Probabilistic Techniques, GIS and Remote Sensing in Landslide Hazard Mitigation: A case Study from Sikkim Himalayas, India

A Ramakrishnan, M.K. Ghose, R. Vinu Chandran and **A. Jeyaram**
Regional Remote Sensing Service Centre
ISRO, IIT Campus, Kharagpur- 721302
India

Abstract

Landslides form one of the most devastating natural hazards in the Sikkim State of India. This natural hazard alone causes severe damage to properties and human lives. The identified conditioning factors include adverse rock types (mica schist, phyllite, granite gneiss and calc schist), multiple joint sets, active tectonism, and very high annual precipitation (3539 mm). The triggering factors are mainly rapid pore pressure built up, seismic activity and anthropogenic interference. This paper evaluates the landslide hazard zones using information theory (1) and regression analysis (R) in GIS environment. In all, 14 variables are identified as conditioning and triggering factors and accordingly probabilistic prediction maps are prepared individually by both the methods. The results thus generated are compared and classified into three slope instability zones viz. low ($-0.143 < Ij < 0.02$ & $0.38 < Rj < 0.55$), medium ($-0.02 < Ij < 0.103$ & $0.55 < Rj < 0.73$) and high ($Ij > 0.13$ & $Rj > 0.73$) on the basis of histogram distribution. Further, these probabilistic prediction maps are compared with the actual landslide map generated from recent satellite data (January 2002) for the accuracy of prediction. The generated hazard maps agree with the observed landslide incidences. Thus, the proposed methodology can be utilized effectively in landslide hazard zonation studies.

Introduction

Similar to other Himalayan terrains, the Sikkim state of India also has formidable physical features. The lofty hill ranges, steep valleys, cliffs, gorges owe their origin to complex physical, geologic and tectonic processes. Being a part of the Himalayan orogenic belt, the natural hazards (landslides, earthquakes) also form an integral part of the study area. In the investigated area, landslides are by far the most significant natural hazard in terms of damage caused to lives and properties (Bhasin *et al*. 2002). This is best exemplified by the death toll of 3 3,000 human lives in a single year (1968). Landslides in Sikkim are triggered both due to natural phenomena (high rainfall, seismicity) and anthropogenic activities (road cutting, deforestration). Commonly observed slope failures include block slide, debris slide and earth creep. Thus, mitigation and management of the landslide hazard in this area is one of the foremost requisites for landuse planners.

Factors contributing to slope failures at a specific site are generally specific with respect to conditioning and triggering factors (Harp and Jibson, 1996; Jibson, *et al*., 1994; 1998; 1999). Hence, hazard maps representing the susceptibility of slope failures due to different conditioning and triggering factors (variables) could be a better choice in preparing hazard zonation maps.

The aim of this paper is to develop a methodology that could produce a hazard map over a large area with higher degree of accuracy in a GIS environment. Several probabilistic methods (quantification theory, multiple regression, discriminant analysis, monte-carlo simulation, etc.) were attempted (Hayashi,1952; Carrara,1983; Haruyama and Kitamura, 1984; Kawakami,1984; Yin and Yan, 1988; Jade and Sarkar, 1993; Jibson *et al*. 1999, Luzi *et al*. 2000) in the past to derive a probabilistic zonation map for landslide hazard. This paper pertains to utility of information theory and regression analysis in landslide hazard zonation. This method is preferable over other methods by virtue of its accuracy and field applicability. The generated probabilistic hazard maps are cross-validated with recent satellite data for accuracy.

The study area

The investigated area (Figure 1) forms a part of East Sikkim district and lies between the northern latitudes 27º 15'- 27º 30' and eastern longitudes 88º 30'- 88º 45'. The geomorphology of the area owes its origin to tectonic, glacial and fluvial processes. The elevation in the region ranges from 600 in to 4500 in. The area is characterised by very high average annual rainfall (3539 mm). It is significant to mention that occasional cloud burst resulting in very high rainfall (500 mm in 24 hours) is a common phenomenon.

