Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2016, Article ID 1612901, 12 pages
http://dx.doi.org/10.1155/2016/1612901

Research Article

Application of the Multitype Strauss Point Model for Characterizing the Spatial Distribution of Landslides

Iswar Das[1] and Alfred Stein[2]

[1]*National Remote Sensing Centre, Department of Space, Government Of India, Hyderabad 500 037, India*
[2]*Department of Earth Observation Science, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE Enschede, Netherlands*

Correspondence should be addressed to Iswar Das; das@alumni.itc.nl

Received 16 July 2015; Revised 30 March 2016; Accepted 20 April 2016

Academic Editor: Filippo Ubertini

Landslides are common but complex natural hazards. They occur on the Earth's surface following a mass movement process. This study applies the multitype Strauss point process model to analyze the spatial distributions of small and large landslides along with geoenvironmental covariates. It addresses landslides as a set of irregularly distributed point-type locations within a spatial region. Their intensity and spatial interactions are analyzed by means of the distance correlation functions, model fitting, and simulation. We use as a dataset the landslide occurrences for 28 years from a landslide prone road corridor in the Indian Himalayas. The landslides are investigated for their spatial character, that is, whether they show inhibition or occur as a regular or a clustered point pattern, and for their interaction with landslides in the neighbourhood. Results show that the covariates lithology, land cover, road buffer, drainage density, and terrain units significantly improved model fitting. A comparison of the output made with logistic regression model output showed a superior prediction performance for the multitype Strauss model. We compared results of this model with the multitype/hard core Strauss point process model that further improved the modeling. Results from the study can be used to generate landslide susceptibility scenarios. The paper concludes that a multitype Strauss point process model enriches the set of statistical tools that can comprehensively analyze landslide data.

1. Introduction

Landslides are defined as the movement of a mass of rock, debris, or soil along a downward slope, due to gravitational pull. The inherent properties of the Earth material, encompassing various geoenvironmental factors, can make a particular area susceptible to landslides. Landslides are among the most common natural hazards. They exhibit themselves in different mass movement processes and are considered as complex natural hazards occurring on the Earth's surface [1]. Although individual landslides are not as spectacular or damaging as earthquakes, floods, and hurricanes, they are widespread and frequently occurring. Over the years they have caused great loss of life and property and their effects on the economy are growing at a rapid pace in many countries.

Spatial zonation of landslide occurrences, also known as landslide susceptibility mapping, aims to differentiate a land surface into homogeneous areas according to their degree of failure caused by mass movement at specific locations [2]. It relies on understanding complex mass movement processes and their controlling factors [3]. Approaches to the spatial modeling of landslides can broadly be divided into two groups [4]. The first approach consists of deterministic, dynamic modeling of the physical mechanisms that control slope failure, using mathematical methods. This approach is highly localized because of the detailed data requirements. The second approach uses the relation between the locations of previous landslides and geoenvironmental variables, to predict areas of different landslide susceptibility, using heuristic or statistical methods. The statistical methods used successfully in landslide susceptibility mapping to date include discriminant analysis [5, 6], multivariate statistics [7], likelihood ratio [8], information value method [9], and logistic regression methods [10]. These methods

allow the analysis of geoenvironmental variables controlling landslide occurrence with respect to previous landslides without looking at the mutual interactions of landslides and their distribution patterns. Commonly applied generalized linear modeling uses a maximum likelihood estimation that results into point parameter estimates with standard errors [11]. As individual landslides cover only a small fraction of an unstable area, landsliding can be considered as a spatial point process that is controlled by a number of surface and subsurface spatial variables.

A spatial point process underlies a pattern of spatial point data within a region. Spatial point patterns are characterized by the 1st- and 2nd-order effects of a point process [12], specifying the intensity and point interactions, respectively. Nearest neighbor functions between pairs of points commonly model the 2nd-order effects as a function of positions of points and their distances. Such functions usually consider the relative position of two points in a bounded region and are taken as a function of distance only [13]. Spatial point processes play a fundamental role in spatial statistics and exhibit an active area of research. Disciplinary applications occur in forestry addressing positions of trees in a forest and log-landing sites [13–15], ecology addressing locations of bird's nests [13], seismology addressing earthquake epicentres [16, 17], astrophysics addressing locations of stars in a constellation [18], and environmental modeling for peak concentrations of a pollutant in a geographical region [19]. In this study we use a spatial point process model to analyze the landslide data for inferring properties of the spatial distribution pattern.

The objective of this study was to identify significant factors for landslide susceptibility by applying a multitype Strauss point model. In this way the landslide occurrence patterns could be better understood. The point process model used information at the level of detail offered by the landslide data. That information was combined with geoenvironmental covariates to explain the underlying process for determining landslide susceptibility within an area. The model output was compared with the output of logistic regression model. The study was applied to a landslide prone area in the northern Himalayas, India.

2. Methods

A landslide distribution pattern can be considered as a collection of n point data spread irregularly in an essentially planar region. A basic assumption for the analysis is that the data can be regarded as a realization of a stochastic point process [13]. This process is characterized by the intensity at location x, $N(x)$, defined as

$$N(x) = \lim_{dx \to 0} \left\{ \frac{E(n(dx))}{|dx|} \right\}, \tag{1}$$

where $|dx|$ is the area of a small region dx, E is expectation operator, and $n(dx)$ is number of points in the region dx.

Landslide occurrences, considered as a spatial point pattern, show variation in the relative frequency as a function of the distance between positions

$$\gamma(x_i, x_j) = \lim_{dx_i, dx_j} \left\{ \frac{E\left(n(dx_i)\, n\left(dx_j\right)\right)}{\left|dx_i\right| \left|dx_j\right|} \right\}. \tag{2}$$

Assuming stationarity, (2) depends only upon the relative position of the two landslides between positions x_i and x_j, that is, on their distance [13]. A landslide process is second-order stationary if its intensity is independent upon translation over R, so that $N(dx) = N$ and the second-order intensity

$$\gamma(x_i, x_j) = \gamma\left(dx_i - dx_j\right) = \gamma(d) \tag{3}$$

depends only upon the distance vector d between x_i and x_j and not on their locations.

2.1. Conditional Intensity and the Gibbs Model. For a landslide process that exhibits inter-landslide interaction the pairwise interaction models define the intensity in the form of probability densities:

$$p(x) = \alpha \left[\prod_{i=1}^{n(x)} b(x_i) \right] \left[\prod_{i<j} c\left(x_i, x_j\right) \right], \tag{4}$$

where α is a normalizing constant, $b(x)$, $x \in W$ is the intensity or the first-order term, and $c(x_i, x_j)$, $x_i, x_j \in W$, is the pairwise interaction or second-order term in a bounded window W. Pairwise interaction models are a special case of Markov point process, called Gibbs point process models. Analysis of a Gibbs point process model is based on its conditional intensity [20]. The probability of the occurrence of a landslide $N(x)$ at location x is determined by the conditional intensity defined by $N(x, u)$. For a landslide process in a bounded area W the conditional intensity is related to the probability density p by

$$N(x, u) = \frac{p(u \cup \{x\})}{p(u)}. \tag{5}$$

For the general pairwise interaction process the conditional intensity is

$$N(x, u) = b(x) \prod_{i=1}^{n(u)} c(x_i, u). \tag{6}$$

Landslides of a particular zone with a specified radius of influence, however, might have a homogeneity condition with respect to each other. A Strauss process emphasizes such homogeneity conditions for deriving the relationship among the events. For the Strauss process, a simple model of dependence between landslides has the conditional intensity

$$N(x, u) = \beta \gamma^{t(x,u)}, \tag{7}$$

where $t(x, u)$ is the number of points of the landslides u that lie within a distance r of the location x. The conditional

intensity is useful in modelling because the two distinct components in the functional form represent the interaction of landslides that can be interpreted in a straightforward way, following Baddeley [20].

