Communality from equal weight for GIS-based landslide susceptibility mapping

Sandip Som (✉ sksom@rediffmail.com)  
Geological Survey of India  https://orcid.org/0000-0002-1134-4736

Saibal Ghosh  
Geological Survey of India

Soumitra Dasgupta  
Geological Survey of India

Thrideep Kumar  
Geological Survey of India

J. N. Hindayar  
Geological Survey of India

Murali Mohan  
Geological Survey of India

P. Dasanwar  
Geological Survey of India

Methodology

Keywords:

DOI: https://doi.org/10.21203/rs.3.rs-26084/v1

License: ☇️ 📧 This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Modeling landslide susceptibility is one of the important aspects of land use planning and risk management. Several modeling methods are available based either on highly specialized knowledge on causative attributes or on good landslide inventory data to use as training and testing attribute on model development. Understandably, these two criteria are rarely available for local land regulators. This paper presents a new model methodology, which requires minimum knowledge of causative attributes and does not depend on landslide inventory. As landslide causes due to the combined effect of causative attributes, this model utilizes communality (common variance) of the attributes, extracted by exploratory factor analysis and used for calculation of landslide susceptibility index. The model can understand the inter-relationship of different geo-environmental attributes responsible for landslide along with identification and prioritization of attributes on model performance to delineate non-performing attributes. Finally, the model performance is compared with the well established AHP method (knowledge driven) and FRM method (data driven) by cut-off independent ROC curves along with cost-effectiveness. The model shows it's performance almost at par with the established models, involving minimum modeling expertise. The findings and results of the present work will be helpful for the town planners and engineers on a regional scale for generalized planning and assessment.

1 Introduction

One of the main criteria for land use decisions is the effect of geoscientific attributes on the stability of land, which are rarely considered to be of importance (Lamelas et al., 2009), resulting from massive economic loss. Goal 11 of United Nation's 2030 agenda for sustainable development emphasizes resilience to disaster, develop and implement, in line with the Sendai Framework for Disaster Risk Reduction 2015-2030 (http://www.sustainabledevelopment.un.org). Like many other disasters, gravitational mass movement along downslope due to a complex interplay of geoscientific attributes causes landslides, triggered may due to natural causes (rainfall, earthquake, stream undercut, etc.) or unplanned human-modified slope instability. Between 1998-2017, landslides affected an estimated 4.8 million people and cause more than 18,000 deaths (http://www.who.int). This needs spatial prediction of stable to unstable land surfaces encompassing both visited and unvisited locations by interpolating spatially continuous surfaces, using different modeling methodologies and represented in the form of maps. These maps are indispensable for planning and decision making in land management, although usually not readily available and often difficult and expensive to acquire, especially for mountainous regions (Li et al., 2011).

Landslide susceptibility mapping is carried out either by knowledge driven methods (qualitative) or data driven methods (quantitative). Qualitative methods involve a combination of several geo-environmental attributes with predetermined weights to generate susceptibility map, mainly based on expert knowledge of the geoscientist (Anbalagan, 1992; Moon, 1990). The limitation in this approach is on subjectivity in decision rule which is highly personal and may include some virtual admission (Yalcin et al., 2011). Moreover, for the existing knowledge driven models, one should have good knowledge of the influence of
Geo-environmental attributes on slope instability, which may differ from terrain to terrain. Statistical methods are quantitative with numerical expressions (Aleotti and Chowdhury, 1999) such as logistic regression, frequency ratio, fuzzy logic, artificial neural networks, etc (Carrara et al., 2008; Carrara and Pike, 2008; Ercanoglu and Gokceoglu, 2004; Ghosh et al., 2011; Komac, 2006; Lee and Pradhan, 2007; Lee et al., 2004; Suzen and Doyuran, 2004; van Westen et al., 2008) all of which depends on apriori knowledge of landslide events. Understandably, it is quite difficult to assure the completeness of landslide inventory due to non-approachability at all sites on rugged mountainous terrains, rapid land-use change and obliteration of historic landslide signatures (Gariano and Guzzetti, 2016; Ghosh et al., 2012; Rowbotham and Dudycha, 1998; Westen et al., 2013) or the area might not have any previous history of the landslide (ex. Malin landslide in India, claimed 160 lives (Ering and Babu, 2016)).