Geologically, Precambrian rocks constitute the major

Geocarto International, Vol. 20, No. 4, December 2005
Published by Geocarto International Centre, G.P.O. Box 4122, Hong Kong.
E-mail: geocarto@geocarto.com
Website: http://www.geocarto.com

portion (Figure 2) of the study area (Raina and Srivastava, 1981). The Sikkim group of rocks comprises lower metamorphic grades such as phyllite, chlorite schist, and quartz schist. The Chungthang group of rocks is gneissic in nature. Besides these, pockets of sandstone, shale and conglomerate (Gondwana group) are also exposed in the western parts of the study area. The flood plain deposits represent the Quaternaries.

Akin to other parts of the Himalayas, Sikkim is also tectonically very active. Evidences of neotectonic signatures (active faults, earthquakes) are omnipresent. Salient among the faults/thrusts in and around the study area include the Main Boundary Thrust (MBT), Main Central Thrust (MCT) and Yungthang Thrust (YT). Besides these major thrusts, several other minor thrusts and faults can be evidenced in both satellite imagery and in the field.

Methodology

The adopted methodology (Figure 3) includes a four phased approach viz. identification of conditioning and triggering factors (variables), data processing for GIS environment, modeling, and validation of results with satellite / field data. Data on lithology, seismicity, structure are collected from the field, satellite data and existing sources. The slope map is generated using GPS data from the field in conjunction with the information from Survey of India topographic sheets. A satellite data set (IRS-LISS-3) for the period December 1998 is used to map landslide events of the past. A similar landslide event map for the period January 2002 is used for comparing and validating the probable hazard zones generated using information theory and regression analysis.

Raster data are rectified and registered using ERDAS-IMAGINE version 8.5. The vector layers are generated in ARC/INFO version 7. 1. The programme for statistical analysis is written in Visual Basic with Arc Macro Language (AML) as back end.

Results

Input thematic layers

This requires identification of the total number of polygon elements associated with the given area (N) and terrain specific conditioning and triggering factors (variables).

Dividing the study area into 172 small blocks/elements on the basis of slope, aspect and watershed divide (Figure 4) derives polygon elements. From the field investigations, the following variables (M) are identified as key conditioning and triggering factors.

(A) Lithology: Lithology of the investigated area is comprised of four major rock types with primary and secondary planes of discontinuity. Accordingly four classes viz.

.if Mica Schist	then Xl = 1 else 0
.if Phyllite	then X2= 1 else 0
.if Granite	then X3= 1 else 0
.if Gneisses	then X4= 1 else 0

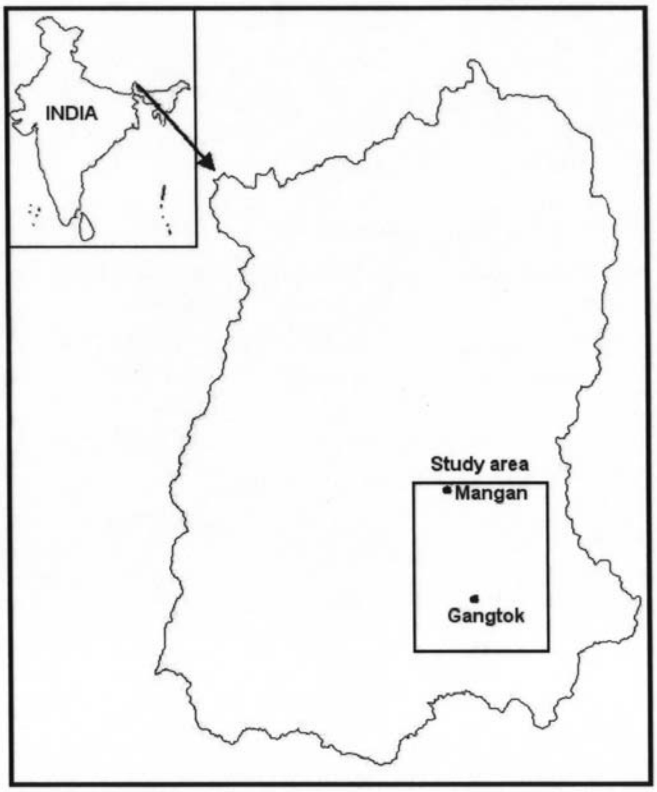

Figure 1 Location map of the study area.

