Weights of evidence modelling for landslide hazard zonation mapping in part of Bhagirathi valley, Uttarakhand

John Mathew[1,*], V. K. Jha[2] and G. S. Rawat[3]

[1]Regional Remote Sensing Service Centre (ISRO), Dehradun 248 001, India
[2]Indian Institute of Remote Sensing (ISRO), Dehradun 248 001, India
[3]Department of Geology, HNB Garhwal University, Srinagar 246 174, India

Weights of evidence method, which is basically the Bayesian approach in a log-linear form, using the prior probability of occurrence of an event like landslide, helps to find out its posterior probability based on the relative contributions of evidential themes which are influential in creating slope instability. In the present study, this method has been used to find out the probability of occurrence of landslides for unique combinations of evidential themes and to prepare a landslide hazard zonation map of part of Bhagirathi valley, Uttarakhand, within a Geographic Information System environment. Lithology, structure, slope, slope aspect, land use/land cover, drainage and distance to road are the evidential themes considered in the study. The model has been further validated using receiver operator characteristic curve analysis, which shows an accuracy of 84.6%.

Keywords: Geographic Information System, landslide hazard zonation, receiver operator characteristics, weights of evidence.

LANDSLIDES are one of the most common natural hazards in the Himalayan terrain, causing widespread damage to property and infrastructure, besides loss of human lives almost every year. Appropriate management measures taken at the right time will reduce the risk from potential landslides. In order to prioritize the area for hazard-mitigation efforts, it is beneficial to have a Landslide Hazard Zonation (LHZ)[1] map prepared depicting the ranking of the area based on actual and/or potential threat of slides in future. Any approach towards LHZ would require identification of the conditions leading to slope failure, their systematic mapping and evaluation of their relative contributions to landsliding in the area (M. L. Süzen, unpublished).

LHZ mapping is being carried out using qualitative or quantitative approaches. The qualitative methods essentially depend on expert opinion in dividing an area into different zones of varying landslide susceptibility. Using an inventory of existing landslides, the expert can assess the hazard of the area by identifying regions of similar geological and geomorphological conditions[2]. Keinholz[3] and Rupke *et al.*[4] used geomorphological and qualitative methods for assessing landslide hazard. Gee[5] suggested the term blind-weighting method for the qualitative/semi-quantitative approach in which the expert gives subjective weights for the contributing factors and then integrates these to derive a cumulative influence factor representing the degree of hazard. Pachauri and Pant[6] demonstrated a weighted landslide hazard mapping procedure in the Aglar catchment of Himalaya. Gupta *et al.*[7] and Saha *et al.*[8] used parameter-weighting method for LHZ mapping in part of Bhagirathi valley of Garhwal Himalaya. NRSA[9] adopted the Analytical Hierarchy Process methodology for preparing LHZ maps along the corridors of the major pilgrimage routes in Uttarakhand, Uttar Pradesh and Himachal Pradesh Himalaya. Qualitative methods are subjective and offer less reproducibility of the output[10].

The quantitative methods in LHZ involve statistical, geotechnical and artificial neural network methods. Geotechnical approaches include the physical processes involved in landsliding and express the hazard in the form of safety factors.

Statistical methods are based on the mathematical relationship between the observed landslides and their controlling factors. Such methods reduce the subjectivity and ensure better reproducibility of the hazard zonation processes and their outputs. The model developed for one area may not exactly give the same type of result in a different area. The statistical methods involve both bivariate as well as multivariate techniques for LHZ mapping. The commonly used multivariate statistical methods in landslide hazard assessment are linear regression, discriminant analysis and logistic regression. The bivariate statistical methods utilize the normalized landslide densities derived using the landslide occurrence in each parameter class to arrive at the hazard map. Information value method and weights of evidence modelling are two common bivariate methods applied in LHZ mapping process.

The objective of this study is to assess the utility of weights of evidence model for landslide susceptibility mapping in an area, which is prone to landslides. If suc-

*For correspondence. (e-mail: john_isro@yahoo.com)

cessful, this model can be extended to other areas of similar geological and/or geomorphological set-up. The LHZ map thus developed can be used for prioritizing the areas for timely management measures. Weights of evidence model is relatively straightforward and can handle the problem of missing data in analysis. The only requirement is the conditional independence of input variables.

Study area and data used

The present study area is in the catchment of River Bhagirathi, which is a tributary of River Ganga. Bhagirathi originates from the Gangotri glacier, which is a place of pilgrimage and the approach road to Gaumukh runs side by side the Bhagirathi river. Severe landsliding happens every year along this route, especially during the rainy season, causing a lot of damage to property and infrastructure, and inconvenience to the pilgrims. Keeping this in mind, the study area is selected along the corridor of the Rishikesh–Gaumukh road, covering about 102 sq. km.

The Indian Remote Sensing (IRS) 1D satellite data (LISS III and PAN) of 2 January 2000 were used in the study. The LISS III data were fused with PAN data for better spatial resolution and for the preparation of landslide inventory map. In addition, Survey of India topographic maps were also used, where the mapping scale was kept at 1 : 50,000. Selected fieldwork was carried out in the year 2000. The location map of the study area, along with the LISS III False Colour Composite is given in Figure 1.