To overcome these shortcomings, here we introduce a new mixed modeling method using both knowledge and statistics, which can fairly classify the area at a regional scale with simple judgments on geo-environmental attributes, using a widely used statistical tool and without any landslide inventory data. We present the model with an aim to address the following questions:

a. Can we understand the inter-relationship of different geo-environmental attributes responsible for landslide by this modeling technique?

b. Can we identify and prioritize the contribution of attributes on model performance and delineate non-performing attributes?

c. What are the proposed model performance and cost-effectivity concerning other widely-used models.

We coin this model as Equal Weight Method (EWM), as it starts with assuming the equal influence of all attributes causing landslides and ends up with a formulation of composite indicator to delineate the landslide susceptible zones.

2 Equal Weight Method

The weight assigned to an attribute cannot be directly related to the importance of that attribute in defining the composite indicator as the importance of the attribute depends on its correlation and variances within the system. To develop a statistically sound index, weights can be used as “scaling coefficients” by taking values between 0.5 to 1.0 (Becker et al., 2017). Assigning equal weights to different attributes are commonly employed for Resource Governance Index (Quiroz and Lintzer, 2013), Global Innovation Index (Dutta et al., 2014), Good Country Index (Anholt and Govers, 2014), as well as, for multi-model atmospheric forecasting (DelSole et al., 2013). The present model Equal Weight Method (EWM) is based on understanding the correlation between the geo-environmental attributes, or in other words influence of one attribute over the other, as landslides result from the interplay of different attributes (Carrara and Pike, 2008). In this method, initially, all attributes are considered to have a similar influence on landslide production and given equal weight (say 1). The weight of the parent attribute (1) is equally and systematically distributed to the sub-attributes considering their possible role with landslide hazards. At this stage, the modeler needs to make a simple judgment to rank the importance of sub-
attributes concerning landslide. For example, relative relief is an important attribute during landslide activity in mountainous terrain. We can divide relative relief into three sub-attributes, viz. Low (< 100 m), Medium (100 m-300 m) and High (> 300 m) and accordingly we can scale the influence of relative relief on landslide as low relative relief < medium relative relief < high relative relief. For giving weights to each of the sub-attribute, we distribute the original weight to three equal parts (i.e. 0.33) and add cumulatively from the lowest influence sub-attribute to highest vulnerable sub-attribute. In this process, low relative relief will have a weight of 0.33 and high relative relief will have 1.0 weight. This weighting scheme is applied to all attributes. The process will generate multivariate data, which we will take as input to find out the relationship between the attributes.

As landslides are caused due to complex interplay of the attributes, we should know the interrelation between attributes. To understand the interrelation in multivariate data by providing a parsimonious and meaningful explanation for the observed correlation in different attributes, we performed factor analysis to reduce variable complexity (Kerlinger, 1979). The basic idea of factor analysis is to reduce the dimensionality of observed attributes to unobservable latent attributes that share a common variance. An important postulate of factor analysis is that the internal attributes, which cannot be directly measured but their effects are reflected when one obtains measures on measurable attributes. These internal attributes are referred to as factors or latent attributes.