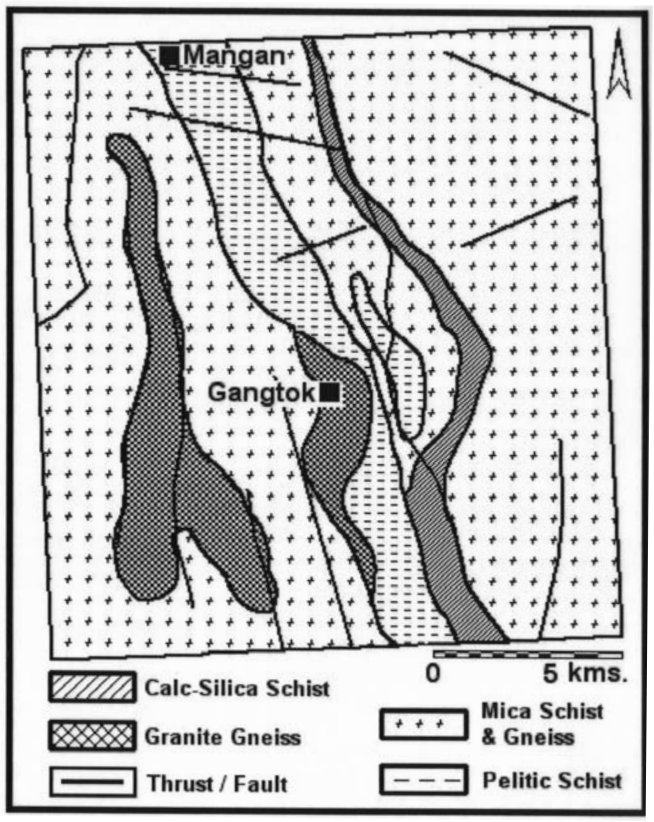

Figure 2 Geological setup of the investigated area

(B) Seismicity / earthquakes: In the investigated area, occurrences of seismic events are so far confined to major thrusts / faults. Further, the intensity of these shocks seldom exceeds 6 on the Richter scale. Hence, buffer zones up to 50 kms from the epicenter viz.

.if earthquakes (with Mb>2) epicentre is < 10kms
then X5 = 1, else = 0
.if earthquakes (with Mb>2) epicentre is < 25kms
then X6 = 1, else = 0
.if earthquakes (with Mb>2) epicentre is < 50kms
then X7 = 1, else = 0 is considered

(C) Slope Angle: Five classes are made on the basis of field observation between slope amount and incidence of landslides. Accordingly,