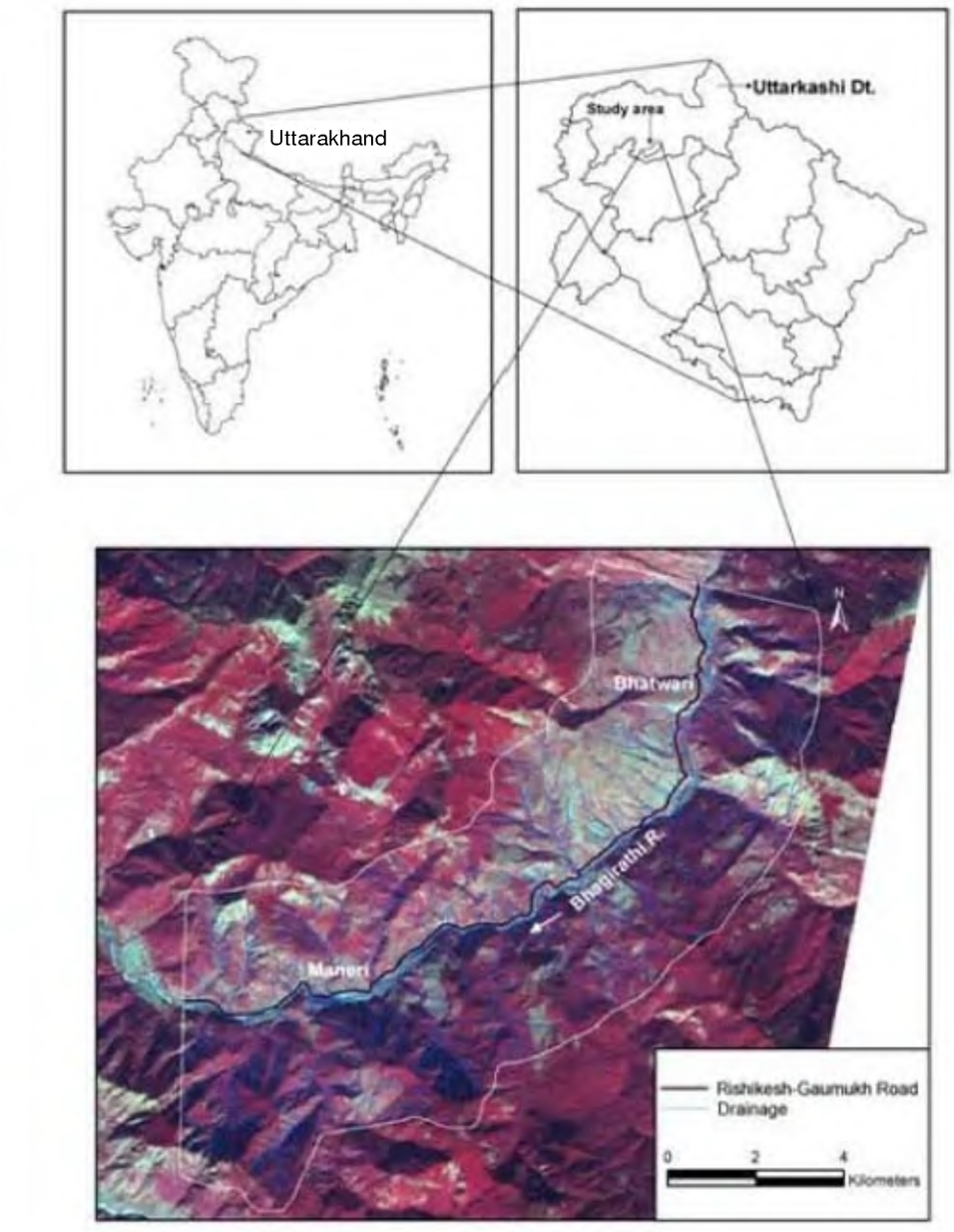

Figure 1. Location map of the study area.

Methodology

In the present study, weights of evidence modelling has been applied for landslide susceptibility mapping. Weights of evidence is a data-driven method which is basically the Bayesian approach in a log-linear form using the prior and posterior probability and is applied where sufficient data are available to estimate the relative importance of evidential themes by statistical means[11]. This method had been initially applied to non-spatial, quantitative, medical diagnosis to combine evidence from clinical diagnosis to predict diseases. The weights of evidence model was used for mineral potential mapping[12–15] by implementing it in a Geographic Information System (GIS) framework. This method was also applied for LHZ mapping in Mussoorie hills (S. V. Panickar, unpublished).

The spatial association between a set of evidential themes and a set of known landslide locations, which are expressed as the weights of evidence, is combined with the prior probability of occurrence of landslides to derive the posterior probability of occurrence of landslides, provided the evidential themes are conditionally independent with respect to the slides[16]. This method consists of reducing each set of landslide-related factors on a map to a pattern of a few discrete states. In its simplest form, the pattern for a feature is binary, representing its presence or absence within a pixel (M. L. Süzen, unpublished).

The posterior odds of the landslide (S), given the presence or absence of any binary pattern (B_i or $\overline{B}_i$) is given by:

$$O(S \mid B_i) = O(S)[P(B_i \mid S)/P(B_i \mid \overline{S})],$$

and

$$O(S \mid \overline{B}_i) = O(S)[P(\overline{B}_i \mid S)/P(\overline{B}_i \mid \overline{S})],$$

where $O(S)$ is the prior odds of the slide, $[P(B_i \mid S)/P(B_i \mid \overline{S})]$ is called the sufficiency ratio (LS) and $[P(\overline{B}_i \mid S)/P(\overline{B}_i \mid \overline{S})]$ is called the necessity ratio (LN).

Bonham-Carter[11] defined positive and negative weights (w_i^+ and w_i^-) that combine these conditional probabilities as:

$$w_i^+ = \log_e[P(B_i \mid S)/P(B_i \mid \overline{S})]$$

and

$$w_i^+ = \log_e[P(\overline{B}_i \mid S)/P(\overline{B}_i \mid \overline{S})].$$

In GIS, this method can be implemented easily. By cross-tabulating the landslide map with the binary variable map, the following combinations (Table 1) are possible.

From Table 1, the weights can be calculated as:

$$w_i^+ = \log_e[(N_{\text{pix1}}/(N_{\text{pix1}} + N_{\text{pix2}}))/(N_{\text{pix3}}/(N_{\text{pix3}} + N_{\text{pix4}}))],$$

$$w_i^- = \log_e[(N_{\text{pix2}}/(N_{\text{pix1}} + N_{\text{pix2}}))/(N_{\text{pix4}}/(N_{\text{pix3}} + N_{\text{pix4}}))].$$

Table 1. Cross-tabulation of landslide and binary variable maps

	Variable class binary pattern	
Landslide	Present	Absent
Present	N_{pix1}	N_{pix2}
Absent	N_{pix3}	N_{pix4}

If there are n binary patterns, then the weights can be added to find the natural logarithm of posterior odds of the slides as given by:

$$\log_e O_{\text{posterior}}(S) = \sum_{i=1}^{n} w_i + \log_e O_{\text{prior}}(S).$$

The prior odds of the slides ($O_{\text{prior}}(S)$) can be calculated from the prior probability of the slides. If any map pattern is absent, w^+ will be replaced by w^-. If data are absent or missing for any location, then the weight is set to zero. Combining predictor maps results in unique conditions, which are pixels or groups of pixels with the same combination of spatial evidence. For the kth (k = 1, 2, ..., m) unique condition, posterior odds of the slides are converted to posterior probability (P_k) as

$$P_k = \frac{e^{\sum_{i=1}^{n} w_i^k + \log_e O_{\text{prior}}(S)}}{1 + e^{\sum_{i=1}^{n} w_i^k + \log_e O_{\text{prior}}(S)}},$$

where w_i^k denotes the weights (w_i^+ or w_i^-) contributed by binary pattern B_i in the kth unique condition.