Mathematically, let A be the number of variables (X₁, X₂, X₃,……Xₐ) and B be the number of factors (F₁, F₂, F₃,……F₉), then the model assumes that each variable is a linear function of all the factors reproducing maximum correlation. Now, if Xᵢ is a variable, then;

\[ Xᵢ = fᵢ₁F₁ + fᵢ₂F₂ + fᵢ₃F₃ + ………+fᵢ₉F₉ \] ........................(1)

Where, \( fᵢ₁, fᵢ₂, fᵢ₃,……fᵢ₉ \) are factor loadings which give an idea that how much the variable contributed to each factor and it can vary from +1 to -1 (Dragon, 2006). Factor analysis used matrix algebra where the basic input is the correlation coefficient of the variables. Once the correlation matrix of variables are computed, factor loading can be computed by (Rummel, 1988);

\[ R_{BXB} = F_{BXA}F'_{AXB} + U^2_{BXB} \] ........................(2)

Where \( R_{BXB} \) is the correlation matrix, \( F_{BXA} \) denotes common factor loading (\( F'_{AXB} \) is the transpose) and \( U^2_{BXB} \) is a unique variance. The variance associated with any variable arises from three sources. The variance due to common factors is referred to as common variance or communality. The variance associated with a specific factor is referred to as a specific variance. Other than these two types of variances, another variance occurred due to error of measurement, referred to as error of measurement variance. The specific variance is commonly combined with an error of measurement variance to form the unique variance or uniqueness. The difference between \( R_{BXB} \) and \( U^2_{BXB} \) is the communality, which can be readily computed for each variable by adding a square of factor loadings against that variable. Communality coefficients are specific to measured attributes.
As we do not have any clear prior hypotheses regarding the factor structure underlying the attributes, the exploratory method of factor analysis is employed (Yong and Pearce, 2013). In this method, there are no apriori restrictions placed on the pattern of relationships between the observed attributes and latent attributes. As we want to keep the maximum variance in the solution and as we have 7 number of attributes, we extracted 6 factors, explaining about 92% of the total variance which is nearly equivalent to required variance as per Hair et al (1995) (Table-1). Factor-1 extracts maximum variance followed by other factors. A close look in factor loadings of attributes indicates that slope, lithology, relative relief and road have important contributions for factor-1. Out of these four attributes, lithology and road are positively correlated and slope and relative relief are negatively correlated with factor-1. This also indicates good interrelationship within these attributes and has maximum influence on the factor model. The other three attributes (viz. LULC, drainage, and lineament) have poor correlation with any factors indicating their interrelationships with other attributes are poor and have insignificant contributions for the factor model.

In the next stage, extracted communality of each variable (Table-1) is multiplied with the original distributed values of sub-attributes to obtain weighted influence (Table-2) of each variable in the mapping unit (grid cell). Summing up all weighted values of attributes gives the Landslide Susceptibility Index (LSI) of that particular cell. Higher values indicate higher landslide susceptibility and lower values indicate lower the susceptibility to landslide occurrence.

\[ \text{LSI} = C_1V_1 + C_2V_2 \ldots \ldots \ldots + C_nV_n \]  

Where, \( C_1, C_2, \ldots C_n \) is communality of \( V_1, V_2 \ldots \ldots V_n \) is attributes.

Now the question is how to fine-tune the model performance, i.e. to exclude the attributes from the model having minimum contribution in model performance. From the above discussion, we have seen that the main pillar of the EWM model is a common variance or communality which is specific to the measured attribute. This means that the attribute having more communality has more influence on model performance. This can be indirectly measured by stepwise standard deviation cumulative plot of LSI along Y-axis and cumulative no of attributes along the X-axis starting with the attribute of highest or lowest communality or directly plotting cumulative communality starting with the highest or lowest communality. The attributes having a minimum influence on model performance will form a break in slope and will trend nearly asymptotically with the X-axis due to their lower contribution of common variance.

3 Application Of Wqual Weight Method

3.1 Study area

Although the model generates a composite map giving information for landslide susceptibility of an area, it does not require any spatial structure information and hence it can be applied to any geographical area.
As mountainous terrains are more prone to land sliding and for any hilly terrain, road corridors are the main locales for the development activities, we have selected a well studied major road corridor of the eastern Himalaya in Sikkim state of India (Kumar N et al., 2017) connecting several densely populated townships and major hydroelectric projects which are affected by the frequent occurrence of landslides (Fig-1).

