LANDSLIDE SUSCEPTIBILITY MAPPING IN A HIMALAYAN ROAD CORRIDOR USING BAYESIAN LOGISTIC REGRESSION MODELS

Iswar Das,[ab] Alfred Stein,[a] Norman Kerle[a] and Vinay K. Dadhwal[b]

[a]International Institute for Geo-Information Science and Earth Observation (ITC), 7500 AA, Enschede, The Netherlands

[b]Indian Institute of Remote Sensing, 4-Kalidas Road, Dehradun, India

Introduction

Landslide susceptibility mapping aims to differentiate a land surface into homogeneous areas according to their probability of failure caused by mass-movement at specific locations. During last two decades various statistical methods were used in landslide susceptibility mapping. All these methods analyze the geo-environmental variables controlling landslide occurrence with respect to previous landslide events, either by means of bivariate or of multivariate frequentist statistics [3]. These frequentist methods, however, result in discrete parameter estimates with distribution properties based on the assumption of normality [5]. Bayesian methods, on the other hand, result in a probability distributions of the parameter estimates based on the data and on prior knowledge, thus facilitating uncertainty estimation procedures [4].

The aim of this study is to develop and apply a Bayesian logistic regression (BLR) model for landslide susceptibility mapping. To do so we use the Bayesian paradigm to modify the commonly applied frequentist logistic regression (FLR) model. Uncertainty in parameter estimates is quantified by means of the posterior density distributions. The model consists of two steps. First, parameter estimates are obtained for each variable by means of FLR. The same variables are then used in a Bayesian framework for parameter estimation and the significance of each estimate is evaluated by means of the posterior density function obtained from the Bayesian analysis. Simulation of the model is done using MCMC methods. The methodology is applied to a landslide-prone road corridor in the northern Himalayas, in India. In the present study, nine different landslide influencing geo-environmental factors have been analyzed in their relation to the landslide occurrence data. These maps were derived from high resolution Cartosat-1 and Resourcesat-1 data (resolutions of 2.5m and 5.8m, respectively) along with auxiliary data like published maps and reports and field checks.

Research Methods and Model

In landslide studies a logistic regression model incorporates the occurrence of landslides as a discrete and dichotomous response variable, and the geo-environmental factors that influence it as explanatory variables. The logistic regression model for k explanatory variables is formulated as

$$p_i = \Pr(Y_i = 1) = \exp(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}) / \left[\exp(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}) + 1 \right] \quad (1)$$

where Pr (Y_i = 1) is the probability of the occurrence of a landslide, constrained to lie between 0 and 1, the x_{ij} denote the different categories of the geo-environmental factors and the β_j, j = 0,…,k are unknown regression coefficients.

A Bayesian framework contains three key components associated with parameter estimation: the prior distribution, the likelihood function and the posterior distribution. A simple Bayesian equivalent of the frequentist logistic model was constructed after Clark et al. [2]

$$y_i \sim Bernoulli(\mathrm{logit}^{-1}(\eta_i)) \quad (2)$$

$$\eta_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$$

$$\beta_j \sim N(0, 0.00001), j = 0, \ldots, k$$

where y_i represents the response variable, the β_j's are coefficients having independent normal prior distributions with a very high variance, x_{ij} represents the value of the j^th^ variable at i^th^ location and η_i is the linear predictor.

Using the Bayes formula, the posterior distribution of the parameters β under this model is given by:

$$\pi(\beta \mid y, X) \propto \prod_{j=0}^{k} \Pr(\beta_j) \times \prod_{i=1}^{n} \Pr(y_i \mid \eta_i) \quad (3)$$

where, $\beta = (\beta_0, \beta_1, ..., \beta_k)$, $y = (y_1, y_2, ..., y_n)$ and $X = \lfloor x_{ij} \rfloor, i = 1,2,...,n, \; j = 1,2,...,k$.

This is an extension of the Bayesian formula $f(\theta \mid y) \;\alpha\; g(\theta) \times L(y \mid \theta)$, that relates the posterior distribution as proportional to the product of the prior distribution and the likelihood function.