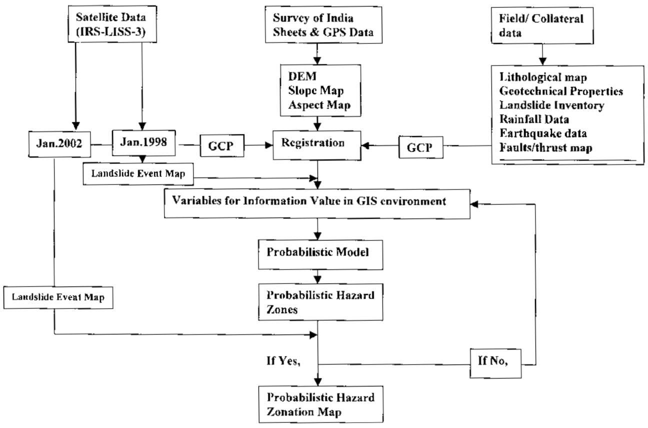

Figure 3 Flow chart depicting the adopted methodology

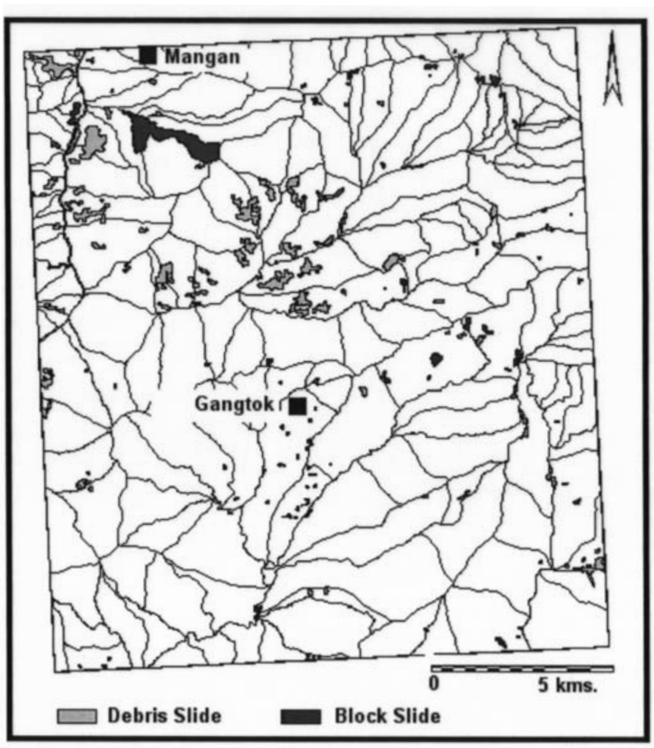

Figure 4 Discretized polygon elements and historic landslide activity map

.if slope > 46 then X8=1 else 0
.if slope 35 - 45º then X9= 1 else 0
.if slope 25 - 34º then XI0=1 else 0
.if slope 15-24º then XI 1=1 else 0
.if slope < 15º then X 1 2= 1 else 0

(D) Active Faults/Thrusts: Since the study area is an integral part of Himalayan tectonics and traversed by major active thrusts and faults, presence of tectonic elements within a buffer area of 10 kms is incorporated as a triggering factor. Accordingly,

if fault/thrust is present within a buffer distance of 10 kms, then XI 3 = 1, else = 0

(E) Historic Landslide activity: Presence or absence of landslide events (Figure 4) signifies the tendency for new / reactivation of landslides and hence need to be studied with an emphasis on their susceptibility. This parameter is considered as a dependant variable and hence, the information theory and regression methods are evaluated.

Additional information on geotechnical parameters, weathering, hydrogeological conditions can be incorporated depending upon the field knowledge and terrain conditions.

All the different input layers are intersected in ARC/ INFO and the statistical analyses are done on the resultant layer in Visual Basic front end. Result from this analysis will be in tabular form. The output will comprise details pertaining to the information value and regression coefficient of variables, information value and regression value of each element, minimum and maximum information value and regression value, element number and grades of instability associated with each element. This data can be further classified into different grades of instability based on the range of information and regression value and the number of elements in each of the instability classes.

Programme Structure for Information Value (IV) Method

The analysis used for the landslide hazard is the Information value method based on probability theory and is summarized as below:

Suppose there are N potential factors/ variables that affect the slope instability, then the degree of potential hazard in an area can be estimated on the basis of number of fatigue factors and their severity and interactions. However, the main objective is to predict the areas of various degrees of landslide susceptibility. For this, first a given area is divided analytically into a number of polygon elements by considering the micro-watershed boundaries. As per the law of Information theory, every element j (j =1, 2................N) can be defined stable or unstable on the basis of the information value (I_j) of that element. Higher the value of Ij more unstable the element j is, within the slope.

The total information value in the element j can be calculated as:

