# **GIS Based Spatial Data Analysis for Landslide Susceptibility Mapping**

#### **S. Sarkar**

*Corresponding author: Scientist, Central Building Research Institute, Roorkee – 247667, India e-mail:shantanu\_cbri@yahoo.co.in* 

#### **D.P. Kanungo**

 *Scientist, Central Building Research Institute, Roorkee – 247667, India* 

#### **A.K. Patra and Pushpendra Kumar**

*Research Fellow, Central Building Research Institute, Roorkee – 247667, India*

**Abstract:** Landslide susceptibility map delineates the potential zones for landslides occurrence. The paper presents a statistical approach through spatial data analysis in GIS for landslide susceptibility mapping in parts of Sikkim Himalaya. Six important causative factors for landslide occurrences were selected and corresponding thematic data layers were prepared in GIS. Topographic maps, satellite image, field data and published maps constitute the input data for thematic layer preparation. Numerical weights for different categories of these factors were determined based on a statistical approach and the weighted thematic layers were integrated in GIS environment to generate the landslide susceptibility map of the area. The landslide susceptibility map classifies the area into five different landslide susceptible zones i.e., very high, high, moderate, low and very low. This map was validated using the existing landslide distribution in the area.

**Keywords:** Landslide susceptibility; GIS; Sikkim Himalaya; statistical approach; Himalaya

# **Introduction**

Landslide occurrences are very common

**Received:** 23 August 2007 **Accepted:** 25 October 2007

phenomenon in the Himalayas which sometimes cause loss of life and property. Though the losses due to earthquakes and floods in India are much more than landslides, however, landslide occurrence being more frequent is considered to be a major geological hazard. Recently few landslide disasters in Himalaya have made tremendous impact on the society. As a result of increasing urbanization, hill slopes are being disturbed due to various construction activities particularly the road and building construction. It is therefore necessary to know the landslide prone zones before any construction activity begins so that adequate control measures can be adopted well in time. Landslide susceptibility mapping, which delineates the potential landslide zones, is useful for such purpose.

Several works have been carried out all over the world for landslide hazard zonation and susceptibility mapping. There are different approaches adopted by different workers (Pachauri and Pant 1992, Anbalagan 1992, Juang *et al.* 1992, Jade and Sarkar 1993, Sarkar *et al*. 1995, Chung *et al*. 1995, Mehrotra *et al*. 1996, Gupta and Joshi 1990). The basic difference among these approaches lies in the assignment of numerical weights to the landslide causative factors. Over the past few years, Geographic Information System (GIS) has gained significant importance for spatial

data analysis. It has been proved to be a very powerful tool for landslide study. In recent years, GIS has been employed globally for spatial data analysis for landslide hazard zonation mapping (Westen 1994, Carrara *et al.* 1991, Kingsbury 1992, Nagarajan *et al.* 1998, Dhakal *et al.* 2000, Saha *et al.* 2002, Sarkar and Kanungo 2004, Kanungo *et al.* 2006).

The present paper deals with thematic data layer generation and their spatial analysis in GIS environment for landslide susceptibility mapping in parts of Sikkim Himalaya. Landslide occurrences are quite common in the Sikkim Himalaya, and the magnitude of damages caused every year in various parts of the state is quite large (Bhasin *et al.* 2002).

#### **1 Study Area**

The Sikkim Himalaya rises abruptly from the alluvial plains of North Bengal and attains a

maximum elevation of about 8500 m. The state of Sikkim is bounded by Nepal in the west, Bhutan in the east and China in the north. The present study is focused in the East District of Sikkim, which lies within the latitude  $27^{\circ}10^{\prime}$ N ~  $27^{\circ}23^{\prime}$ N and longitude  $88^{\circ}25'$  E ~  $88^{\circ}45'$  E, and covers an area of about 549 km2 (Figure 1). The maximum elevation in the study area is 3250 m in the extreme north eastern part. The area is bounded by the rivers Rangpochu in the south, Tista in the west and Dikchu in the north. The annual average rainfall at Gangtok, the capital town of Sikkim state, is of the order of 3500 mm (Bhasin *et al*. 2002). The maximum rainfall occurs during the monsoon months from April to September. Topographically the area is traversed by many ridges and valleys and maximum dissection is towards north east and eastern part. The area is dominated by the slopes ranging between 15°~45° while steep slopes of >45° occupy much smaller area. The gentle slopes of <15° are found on the ridges and valleys.