The elevation of the area varies between 320 meters to 3680 meters and there is a general increase in elevation from south to north. Due to high elevation differences, climatic conditions vary from tropical to alpine types from south to north with average annual rainfall between 300 to 400 cm. Monsoon generally sets in June and continues for about four months. The Tista river is one of the major rivers of Sikkim, originates from a Tista Khangse glacier in the north, passes through the entire study area from Chungthang up to Rangpo. Geologically, the area falls in the Lesser Himalayan and Central crystalline zone of the Sikkim-Darjeeling Himalayas. The rock types found in the area are represented by alternate meta-pelites and meta-psammites of different metamorphic grades belonging to the Daling group, high-grade gneiss and schist belonging to Darjeeling Gneiss Formations. We have mapped along 100 km. National Highway (NH-34A) and 44 km. of village roads (made up of gravels and often mixed with tar), delimiting the lateral boundaries between the Tista river on one side and water divide in the opposite hillside. For up to date landslide inventory, we have taken the help of a previous inventory catalog (Paul and Ghosal, 2009), remote sensing (Google Earth Imagery) and field verification using GPS survey. All landslides are debris slide (Varnes, 1978), mostly triggered due to heavy precipitation, favored by oriented rock mass discontinuities, joints, faults, or schistosity. A total of 55 number of landslides were mapped within an area of about 149 sq.km.

The Himalaya is a tectonically active region, geological structure (fault-lineament) and lithology are important controlling factors. Moreover, as landslide is a gravitational process, terrain slope and relative relief play important roles. Similarly, vegetation cover and distance from drainage are also found to play pivotal roles in landslide occurrences (Pradhan and Lee, 2010). In addition to natural controlling factors, anthropogenic activities, viz. road construction, land use-landcover, etc. also promote landslide. Based on the above importance, for the present study, we have selected lithology, land use-landcover, slope, relative relief, lineament-fault, drainage and road as causative geo-environment attributes responsible for landslides. We used the ASTER Digital Elevation Model (DEM) for the derivation of slope angle and local relative relief. Since ASTER DEM does not pick up water bodies and low in vertical resolution, we rectified the DEM by using the topographical map of Survey of India for water bodies and high-resolution field GPS data for vertical correction. Excluding drainage, road and lineament-fault layers, all other layers were classified following code provided by the Bureau of Indian Standard (BIS, 1998) which is widely used in the Indian sub-continent. For drainage, road and lineament-fault layers, we calculated systematic buffer intervals by using a GIS platform. The geo-environmental attributes were converted to a raster grid with 50 × 50 m cell size and used as a mapping unit with an assumption that each grid cell represents a spatially homogenous domain for application to the susceptibility modeling. The area grid consists of 61667 numbers of cells, out of which 55 cells occupy landslides. Details of the geo-environmental attribute are given by (Kumar N et al., 2017).
3.2 Landslide susceptibility map

The proposed model is used to prepare a landslide susceptibility map which can demarcate the areas having a likelihood or probability of landsliding. In the present work we have used natural breaks for five classes (viz. Very low, Low, Moderate, High and Very high) in GIS platform to produce the susceptibility map from the calculated LSI by EWM. We have calculated LSI taking into account all attributes i.e. total communality, derived for the model and described as EWMTC map (Fig-2a) and compared with the susceptibility map considering partial communality of the model (EWMPC, Fig-2b) by the exclusion of some attributes. The decision on exclusion of attributes from the model is taken by plotting cumulative standard deviation as described earlier (inset in fig-2b). This process excludes LULC, drainage, and lineament which individually represents less than 10% of total communality. Distribution of landslide shows the presence of 79% and 84% landslide per sq. km. in high and very high zones of EWM-TC and EWM-PC maps respectively within an aerial coverage of about 32% of the total mapped area, while very low and low susceptible areas although having cumulative coverage of about 42% and 45% for EWM-TC and EWM-PC maps respectively, the very low susceptible area does not contain any landslide and low susceptible area contain only 7% and 4% landslides per sq.km respectively. (Table-3).

