International Journal of Civil and Structural Engineering– IJCSE
Volume 2 : Issue 1 [ISSN : 2372-3971]
Publication Date: 30 April, 2015

Estimation of Shallow Landslide Susceptibility Using GIS Integrated Support Vector Regression

G Antherjanam, S Chandrakaran, MC Philipose

Abstract— This paper proposes an effective method for susceptibility estimation of shallow landslides integrating the geographical information system based landslide susceptibility estimation model and a data driven paradigm. The study incorporates geotechnical properties of soil in modeling exercise along with the traditional geospatial landslide causative factors such as landuse and slope angle. The entire database is applied in SINMAP (stability index mapping) platform in the GIS environment to compute the susceptibility indices of the concerned study area in a multi-calibration mode. Then the geotechnical properties are extracted using kriging interpolation to use them as predictor variables to develop a regression model using support vector machine (SVM) and the prepared model is validated statistically. The methodology is demonstrated by applying it in Aruvikkal basin in Kerala state in India and the model is suitable for landslide susceptibility prediction problems in Western Ghats.

Keywords: landslide, GIS, support vector machines, staibility index

I. Introduction

Landslides are one of the most damaging natural disasters which often draw attention of geotechnical investigators. The Western Ghats in Indian subcontinent are identified to be landslide prone by many researchers recently [1,2]. The landslide susceptibility zonation is one of the most important tasks in landslide risk assessment. The different approaches for landslide susceptibility modeling includes heuristics (eg., index-based approach and an analytical hierarchical process approach), statistical (statistical index, certainty factor, probability based methods, weight of evidence modeling, multiple linear regression and logistic regression analysis), process-based or deterministic modeling (slope stability factor) [3]. In the Western Ghats area, the first reported study on landslides was conducted by [4]. Many researchers in the past identified a variety of causal factors and their importance in landslide susceptibility which is a crucial step in accurate landside predictions and their method of estimation also falls in one among the aforementioned approaches of modeling [1,2, 5-8]. Recently, the use of GIS has got considerable attention among the research community and the use of a physically based landslide susceptibility model assisted by GIS are credible tool for estimation of landslide susceptibility. Recently, the use of data driven tools along with GIS is gaining popularity in landslide susceptibility zonation at different parts of the world [9-11]. However the use of a GIS integrated physical model in conjunction with data driven tools is still rare in literature. Also it is a well-known fact that geotechnical parameters are often easily measurable and have strong association with proneness of landslides. But most of the studies in the past have't given enough attention to the geotechnical parameters in the estimation of landslide susceptibility. Therefore, this paper presents a hybridized approach by using a physically based landslide susceptibility model along with data driven paradigm to build the user friendly regression model considering geotechnical parameters as the primary inputs.

GOURI ANTHERJANAM
TKM College of Engineering Kollam
India

Dr S CHANDRAKARAN
NIT calicut
India

Dr M C PHILIPOSE
St Gits College of Engineering Kottayam
India

II. Methodology

Stability Index and Mapping (SINMAP) is a GIS integrated landslide susceptibility estimation tool proposed by Pack et al., [12]. The first step in the proposed approach involves the use of geo-spatial database as input to compute stability index of the study area using SINMAP with GIS interface. Then geo spatial data and geotechnical data and stability indices pertaining to different locations of the study area are extracted to create the input dataset for the data driven based modeling. In the data based modeling phase, a non-linear regression model is built and validated. This model is suitable for landslide susceptibility estimation at different regions with similar geographical characteristics. The step by step procedure of proposed methodology is presented in the form of a flow chart in Fig 1.