The variance of posterior probability can be estimated based on variance of the weights[17]:

$$s^2(P_k) = \left[\frac{1}{N_{\text{pix}}(S)} + \sum_{i=1}^{n} s^2(w_i^k)\right] \times P_k^2,$$

where $s^2(w^+) = [1/N_{\text{pix}}(B_i \cap S)] + [1/N_{\text{pix}}(B_i \cap \overline{S})]$, and $s^2(w^-) = [1/N_{\text{pix}}(\overline{B}_i \cap S)] + [1/N_{\text{pix}}(\overline{B}_i \cap \overline{S})]$.

The contrast ($w^+ - w^-$) gives a useful measure of the correlation between the variable maps and landslide occurrence. The contrast (C) is a useful measure to convert continuous evidential themes into binary patterns. In cases where the contrast does not show a clear maximum for deciding the cut-off for this type of conversion, the studentized contrast (Stud(C) = C/σ_C) can be used[11].

The weights of evidence model is generally applied using binary evidential themes. Real-world geospatial data are usually multi-class or continuous and the conversion of such data into binary type will result in the loss or distortion of valuable information. The extended weights of evidence

model using multi-class predictor variables is preferable to binary weights of evidence model[16]. In the present study, weights of evidence has been utilized with multi-class as well as binary evidential themes using ArcGIS Spatial Analyst and Arc-SDM[18].

Landslide sample locations

Though it was difficult to identify the old stabilized slides directly on the satellite image, the locations were confirmed through interaction with local residents during field check. Twenty-five slides were mapped which vary in area from 0.08 to 15 ha. Most of the slides are close to the Bhagirathi river. The centres of the raster cells representing the landslide locations have been converted to a point-shape file which has been considered as the sample location input. There are 1234 such points representing the sample locations.

Evidential themes and weights calculation

Geology: The study area falls partly under the Garhwal Lesser Himalaya and partly under Higher Himalaya. The Main Central Thrust (MCT) separates the two and passes through the study area. The major lithological units in the study area are quartzites, amphibolites, epidiorites, granitic gneisses, migmatites and schists. In addition, minor amount of metavolcanic rocks are also present[19].

It has been observed that landslides are common in the area dominated by gneisses, migmatites and schists. Here, the presence of weak planes in gneisses and schists makes the rocks weaker and thereby facilitates slope instability. The amphibolite and epidiorite areas show high degree of weathering and have clay-rich soils. Landslides occur in such areas also, but are not as frequent as they are in the gneiss/migmatite/schist areas. In the quartzitic areas a few landslides occur where there is jointing or where there are alternate layers of quartzite and slate/chlorite schist. The geological map of the study area is given in Figure 2. The weights have been computed for the different lithological units and are given in Table 2.

Lineament analysis: Traces of faults/fractures which appear as linear to curvilinear features (lineaments) on the satellite image are considered as important geological

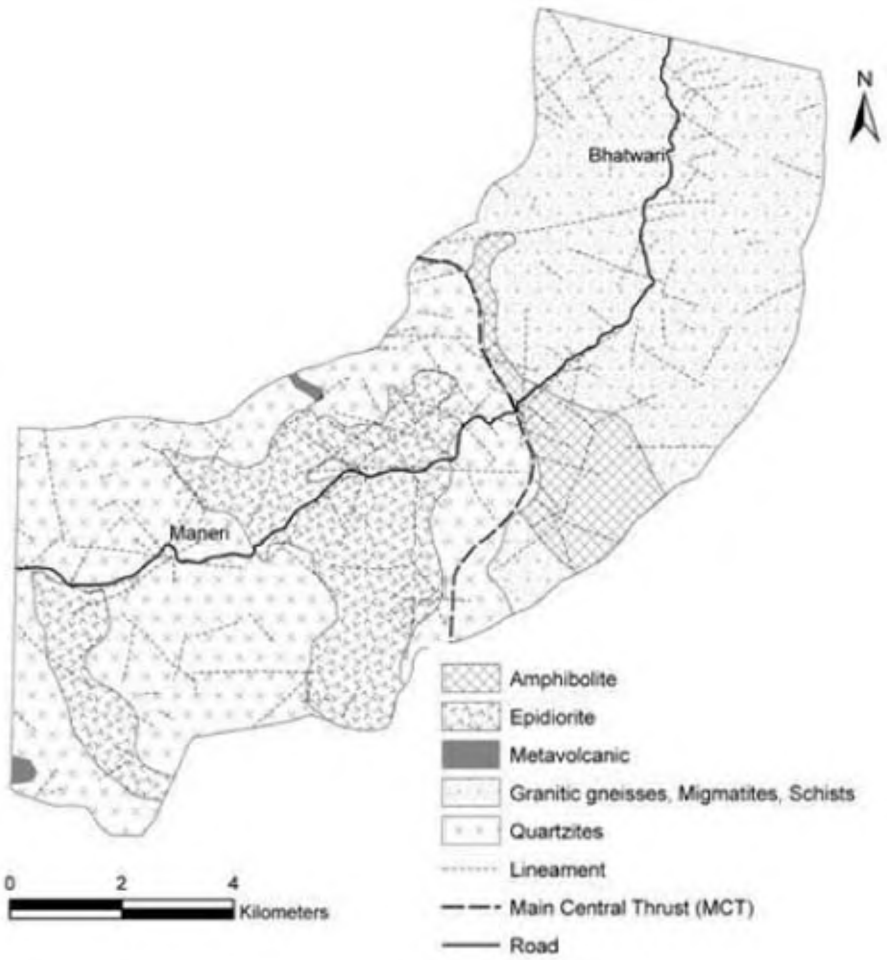

Figure 2. Geological map showing MCT and other lineaments in the study area (after Agarwal and Gopendra Kumar)[19].