To compare the performance of EWM based landslide susceptibility map, we have selected a well-referred Analytic Hierarchy Process (Saaty and Vargas, 2001) and the Frequency Ratio Method (Lee and Pradhan, 2007). The former method (AHP) is full knowledge driven based primarily experts opinion without taking into account known landslide incidences, while the later one (FRM) is fully data driven, in which the model solely depends on landslide inventory data. Details of model results by these two methods are given in table-2.

Figure-3a shows the landslide susceptibility map after AHP method with about 69% and 17% landslides per sq.km respectively for very high and high susceptible zones, covering about 26% of the mapped area while about 37% of the mapped area is demarcated as very low and low susceptible for landslides with a concentration of about 4% and 6% landslide per sq.km (Table-3). On the other hand, FRM based landslide susceptibility map (Figure-3b) delineate about 13% of the mapped area as very high to highly susceptible for a landslide with about 80% and 12% of landslides per sq.km respectively while about 73% of the mapped area is demarcated as very low and low susceptible for landslides with a concentration of about 1% and 4% landslide per sq.km (Table-3).

Comparison plots of all the three models (Figure-4) show a sharp break in slope of landslides per sq.km from high onwards, indicating all three models almost equally picked up the high and very high susceptible areas, but from medium to very low susceptible areas the curves for AHP and FRM models are almost parallel to X-axis, indicating low resolution of these models to clearly distinguish the classes of comparatively low landslide susceptibility, while both the models of EWM clearly distinguishes the very low susceptible area without any reported landslide incidences and maintains a low angle smooth gradient till the high susceptible zone reaches. Area percentage plots show high positively skewed area distribution pattern for the FRM model which is unrealistic, indicating the probability of more
misclassification of the areas to demarcate as low landslide-prone. AHP model shows a nearly Gaussian area distribution pattern, but here the problem is with a high proportion of medium landslide-prone areas, the class, which is difficult by any planner to decide on future developments. On the other hand, neither skewness nor high medium class is observed for the EWM model. EWM-TC nearly equally distributes the different classes, while fine-tuning to EWM-PC has the potential to make a bimodal distribution pattern separating high and low susceptible areas and comparatively low medium classes than AHP.

### 3.3 Statistical comparison

The computational power of a model depends on its sensitivity, or in other words its ability to minimize misclassification and constraining the attributes having a minimum effect on model results, at least for modeling (Douglas-Smith et al., 2020). The performance of the models is judged by ROC (Receiver Operating Characteristic) curves with a binary classification of stable and unstable units and is considered as one of the best performance evaluation techniques. The confusion matrix in a two-class classifier system is built up with rows representing observed class and the columns represent predicted class (Table - 4). Four conditions can arise in a confusion matrix; (i) mapping unit modeled as stable does not have any landslide (true negative, TN), (ii) mapping unit modeled as unstable does not have any landslide (false positive, FP), (iii) mapping unit modeled as stable have landslide (false negative, FN) and (iv) mapping unit modeled as unstable have landslide (true positive, TP). Dividing the entries in each row by the row total gives the classifiers probability of prediction. These are known as the true positive rate (TPR), false-positive rate (FPR), true negative rate (TNR) and false-negative rate (FNR). The ROC curve deals with the plotting of true positive rate (TPR) or sensitivity against false positive rate (FPR) showing all possible combinations of misclassifications and class distributions and it not only preserves the performance of the classifier but also allows to understand the relationship of the classifiers by visual inspection. ROC curves are prepared by standard curve fitting of several points representing confusion matrices calculated at each point for a range of probabilities. Area Under Curve (AUC) of ROC is used as a measure for model performance. The accuracy of tests with AUC between 0.50 and 0.70 is low; an accuracy between 0.70 and 0.90 is moderate, while an AUC over 0.90 indicates high accuracy (Streiner and Cairney, 2007).