53

## **2 Landslide Distribution**

In this study, IRS LISS II image was used for mapping the landslide distribution. Landslide detection on satellite image largely depends on the contrast that results from the spectral difference between the landslide and its surroundings. Landslides show high reflectance as these areas in general are bare of any vegetation. The circular to elliptical shape of landslides also helps in identifying these on satellite image. A landslide distribution map (Figure 2) was prepared, which shows 144 landslides. Hence the landslide density of the area is 0.262 per km2. A few of these landslides were cross-checked in the field. Majority of the landslides in this area are debris slide (Figure 3). The causative factors as observed in the field are weathered gneisses, steep slope and erosion along the stream channels. Landslides are distributed in the whole area; however, major concentrations are in the central and north-eastern part.



**Figure 2** Landslide distributions in the area

#### **3 Spatial Data Layers**

Landslide occurrences are governed by several factors. However it is not always possible to obtain all the data required for landslide susceptibility mapping. The most common factors are geology, topography, land use, structure, hydrology and climate. In the present study, six factors namely, lithology, fault, drainage, slope, slope aspect and land use were considered for analysis. Though rainfall and earthquakes are the two important triggering factors for landslides, these could not be considered in this study due to non-availability of sufficient data. The factors considered here are essentially the preparatory factors which can be collected from the field and available information. The thematic data layers for each of these factors were prepared in GIS. The data used to prepare these maps were collected from the topographic maps (1:50,000 scale), remote sensing data (IRS LISS II with resolution 36.25 m), published literature and field investigation. A brief description of these layers is given below.



**Figure 3** Landslides in Sikkim: (i) Debris slide on Jawaharlal Nehru road (ii) Debris slide along with subsidence on Rangpo-Rorathang road

# **3.1 DEM and its derivatives**

A digital elevation model (DEM) is the digital representation of a topographic surface with the

elevation above a geodetic datum. The DEM of the study area (Figure 4) was prepared from the contour information given in the topographic maps using TIN module of Arc View 3D Analyst. TIN is a DEM with a network of triangles at randomly located terrain points.

The slope map of the area was derived from the DEM with a 25 m cell size. The map was classified into 5 classes as per the slope classification of earlier workers for such studies (Dhakal *et al*. 2000, Sarkar and Kanungo 2004). The map shows that maximum area was occupied by the slope classes of  $15^{\circ} \sim 25^{\circ}$  and  $25^{\circ} \sim 35^{\circ}$ , followed by  $35^{\circ} \sim 45^{\circ}$ . A slope aspect map, which shows the direction of slope, was also derived from the DEM with same cell size.



**Figure 4** DEM of the study area

#### **3.2 Lithology**

The major geological formations of the area belong to Chungthang Formation, Sikkim Group and Darjeeling Group. The lithology map (Figure 5) was derived from the geological map (Raina and Srivastava 1981) and field investigations. There are six different rock types present in the area. These are gneiss, schist, quartzite schist, quartzite phyllite, quartzites and phyllite. The majority of the area is covered by schists. At higher elevation the schists become coarser and the foliation becomes more prominent. Around north of Gangtok the rocks are more gneissic occurring at higher elevation on the north-eastern part of the area. Quartzites alternating with schists are found in the southwest of Gangtok town. Phyllites are exposed in the south while quartzites alternating phyllites

predominate along the western boundary of the study area. The quartzites and quartzites alternating with phyllites are stronger than the other rocks in the area. However, all the rocks are subjected to tectonic activity and show high weathering along the drainage channels and near the thrust/fault.

#### **3.3 Faults**

A major thrust, known as Chungthang thrust, lies in the north-eastern part of the area (Figure 5). Besides this, a major fault trending N-S is also present. For our analysis a fault buffer map was generated in GIS with buffer width of 1km from these two faults. It has been found in the literature that thrust/faults have an effect on landslide occurrences up to few kilometers. It was assumed that the effect of thrust and major fault on landslide occurrence may extend up to about 3 km. Thus, the fault buffer map with four different buffer zones i.e.,  $0 \sim 1$  km,  $1 \sim 2$  km,  $2 \sim 3$  km and  $> 3$ km was prepared.