A major restriction of Gibbs models is that the parameters cannot be estimated using maximum likelihood estimation, and hence a maximum pseudolikelihood is returned for each model [20]. The unknown scaling factor α is intractable; hence the calculated pseudolikelihood does not involve any unknown factor and it is easier to use when estimating the parameters.

2.2. Nearest Neighbour G-Function. Methods based upon the distances between landslides can be used for investigating inter-landslide interactions, for example, to identify second-order effects of the landslide pattern data. These second-order properties are specified by the pair correlation function that is assessed using the inter-landslide interaction methods like the K-and G-functions [20]. In this way, the nature of the departure from complete spatial randomness (CSR) can be identified. This, in turn, is useful in determining the kind of interaction and interaction distances between the landslides [21]. Alternatively, Ripley's K-function can be used for detecting deviations from spatial homogeneity. The shape of this function indicates the specific type of pattern displayed by the data, that is, indicating whether the landslides show inhibition or occur as a regular or a clustered pattern.

The G-function quantifies the distance distribution of a landslide to the nearest other landslide. It is expressed as

$$\widehat{G}(r) = \sum_{i=1} e(x_i, r)\, I_{\{t_i \le r\}}, \tag{8}$$

where r is the radius of a disk centered at the location of the ith landslide x_i, $e(x_i, r)$ is an edge correction weight such that $\widehat{G}(r)$ is approximately unbiased, $t_i = \min_{j \ne i} \|x_i - x_j\|$ is the distance of each landslide location to its nearest neighbor, and $I_{\{t_i \le r\}}$ is the indicator function equal to 1 if $t_i \le r$ and it is 0 otherwise.

An estimate of G derived from a spatial landslide pattern dataset can be used in exploratory data analysis and formal inference about the pattern [22, 23]. The shape of this function provides information about the way the landslides are clustered in a particular area. If the landslides are clustered, G increases rapidly at short distances, whereas for landslides that are evenly spaced, G increases slowly up to the distance at which most events are spaced, and only then it increases rapidly [24]. For a homogeneous Poisson point process of intensity λ, the nearest neighbor distribution distance function for landslide distribution equals

$$G_{\text{pois}}(r) = 1 - \exp\left(-\lambda \pi r^2\right). \tag{9}$$

If $G(r) > G_{\text{pois}}(r)$ then the landslide pattern is clustered, whereas if $G(r) < G_{\text{pois}}(r)$ the landslide pattern is regular.

2.3. Strauss Point Process for Marked Point Pattern Analysis of Landslides. A multitype pairwise interaction process is a Gibbs process which assumes symmetric interactions of landslides have the probability density of the form

$$p(x) = \alpha \left[\prod_{i=1}^{n(x)} b_m(x_i)\right] \left[\prod_{i<j} c_{m,m'}\left(x_i, x_j\right)\right], \tag{10}$$

where α is a normalizing constant, $b_m(x_i)$ is a function determining the first-order trend for landslides $n(x)$ of the mth type of patterns, indicating whether the landslides show inhibition or occur as a regular or a clustered pattern, and $c_{m,m'}(x_i, x_j)$ are symmetric functions that describe the interaction between a pair of landslides x_i and x_j of given types m and m'; that is, $c_{m,m'}(x_i, x_j) = c_{m,m'}(x_j, x_i)$ and $c_{m,m'} \equiv c_{m',m}$ [20]. Thus, the conditional intensity of two categories of landslides *Small* and *Large* as defined in (11) for a multitype Strauss process is given by

$$\lambda((u, m); x) = b_m(u) \prod_{i,j=1}^{n(x)} c_{m,m'}\left(x_i, x_j\right). \tag{11}$$

The multitype Strauss process has pairwise interaction terms

$$c_{m,m'}\left(x_i, x_j\right) = \begin{cases} 1 & \text{if } \left\|x_i - x_j\right\| > r_{m,m'} \\ \gamma_{m,m'} & \text{if } \left\|x_i - x_j\right\| \le r_{m,m'}, \end{cases} \tag{12}$$

where $r_{m,m'} > 0$ are interaction radii as above and $\gamma_{m,m'} \ge 0$ are interaction parameters [20].

To fit a multitype Strauss process model to landslides, the matrix of interaction radii $r_{m,m'}$ between individual landslides is specified based on field conditions and distribution patterns. Model fitting generates values for the interaction parameter $\gamma_{m,m'}$ and the model coefficients describing spatial inhomogeneity and inter-landslide interactions.

2.4. The Multitype/Hard Core Strauss Point Process Model. The multitype/hard core Strauss point process is a hybrid of the multitype Strauss process and the hard core process, for example, the case that $\gamma_{m,m} = 0$, $\gamma_{m,m'} = 0$, or $\gamma_{m',m'} = 0$ of the Strauss process. A pair of points of types m (*Small*) and m' (*Large*) must not lie closer than $h_{m,m'}$ units apart; if the pair lies more than $h_{m,m'}$ and less than $r_{m,m'}$ units apart, it contributes a factor $\gamma_{m,m'}$ to the probability density similar to the m, m pairs and the m', m' pairs. For landslides, this extension makes sense, as the large landslides are usually well separated and small distances do not occur; the same applies to the smaller landslides, although to a lesser degree as they can be closer. Moreover, a small landslide does not occur below a large one, whereas, if the location of a landslide is indicated as a point, the landslides cannot occur immediately close to each other.

2.5. Goodness of Fit. Akaike's Information Criterion (AIC) is a measure of the goodness of fit of an estimated statistical model. The AIC is defined as $\text{AIC} = -2\ln(L) + 2k$, where L is the maximum likelihood value for the estimated model and k is the number of parameters in the model.

It can be interpreted as the trade-off between bias and variance in model construction indicating that of accuracy and complexity of the model [25].

The study system is based on the point process modeling of landslide data along with the geoenvironmental covariates influencing landslide such as lithology, topography, or geology. The models are selected based on the AIC. The study makes use of the spatial distribution of landslide to make the exploratory data analysis, their interactions using G- and K-functions, and the possible susceptibility intensity in the area using the Strauss point model. The AIC is a test between models and hence it may serve as a tool for model selection. Given a dataset, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best [25].

3. Site Characteristics and Data Description

3.1. Study Site. The study area lies between $30^\circ 47' 29''$N and $30^\circ 54' 45''$N latitude and $78^\circ 37' 41''$E and $78^\circ 44' 03''$E longitude in the northern Himalayas, India, in the catchment of the river Bhagirathi, a tributary of the river Ganges (Figure 1). This study area of a 12 km long road corridor with a total area of 8.88 km^2 was selected judiciously with corroboration that any landslide that occurs in the area affects the road. In the Himalayan terrain rock strength and geological structures play a major role in the landslide activity. The dominant rock types in the area include low grade metamorphic rock such as chlorite schist, schistose quartzite, and quartz mica schist along with high grade migmatites and gneisses. Rock mass properties, such as intact rock strength (IRS) computed for the area, varies between 50 and 200 MPa and corresponding cohesion of rock mass varies between 9 and 29 KPa [26]. Detailed assessment showed that the IRS varies due to compositional changes, the spacing and orientation of the joints present in these rocks, and the degree of weathering in each rock type. Elevation in the area ranges between 1550 and 2100 m with a high relative relief; average elevation of the area is around 1900 m.