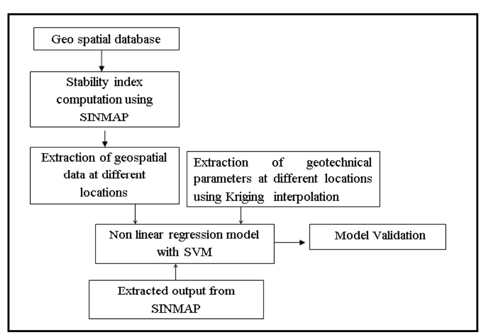

Figure 1 Flowchart of the methodology

International Journal of Civil and Structural Engineering– IJCSE
Volume 2 : Issue 1 [ISSN : 2372-3971]
Publication Date: 30 April, 2015

III. Support Vector Machine

Support Vector Machine (SVM) is a relatively recent addition to the family of data driven techniques evolved from the concept of statistical learning theory which perform non-linear regression problems by structural risk minimization [13]. Considering a set of input-output pairs as training dataset, $[(x_1, y_1), (x_2, y_2)\dots(x_l, y_l)]$ $x \in R^N$, $y \in r$, where, x is the input, y is the output, R^N is the N dimensional vector space and r is the one dimensional vector space. In this problem x = [c, ϕ, γ, k, β, n] and y = [SI] where c is the cohesion, ϕ is the angle of internal friction, γ is the bulk density, k is the coefficient of permeability, β is the slope angle, n is the porosity and SI is the stability index.

The intension of SVM is to fit a function that can approximately predict the value of output on supplying a new set of predictors (input variables).The ε-intensive loss function can be described as follows:

$$L_\varepsilon(y) = 0 \quad for \ \left| f(x) - y \right| \le \varepsilon \qquad (1)$$

otherwise,

$$L_\varepsilon(y) = \left| f(x) - y \right| - \varepsilon \qquad (2)$$

This defines an ε-tube (a tolerance margin/band) so that if the predicted values is within the tube, the loss is zero; otherwise the loss is equal to the absolute value of the deviation minus ε. In SVM an attempt is made to find a function $f(x)$ that gives the deviation of 'ε' from the actual output is as flat as possible.

Considering a linear function of the form,

$$f(x) = (w.x) + b \qquad w \in R^N, b \in r \qquad (3)$$

where, w is an adjustable weight vector and b is the scalar threshold. Fitness means the search for a small value of 'w'. It can be represented as a minimization problem with an objective function comprising the Euclidian norm incorporating two slack parameters ξ and ξ* to penalize the samples with error more than 'ε'. Thus the infeasible constraints of the optimization problem are eliminated and the modified formulation takes the form:

Minimize : $$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \qquad (4)$$

Subject to $$y_i - [(w.x_i) + b] \le \varepsilon + \xi_i, \quad i = 1, 2, 3....l \qquad (5)$$

$$[(w.x_i) + b] - y_i \le \varepsilon + \xi_i^*, \quad i = 1, 2, 3....l \qquad (6)$$

$\xi_i \ge 0$ and $\xi_i^* \ge 0$, i=1,2,3$\cdots$.l

The constant $0 < C < \infty$ determines the trade-off between the flatness of $f(x)$ and the amount upto which the deviations larger than 'ε' are tolerated [14]. The above optimization problem is solved by Vapnik [15] using Lagrange Multiplier method. The solution is given by:

$$f(x) = \sum_{i=1}^{M}(\alpha_i - \alpha_i^*)(x_i.x) + b \qquad (7)$$

where, $$b = -\left(\frac{1}{2}\right)w.(x_r + x_s) \qquad (8)$$

where, x_s and x_r are known as support vectors and M is the number of support vectors.

Some Lagrange multipliers (α_i, α_i^*) will be zero, which implies that these training solutions are irrelevant to the final solution (known as sparseness of the solution). The training objects with non-zero Lagrange multipliers are called 'support vectors'. When linear regression is not appropriate, input data have to be mapped into a high dimensional feature space through non-linear mapping and the linear regression needs to be performed in the high dimensional feature space [16]. The mapping of input data onto the feature space can be done by the use of Kernel functions which satisfies Mercer's theorem [16]. Polynomial functions, radial basis function (RBF), and splines are the most commonly used functions for data fitting using SVM. Recently, SVM is successfully applied to many problems landslide susceptibility estimation [17-19]. But in the context of shallow landslide susceptibility estimation, the integration of a physically based model with SVR is an innovative step in landslide hazard zonation.