$$\mathbf{I_j = \sum_{i=1}^{M} I_i X_{ji}} \tag{1}$$

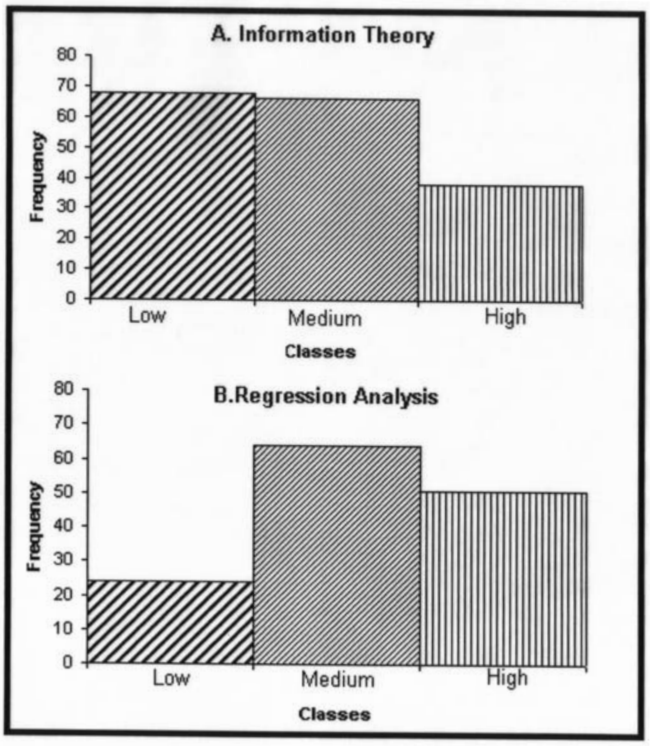

Figure 5 Histogram distribution of slope instability classes
A. information value method
B. Regression analysis method

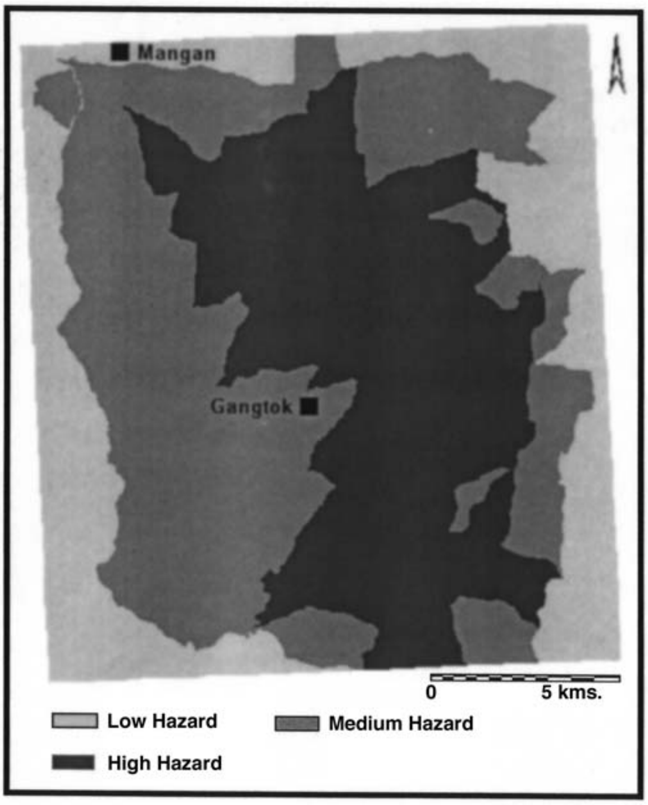

Figure 6 Landslide hazard zonation map by Information theory method.

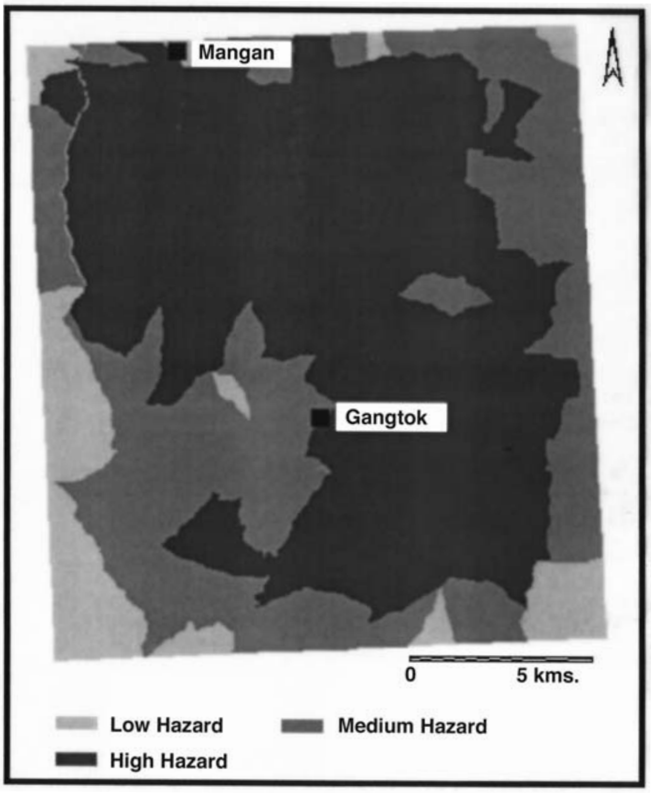

Figure 7 Landslide hazard zonation map by Regression analysis method.