Figure-5 shows ROC curves derived for all models. Calculated AUC of ROC curves for AHP, FRM, EWMTC and EWMPC models are 0.755, 0.809, 0.74 and 0.774 respectively. Considering the AUC utility in determining the model performance all of the above-tested models have moderate accuracy and slightly vary from each other (Streiner and Cairney, 2007). Although AUC can compare the different classification schemes directly, it has some drawbacks also, when the ROC curves cross over one another. When a classifier crosses over the other classifier, it means both the classifiers are the best performer in a certain range of points (Drummond and Holte, 2006) which we can see from the figure-5, where both EWMTC and EWMPC crosses over AHP and FRM at different points indicating they are superior to AHP and FRM at certain ranges. Moreover, all the models have definite misclassification as are evidenced from the convex hull of ROC curves (otherwise ROC curve would have merged with co-ordinate axes ). For any land classification scheme, identifying the stable land is important from the economic point of view because
the unstable land will be restricted in use (Frattini et al., 2010), which means the model should have a minimum false-negative count (type-II error). Higher false positive count (type-I error) will lower the extent of stable areas, but it will not be vulnerable as that of type-II error. A comparison of the ratio of misclassification counts (type-II/type-I) of all four models shows that EWMPC performs better than FRM and AHP from 40% and 85% probability onwards. At 100% probability, this ratio is much lower for EWMPC than that of AHP, FRM, and EWMTC indicating EWMPC is a comparatively better performer (figure-6). However, at this stage, we do not know the misclassification cost and class probabilities to choose the best performer. Hence it is important to identify a model with minimum misclassification cost having a definite probability range.

To achieve this goal, (Drummond and Holte, 2006) proposed cost curve analysis, where a point in the ROC curve \((X, Y)\) corresponds to a line segment in the cost curve that has \(Y = FPR\) when \(X = 0\) and \(Y = FNR\) when \(X = 1\). Equation of this line is given by:

\[
Y = (FNR - FPR) \times X + FPR \quad \text{…………………………(i)}
\]

Where \(X, Y\) represents probability and normalized expected cost respectively. In the cost analysis, the minimum cost is involved for correctly classifying and misclassification cost is always higher.

However, we can not use directly the above equation because the equation is meant to evaluate models having the capability to measure maximum positive cases. This type of analysis is mainly performed in bio-medical, signal processing, etc. where positive cases are the goal sought for. Hence the model should have predictability power with minimum false positive cases as false negative is less vulnerable, although having its importance in total misclassification cost. But for landslide susceptibility we are mainly interested in stable areas i.e, areas with no or minimum landslide with a different representation of confusion matrix as shown in table-4 (Frattini et al., 2010) than that of (Drummond and Holte, 2006). In this case, false negative (area demarcating stable but not stable) is more important than false positive (area demarcating unstable but stable). In this case equation (i) has to be rewritten with \(Y = FNR\), when \(X = 0\) and \(Y = FPR\), when \(X = 1\) and can be written as:

\[
Y = (FPR - FNR) \times X + FNR \quad \text{…………………………(ii)}
\]

\(X\) and \(Y\) remain the same as stated earlier.

Figure-7 shows cost analysis of all the models having a maximum bounding cost of 0.4 which indicates that all models are cost-effective. Out of the four models, FRM is more cost-effective followed by AHP and EWM. The maximum cost difference between FRM and EWM is only 0.13 (A-B line in figure-7). However, it can be seen that at < 0.25 probability, all four models behave almost similarly. Within a probability range of 0.25 to 0.5, FRM outperforms, followed by AHP and both EWM models respectively. From 0.5 to 0.78, FRM, AHP, and EWMPC behave similarly and outperform to that of EWMTC. At > 0.78, all models behave similarly. This indicates that EWM models can perform equally well with that of other
well-established models for at least three fourth of the entire probability range with a competitive bounding cost.