IV. Study Area

The study area belongs to the Western Ghats region of South India. The region is famous for large amount of shallow landslides/debris flows. Aruvikkal is a tributary of Tikovil river, originates on the facets of plateau margin and mountain flanks situated in Kottayam district in Kerala state India. Fig. 2 presents the study area where Tikkovil and Aruvikkal watershed respectively covers approximately 56 and 9.4 sq.km of area. The Aruvikkal basin can be considered as a representative of the midland area of Kerala as it starts at an altitude of about 55 meter, close to Tikovil River confluence and extended upto 1120 meters towards north. The study area consist the ideal environment to test physically based dynamic landslide model due to unique environment, under laying geology and frequently occurring shallow landslides/debris flows [2]

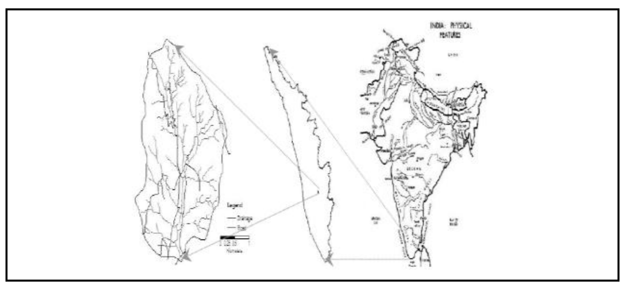

Figure 2 Location map of the study area

V. SINMAP implementation for landslide susceptibility estimation

The deterministic slope stability model namely Stability INdex MAPping (SINMAP), developed by Pack et al. [11] is used in this study to assess the instability conditions and to establish a landslide hazard zonation map. SINMAP is a raster based slope stability predictive tool based on coupled hydrological-infinite slope stability model implemented in ArcView 3.3 platform. This approach applies to shallow translational land sliding phenomena controlled by shallow ground water convergence and more details on SINMAP can be found in [12]. The data requirement include inventory of past landslides, digital elevation model (DEM), geotechnical data such as soils strength properties, thickness of soil above the failure plane, and hydrological data such as soil hydraulic conductivity and the rainfall. All spatial data had a resolution of 20 m by 20 m. Then a calibration regions to be created in single or multi calibration framework by supplying lower bound and upper bound calibration parameter values of wetness index (*T/R*) where, *T* is the transmissibility, *R* is the recharge, cohesion index (c), and friction angle (φ). In this study, a multi-calibration theme involving eight calibration regions based on land use type, was applied. For all land use type, T/R ratio ranges between 2000-3000, cohesion ranges between 0-0.25 while, φ ranges between 30-45 are adopted in this study. The primary output of the model is a stability index, which is the probability that a location is stable assuming uniform distribution of the parameters over their uncertainty, is used to classify the terrain stability for each grid cell of the study area. This SI value ranges between 0 (most unstable) and 1 (least unstable). In this study, the causative factors like saturation, inclination of slope, bulk density, porosity, angle of internal friction and cohesion are the input to predict stability index of the given basin.

V Results and Discussion

The first sub section discuss the results of single and multi-calibration modes followed by the susceptibility estimation using the support vector regression (SVR). The SI Map derived is helpful for preparing a database for the regression model connecting SI and its causative factors.

A . Stability Index Grid Theme using SINMAP

The Stability Index Grid theme of Aruvikkal basin(landslide hazard zonation map) is presented in Fig 3. Also a slope area chart (SA plot) is prepared and a statistical summary for each region of the study area to aid in the data interpretation and parameter calibration. The SA Plot provides a view of study data in slope area space. The resulted SA plot is shown in Fig. 4.

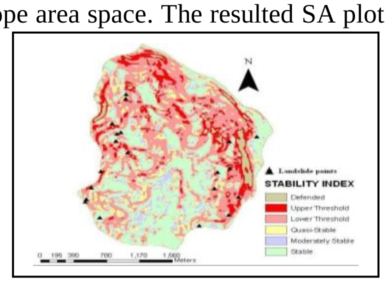

Figure 3 Susceptibility to failure predicted by SINMAP overlaid with the landslide inventory

Statistical results obtained from the slope-area plot charts for one region of the basin are shown in Table 1.