Prior distribution Pr (β_j). In this study noninformative priors are considered, because the aim is to show the advantage of a Bayesian method over ordinary frequentist method using normal priors without using additional knowledge.

Likelihood function Pr(y_i|η_i). Preliminary inference is derived for all variables from the FLR model. The same variables are used in the BLR model for better comparison of the output maps and the significance of a Bayesian statistical data analysis framework. The Bayesian analysis is employed in this study using the Gibbs sampler, an iterative algorithm that achieves posterior estimation via Markov Chain Monte Carlo (MCMC) integration for estimation of the posterior distributions [1].

Results

The mapped landslides cover an area of 0.45 km^2, corresponding to 5.6% of the total area (minimum 144 m^2, maximum 0.052 km^2, mean 7261.3 m^2 and standard deviation 9242.3 m^2). The results of the modeling showed that the posterior parameter estimates obtained for BLR model are similar to the estimates generated with FLR analysis. However, the posterior density functions are useful to reveal the significance of the different explanatory variables for their contribution towards the landsliding (Figure 1).

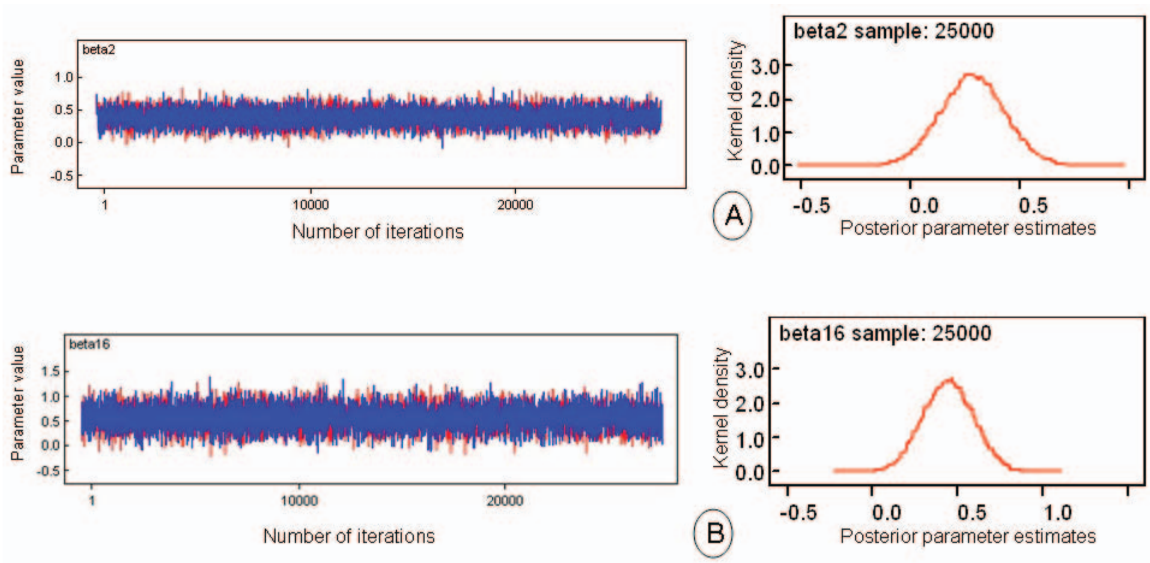

Figure 1. History of trace plots and density distribution of the corresponding posterior parameter estimates (beta's) for 2 selected variables (A & B). History of trace plots indicates the parameter value after 25,000 iterations for convergence of simulation.