# **3.4 Drainage**

The drainage network of the area was mapped in GIS from the topographic maps on 1:50,000 scale. The drainage lines were digitized and the map was prepared in GIS. The drainages were then classified based on stream orders (Figure 6). Since landslides in this area are mostly associated with 1<sup>st</sup> and 2nd order drainages, a drainage buffer map was prepared in GIS considering the 1st and 2nd order drainages only. The buffer widths of 0~50 m, 50~100 m and >100 m were considered.



**Figure 5** Lithology map of the area



**Figure 6** Drainage map showing stream orders

#### **3.5 Land use**

The land use map of the area was prepared from the remote sensing data. Ground truth data were collected from the field and also perceived from the topographical maps. IRS LISS II satellite image was interpreted for various land use in terms of vegetation cover. The land use map, thus prepared, classifies the area into thick forest, moderate forest, sparse forest, agriculture land and barren land (Figure 7). It can be observed from the map that the maximum area is covered by moderate and sparse forest categories followed by agriculture land and thick forest respectively. The barren land which occupies the least area is predominant at higher elevation in the northeastern part.



**Figure 7** Land use map of the area

# **4 Methodology**

To perform spatial data analysis in GIS the factor categories have been assigned numerical weights according to their importance for landslide occurrence. There are basically two different ways to determine the weights for the various categories of causative factors. The one is based on subjective approach where a person assigns weights based on expert opinion or his own experience on the subject and the area (Anbalagan 1992, Sarkar and Kanungo 2004) and the other is the statistical approach where factors are correlated with existing landslides to determine the weights (Carrara 1983, Yin and Yan 1988, Jade and Sarkar 1993, Sarkar and Gupta 2005). In the present study, the weights for the categories were determined using a statistical approach known as information value method. The method is based on the frequency distribution of landslides in different categories of factors. The landslide susceptibility map was prepared by integrating the information values obtained for different categories of the factors. The map was then validated. The outlay of the methodology is shown in the flow diagram (Figure 8).



**Figure 8** Flow diagram showing the methodology

## **5 Information Model**

Information model is a statistical method for spatial prediction of an event based on the parameter and event relationship. It has been very useful method for landslide susceptibility mapping by determining the influence of parameters governing landslide occurrence in an area and was used by several workers (Yin and Yan 1988, Jade and Sarkar 1993).

The information value  $I_i$  for a parameter  $i$  can be expressed as:

$$
I_i = \log \frac{S_i/N_i}{S/N} \tag{1}
$$

in which:

is

*N* = total number of grid cells

*S* = number of grid cells with landslide

 $S_i$  = number of grid cells with the parameter *i* and containing landslide

 $N_i$  = number of grid cells with the parameter *i* 

Now the total information value in a grid cell *j* 

$$
I_j = \sum_{i=1}^{M} X_{ji} I_i
$$
 (2)

where,  $X_{ii}$  is value of parameter *I*;  $j = 1, 2, \ldots, N$  and *i*  $= 1, 2, \ldots, M; X_{ji} = 1$ , if parameter *i* exists in grid cell *j*; *Xji* = 0, if parameter *i* does not exist in grid cell *j; M* = number of parameters considered.

The above model was used to determine the information values for different categories of the factors and the total information value of each cell of the area. The more the total information value, the more is the degree of landslide susceptibility.

# **6 GIS Analysis for Landslide Susceptibility Mapping**

The six thematic data layers and the landslide distribution layer were the input data for spatial analysis in GIS. The cell size selected for the analysis was 25 m×25 m. The total number of cells in the area is 878077. The number of cells occupied by each category of all the factors was calculated in GIS. Then each layer was integrated with the landslide layer to determine the number of cells in each category of factor containing landslides. From the above data using equation (1), information values for all the categories were computed. The categories of the six factors and their information values are given in Table 1.