Table 2. Weights and contrast values for lithological classes

Lithounit	Area (sq. km)	No. of slide points	w^+	w^-	Contrast	Stud (C)
Quartzite	38.2675	332	–0.3320	0.156596	–0.489	–7.5879
Amphibolite	6.4244	7	–2.4115	0.059655	–2.471	–6.5172
Epidiorite	18.6719	224	–0.0059	0.001304	–0.007	–0.0966
Metavolcanic	0.2913	0	–5.8698	0.002866	–5.873	–0.5873
Gneisses/migmatites/schists	38.6106	671	0.3682	–0.31267	0.6808	11.865

Table 3. Weights and contrast values for lineament density classes

Lineament density class	Area (sq. km)	No. of slide points	w^+	w^-	Contrast	Stud (C)
I	6.0000	64	–1.95901	0.405347	–2.36436	–18.4031
II	41.0000	496	0.036909	–0.02406	0.060971	1.0461
III	18.0000	366	0.527039	–0.15934	0.686375	10.95302
IV	5.0000	151	0.928885	–0.08098	1.009869	11.52369
V	2.0000	157	1.915393	–0.11717	2.032558	23.26484

Table 4. Weights and contrast values for distance to lineament

Distance to lineament (m)	Area (sq. km)	No. of slide points	w^+	w^-	Contrast	Stud (C)
< 85	29.0388	718	0.737208	–0.54565	1.282856	22.12678
> 85	72.9612	516	–0.54565	0.737208	–1.28286	–22.1268

structures which influence landslides. The surface trace of the MCT is one of the most important lineaments in the area. All the lineaments (with bearing on structure) have also been mapped on satellite data using digital image enhancement techniques.

It has been observed that major lineaments/planar structures are important factors governing the stability of slopes in the study area. The faults, fractures and joints not only tend to destabilize the area through deterioration of the strength of the rocks, but also accelerate the weathering process. The probability of landslides is higher close to the lineaments as well as in areas where lineament density is high. In order to accommodate these two factors, the distance to lineament and lineament density maps have been prepared. The lineament density values vary from zero to 98 km per sq. km and have been classified into five classes (Class I, < 15; Class II, 15 to 25; Class III, 25 to 35; Class IV, 35 to 50 and Class V, > 50 km/sq. km). The estimated weights for the classified lineament density classes are given in Table 3. The distance to lineament is a continuous data and has to be made either categorical or binary to be used in the model. The weights have been determined on a cumulative incremental basis for the distance to lineament and the studentized contrast (Stud(C)) is plotted against the distance to lineament (Figure 3) to find the cut-off to convert it into binary pattern. The threshold is defined where the studentized contrast is maximum and is found to be 85 m. Using this threshold, the distance to lineament map is converted to a binary map and the weights have been recalculated (Table 4).

Geomorphology: The study area falls in an active, young, tectonic mountainous terrain. As the valley side slopes are generally steep and extend down into the Bhagirathi river, no piedmont zone is observed in the area. The selected area has more or less uniform morphology, dominated by hills and valleys. The area has uneven topography in the Bhagirathi river valley. The Bhagirathi river in the area flows predominantly in ENE–WSW direction and has a constricted V-shaped valley together with steep gradient and high run-off. This indicates a youthful geomorphological nature of the region[7]. The elevation of the area varies from about 1200 to 3000 m amsl.

The joints, fractures and faults are responsible for dissecting the hills. As the dissection increases, weathering and soil formation processes also increase and progressively facilitate instability of the slopes. Based on the dissection intensity, the area has been further classified into denuadational hills of high, moderate and low dissection, which exhibit high, medium and low lineament densities respectively. Since the influence of lineaments on slope destabilization has already been taken care of by the lineament density and the distance to lineaments factors, the geomorphological theme where the basis of major subclassification is 'discontinuities' has been omitted from the

Table 5. Weights and contrast values for slope classes

Slope class	Area (sq. km)	No. of slide points	w^+	w^-	Contrast	Stud (C)
I	4.3306	43	–0.1963	0.0079	–0.2042	–1.31112
II	6.3200	59	–0.25834	0.0149	–0.2733	–2.04205
III	57.6631	617	–0.12105	0.1377	–0.2588	–4.52817
IV	33.7156	515	0.237785	–0.1411	0.3789	6.536597
V	0.2363	0	–5.66029	0.0023	–5.6626	–0.56625

Table 6. Weights and contrast values for aspect classes

Aspect class	Area (sq. km)	No. of slide points	w^+	w^-	Contrast	Stud (C)
I	10.93375	26	–1.63028	0.09251	–1.72279	–8.68443
II	11.61125	37	–1.33708	0.090793	–1.42787	–8.54486
III	15.7775	68	–1.03439	0.111776	–1.14617	–9.1736
IV	12.103125	206	0.347094	–0.0571	0.404199	5.268337
V	10.440625	446	1.283663	–0.34302	1.626685	27.1888
VI	11.598125	112	–0.2243	0.025421	–0.24972	–2.51224
VII	15.499375	100	–0.6296	0.080478	–0.71008	–6.79212
VIII	14.03625	239	0.347507	–0.06814	0.415643	5.7412

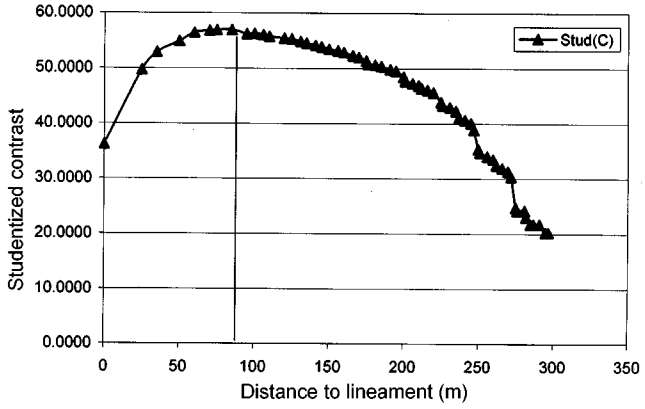

Figure 3. Plot of cumulative incremental distance to lineament against its studentized contrast.

model. Being an active tectonic terrain, which is continuously undergoing uplifting, former riverbeds of Bhagirathi occur as terraces and are mapped at many places in the study area.