The last three decades of rainfall information between 1982 and 2009 showed that the highest (1900 mm) and lowest (600 mm) annual rainfall occurred in years 2003 and 1991, respectively, with an annual average of approx. 1200 mm [27]. The area receives heavy precipitation during the summer months starting from mid of June to mid of October and moderate rainfall during the winter months from January to March. However, the rainfall is uniform in the area and no variation of rainfall was observed spatially in this road corridor. In the Himalayan region, landslides are recurring annually and are prominent during the summer months between June and October when the seasonal monsoon occurs. Landslides in this area were the result of a combination of geotectonics, adverse natural topography, such as steep slopes, weathered rocks and soils, human influences on the topography, and high rainfall [28, 29].

3.2. Spatial Data Quality. Spatial data quality plays an important role for the precise landslide identification and accurate landslide mapping. Landslide data were thus collected from reliable sources. The major organizations which keep the updated record of landslides in the Indian Himalayan terrain are Border Road Organization (BRO) and Geological survey of India (GSI). The historical landslide records of BRO during 1982–2009 was used in this study for preparing the inventory. A total of 178 active landslides were mapped at the 1 : 10,000 scales. Areas affected by these landslides were clearly recognizable from the remote sensing images. They were correlated with BRO records for the road corridor occurring along the cut slopes as well as in the natural slopes of the road corridor. The mapped landslides covered an area of 0.45 km^2, corresponding to 5.6% of the total area (min. 125 m^2, max. 41,000 m^2, median 1884 m^2, and mean 3967 m^2). For the point pattern analysis the centroids of landslide points were marked. Each landslide was attached with the attribute, that is, the area of the landslide. The area was a proxy for the area of influence of a landslide, directly reflecting its intensity. The landslide data were converted into the multitype marked point pattern by classifying the landslides into *Small* and *Large* according to their size. Categorization was done by considering the mean area of the landslides. Therefore, for our study any landslide with size less than 4000 m^2 (mean of the distribution) was *Small* and more than 4000 m^2 was *Large*. To investigate the nature of distribution of the landslides the nearest neighbor G-function was calculated for the landslide spatial patterns whether they follow a random, normal, or clustered distribution. A multitype Strauss model was fitted and the AIC was determined by applying the optimum interaction radii. To improve model fitting, various geoenvironmental factors like lithology, slope, aspect, lineament density, drainage density, weathering, soil depth, terrain units, road buffer, and land cover were considered as covariates [30]. Lithology, road buffer, terrain units, land cover, and drainage density were helpful in significantly reducing the AIC. The final susceptibility map was created by including these as explanatory variables for the landslide spatial pattern. In this study the Spatstat module was used in the R software for spatial statistical modeling [20].

4. Results

4.1. Distance Curves. The temporal distribution of *Large* and *Small* landslides in the area during 1982–2009 is shown in Figure 2(a). The area apparently contained more *Small* landslides in the early years than in recent years. This indicated that major landslides were subsequent to the occurrence of *Small* landslides as could be confirmed during the field survey. However, the occurrence of *Large* landslides could also be due to the regional geomorphological settings and modification of slopes in later years in this area. Figure 2(b) indicated the spatial distribution of the multitype point pattern of *Small* (represented by circle) and *Large* (represented by triangle) landslides in the road corridor resembling a clustered pattern.

The intensity was inhomogeneous throughout the road corridor as shown in Figure 3. The estimated average intensity of the landslide was 2.02×10^{-5} landslides m^{-2} with a

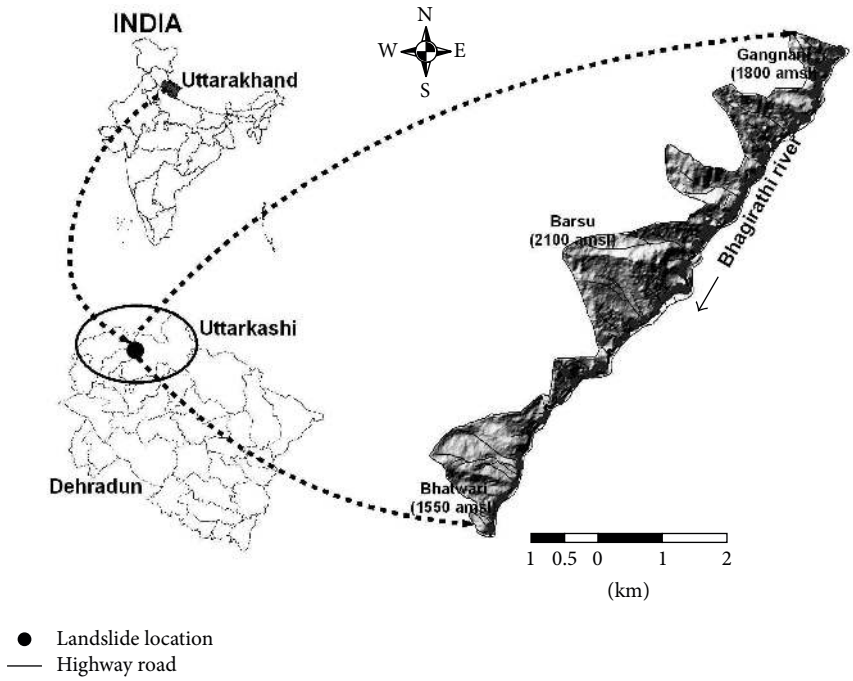

Figure 1: Location map of study area. The study is carried out in the National Highway Corridor of NH-108 between Bhatwari and Gangnani.

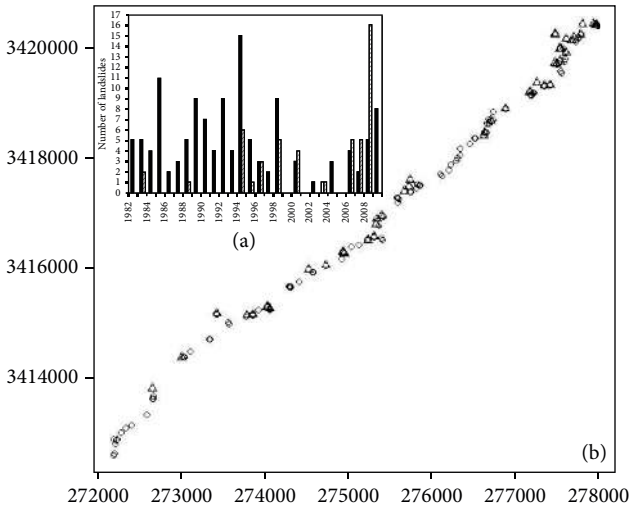

Figure 2: (a) Temporal distribution of *Small* (solid bar) and *Large* (dashed bar) landslides in the study area between 1982 and 2009 and (b) the spatial distribution of *Small* (represented by circles) and *Large* (represented by triangles) landslides along the road corridor with UTM coordinates.

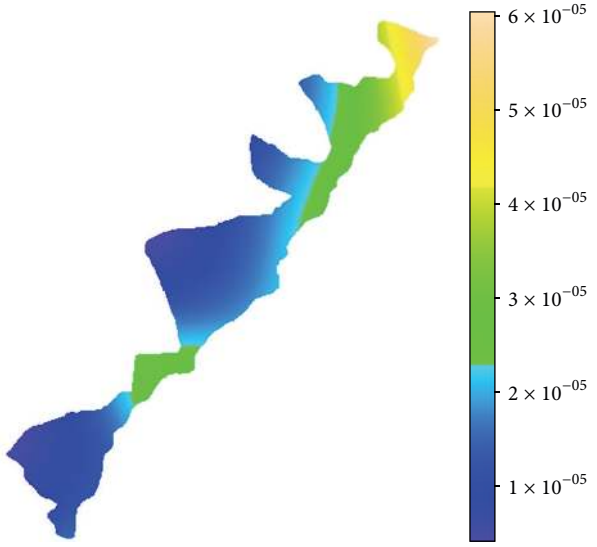

Figure 3: Landslide intensity for the road corridor. Average intensity equals 2.02×10^{-05} landslides per m^2.

maximum intensity of 6.00×10^{-5} landslides m^{-2} and a minimum intensity of 0.5×10^{-5} landslides m^{-2}. The single order intensity clarified that the area was prone to landslides at a given point of time though the intensity varied from south-east to north-west. High concentrations of landslides occurred in the north-eastern region.