As observed from the table, moderate to steep slopes have higher information values than the gentle slopes. This has also been reported by the other workers (Sarkar & Gupta 2005, Kanungo et al. 2006) in other parts of Himalayas. The slope aspect has an indirect influence on slope instability. In general, south facing slopes have lesser vegetation density as compared to north facing slopes and hence, the erosional activity is relatively more in former case (Sinha et al. 1975). Based upon the landslide distribution, south and east facing slopes were considered susceptible to landslides by Dhakal et al. 2000. According to Lin and Tung (2003), the southeast aspect displayed higher potential landslide risk than other directions. The present study also shows high information values for south and southeast facing slopes. In case of lithology, the gneissic rocks are highly weathered and pulverized which may be due to the presence of a major thrust in that area. This part of the area also receive snow fall in every winter season and thus affected by water infiltration causing more weathering. The area occupied by the gneisses i.e. the northeastern part is also steep with higher relief. So all this factors probably contributed towards more landslide occurrence in the gneissic rocks and hence show high information value. It is also observed from Table 1 that the fault has a major effect on landslide occurrences up to 1 km and the effect diminishes as we go away from the fault. The maximum information value observed for the 50~100 m category of drainage buffer may be due to the fact that the landslide initiation and activities are observed in the field mostly beyond 50 m distance from drainage. Within 50 m distance from drainage, it is the zone of deposition of landslide debris.

It is inferred that out of 33 categories belonging to six factors the top ten important categories with respect to information values in descending order are the  $25^{\circ}$   $\sim$  35° slope, gneiss rock, barren land use, 60°~90° slope, southeast aspect,  $45^{\circ}$ ~60° slope, southwest aspect, 1~1000 m fault buffer, agriculture land use and  $35^{\circ}$   $\sim$  45° slope. Hence it could be inferred that the slope, lithology and land use seem to be the major controlling factors for landslide incidence in the area.

Assigning these information values as weights to the factor categories, the weighted thematic data layers were generated for further spatial analysis.

The weighted data layers were integrated to obtain the total information value as given in equation (2) for each cell using Spatial Analyst Model of ARC View GIS. The arithmetic overlay method of data integration in raster mode was performed on the weighted data layers. The total information value obtained for each cell represented the landslide potential index (LPI) of the cell. The range of LPI values varied from -1.514 to 0.916 for the 878077 cells of the area. These values need to be classified into different classes to produce landslide susceptibility map.





To classify the LPI values into different classes, the LPI values of all the cells were plotted with their frequency. At the first instance the graph showed many oscillations and was difficult to

classify. Hence the graph was smoothened by moving average method with averaging window lengths 5, 7 and 9. It is to be noted that the window length 5 means the frequency value at any point is an average of the five consecutive values centered at that point. The graph with window length 9 showed the best segmentation for classification. The composite graph with window lengths 5, 7 and 9 is shown in Figure 9. After interpreting the graphs carefully, the class boundaries were drawn at significant changes in the gradients of the curves. Applying these class boundaries to the LPI values, the area was classified into very high susceptible (VHS), high susceptible (HS), moderate susceptible (MSZ), low susceptible (LSZ) and very low susceptible (VLS) zones (Table 2). The landslide susceptibility map is shown in Figure 10. The northeastern part of the map shows a major area of very high and high landslide susceptible zone which has gneissic rocks in the vicinity of Chungthang thrust on barren steep slopes having less vegetation. There are also few scattered zones of high susceptibility class in the southern and central portions of the area.

**Table 2** Landslide Potential Index (LPI) and susceptibility zones

Susceptibility zones	<b>LPI</b>
Very low	$<-0.65$
Low	$-0.65 \sim -0.13$
Moderate	$-0.13 \sim 0.15$
High	$0.15 \sim 0.42$
Very High	> 0.42



**Figure 9** Frequency distribution of landslide potential index

# **7 Map Validation**

For map validation the landslide susceptibility map was physically verified in the field at few locations particularly at the very high and high susceptible zones. It was observed that the areas of very high and high susceptible zones are marked by slope instability signatures such as landslides, slope erosion, road subsidence etc. Further, two different methods were used for a quantified validation of the map.