All of the above metrics indicate the preamble on equal weight method fulfills pragmatically the desire envisaged on a sensible real-world constraint. Land regulatory bodies may not have highly specialized knowledgeable persons to categorize stability of land-based on his experience or the area may not have sufficient landslides to use for training and testing with standard data driven methods. This shows the robustness of proposed variance-based EWM methodology, in which the modeler classifies the area with minimum knowledge on attributes intrinsic properties and without any landslide checks, even with the opportunity for fine-tuning the model performance by the exclusion of insignificant attributes for that particular area.

5 Conclusion

Landslide susceptibility mapping lays the foundation for land hazard management and is of great help to planners and engineers in choosing suitable locations for development. With the available techniques, a good landslide susceptibility map can be produced either by a person having sound knowledge on effects of Geo-environmental attributes of the studied area or the area to be mapped should have a nearly complete inventory data. The proposed Equal-Weight Method depends on the mapper’s preliminary terrain specific knowledge and it does not depend on landslide inventory data. Common variance (communality) is extracted by factor analysis technique from the data assigned by the modeler for each attribute. This communality is used to weight the original data and finally, LSI is calculated by summing up the individual weighted attributes.

Factor loadings imply a correlation between attributes and factors. Factor analysis of the proposed model indicates a good correlation between lithology, relative relief, slope and road with the latent variable, factor-1, indicating good intercorrelation between these variables which has a major influence on the EWM model. Other attributes (LULC, drainage, and lineament ) have insignificant correlations with any factors, which also indicate their poor interrelationships with other attributes.

Landslide is a result of the cumulative effect of different attributes and in turn dependent on the communality of individual attributes. To prioritize the contribution of attributes on model performance and to delineate non-performing attributes, cumulative stepwise LSI plot or communality plots are suggested which can indicate the non-influential attributes in model performance. We have generated two models by using EWM. The first one takes into account of all communality weighted attributes (EWMTC) and the second one with selective deletion of attributes (EWMPC) based on the above criterion.

To understand the acceptability of EWMTC and EWMPC models, we carried out ROC and cost analysis and compared them with well-referred knowledge driven (AHP) and data driven (FRM) models. AUC of ROC curves for AHP, FRM, EWMTC, and EWMPC models are 0.755, 0.809, 0.74 and 0.774 respectively indicating little performance difference between models. However, as the ROC curves cross each other, it does not indicate model performances sensu stricto. The ratio of misclassification counts indicates the
best performance of EWMPC for a wider probability range. Cost effectivity analysis indicates that EWM models can perform equally well with that of other well-established models for at least three fourth of the entire probability range with a competitive bounding cost.

All of these indicate the robust predictive power of the EWM method for delineating the landslide susceptible. The findings and results of the present work will be helpful for the town planners and engineers on a regional scale for generalized planning and assessment. This type of map can be prepared without taking help from persons with specialized knowledge of environmental attributes and without any landslide inventory data defining a sensible real-world constraint. However, the application of the proposed methodology for local and site-specific studies needs to be tested.

**Declaration**

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request. There is no competing interest. The work was funded by Geological Survey of India. SKS contributed mainly on the research work by developing model for landslide study and prepared 1st draft of the manuscript. GIS inputs and model comparisions are from SG and SDG. TK, JNH, MM and PD were involved in field data generation and compilation.

The authors are grateful to National Mission Head, III, GSI for kind permission to publish the paper. The authors are also indebted to their colleagues for helpful suggestions to upgrade the contribution.

**References**

Aleotti, P., Chowdhury, R. 1999. Landslide hazard assessment: summary review and new perspectives. *Bulletin of Engineering Geology and the Environment* 58(1): 21-44.

Anbalagan, R. 1992. Landslide hazard evaluation and zonation mapping in mountainous terrain. *Engineering Geology* 32(4): 269-277.

Anholt, S., Govers, R. 2014. The Good Country Index: Technical report. *The Good Country Party*.