For the second-order characterization of the landslide data the pairwise distance K-functions were plotted to ascertain the nature of distribution of the landslides. The simulated curves for different estimates of K with edge corrections were plotted well above the theoretical line indicating clustering of landslides (Figure 4). In addition, it was clear from Figure 4 that the *Small* landslides were spatially correlated to large landslides within the distance of 1 km.

Next, the G-functions were used for the multitype marked point patterns of landslides. Figure 5 presented a graphical interaction of the *Small* and *Large* landslides. Figure 5(a) showed the cumulative distribution of the nearest neighbor distance of each *Small* landslide from a landslide of the same

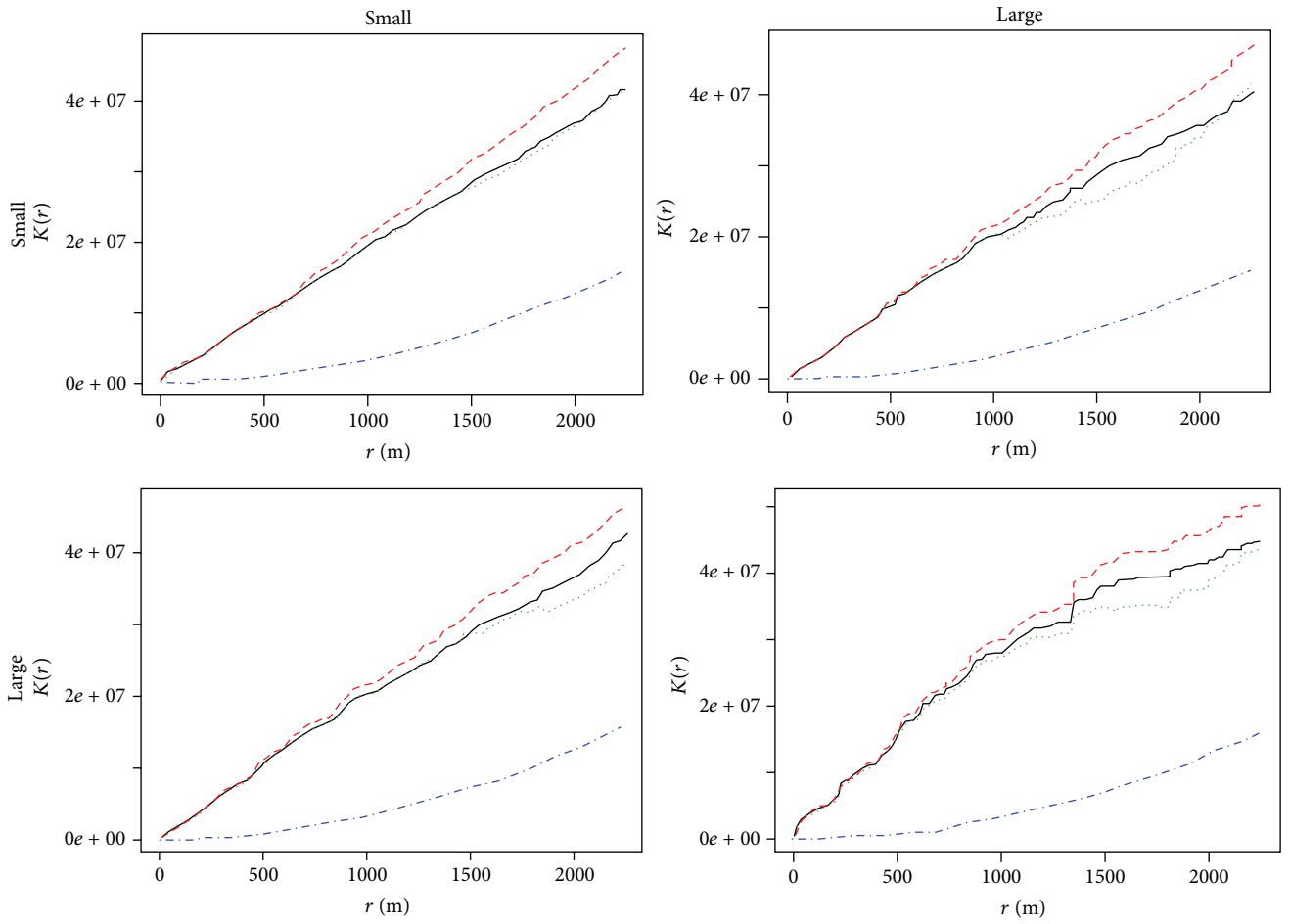

FIGURE 4: K-functions for interaction of *Small* and *Large* landslides with theoretical values for clustered distribution (blue line) lying below the simulated curves (black, green, and red lines).

type. The estimated curve for the nearest neighbor distances laid well above the curve of the Poisson process, indicating clustering. The maximum interaction radius as shown in the figure was 80 m, whereas a sharp rise of the G-function occurred at distances above 30 m. These indicated more landslides in a close vicinity. Figure 5(b) revealed the observed cumulative distribution pattern of the distances from *Small* landslide to its nearest neighbor landslide of type *Large*. Here also clustered pattern was observed as all the curves are above the curve of the Poisson process. Figure 5(c) showed clustering of the *Large* landslides around the *Small* landslide within distances of 80 m. Clustering of *Small* around *Large* landslides was different from the clustering pattern of *Large* around *Small* landslides. Apparently, the occurrence of *Small* landslides around and following large landslides differed from that of large landslides around and following *Small* landslides. *Large* landslides were tightly clustered around *Small* landslides.

Figure 5(d) indicated the clustered pattern of *Large* landslide as the plotted G-function is above the curve of the Poisson process. Many large landslides however were tightly clustered within radii of 80 m. Therefore, the G-function plotted in Figure 5 clearly indicated the clustering of the landslides along the road corridor.

4.2. Model Fitting. It is a common practice to determine the interaction radii by observing G-function values between the different pairs of points. There is no single optimal way to determine the interaction radii for the multitype Strauss point pattern that fits to a dataset [20]. The following matrix was determined interactively for the interaction radii that satisfy the best model fit:

$$r_{ij} = \begin{bmatrix} 10 & 19 \\ 19 & 105 \end{bmatrix}. \tag{13}$$

The values in (13) suggested that the distance for the optimum interaction of *Small* and *Large* landslides in the study area equaled 19 m. Estimation of the interaction radii is one of the requirements for fitting multitype Strauss point models. After trying the different Strauss model for different interaction terms, the model in which the intensity was log linear function of the Cartesian coordinate was selected as the appropriate one. Since landslide is a natural phenomenon of slope failure the underlying cofactors practically follow a log linear relationship with the landslide occurrence data. The model output shown in Figures 6(a) and 6(b) represented a log linear model with a multitype Strauss process of *Small* and *Large* landslides with AIC equal to 1416 (Table 1). Figures 6(a)