# **7.1 Map validation based on landslide density**

Landslide density is defined as the ratio of the existing landslide area to the area of each landslide susceptibility zone, and is calculated here on the

basis of the number of pixels. Landslide density values for each of the susceptibility zones (i.e., VHS, HS, MS, LS and VLS) were calculated separately (Table 3). Usually, an ideal landslide susceptible map should have the highest landslide density for VHS zone as compared to other zones and there ought to be a decreasing trend of landslide density values successively from VHS to VLS zone.

From the table it could be observed that landslide density values for very high and high susceptibility zones are 0.015 and 0.01 respectively which are remarkably higher than the other zones. Further, there is a gradual decrease in density values from very high to very low susceptible zone and there is also considerable separation in these values. This reflects the validity of the landslide susceptibility map with existing slope instability conditions.

**Table 3** Landslide density in different landslides susceptibility zones



#### **7.2 Map validation using success rate curve**

Another way for landslide susceptibility map validation is by using success rate curve method. Success rate is defined as percentage of landslide occurrence in any susceptibility zone. The suitability of a map can be judged by the fact that more percentage of landslides must occur in VHS zone as compared to other zones. Therefore, the cumulative percentage of landslide occurrences in various susceptibility zones ordered from very high susceptibility to very low susceptibility were plotted

against the cumulative percentage of area of the susceptibility zones for the landslide susceptibility map (Figure 10). This curve (Figure 11) has been defined as the success rate curve (Chung and Fabbri 1999, Lu and An 1999, Lee et al. 2002b).

It could be observed from the figure 11 that 10 % of the area in higher susceptibility zone contains 23 % of existing landslides and 20 % of the area contains 40 % of existing landslides. This also reflects the validity of the landslide susceptibility map with existing slope instability conditions.



**Figure 10** Landslide susceptibility map of the area

#### **100 90** (%) **Landslides in cumulative percentage (%)** andslides in cumulative percentage ( **80 70 60 50 40 30 20 10 0 0 10 20 30 40 50 60 70 80 90 100 VHS LSZ area in cumulative percentage (%) VLS**

**Figure 11** Success rate curve

# **8 Conclusions**

The present work is an attempt towards application of GIS for landslide susceptibility mapping based on statistical approach. Six factors were selected and classified into various categories in the present study. Application of GIS involved generation of thematic data layers and their spatial analysis to determine the numerical weights of the categories of the factors in order of their influence for landslide occurrence. The information model was used for determination of weights of the factor categories and was found to be a useful method for landslide susceptibility mapping. All the data analysis was carried out in GIS which is a powerful tool for data storage and retrieval, map preparation and performing complex operations of voluminous data.

The landslide susceptibility map shows five

relative classes of landslide susceptible zones. The map was validated by determining landslide density and also using success rate curve method for different susceptible classes and was found to be in coherence with the ground instability conditions. Very high and high landslide susceptible zones need to be investigated further in detail before implementation of any hill developmental project. Such landslide susceptibility maps are very useful for planners for selecting suitable locations for developmental activities in hilly regions.

## **Acknowledgements**

Authors are grateful to the Director, CBRI for his kind permission to publish the work.

#### **References**

- Anbalagan R. 1992. Landslide hazard evaluation and zonation mapping in mountainous terrain. *Engineering Geology*, **32**:269 **~** 277.
- Bhasin R., Grimstad E., Larsen J. O., Dhawan A. K., Singh R, Verma S. K. and Venkatachalam K. 2002. Landslide hazards and mitigation measures at Gangtok, Sikkim Himalaya. *Engg. Geology*, **64**:351 **~** 368.
- Carrara A. 1983. Multivariate models for landslide hazard evaluation. *Mathematical Geology*, **15**(3):403 ~ 427.
- Carrara A., Cardinali M., Detti F., Guzzetti F., Pasqui V. and Reichenbach P. 1991. GIStechniques and statistical models in evaluating landslide hazard. *Earth Surface Processes and Landforms*, 16(5):427 ~ 445.
- Chung C. F. and Fabbri A. G. 1999. Probabilistic Prediction Models for Landslide Hazard Mapping, *Photogrammetric*

*Engineering & Remote Sensing*, **65**(12), 1389 ~ 1399.