Slope and slope aspect also play a crucial role in governing the stability of the terrain in the Himalayan region. As the slope increases, the chances of failure also increase. It has been observed in the study area that slopes steeper than 60° usually stand out with barren-rock exposures and are more susceptible to rock falls than landslides. At many places in the study area, the slopes are in the range of 35–60° and are found to be prone to landslides. This may be because of the fact that the colluvial accumulations are formed around 35° of slope. Slope aspect is another factor influencing the stability of a terrain. In the Himalaya, it has been found that the southern slopes are usually drier and are devoid of much vegetative cover, whereas the northern slopes are more moist and support better growth of vegetation. In the present study, slope aspect has also been considered as an independent variable in building the model.

Using the DEM as input, the slope and aspect have been derived. The slope values vary from 1 to 70°. These have been classified into five classes (Class I, < 5°; Class II, 5 to 15°; Class III, 15 to 35°; Class IV, 35 to 60° and Class V, > 60°) to get a classified slope map. The slope aspect values have also been grouped into eight classes (N, NE, E, SE, S, SW, W and NW) to create a classified aspect map. The weights estimated for the slope and aspect classes are given in Tables 5 and 6, respectively.

Drainage: The presence of streams greatly influences the stability by toe erosion or by saturating the slope material or both[20]. In this context, the distance to drainage and drainage density have been considered as influential evidential themes.

The drainage pattern is sub-dendritic to sub-parallel in the area. The drainage map has been prepared from the toposheet and updated with the help of the merged LISS III and PAN satellite data. The drainage density has been found to vary between zero and 95 km per sq. km and has been further classified into five classes (Class I, < 15; Class II, 15 to 25; Class III, 25 to 35; Class IV, 35 to 50 and Class V, > 50 km/sq. km) to get a classified drainage density map. The estimated weights for these classes are given in Table 7. The distance to drainage theme which is a continuous one has been converted to a binary pattern

Table 7. Weights and contrast values for drainage density classes

Drainage density class	Area (sq. km)	No. of slide points	w^+	w^-	Contrast	Stud (C)
I	15.0000	64	–1.0445	0.106204	–1.15071	–8.95052
II	21.0000	56	–1.5155	0.184939	–1.70045	–12.4205
III	32.0000	462	0.1809	–0.09442	0.275301	4.661098
IV	28.0000	652	0.6645	–0.43428	1.098768	19.17637
V	6.0000	0	–8.8949	0.060928	–8.95587	–0.89558

Table 8. Weights and contrast values for distance to drainage

Distance to drainage (m)	Area (sq. km)	No. of slide points	w^+	w^-	Contrast	Stud (C)
< 75	54.4794	858	0.279518	–0.44241	0.721925	11.63522
> 75	47.5206	376	–0.44241	0.279518	–0.72193	–11.6352

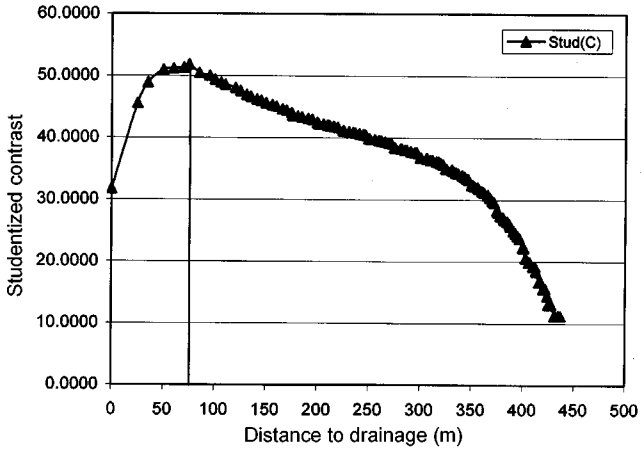

Figure 4. Plot of cumulative incremental distance to drainage against its studentized contrast.

with the help of the plot of studentized contrast against distance to drainage (Figure 4). Here the threshold has been found as 75 m. The weights have been recalculated for this binary theme and are given in Table 8.

Land use/land cover: The land use/land cover map of the area has been prepared with the help of IRS LISS III data. The cell size is kept at 25 m × 25 m. Supervized classification using maximum likelihood estimation method has been used for classifying the LISS III data with appropriate training sets for different classes. Post classification contextual refinement has been applied on the classified data to get the final land use/land cover map. Dense forest, open forest, scrub, barren land and agriculture are the land use/land cover classes in the study area.

Presence of vegetation is crucial in slope stability due to better bonding of the slope material. Thus slopes with dense vegetation should be less prone to the occurrence of shallow landslides than barren slopes, while all other factors remain constant. In order to assess the contribution of different land use/land cover classes to slope destabilization, this factor has also been taken into account. The estimated weights for different land use/land cover classes are given in Table 9.

Distance to road: It has been observed that many landslides occur close to the road. It is possible that the slope destabilization has been caused either by the uncontrolled or controlled blasting and widening of the roads, or by the loss of support due to removal of material from the lower portion of the slopes during road construction. In order to accommodate the effect of this anthropogenic activity, distance to road has been taken as an evidential theme. The continuous data have been converted to a binary pattern using the plot of distance to road against the studentized contrast. The cut-off is taken at 500 m where the studentized contrast is found to be maximum (Figure 5). The recalculated weights for the binary pattern of distance to road is given in Table 10.

Analysis and interpretation

The evidential themes can be integrated together to find the combined influence of the different input parameter classes. The posterior probability of each unique combination of input parameter classes can then be estimated, provided the themes are conditionally independent with respect to the occurrence of landslides. The evidential themes have been combined pair-wise and the conditional independence (CI) has been tested in each pair using the Omnibus Test, which is the ratio of observed number of slide points (n) to the predicted number of slide points (T). The predicted number of slided cells can be estimated as:

$$T = \sum_{k=1}^{m} P_k N(S)_{\text{observed}},$$

Table 9. Weights and contrast values for land use/land cover classes

Land use/ land cover	Area (sq. km)	No. of slide points	w^+	w^-	Contrast	Stud (C)
Dense forest	24.6775	60	–1.6080	0.2282	–1.8363	–13.8605
Open forest	50.7762	539	–0.1290	0.1130	–0.2421	–4.2018
Barren land	15.1962	239	0.2673	–0.0548	0.3221	4.4504
Scrub	5.6231	247	1.3123	–0.1679	1.4803	20.5624
Agriculture	5.1156	149	0.8920	–0.0779	0.9699	11.0079
Water	1.5194	0	–7.5215	0.0151	–7.5366	–0.7537