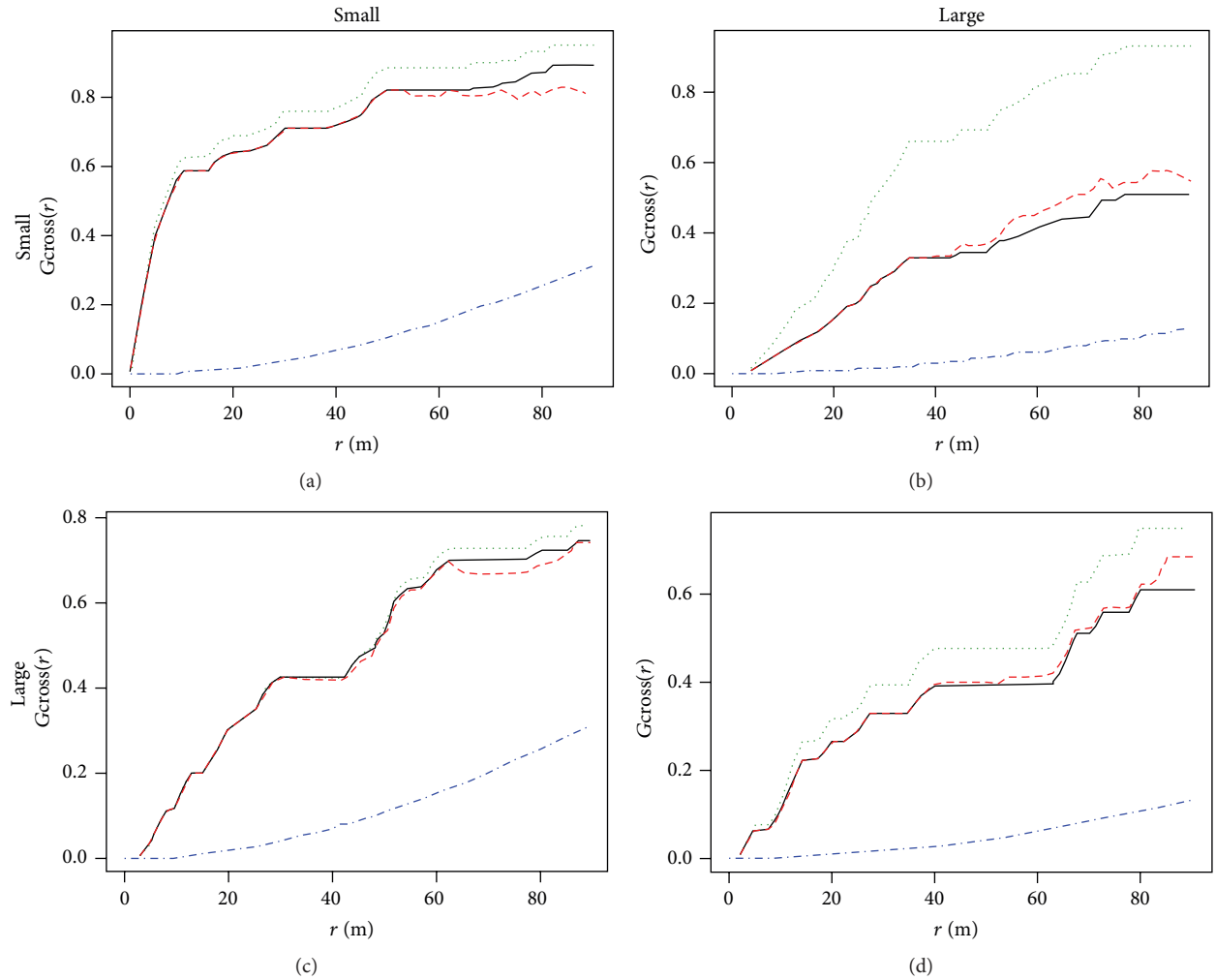

FIGURE 5: Nearest neighbor G-function for interaction of *Small* and *Large* landslides showing the distribution of landslide data. The theoretical values (blue line) lies well below the simulated curves (black, green, and red lines) indicating clustering of the data.

and 6(b) depicted that the influence of both *Small* and *Large* landslides and their intensity was higher towards the valley portion in the upper half of the road corridor.

Next, we fitted the model for various numbers of covariates adding sequentially and monitored the AIC. Two models are significantly different if the difference of their AIC is at least 2 [31]. Various covariates were explored during model fitting in order to reduce the AIC to identify the best model. The covariates that contributed to a significant decrease were maintained. The covariates that were finally selected in the model were the causal factors for the landslides in the area. Lithology, land cover, road buffer, drainage, and terrain significantly reduced the AIC. The resultant model including these covariates along with the landslide density reflected the susceptibility of the area to landslides. The model was implemented as

$$\begin{aligned} Model &= ppm\left(X, \sim x + y + lithmap + lcmap\right. \\ &\quad + roadmap + drainmap + geomap, covariates \\ &\quad = list\left(lithmap = lith, lcmap = lulc, roadmap\right. \\ &\quad = roadbuff1, drainmap = drain, geomap \\ &\quad = geom\left.\right), \\ &\quad MultiStrauss\left(c\left(\text{“}aSmall\text{”}, \text{“}bLarge\text{”}\right), r\right)\left.\right). \end{aligned} \tag{14}$$

Fitting this model we obtained the corresponding parameter values listed in Table 1 with AIC value equal to 1395.

The estimated interaction parameters γ_{ij} for the models were 6.16, 0.56, and 1.86 for the *Small-Small*, *Small-Large,* and *Large-Large* interactions, respectively. Interaction parameters of the *Small-Small* and *Large-Large* landslides had values larger than 1, thus showing clustering. *Small-Large* landslides had an intensity value below 1 suggesting inhibition. The detail estimated parameters of the fitted intensity are presented in Table 1 and the fitted density outputs with all significant covariates are presented in Figures 6(c) and 6(d). The covariate analysis showed that lithology, road buffer, drainage density, and terrain units positively influenced the outcome of the model whereas land cover had a negative

