr_t, HRDEM*</td>
<td>47.8</td>
</tr>
<tr>
<td>5</td>
<td>LRtree, LRr_n, HRr*, LRDEM</td>
<td>35.7</td>
</tr>
<tr>
<td>6</td>
<td>LRtree, HRr_n*, LRr_n, LRDEM</td>
<td>31.4</td>
</tr>
<tr>
<td>7</td>
<td>HRtree*, LRr_n, LRr_t, LRDEM</td>
<td>29.3</td>
</tr>
</tbody>
</table>

* Data aggregated to 25 m x 25 m support

The results in Table 7.2 show that replacing LRtree by the aggregated HRtree (scheme 7) resulted in a MSE of 29.3, which is lower than the initially produced MSE by scheme 3 (see also Fig. 7.6a). Substitution of LRr_t by the aggregated HRr_t (scheme 5) increased the MSE from 31.3 to 35.7. Table 7.2 furthermore shows that substituting LRr_n by the aggregated HRr_n (scheme 6) resulted in a negligible change in the MSE, which indicates that the net effect of the r_n on the simulation results is small.

A remarkable result is that the substitution of the LRDEM by the aggregated HRDEM (scheme 4) did not decrease the MSE. On the contrary, this substitution produced a strong increase of the MSE from 31.3 to 47.8. The scatter plot of the result of simulation scheme 4 is shown in Figure 7.6b. This plot shows that this simulation scheme strongly overestimated the lower observed values and strongly underestimated the higher observed values.
7.5.2 Causes of simulation errors

The 'intermediate' simulation scheme 4 indicated that the aggregated HRDEM was mainly responsible for the increase in error (Table 7.2). An explanation for this is that the transport channel, which is shown in Figure 7.2 and in the HRDEM and HRtree in Figure 7.3, is averaged-out to a certain extent in the aggregated HRDEM, but still present. As a result, the fall directions calculated on the basis of the aggregated HRDEM were generally towards this channel. This led to a concentration of falling rocks in the right side of the raster. As a consequence, the number of impacts is higher at that side, which is shown in Figure 7.7 (scheme 4). This effect is reinforced by the fact that the transport channel is almost free of trees. Therefore, hardly any rock impacts against trees occur in the channel in reality. However, when using the aggregated HRtree, the forest structure in the channel as observed in the field and in HRtree, which is directly derived from orthophotos, is completely lost (i.e. compare Fig. 7.2 and Fig. 7.3). As a result, the number of trees in the fall track of the simulated rocks as represented by the aggregated HRtree is greatly overestimated, although the number of trees in the channel is still smaller than in the surrounding areas. Therefore, the number of impacts in the channel is also overestimated when compared to the other simulation schemes, as shown in Figure 7.7 (schemes 1, 2 and 3). In simulation scheme 3, this error occurred to a lesser extent, since in the LRDEM the channel was completely 'smoothed-out'. Accordingly, a more uniform distribution of rock impacts was produced.

The squares with an overestimation of 10.7% and an underestimation of 7.9% in Figure 7.7 (scheme 1) indicate a potential spatial mismatch between model results and field observations. Schemes 2 and 3 underestimated the observed value in this area with approximately 12% (Fig. 7.7). The errors in the validation squares on the lowest part of the
hillslope show that simulation schemes 1, 2 and 3 modelled the maximal extent of rockfall runout zones quite well (see also Fig. 7.4), as the error values in those parts are smaller than 5%. For all simulation schemes no rocks reached the bottom part of the hillslope, which was also observed in reality. As mentioned before, the modelled rockfall impacts on tree stems were not very accurate, as the MSE produced by all schemes was quite high.

7.5.3. Rockfall simulation at different scales
Overall, the results indicate that the GIS-based distributed model used in this study, which was developed for rockfall assessment at a slope scale, could be used for rockfall assessment at a regional scale. As expected, the simulation schemes analysed in this study indicated that the use of input data with 25 m × 25 m support increased the mean squared error (MSE) compared to the use of input data with 2.5 m × 2.5 m support. However, the simulated maximum extents of rockfall runout zones were quite similar for simulation schemes 1, 2 and 3. In addition, these runout zones corresponded with those observed in reality. This shows that rockfall modelling at regional scale is feasible and realistic. The simulated rockfall impacts on tree stems using input data with 25 m × 25 m support were not accurate, however, as the MSE produced by simulation schemes 2 and 3 were significantly higher than the MSE produced by scheme 1. Using tree distribution data with higher quality could reduce the MSE produced by scheme 3 with approximately 18% as shown by the ‘intermediate’ simulation scheme 7. The accuracy of the results produced by scheme 2 and 3 indicates that simulating damage on tree stems caused by rockfall using input data with 25 m × 25 m support is not realistic. For such a modelling objective, high quality data with small support is required. More specifically, this requires at least a quality and support similar to the ones used in simulation scheme 1.
Figure 7.7. Visualisation of the validation squares where $|P_i - O_i| > 5$ for simulation schemes 1, 2, 3 and 'intermediate' scheme 4; black filled squares indicate underestimation by the model and white filled squares indicate overestimation by the model. The white outlined squares indicate an absolute error smaller than 5.
As anticipated, the model using data with a support of 25 m × 25 m produces a larger error than the model using data with a support of 2.5 m × 2.5 m. The hypothesis was that poorer data quality is a more important cause of a larger model error rather than the fact that the model simulates a similar process at a larger support. This turned out not to be true, as simulation scheme 2, which used data with higher quality, produced a larger error than simulation scheme 3. Here it was interesting to observe that due to aggregation the loss of important spatial structure in the input data (i.e. the rockfall channel represented in the slope map and in the tree distribution map) resulted in a larger model error than with the use of a more general landscape representation. The results of this study also indicate that the use of a regional DEM with high quality requires data on forest structure with higher quality than the use of a regional DEM with poorer quality, but this is not necessarily true in every case.

7.6 Conclusions

In this chapter the relationship between the aggregation level of the input data and the accuracy of the model output was investigated. The results showed that rockfall simulation with a GIS-based distributed model using input data with 25 m × 25 m support is feasible and realistic for simulation of rockfall runout zones, but not for simulation of tree damage caused by rockfall. The latter was mainly caused by the fact that collisions of rocks against tree stems cannot be simulated accurately in the case data with a relatively poor quality and a large support is being used.

Goodchild (2001) stated that the input data for a model should be of sufficient detail to capture the spatial variations that are essential for the process or pattern being modelled. The results of this study confirm this statement, because it shows that for modelling tree damage caused by rockfall, a detailed tree distribution is essential.