- Chung C. F., Fabbri A. G. and van Westen C. J. 1995. Multivariate regression analysis for landslide hazard zonation. *Geographical Information Systems in Assessing Natural Hazards* (A. Carrara, and F. Guzzetti, editors), Kluwer Academic Publishers, Dordrecht, pp.107 ~ 134.
- Dhakal A. S., Amada T. and Aniya M. 2000. Landslide hazard mapping and its evaluation using GIS: An investigation of sampling schemes for a grid-cell based quantitative method. *Photogrammetric Engg. & Remote Sensing*, **66**(8):981 ~ 989.
- Gupta R. P. and Joshi B. C. 1990. Landslide hazard zoning using the GIS approach – A case study from Ramganga catchment, *Himalaya. Engg. Geol*, **28**:119 ~ 131.
- Jade S. and Sarkar S. 1993. Statistical model for slope instability classifications. *Engineering Geology*, **36**:71 ~ 98.
- Juang C. H., Lee D. H. and Sheu C. 1992. Mapping slope failure potential using fuzzy sets. *Geotechnical Engineering*, **118** (3):475 ~ 494.
- Kanungo D. P., Arora M. K., Sarkar S. and Gupta R. P. 2006. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Engineering Geology*, Vol. **85**, pp. 347 **~** 366.
- Kingsbury P. A., Hastie W. J. and Harrington A. J. 1992. Regional landslip hazard assessment using a GIS. *Proceedings of the 6th International Symposium on Landslides*, Christchurc. **2**:995 ~ 1000.
- Lee S., Choi J. and Min K. D. 2002b. Landslide Susceptibility Analysis and Verification using the Bayesian Probability Model, *Environmental Geology*, **43**, 120 ~ 131.
- Lin, M-L. and Tung, C-C. 2003. A GIS-based Potential Analysis of the Landslides induced by the Chi-Chi Earthquake, *Engineering Geology*, **71**, 63 ~ 77.
- Lu P. F. and An P. 1999. A Metric for Spatial Data Layers in Favorability Mapping for Geological Events, *IEEE Transaction on Geosciences & Remote Sensing*, **37**, 1194 ~ 1198.
- Nagarajan R., Mukherjee A., Roy A. and Khire M. V. 1998. Temporal remote sensing data and GIS application in landslide hazard zonation of part of Western Ghat, India. *Remote Sensing*, **19**:573 ~ 585.
- Pachauri A. K and Pant M. 1992. Landslide hazard mapping based on geological attributes. *Engineering Geology*, **32**:81 ~ 100.
- Raina V. K and Srivastava B. S. 1981. Tectonic evolution of the Sikkim Himalaya. *Contemporary Geoscientific Reseraches in Himalaya*, 1:29-40
- Saha A. K. Gupta R. P and Arora M. K. 2002. GIS-based landslide hazard zonation in the Bhagirathi (Ganga) valley, Himalayas. *Int. Jour. of Remote sensing*, **23**(2):357 ~ 369.
- Sarkar S., Kanungo D. P. and Mehrotra G. S. 1995. Landslide hazard zonation : A case study in Garhwal Himalaya, India. *Mountain Research and Development*, **15**(4):301 ~ 309.
- Sarkar S. and Kanungo D. P. 2004. An integrated approach for landslide susceptibility mapping using remote sensing and GIS. *Photogrammetric Engineering and Remote Sensing*,  $70(5):617 \sim 625.$
- Sarkar S. and Gupta P. K. 2005. Techniques for landslide hazard zonation – application to Srinagar-Rudraprayag area of Garhwal Himalaya. *Journal of Geological Society of India*, **65**:217 ~ 230.
- Sinha, B.N., Varma, R.S. and Paul, D.K., 1975. Landslides in Darjeeling district (West Bengal) and adjacent areas, *Bulletins of the Geological Survey of India*, Series B, No.36, 45 p.
- Westen C. J. 1994. GIS in landslide hazard zonation: a review, with examples from the Andes of Colombia. *Mountain environments and geographical information systems* (M.F. Price and D.I. Heywood, editors), Taylor & Francis Publishers, London, pp.  $135 \sim 165$ .
- Yin K. L. and Yan T. Z. 1988. Statistical prediction models for slope instability of metamorphosed rocks. *Proceedings of the 5th International Symposium on Landslides*, Lausanne, **2**:1269 ~ 1272.