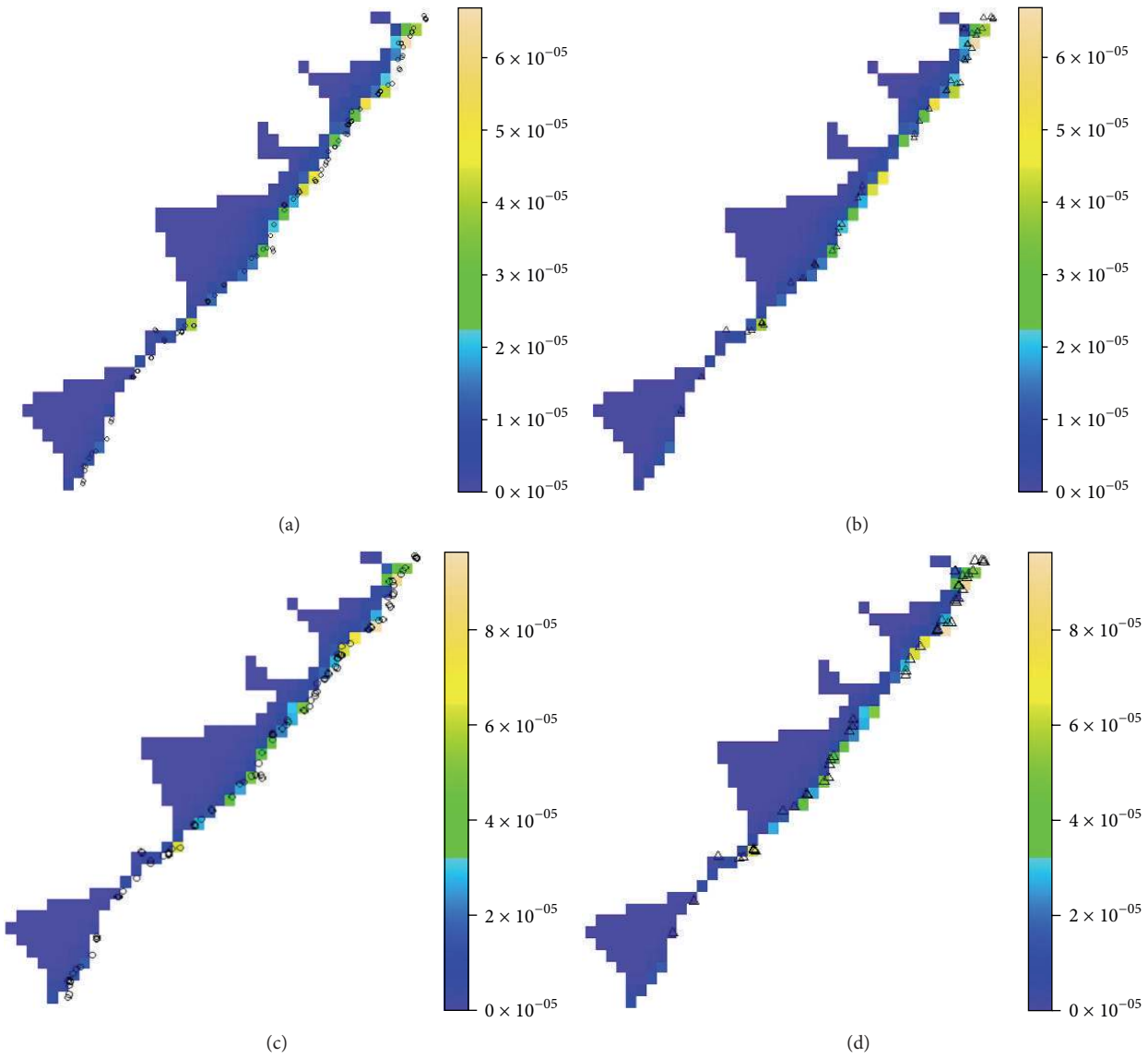

Figure 6: Map showing the multitype Strauss model fitted to the landslide data for (a) *Small* (represented by circles) and (b) *Large* (represented by triangles) landslides using a glm function and fitted density functions to landslides along with the significant covariates data for (c) *Small* and (d) *Large* landslides to show the spatial variation of landslide susceptibility within the study area.

influence in the final model. Further, the overall intensity was increasing because of the influence of the covariates that act as causal factors for the occurrence of landslides.

4.3. Comparison of the Model Output. The evaluation of the model was done by comparing the Receiver operation characteristics (ROC) curves generated for the logistic regression model and the Strauss point process model. A ROC curve, also called success rate curve, was generated to analyze the success of the developed model for landslide susceptibility. By using the ROC curve, success was assessed by comparing the calculated probability values vis-à-vis their actual present condition. The area under the ROC curve (AUC) characterizes the quality of a prediction system by describing the system's ability to anticipate correctly the occurrence or nonoccurrence of predefined "events" [32]. True positive rates (sensitivity) are compared against false positive rates (1−specificity) to assess the prediction accuracy by the models. The ROC curve is shown in Figure 7.

The area under the curve (85.0% for Strauss model and 79.6% for LR model) implied a higher success rate for Strauss model, even with same sets of variables. Further, the ROC curve for the Strauss model showed a stable increase in the sensitivity as compared to the LR model finally achieving the higher accuracy. This indicated the higher sensitivity of the Strauss model to the output probability.

4.4. The Multitype/Hard Core Strauss Point Process Model. Table 2 shows the fitted coefficients and the AIC values for the multitype/hard core Strauss point process model. As

Table 1: Coefficients and AIC for the fitted models multitype Strauss model.

	Intercept	x	y	Covariates					AIC
				Lithology	Land use	Road buffer	Drainage density	Terrain units	
Model	1.03×10^4	4.83×10^{-3}	-3.43×10^{-3}	—	—	—	—	—	1416
Model + 1 covariate	1.14×10^4	5.19×10^{-3}	-3.78×10^{-3}	1.82×10^{-1}	—	—	—	—	1414
Model + 2 covariates	1.11×10^4	5.02×10^{-3}	-3.67×10^{-3}	1.85×10^{-1}	1.28×10^{-1}	—	—	—	1412
Model + 3 covariates	8.92×10^3	4.15×10^{-3}	-2.95×10^{-3}	1.73×10^{-1}	-1.14×10^{-1}	8.23×10^{-1}	—	—	1409
Model + 4 covariates	8.64×10^3	4.06×10^{-3}	-2.86×10^{-3}	6.98×10^{-2}	-1.15×10^{-1}	7.40×10^{-1}	7.79×10^{-1}	—	1401
Model + 5 covariates	1.13×10^4	5.07×10^{-3}	-3.72×10^{-3}	3.18×10^{-2}	-1.14×10^{-1}	5.48×10^{-1}	4.30×10^{-1}	1.66×10^{-1}	1395

Table 2: Coefficients and AIC values for the fitted models multitype Strauss hardcore model.

	Intercept	x	y	Covariates					AIC
				Geomorphology	Drainage	Land use	Drainage density	Terrain units	
Model	1.20×10^4	5.42×10^{-3}	-3.96×10^{-3}	—	—	—	—	—	1379
Model + 1 covariate	1.51×10^4	6.56×10^{-3}	-4.95×10^{-3}	2.29×10^{-1}	—	—	—	—	1374
Model + 2 covariates	1.43×10^4	6.26×10^{-3}	-4.68×10^{-3}	1.79×10^{-3}	4.68×10^{-1}	—	—	—	1361
Model + 3 covariates	1.47×10^4	6.40×10^{-3}	-4.81×10^{-3}	1.93×10^{-1}	3.79×10^{-1}	-1.57×10^{-1}	—	—	1356
Model + 4 covariates	1.26×10^4	5.57×10^{-3}	-4.14×10^{-3}	1.67×10^{-1}	4.37×10^{-1}	-1.59×10^{-1}	6.67×10^{-1}	—	1355
Model + 5 covariates	1.28×10^4	5.66×10^{-3}	-4.21×10^{-3}	1.66×10^{-1}	4.13×10^{-1}	-1.61×10^{-1}	6.42×10^{-1}	1.08×10^{-1}	1357

in the multitype Straus point process model, initially the inclusion of covariates improved the model, as the AIC takes significantly lower values at each step. The last two steps, however, show that the model with three covariates is to be preferred above the model with four and five covariates. The best fitting model thus has geomorphology, drainage density, and land use as the significant covariates. As both geomorphology and land use are categorical variables, not much value can be assigned at present to the positive or negative values observed.

5. Discussion

Landsliding is a geomorphic slope failure process triggered by natural as well as anthropogenic factors and is controlled by unfavorable terrain conditions that act as causal factors. To understand the landslide mechanism in an area and to identify the unknown factors affecting their occurrences, several geoenvironmental variables are included in the analysis [32]. Therefore, the problem of spatial zonation of landslides lies in the landslide inventorization, as well as in their integration with causal factors in a conceptual framework. Various statistical methods have been applied successfully to model the landslide susceptibility mapping [26, 33–37]. All these methods handle the landslide data as spatially independent locational variables, whereas a point pattern analysis allows preserving the level of detail offered by the landslide data itself. This is achieved through inter-landslide interactions and variation in the relative frequency of the pairs of landslides as a function of their position.