Table 10. Weights and contrast values for distance to road

Distance to road (m)	Area (sq. km)	No. of slide points	w^+	w^-	Contrast	Stud (C)
< 500	18.8081	853	1.3452	–0.9767	2.321906	37.47587
> 500	83.4575	381	–0.9767	1.3452	–2.32191	–37.4759

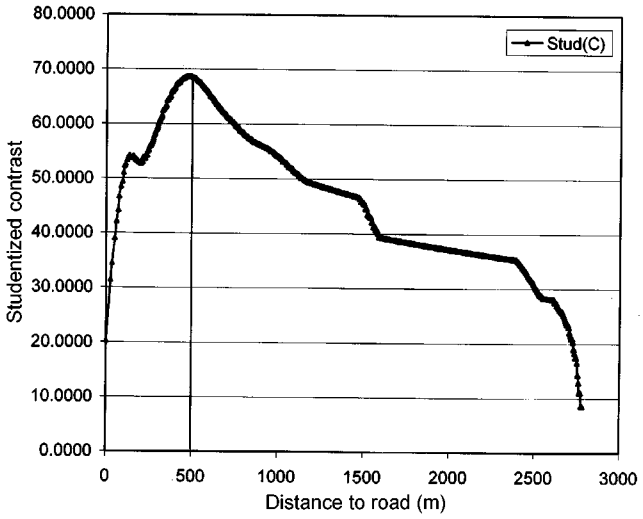

Figure 5. Plot of cumulative incremental distance to road against its studentized contrast.

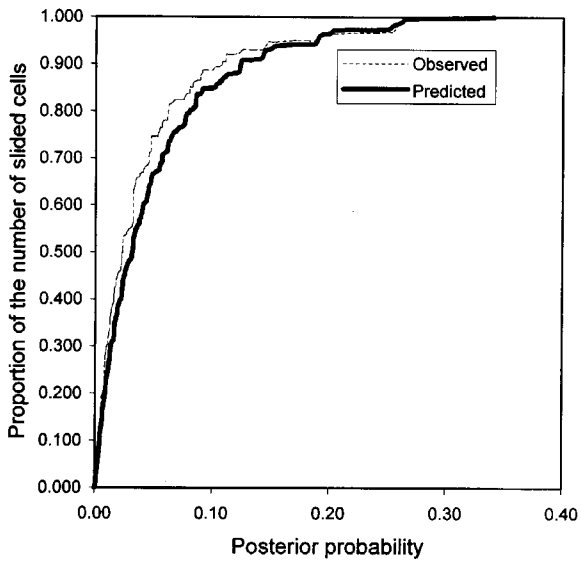

Figure 6. Plot of observed and predicted cumulative frequencies of landslide cells against posterior probability.

where P_k is the posterior probability of the kth unique condition.

If n/T is less than 0.85, then CI is violated. CI values for the pairs of evidential themes are given in Table 11. It can be inferred from Table 11 that the drainage density and lineament density themes show problem of conditional dependency with one of the themes (distance to road) and hence have been omitted from the final model. Thus the final model for hazard zonation retains seven of the nine themes (distance to drainage, distance to lineament, distance to road, lithology, slope aspect, land use/land cover and slope). Posterior probability has been estimated for the unique conditions resulting out of the combination of these seven themes. This seven-parameter model has a CI ratio of 0.87, which is above the acceptable limit of CI, indicating that the model satisfies the condition of CI among the parameters. There are 2317 unique combinations and the posterior probability values for these unique combinations vary from 0 to 0.34.

The observed and predicted cumulative frequencies of slided cells are plotted against the posterior probability (Figure 6). The cumulative posterior probability percentage values have been plotted against cumulative percentage of the area (Figure 7) to decide the boundary values for classifying the posterior probability values. The slope of the curve turns from steep to moderate and from moderate to flat at the inflection points, where the posterior probability values are 0.0088 and 0.0343 respectively. These values have been used to classify the posterior probability values into low, medium and high hazard classes. The resultant LHZ map showing locations of the existing landslides is given in Figure 8.

Validation

The model has been validated by comparing the calculated probability values for different cells and their actual

Table 11. Pair-wise conditional independence test values using Omnibus Test

Evidential theme	Distance to drainage (m)	Distance to lineament (m)	Distance to road (m)	Lithology	Slope aspect	Drainage density	Lineament density	Land use/ land cover	Slope
Distance to drainage (m)		0.954	0.973	0.986	0.985	0.966	0.982	0.978	0.997
Distance to lineament (m)	0.954		0.895	0.977	0.996	0.931	0.848	0.996	0.994
Distance to road (m)	0.973	0.895		0.970	0.869	(*0.809*)	(*0.695*)	0.857	0.979
Lithology	0.986	0.977	0.970		0.985	0.942	0.904	0.954	0.994
Slope aspect	0.985	0.996	0.869	0.985		0.928	0.915	0.864	0.992
Drainage density	0.966	0.931	(*0.809*)	0.942	0.928		(*0.809*)	0.958	0.993
Lineament density	0.982	0.848	(*0.695*)	0.904	0.915	(*0.809*)		0.948	0.991
Land use/land cover	0.978	0.996	0.857	0.954	0.864	0.958	0.948		0.987
Slope	0.997	0.994	0.979	0.994	0.992	0.993	0.991	0.987	

Numbers in italics given in parentheses show violation of conditional independence.

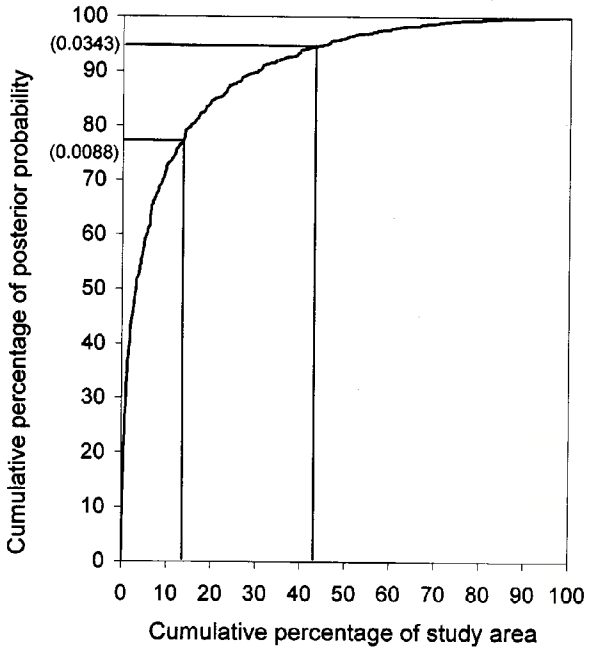

Figure 7. Plot of cumulative percentage of posterior probability against cumulative percentage of the study area.