Table 1: Summary of the analytical data resulted from SINMAP analysis for one of the region (S-Stable; MS-Moderately stable; QS-Quasi Stable; LT-Lower Threshold; UT-Upper Threshold; Defended-; Total-)

	S	MS	QA	LT	UT	D	T
Grass & Rock							
Area (km^2)	0.2	0.1	0.2	0.5	0.2	0.1	1.3
% of region	16.5	7.8	12.6	39.2	15.2	8.6	100
Number of landslides	0	0	0	1	0	0	1
% of Slides	0	0	0	100	0	0	100
LS Density	0	0	0	1.9	0	0	0.8

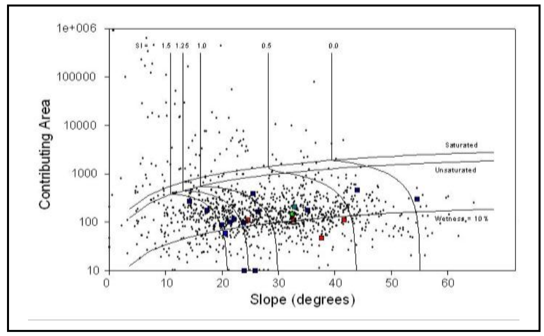

Figure 4 Graphical representation of the prediction of susceptibility to failure by SINMAP
(Triangles are landslide locations in the parameter space)

B. SVM based landslide susceptibility model

The Stability Indices obtained for 153 different locations are extracted along with corresponding input parameters using Kriging interpolation[20]. Out of the153 dataset, 70 % data is used for training and 30 % for validation. The statistical properties of the training and validation dataset are shown in Table 2. Then a non-linear model is fitted using SVR. The popular grid search method is used for estimation of SVM control parameters and a Radial Basis Function Kernel (RBF) is adopted for modeling. The parameters *C*=39 and width of Gaussian Kernel =0.28 are found to be optimal as this combination gives minimal error measure during the training stage.

Table 2 STATISTICAL PROPERTIES OF TRAINING AND VALIDATION DATASETS

Statistics	**Training**						
Max	3.000	2.035	1.364	0.537	27.808	10.294	6.488
Min	0.026	0.078	1.229	0.486	24.942	0.529	0.000
Mean	0.411	0.503	1.294	0.512	26.890	5.668	1.275
Standard Deviation	0.648	0.270	0.040	0.015	0.739	2.046	0.816

International Journal of Civil and Structural Engineering– IJCSE
Volume 2 : Issue 1 [ISSN : 2372-3971]
Publication Date: 30 April, 2015

Statistics	Validation						
Max	3.000	1.504	1.365	0.533	27.854	9.540	1.893
Min	0.041	0.222	1.239	0.485	24.042	4.429	0.000
Mean	0.526	0.684	1.288	0.514	26.605	6.858	0.820
Standard Deviation	0.898	0.271	0.0269	0.0100	1.084	1.669	0.452

The scatter plot of SVR based estimation of SI for training and validation dataset are presented in Fig. 5, which indicate that the model is trained reasonably well and able to capture the non-linearity satisfactorily. The statistical performance evaluation was done by computing correlation coefficient (R) and different error measures. root mean square error (RMSE), mean – absolute error (MAE) and root-relative squared error (RRSE) and the results are presented in Table 3. The high R value, less difference between R values of calibration and validation and lower error measures indicate that the model is capable in prediction of stability index for similar catchments

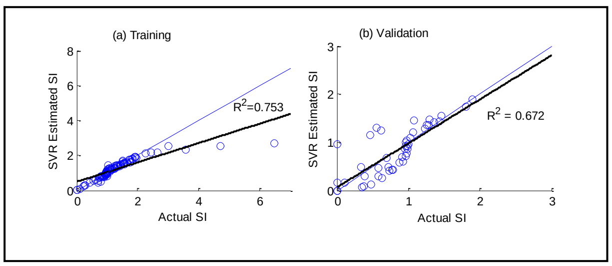

Figure 5 The scatter plot of SVR based estimation of SI for training and validation datasets

Table 3 Statistical performance evaluation for training and validation

Evaluation Criteria	Training	Validation
R	0.869	0.820
RMSE	0.451	0.321
MAE	0.1330	0.222
RRSE (%)	55.38	47.39

VI Conclusions

This paper applies Stability Index Mapping to model for estimating shallow landslide susceptibility of Aruvikkal basin Kerala, India in multi-calibration mode. Then the geotechnical parameters of the study area (and corresponding stability indices) are extracted for a large number of data points within the study area using Kriging interpolation. SVM based modeling is performed on the generated database and validated statistically, to present it as a user friendly approach integrating physics based susceptibility estimation and data driven paradigm. The proposed methodology has potential to estimating landslide susceptibility in the form of a simplified algebraic model for similar geographical locations of Western Ghats.