Interpretation of the coefficients from the logistic model is not as straightforward as interpretation of the coefficients in a linear model, and therefore we turn towards their exponential values. For example, BLR model generates the parameter mean value 0.239 with the end points of a 95 percent confidence interval as (0.002, 0.582). Estimation indicates that the changes in the log-odds of landslide for this slope class are between 0.002 and 0.582 with 95% confidence and the parameter estimate is contributing positively to the landslide occurrence. As shown in the Figure 2, the BLR model was conservative in predicting very high susceptibility class (0.75 – 1.0) as the area percentage almost matches the percentage of landslide in the study area (5.25 % predicted versus 5.6 % actual). On the other hand FLR model had a higher prediction percentage i.e. 9.87%. in this class. We conclude that being performed iteratively and accounting for prior information, a Bayesian logistic regression model leads to a refined output of parameter estimates, thereby increasing the success rate of predicted probabilities of susceptibility map.

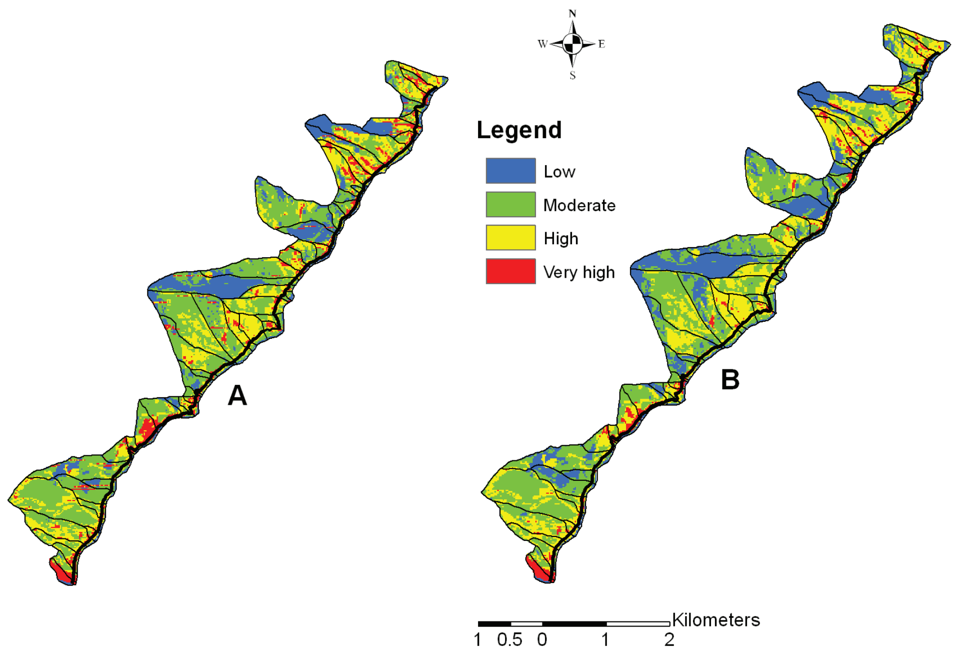

Figure 2. Landslide susceptibility maps. (A) Using frequentist logistic regression. (B) Using Bayesian logistic regression methods. Low (0.00-0.25), moderate (0.25-0.50), high (0.50-0.75) and very high (0.75-1.0).

References

[1] Brooks, S.P., Roberts, and G.O., 1998, "Assessing convergence of Markov Chain Monte Carlo algorithms Statistics and Computing", vol. 8, pp. 319-335.

[2] Clark, T.G., De Iorio, M., Griffiths and R.C., 2007, "Bayesian logistic regression using a perfect phylogeny. Biostatistics", vol. 8, no. 1, pp. 32-52.

[3] Das, I., Sahoo, S., Van Westen, C.J., Stein, A. and Hack, R., 2009, "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, pp. 627-637.

[4] Eckert, N., Parent, E., Naaim, M. and Richard, D., 2008, "Bayesian stochastic modeling for avalanche predetermination: from a general system framework to return period computations. Stochastic Environmental Research and Risk Assessment", vol. 22, pp. 185-206.

[5] Mila, A.L., Yang, X.B. and Carriquiry, A.L., 2003, "Bayesian logistic regression of Soyabean Sclerotinia stem rot prevalence in the U.S. north-central region: Accounting for uncertainty in parameter estimation", Phytopathology, vol. 93, no. 6, pp. 758-763.