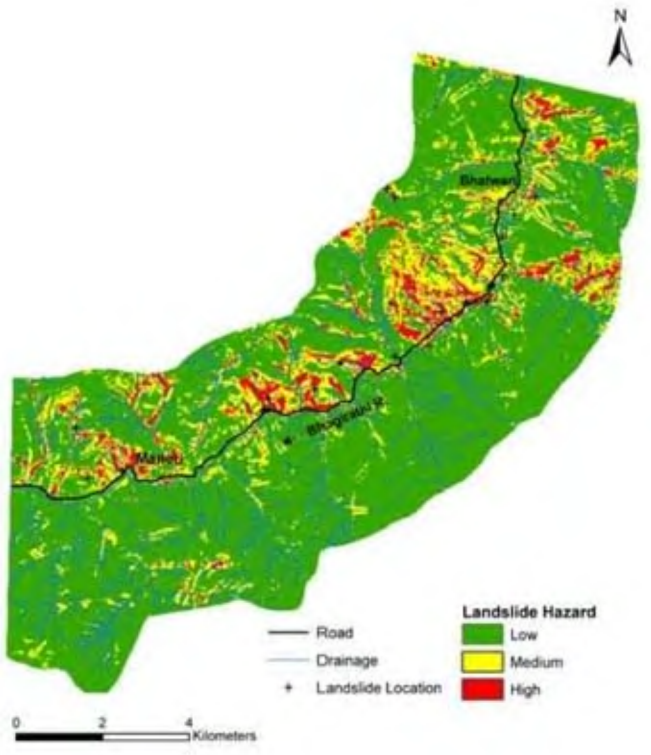

Figure 8. Landslide hazard zonation map of the study area.

present condition. This is achieved using Receiver Operator Characteristic (ROC) curve analysis[21,22]. The ROC curve is a plot of the probability of true positive identified landslides versus that of false positive identified landslides, as the cut-off probability varies. Equivalently, it is a representation of the trade-off between sensitivity and specificity. Sensitivity is the probability that a slided cell is correctly classified, and is plotted on the *y*-axis in an ROC curve. 1-sensitivity is the false negative rate. Specificity is the probability that a non-slided cell is correctly classified. 1-specificity is the false positive rate and is taken along the *x*-axis of the curve. The area under the curve represents the probability that the model-calculated landslide susceptibility value for a randomly chosen slided cell would exceed the result for a randomly chosen non-slided cell. Thus, the area under the ROC curve can be used as a measure of the accuracy of the model.

The ROC curve for the model developed is given in Figure 9. The area under the curve is 0.846 (Table 12), which gives an accuracy of 84.6% for the model developed using weights of evidence. The asymptotic significance is less than 0.05, which means that using the model to predict the landslide is better than guessing.

Discussion and conclusion

The present study demonstrates the application of weights of evidence modelling for landslide susceptibility mapping in part of the Himalaya, which is prone to frequent occurrence of landslides. Remote sensing and GIS have been useful in data preparation and at integration stages.

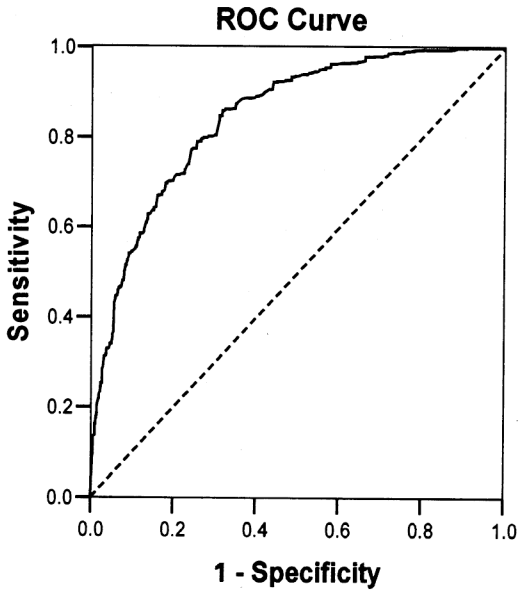

Figure 9. Receiver operator characteristic curve of the developed model. Diagonal segments are produced by ties.

Table 12. Area under the receiver operator characteristic curve

Area	Standard error[a]	Asymptotic significance[b]
0.846	0.008	0.000

Test result variable: Posterior probability. The test result variable has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.
[a]Under the nonparametric assumption.
[b]Null hypothesis: true area = 0.5.

Positive and negative weights and contrast values have been calculated for various classes of parameters used in the study. The contrast values have been used to find the threshold value for the binary conversion of continuous variables like distance to road, distance to drainage and distance to lineament.

Amongst the different lithological classes, gneisses/migmatites/schists have maximum value for studentized contrast. Thus the area occupied by Higher Himalayan crystallines (gneisses/migmatites/schists) shows maximum susceptibility with reference to landslides in the study area. This can be clearly seen as one traverses along the road to Gaumukh. After the MCT, there is widespread occurrence of landslides in these lithologies.

The presence of planar structural features/lineaments greatly reduces the strength of rock-mass units and they contribute positively in making the slopes unstable. It is observed widely in the Himalaya that close to regional thrusts/faults (e.g. MCT), the landsliding phenomenon is severe. This observation is corroborated by the high values of contrast for Class V of lineament density and for the distance to lineament class of < 85 m. The drainage density and distance to drainage parameters also have shown positive influence towards slope destabilization as seen from the contrast values. Slope saturation might be the reason for this phenomenon. Amongst the different land use/land cover categories, scrub land area has shown highest contrast and this area is more prone to landslides due to lack of cohesion of slope material, as in case of dense forest areas. Agricultural areas also shows high contrast values. This could be because of the fact that cultivation has been done on stabilized, old landslides, thereby giving high positive weights in such areas.