Table 1 Information value and regression coefficients of identified variables.

Variable	Information Value	Regression Coefficient
X1	0.001	0.289
X2	0.158	0.132
X3	0.110	0.054
X4	0.061	0.073
X5	-0.037	-0.034
X6	0.020	0.179
X7	0.004	0.117
X8	-0.120	-0.111
X9	-0.008	0.025
X10	-0.001	0.007
X11	-0.004	-0.271
X12	0.001	0.213
X13	0.007	0.270

where:
X_{ji} = value of i^{th}. variable (i =1, 2, ..,M) for the j^{th} - element (j = 1, 2, ...,N);
= 1, if variable i exists in element j;
= 0, if variable i does not exist in element j;
M = number of variables associated with a given area;
I_i = Information value supplied to landslide by variable i
= log [(Si/Ni)/ (S/N)] (2)

where:
N = total number of elements;

S = number of elements with history of landslide occurrence;

S_i = number of elements with history of landslide occurrence involving variable i ;

N_i = number of elements involving variable

Programme Structure Regression Value (RV) Method

Regression is defined as the dependence of variable Y (slope instability) on variable X (conditioning and triggering factors). The regression co-efficient b is a measure of the dependence and a is the constant of regression equation or intercept with x-axis. Mathematically, this can be expressed as:

Y^ = **a+bx** , where Y^ is the estimated deviation of Y corresponding to any x deviation.

If the relationship among the different conditioning, triggering factors and slope failure in a given area is established, then a measure of the instability of that area can be determined. In the present case, the regression value for the occurrence of landslide in an element j 0=1, 2, 3,...... N) is expressed as:

$$\mathbf{R_j = \sum_{i=1}^{M} B_i\, X_{ji}} \qquad (3)$$

where:

R_j = Regression value for the j the element due to i (= 1,2,.......M) factors.

= 1, if the landslides occur in element j.

= 0, otherwise.

M = number of variables;

B_i = regression co-efficient (i=l, 2, 3........M)

The estimates (bi) of the regression coefficients (Bi) are determined by least-square

method using the regression equation as stated above, and substituted to obtain the

estimated values of R_js for each of the element j (=I, 2...., N). A higher value of R_j

indicates that the element j falls within the unstable zone.

The results of the information value and regression coefficient values for the individual variable (Xi) are tabulated in table 1. On the basis of histogram distribution (Figure 5A,

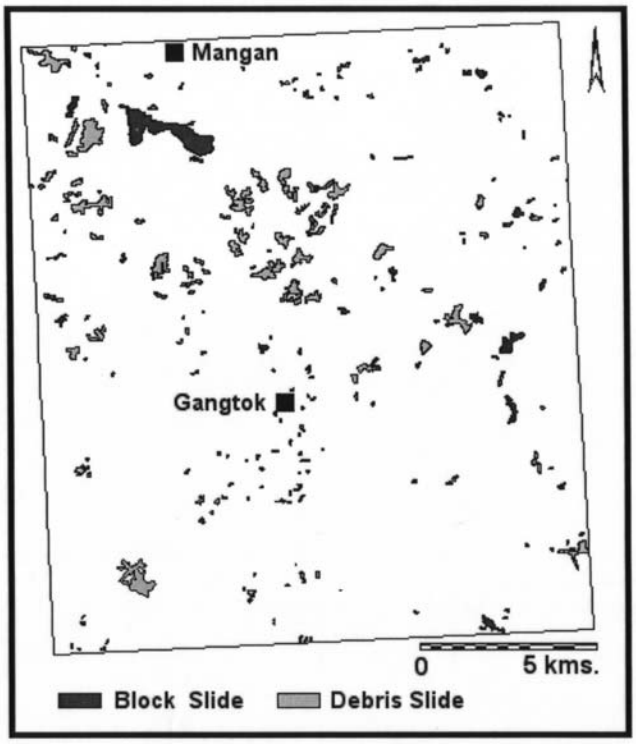

Figure 8 Landslide-event map for the period January 2002.