A multitype Strauss point process model was used in this study with ten explanatory variables, including morphological, lithological, and structural covariates. The G-function was able to classify landslides as a clustered pattern. The AIC of the fitted model was calculated for each of the covariates and significant ones were recorded [38]. Covariates slope, aspect, soil, weathering, and lineament density did not reduce the AIC significantly. Covariates lithology, land cover, road buffer, drainage density, and terrain units significantly reduced AIC of the fitted model and were included in the model to demonstrate the landslide susceptibility of the area. This study showed that the Strauss point process is capable of modeling the clustered pattern of landslide data for identifying causal factors for susceptibility zonation. The landslide database typically used by experts for landslide susceptibility mapping includes a comprehensively prepared landslide inventory map supported by geoenvironmental variables that cause landsliding [30, 39]. The Strauss point

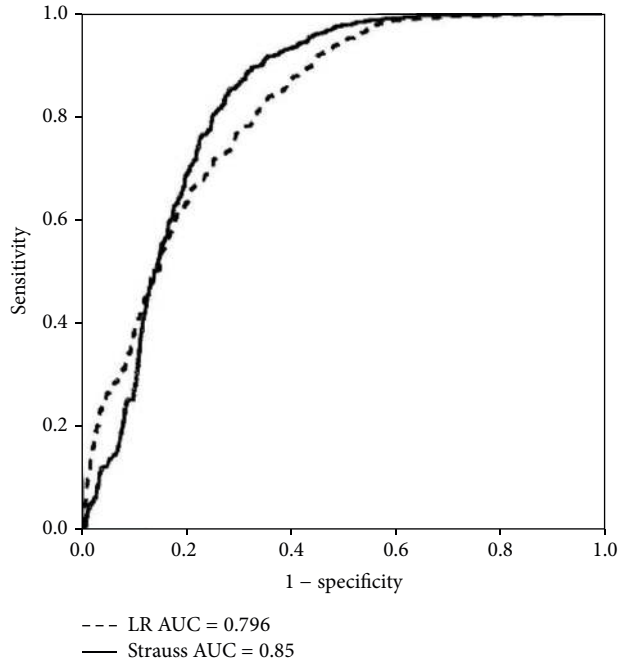

Figure 7: Receiver operation characteristics (ROC) curves representing logistic regression model (dash lines) and Strauss model (solid line). Areas under the curve (AUC) are equal to 0.796 and 0.85, respectively.

model uses such data in a generalized linear modeling framework to show the areas that are susceptible to landslides.

The Strauss point process model discussed here can be applied to any set of landslide database for characterizing its inherent property for susceptibility zonation. This study presented a case for analyzing landslide data as well as identifying significant variables for generating a landslide susceptibility scenario with available information. However, the method is data driven and, therefore, the reliability of the results of modeling is always associated with the quality of the input dataset used in the model development. Thus, accuracy of the outputs invariably depends on the accuracy of the input dataset. For this study the centroid of landslide points were marked which in a way helped to reduce the positional uncertainty of landslides. The landslide data were converted into a point pattern with size (area) of landslide as mark of each landslide making it a marked point pattern. The model used the first-order (intensities) and second-order (interactions) effects of point pattern sequentially for characterizing the landslides. The required parameters like interaction radii were derived iteratively by running the model several times. After model fitting, the corresponding AIC values have been calculated for the fitted model. The model fitting with AIC showed that the covariates like lithology, land cover, road buffer, drainage, and terrain units were significant to the fitted model. All the significant covariates were combined with the landslide data in the multitype Strauss model to generate the map showing areas susceptible to landslides. The nearest neighbor distance analysis carried out for point pattern data of landslides opened an opportunity to include the spatial distribution of landslides in an alternate fashion to the existing methods. Model fitting was carried out by means of GLM functions including a combination of distance correlation functions, model fitting, and simulation which helped in describing and understanding the landsliding pattern in detail.

Several physical and terrain parameters influence the landslide process. To understand the landslide occurrence mechanism, these parameters were analyzed systematically. All the factors invariably had control on the landslides occurring along natural slopes. *Small* landslides occurring exclusively along the cut slopes of road corridor, however, might be controlled mainly by anthropogenic factors rather than by the natural terrain factors. Model fitting through AIC was a good way to address the sensitivity of the fitted model to the landslides data as well as to the significant covariates data. For the Strauss model being a Gibbs model, AIC acted as a goodness-of-fit of the model to the data for global sensitivity assessment of the susceptibility model.

Landslides are spatially discrete events and are controlled by number of geoenvironmental factors that are not straightforward to be easily modeled using statistical methods. Fitted models of the multitype Strauss process to landslide occurrences reflected the nature of model fit to the data. This mathematical model is the model that best fit the data; however, it is not necessarily the best model to serve in situations of practical use. Therefore, it is essential that the fitted model keeps pace with a priori knowledge for consistency. For those reasons we also fitted the multitype Strauss hardcore model. This showed slight differences with the multitype Strauss model, as the geomorphology now more stands out as the most explaining covariate, besides the x- and y-coordinates. The model provided a slightly better fit, that is, lower AIC values.

A point pattern analysis of the landslide data was helpful in understanding the intensity of the landsliding spatially as well as relating the interaction of the geoenvironmental variables in each landslide location with the processes generating the landslide distribution patterns. In addition, an automated module created by considering landslides as point data reduced the bias of sampling errors. In landslide studies more emphasis is given on field work and field data collection and research. Little has been done so far towards dissemination of information and facility for nonexperts. The needs of the disaster mitigating agencies working in remote hazardous front where no geomorphology expert can be made available are critical as also in case of adventure travelers or groups of pilgrims in a hilly terrain. Hence a data driven model can be useful to provide information on susceptibility of landslide in a particular region.

The factors that control landsliding in a particular region largely depend upon the geoenvironmental setup of the area. Point pattern analysis of the landslide can be helpful to urban planners in understanding the distinctive geological and geophysical characteristics at individual landslide location. These provide insight into the processes behind the landslide distribution patterns. The interaction between the landslides and their causal factors derived by means of the point process model as shown in this paper can be useful to address the management practices in hazard prone areas. It was observed

from the area in this study that geoenvironmental factors related to rock types and their weathering conditions, land use patterns, and flow of water, occurring both at the surface and at the subsurface, influenced the occurrence of landslides. The marked point pattern analysis technique had the advantage of including the size of landslides into the analysis, as well as their interactions amongst themselves and with the covariates controlling the landslides. This made a unique combination to address the locations and time of occurrence of landslides in particular area leading to adopt suitable mitigation measures like slope modifications, construction of retaining walls and water channels, afforestation and alternate route alignments, and so forth.

6. Conclusions

A multitype Strauss point process model applied in this study successfully addressed the intensity and spatial interactions of landslides by means of the distance correlation functions, model fitting, and simulation. We concluded that this model enriches the set of statistical methods that can comprehensively analyze landslide data. It expresses properties inherently associated with each landslide and their interactions with landslides in the neighbourhood. As such, point pattern models are advantageous over GLM based models that perform a model fitting and simulation for landslide susceptibility analysis. Thus, our study adds to already existing statistical methods for landslide susceptibility mapping. Model fitting also demonstrated the significance of covariates. In this way, significant causal factors could be extracted that help to better understand the pattern of landsliding in an area.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This paper is the outcome of the research carried out under the IIRS-ITC Joint Research Collaboration. The authors are thankful to Director, NRSC, Hyderabad, India, for allowing them to carry out this research. Encouragement by Director, IIRS (NRSC), ISRO, Dehradun, is highly acknowledged. First author is also thankful to Dr. P. K. Champatiray, Head, Geosciences Division, IIRS, for his constant support and encouragement.

References

[1] F. Guzzetti, P. Reichenbach, M. Cardinali, M. Galli, and F. Ardizzone, "Probabilistic landslide hazard assessment at the basin scale," *Geomorphology*, vol. 72, no. 1–4, pp. 272–299, 2005.

[2] D. J. Varnes, "Slope movement types and processes," in *Landslides: Analysis and Control, Special Report 176*, R. L. Schuster and R. L. Krizek, Eds., pp. 11–33, Transportation Research Board, National Academy of Sciences, Washington, DC, USA, 1978.