The slope category of 35–60° (slope Class IV) and slope aspects of south and northwest are found to be significant contributors with respect to landslides in the study area. The 35–60° slope category is the most unstable in the study area, where most of the colluvial accumulations are found around 35° slope angle. On the other hand, slopes gentler than 30° are found to be stable in the area. It has also been observed that steep slopes (> 60°) stand out with rock exposures and are mostly stable, provided they are devoid of any geological discontinuities. In a major portion of the area, the southern and northwestern aspects receive maximum sunlight because of the predominant ENE–WSW orientation of the main valley and consequent orientations of the tributary valleys. This creates drier slopes in these aspects and results in lesser vegetation growth. Thus these aspects tend to be unstable. Anthropogenic interferences such as slope-cutting for the construction of roads and other developmental activities make the slopes unstable. The 500 m buffer area around the road has shown maximum value for contrast and hence attests to this observation.

The classified posterior probability map depicts 76.8% of the total study area in the low, 17.5% in the medium and 5.6% in the high hazard class. ROC curve analysis has shown 84.6% accuracy for the LHZ analysis model adopted in the present study. Thus weights of evidence modelling can be utilized for estimating the conditional probability of landsliding on a cell-by-cell basis for an area, given the presence or absence of various independent variables which influence slope stability, towards the LHZ mapping.

1. Anbalagan, R., Landslide hazard evaluation and zonation mapping in mountainous terrain. *Eng. Geol.*, 1992, **32**, 269–277.
2. Ayalew, L. and Yamagishi, H., The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda–Yahiko Mountains, Central Japan. *Geomorphology*, 2005, **65**, 15–31.
3. Keinholz, H., Maps of geomorphology and natural hazards of Grindelwald, Switzerland, scale 1 : 10000. *Arct. Alp. Res.*, 1978, **10**, 169–184.
4. Rupke, J., Cammeraat, E., Seijmonsbergen, A. C. and van Westen, C. J., Engineering geomorphology of the Widentobel catchment, Apenzell and Sankt Gallen, Switzerland: A geomorphological inventory system applied to geotechnical appraisal of slope stability. *Eng. Geol.*, 1988, **26**, 33–68.

5. Gee, M. D., Classification of landslide hazard zonation methods and a test of predictive capability. In Proceedings of the 6th International Symposium on Landslides, Vol. 2, Christchurch, New Zealand, 1992, pp. 947–952.
6. Pachauri, A. K. and Pant, M., Landslide hazard mapping based on geological attributes. *Eng. Geol.*, 1992, **32**, 81–100.
7. Gupta, R. P., Saha, A. K., Arora, M. K. and Kumar, A., Landslide hazard zonation in part of Bhagirathi Valley, Garhwal Himalaya, using integrated remote sensing – GIS. *Himalayan Geol.*, 1999, **20**, 71–85.
8. Saha, A. K., Gupta, R. P. and Arora, M. K., GIS-based landslide hazard zonation in the Bhagirathi (Ganga) valley, Himalayas. *Int. J. Remote Sensing*, 2002, **23**, 357–369.
9. NRSA, Landslide hazard zonation mapping along the corridors of the pilgrimage routes in Uttaranchal Himalaya, Technical document, NRSA, Department of Space, India, 2001.
10. van Westen, C. J., *Application of Geographic Information Systems to Landslide Hazard Zonation*, ITC Publication, ITC, Enschede, The Netherlands, 1993, vol. 15.
11. Bonham-Carter, G. F., *Geographic Information Systems for Geoscientists: Modelling with GIS*, Pergamon Press, Oxford, 1994.
12. Bonham-Carter, G. F., Agterberg, F. P. and Wright, D. F., Integration of geological datasets for gold exploration in Nova Scotia. *Photogramm. Eng. Remote Sensing*, 1988, **54**, 1585–1592.
13. Agterberg, F. P., Systematic approach to dealing with uncertainty of geoscience information in mineral exploration. In Proceedings of the 21st APCOM Symposium, Las Vegas, 1989, pp. 165–178.
14. Agterberg, F. P., Bonham-Carter, G. F. and Wright, D. F., Statistical pattern integration for mineral exploration. In *Computer Applications in Resources Estimation Prediction and Assessment for Metals and Petroleum* (eds Gaal, D. F. and Merriam, D. F.), Pergamon Press, New York, 1990, pp. 1–21.
15. Bonham-Carter, G. F. and Agterberg, F. P., Application of a microcomputer based geographic information system to mineral potential mapping. In *Microcomputer Based Applications in Geology* (eds Hanley, T. and Merriam, D. F.), Pergamon Press, Oxford, 1990, vol. 2.
16. Porwal, A., Carranza, E. J. M. and Hale, M., Extended weights of evidence of modelling for predictive mapping of base metal deposit potential in Aravalli Province, Western India. *Explor. Min. Geol.*, 2001, **10**, 273–287.
17. Carranza, E. J. M., Weights of evidence modelling of mineral potential: A case study using small number of prospects, Abra, Philippines. *Nat. Resour. Res.*, 2004, **13**, 173–187.
18. Bonham-Carter, G. F. and Agterberg, F. P., Arc-WofE: A GIS tool for statistical integration of mineral exploration datasets. In Proceeding of the International Statistical Institute, 1999, pp. 497–500.
19. Agarwal, N. C. and Gopendra Kumar, Geology of the Upper Bhagirathi and Yamuna Valleys, Uttarkashi District, Kumaun Himalaya. *Himalayan Geol.*, 1973, **3**, 1–23.
20. Gokceoglu, C. and Aksoy, H., Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image process techniques. *Eng. Geol.*, 1996, **44**, 147–161.
21. Zweig, M. H. and Campbell, G., Receiver Operator Characteristic plots: A fundamental evaluation tool in clinical medicine. *Clin. Chem.*, 1993, **39**, 561–577.
22. Hanley, J. A. and McNeil, B. J., The meaning and use of the area under a receiver operator characteristic (ROC) curve. *Radiology*, 1982, **143**, 29–36.

ACKNOWLEDGEMENTS. J.M. and V.K.J. thank Dr V. Jayaraman, Director, NNRMS/EOS, Department of Space, for support and encouragement to carry out the study. J.M. and G.S.R. thank the Head, RRSSC, Dehradun, for providing the infrastructure to execute the project. J.M. thanks Dr G. Philip, Wadia Institute of Himalayan Geology, Dehradun, for guidance during fieldwork. We thank Dr G. F. Bonham-Carter, Geological Survey of Canada, for providing the software for weight calculation and Dr E. J. M. Carrannza, ITC, The Netherlands, for providing necessary literature on the topic.

Received 27 February 2006; revised accepted 10 September 2006