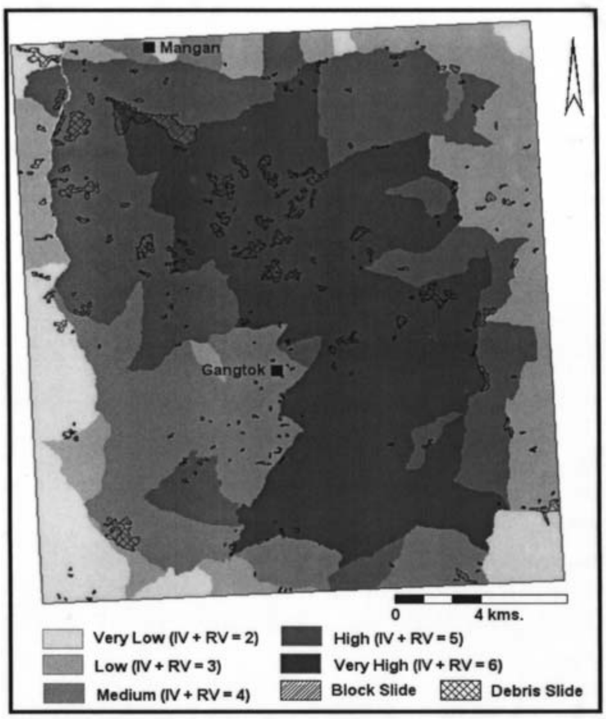

Figure 9 Composite landslide hazard zonation map with landslide event for the period 2002.

Table 2 Probabilistic hazard zones vis-à-vis observed landslide occurences.

Hazard Class	Information Value	Landslide Incidence	Area (km^2)	Regression Value	Landslide incidences	Area (km^2)
Low	-1.0 - -0.55	44	2.96	-0.04 - 0.04	14	1.28
Medium	-0.55 - -0.05	115	8.75	0.04 - 0.15	70	3.69
High	> -0.05	75	8.95	> 0.15	150	15.71

B), the polygon elements are classified into three hazard classes viz. low ($-0.143 < ij < -0.02$ & $0.38 < RJ < 0.55$), medium ($-0.02 < lj < 0.103$ & $0.55 < RJ < 0.73$) and high ($1j > 0.\ 13$ & $RJ > 0.73$) landslide hazard prone zones. On the basis of this information, landslide risk maps are prepared by information theory method (Figure 6) and regression analysis method (Figure 7) in ARC/INFO. Mapping the landslide events for the recent period is done using the IRS-LISS-3 data product for the period January 2002. The satellite data are digitally classified and a landslide event map (Figure 8) is prepared after the field study.

Discussion and Conclusions

The given area has been divided into a number of polygon element units. The factors affecting the slope instability have been expressed as an item. Each item is classified into categories such as geology, slope, seismicity based on field knowledge. These categories are expressed by the variable X_i.

The information and regression value of each polygon element unit have been determined using the equation no I and 3 respectively. Evaluated regression equation is significant at 5% confidence level. On the basis of histogram distribution, they are classified into three major classes. These three grades of instability are defined as low, medium and high depending upon ranges of information and regression value. From the overlay analysis it is apparent that the hazard maps generated (by information theory and regression analysis) correlate well with the landslide event map for the recent period (Figure 9). There is high degree of conformity among the hazard zones vis-à-vis the event map in terms of both landslide incidences and magnitude of an event expressed in terms of area (Table 2). However, the medium and high classes of information theory method do not commensurate well with the landslide event map. Whereas in the case of regression analysis, it commensurates perfectly with the respective landslide hazard zones. Thus, the precision of the hazard map generated by the regression analysis is higher than that of the information theory method. The discrepancy may be attributed to the following reasons.