References

[1] H. Vijith, "Modelling landslide susceptibility on a regional scale using remote sensing and Gis: An assessment of parts of the Meenachil river basin of Western Ghats, India," Ph.D thesis, Mahathma Gandhi university India 2008

[2] N.T. Long Landslide Susceptibility Mapping of the Mountainous Area In A Luoi District, Thua Thien Hue Province, Vietnam, Ph.D thesis, Vrije University Brussels Germany

[3] S.L. Kuriakose "Physically-based dynamic modelling of the effect of land use changes on shallow landslide initiation in the Western Ghats of Kerala, India," PhD thesis, University of Utretcht Netherlands, 2010

[4] M.S. Balasundaram "Landslips near Todupulai, Kottayam district, Kerala State," J. Engng. Geology vol. II(1), pp. 2-11

[5] P. Basak, and N.B Narasimha Prasad "Recurring landslides in the Western Ghats - an environmental hazard," Env. Cons. vol. 16(2), pp. 169-172 1989

[6] M.P. MuraleedharanLandslides in Kerala - a drenched state phenomena in regolith National Seminar on Landslides in Western Ghats. Trivandrum, Center for Earth Science Studies. 1995

[7] R. Krishnanath, and S. Sreekumar Identification and location of slide prone areas in certain hard rock profiles of Western Ghats, Indi. Geotech. J. vol 26(4), 407-416 1996

[8] K.S. Sajin Kumar Geo-informatics in landslide risk assessment and management in parts of Western ghats, Central Kerala South India, Ph.D thesis, Mumbai, India, Indian Institute of Technology – Bombay 2005

[9] X. Yao, L.G. Tham, F. C. Dai "Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China." Geomor. vol. 101, pp. 572–582

[10] L. Ermini, F. Catani, N. Casagli Artificial Neural Networks applied to landslide Susceptibility assessment," Geomorph. vol. 66, pp. 327-343, 2005

[11]. H. Gomez, T Kavzoglu "Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela." Engng Geol. vol. 78, pp. 11 –27, 2008

[12] R.T. Pack, D.G. Torboton, C.N. Goodwin. SINMAP 2.0 - A Stability Index Approach to Terrain Stability Hazard Mapping, User's Manual, Utah state university. 2005

[13] Cortes C. and Vapnik, V. N. "Support vector networks", Machine learning., *Springer*, **20**, pp. 273-297, 1995

[14]. Smola A. J. and Scholkopf, B. "A tutorial on support vector regression", *Statistics and computing*, *Springer*, **14**, Doi: 10.1023/B:STCO.0000035301.49549.88, pp. 199-222, 2004

[15]. V.N. Vapnik "Statistical learning theory", Wiley, New York. 1998.

[16] B.E. Boser, I.M. Guyon, V.N. Vapnik. "A training algorithm for optimal margin classifiers", *In: D. Haussler, ed., 5th Annual ACM Workshop on COLT. Pittsburgh, PA: ACM Press*, pp. 144-152 (1992).

[17] D.T. Bui, B Pradhan, O Lofman and I Revhaug Landslide Susceptibility Assessment in Vietnam Using Support Vector Machines, Decision Tree, and Naïve Bayes Models. Math. Prob. Engng. Article ID 974638, doi:10.1155/2012/974638, 2014

[18] I. Yilmaz Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environmental and Earth Science, vol 61(4), pp.821-836 ,2010

[19]. L. Peng, R. Niu, Bo Huang, X Wu, Y Zhao, R Ye, Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China, Geop=morphology 204(1), pp. 287-301, 2014

[20] Roslee R, TA Jamaluddin, MA Talip, "Integration of GIS using geostatistical interpolation in deterministic models for landslide susceptibility analysis in Kota Kanabalu, Sabha Malaysia." J. Geog. and Geol. vol. 4(1), pp. 18-32, 2012