[3] L. Ayalew and H. Yamagishi, "The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan," *Geomorphology*, vol. 65, no. 1-2, pp. 15–31, 2005.

[4] C. J. Van Westen, T. W. J. Van Asch, and R. Soeters, "Landslide hazard and risk zonation—why is it still so difficult?" *Bulletin of Engineering Geology and the Environment*, vol. 65, no. 2, pp. 167–184, 2006.

[5] C.-T. Lee, C.-C. Huang, J.-F. Lee, K.-L. Pan, M.-L. Lin, and J.-J. Dong, "Statistical approach to earthquake-induced landslide susceptibility," *Engineering Geology*, vol. 100, no. 1-2, pp. 43–58, 2008.

[6] J.-J. Dong, Y.-H. Tung, C.-C. Chen, J.-J. Liao, and Y.-W. Pan, "Discriminant analysis of the geomorphic characteristics and stability of landslide dams," *Geomorphology*, vol. 110, no. 3-4, pp. 162–171, 2009.

[7] A. Nandi and A. Shakoor, "A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses," *Engineering Geology*, vol. 110, no. 1-2, pp. 11–20, 2010.

[8] S. Lee, J.-H. Ryu, and I.-S. Kim, "Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin, Korea," *Landslides*, vol. 4, no. 4, pp. 327–338, 2007.

[9] S. Lee and B. Pradhan, "Probabilistic landslide hazards and risk mapping on Penang Island, Malaysia," *Journal of Earth System Science*, vol. 115, no. 6, pp. 661–672, 2006.

[10] P. M. Atkinson and R. Massari, "Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy," *Computers and Geosciences*, vol. 24, no. 4, pp. 373–385, 1998.

[11] R. Basile, L. Benfratello, and D. Castellani, "Geoadditive models for regional count data: an application to industrial location," *Geographical Analysis*, vol. 45, no. 1, pp. 28–48, 2013.

[12] J. Yang, H. S. He, S. R. Shifley, and E. J. Gustafson, "Spatial patterns of modern period human-caused fire occurrence in the Missouri Ozark Highlands," *Forest Science*, vol. 53, no. 1, pp. 1–15, 2007.

[13] J. Mateu, J. L. Usó, and F. Montes, "The spatial pattern of a forest ecosystem," *Ecological Modelling*, vol. 108, no. 1–3, pp. 163–174, 1998.

[14] D. Stoyan and A. Penttinen, "Recent applications of point process methods in forestry statistics," *Statistical Science*, vol. 15, no. 1, pp. 61–78, 2000.

[15] S. Anwar and A. Stein, "Detection and spatial analysis of selective logging with geometrically corrected Landsat images," *International Journal of Remote Sensing*, vol. 33, no. 24, pp. 7820–7843, 2012.

[16] L. Holden, S. Sannan, and H. Bungum, "A stochastic marked point process model for earthquakes," *Natural Hazards and Earth System Science*, vol. 3, no. 1-2, pp. 95–101, 2003.

[17] M. N. M. van Lieshout and A. Stein, "Earthquake modelling at the country level using aggregated spatio-temporal point processes," *Mathematical Geosciences*, vol. 44, no. 3, pp. 309–326, 2012.

[18] M. Kerscher, I. Szapudi, and A. S. Szalay, "A comparison of estimators for the two-point correlation function," *The Astrophysical Journal*, vol. 535, no. 1, pp. L13–L16, 2000.

[19] C. Walter, A. B. McBratney, R. A. V. Rossel, and J. A. Markus, "Spatial point-process statistics: concepts and application to the analysis of lead contamination in urban soil," *Environmetrics*, vol. 16, no. 4, pp. 339–355, 2005.

[20] A. Baddeley, E. Rubak, and R. Turner, *Spatial Point Patterns: Methodology and Applications with R*, CRC Press, Boca Raton, Fla, USA, 2016.

[21] S. Anwar, *Implementation of Strauss point process model to earthquake data [M.S. thesis]*, University of Twente, Enschede, The Netherlands, 2009.

[22] B. D. Ripley, "Algorithm AS 137: simulating spatial patterns: dependent samples from a multivariate density," *Applied Statistics*, vol. 28, no. 1, pp. 109–112, 1979.

[23] P. J. Diggle, *Statistical Analysis of Spatial Point Patterns*, Arnold, 2nd edition, 2003.

[24] A. Stein and N. J. Georgiadis, "Spatial statistics to quantify patterns of herd dispersion in a savanna herbivore community," in *Resource Ecology: Spatial and Temporal Dynamics of Foraging*, H. H. T. Prins and F. Van Langevelde, Eds., pp. 33–51, Springer, Dordrecht, The Netherlands, 2008.

[25] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[26] I. Das, S. Sahoo, C. J. Van Westen, A. Stein, and R. Hack, "Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India)," *Geomorphology*, vol. 114, no. 4, pp. 627–637, 2010.

[27] K. Vinod Kumar, R. C. Lakhera, T. R. Martha, R. S. Chatterjee, and A. Bhattacharya, "Analysis of the 2003 Varunawat landslide, Uttarkashi, India using Earth observation data," *Environmental Geology*, vol. 55, no. 4, pp. 789–799, 2008.

[28] V. M. Choubey and R. C. Ramola, "Correlation between geology and radon levels in groundwater, soil and indoor air in Bhilangana Valley, Garhwal Himalaya, India," *Environmental Geology*, vol. 32, no. 4, pp. 258–262, 1997.

[29] A. K. Saha, R. P. Gupta, I. Sarkar, M. K. Arora, and E. Csaplovics, "An approach for GIS-based statistical landslide susceptibility zonation-with a case study in the Himalayas," *Landslides*, vol. 2, no. 1, pp. 61–69, 2005.

[30] I. Das, A. Stein, N. Kerle, and V. K. Dadhwal, "Probabilistic landslide hazard assessment using homogeneous susceptible units (HSU) along a national highway corridor in the northern Himalayas, India," *Landslides*, vol. 8, no. 3, pp. 293–308, 2011.

[31] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, *Statistical Analysis and Modelling of Spatial Point Patterns*, Statistics in Practice, John Wiley & Sons, New York, NY, USA, 2008.

[32] E. Yesilnacar and T. Topal, "Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey)," *Engineering Geology*, vol. 79, no. 3-4, pp. 251–266, 2005.

[33] C.-J. F. Chung and A. G. Fabbri, "Probabilistic prediction models for landslide hazard mapping," *Photogrammetric Engineering and Remote Sensing*, vol. 65, no. 12, pp. 1389–1399, 1999.

[34] J. Mathew, V. K. Jha, and G. S. Rawat, "Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method," *Landslides*, vol. 6, no. 1, pp. 17–26, 2009.

[35] I. Yilmaz, "Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat-Turkey)," *Computers and Geosciences*, vol. 35, no. 6, pp. 1125–1138, 2009.

[36] S.-B. Bai, J. Wang, G.-N. Lü, P.-G. Zhou, S.-S. Hou, and S.-N. Xu, "GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China," *Geomorphology*, vol. 115, no. 1-2, pp. 23–31, 2010.

[37] B. Pradhan and S. Lee, "Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling," *Environmental Modelling and Software*, vol. 25, no. 6, pp. 747–759, 2010.

[38] O. Špatenková and A. Stein, "Identifying factors of influence in the spatial distribution of domestic fires," *International Journal of Geographical Information Science*, vol. 24, no. 6, pp. 841–858, 2010.

[39] F. Mancini, C. Ceppi, and G. Ritrovato, "GIS and statistical analysis for landslide susceptibility mapping in the Daunia area, Italy," *Natural Hazards and Earth System Science*, vol. 10, no. 9, pp. 1851–1864, 2010.