1. Certain polygon elements with a given landslide can fall within a low to medium grades of instability; these polygon elements, however, are assigned a high grade of slope instability. This discrepancy of grades arises when factors considered in the analysis have a low weightage in that particular element.
2. The landslide in that element might have occurred due to a factor, which has not been considered in analysis because; it does not have a significant influence on the overall landslide occurrences in the area under study.

In order to increase the accuracy of prediction, results from information value and regression analyses are combined into the following new classes.

$1V_{High}$ and RV_{High} ⇒ Very High
IV_{High} and RV_{Medium} OR IV_{Medium} and RV_{High} ⇒ High
IV_{High} and RV_{Low} OR IV_{Low} and RV_{High} OR IV_{Medium} and RV_{Medium} => Medium
IV_{Medium} and RV_{Low} OR IV_{Low} and RV_{Medium} ⇒ Low
IV_{Low} and RV_{Low} ⇒ Very Low

From the landslide incidences (Figures 8,9) it can now easily be concluded that areas under very high and high classes are vulnerable to landslides.

Acknowledgements

The authors are thankful to Mr. D. Shreshta, Scientific officer, Sikkim Department of Science and Technology for his help during the fieldwork. Authors are thankful to the referee for his comments, which helped to enhance the quality of the paper.

References

Bhasin,R., Grimstad,E., Larsenj.O., Dhawan.A.K., Singh,R., Verma, S.K. and Venkatachalam,K. (2002) Landslide hazard and mitigation measures at Gangtok, Sikkim Himalaya. Intl. Jour. Engineering Geology, vol.64, pp. 351-368.

Carrara, A (1983): Multivariate models for landslide hazard evaluation, Math. Geol., vol. 15, no3, pp 403 - 425.

Harp,E.L., and Jibson,R.W. (1996) Landslides triggered by the 1994 Northridge, California earthquake. Bull. of Seismological Soc. of America, vol. 86 (1B), pp.s319-s332.

Haruyama, M. and Kitamura, R., (1984) An evaluation method by the quantification theory for the risk degree of landslide caused by rainfall in active volcanic area. Int. Symp. Landslides, Toronto, Ont., pp.435-440.

Hayashi, C (1952) On the prediction of phenomena from qualitative data and quantification of qualitative data from the mathematico-statistical point of view., Ann. Inst. Statist. Math., vol.3, pp 69 -98.

Jade,S and Sarkar, S. (1993) Statistical models for slope instability classification, Intl. Jour. of Engineering Geology, vol. 36, pp 91 - 98.

Jibson,R.W., Prentice,C.S., Borissoff, B.A., Rogozhin, E.A. and Langer, C.J. (1994) Some observations of landslides triggered by 29th April 1991 Racha earthquake, Republic of Georgia. Bulletin of the Seismological Society of America, vol. 84(4), pp. 963-973.

Jibson, R.W., Harp,E.L. and Michael, J.A. (1998) A method for producing digital probabilistic seismic landslide hazard maps: An example from the Los Angeles, California, area. U.S. Department of interior, open file report 98113, pp.1-17.

Jibson, R.W., Harp,E.L. and Michael, J.A. (1999) A method for producing digital probabilistic seismic landslide hazard maps. Int. Jour. Engineering Geology, vol 58, pp. 271-289.

Kawakami, H. (1984) Landslide risk mapping by quantification method. Saito, Y. Int. Symp. Landslides, Toronto, Ont., pp 535-540

Luzi, L. Pergalani, F. and Terlien, M.T.J. (2000) Slope vulnerability to earthquakes at subregional scale, using probabilistic techniques and geographic information systems. Int. Jour. Engineering Geology, vol. 58, pp. 313-336.

Raina, V.K., Srivastava, B.S., 19 8 1. Tectonic Evolution of Sikkim Himalaya. Contemp. Geosci. Res. Himalayal, 109-115.

Yin, K.L. and Yan T.Z. (1988) Statistical prediction models for slope instability of metamorphosed rocks. Intl. Symp. Landslides, lausanne, pp. 1269 - 1272.