Becker, W., Saisana, M., Paruolo, P., Vandecasteele, I. 2017. Weights and importance in composite indicators: Closing the gap. *Ecological Indicators* 80: 12-22.

BIS, 1998. Preparation of Landslide Hazard Zonation maps in mountainous terrain-guidelines (IS 14496, Part 2). *Indian Standards*. New Delhi India 1- 19.

Carrara, A., Crosta, G., Frattini, P. 2008. Comparing models of debris-flow susceptibility in the alpine environment. *Geomorphology* 94(3-4): 353-378.

Carrara, A., Pike, R.J. 2008. GIS technology and models for assessing landslide hazard and risk. *Geomorphology* 94(3-4): 257-260.
DelSole, T., Yang, X., Tippett, M.K. 2013. Is unequal weighting significantly better than equal weighting for multi-model forecasting? *Quarterly Journal of the Royal Meteorological Society* 139(670): 176-183.

Douglas-Smith, D., Iwanaga, T., Croke, B.F.W., Jakeman, A.J. 2020. Certain trends in uncertainty and sensitivity analysis: An overview of software tools and techniques. *Environmental Modelling & Software* 124: 104588.

Dragon, K. 2006. Application of factor analysis to study contamination of a semi-confined aquifer (Wielkopolska Buried Valley aquifer, Poland). *Journal of Hydrology* 331(1): 272-279.

Drummond, C., Holte, R.C. 2006. Cost curves: An improved method for visualizing classifier performance. *Machine Learning* 65(1): 95-130.

Dutta, S., Lanvin, B., Wunsch-Vincent, S., Insead 2014. The global innovation index 2014: the human factor in innovation. World Intellectual Property Organization. *Johnson Graduate School of Management*. OCLC: 916514895. Cornell University.

Ercanoglu, M., Gokceoglu, C. 2004. Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey). *Engineering Geology* 75(3-4): 229-250.

Frattini, P., Crosta, G., Carrara, A. 2010. Techniques for evaluating the performance of landslide susceptibility models. *Engineering Geology* 111(1-4): 62-72.

Gariano, S.L., Guzzetti, F. 2016. Landslides in a changing climate. *Earth-Science Reviews* 162: 227-252.

Ghosh, S., Carranza, E.J.M., van Westen, C.J., Jetten, V.G., Bhattacharya, D.N. 2011. Selecting and weighting spatial predictors for empirical modeling of landslide susceptibility in the Darjeeling Himalayas (India). *Geomorphology* 131(1): 35-56.

Ghosh, S., van Westen, Cees J., Carranza, Emmanuel John M., Jetten, Victor G., Cardinali, Mauro., Rossi, Mauro., Guzzetti, Fausto. 2012. Generating event-based landslide maps in a data-scarce Himalayan environment for estimating temporal and magnitude probabilities. *Engineering Geology* 128: 49-62.

Kerlinger, F.N. 1979. Behavioral research:A conceptual approach. *Rinehart&Winston NewYork*. Holt.

Komac, M. 2006. A landslide susceptibility model using the Analytical Hierarchy Process method and multivariate statistics in perialpine Slovenia. *Geomorphology* 74(1-4): 17-28.

Kumar N, T., Hindayar, J.N., Mohan M., Dasarwar, P., Ibrahim, M., Som, S.K. 2017. Population based binary reclassification of Indian standard landslide hazard model. *Journal of the Geological Society of India* 89(2): 175-182.

Lamelas, M.T., Hoppe, A., de la Riva, J., Marinoni, O. 2009. Modelling environmental variables for geohazards and georesources assessment to support sustainable land-use decisions in Zaragoza.
(Spain). *Geomorphology* 111(1): 88-103.

Lee, S., Pradhan, B. 2007. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* 4(1): 33-41.

Lee, S., Ryu, J.-H., Won, J.-S., Park, H.-J. 2004. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. *Engineering Geology* 71(3-4): 289-302.

Li, J., Heap, A.D., Potter, A., Daniell, J.J. 2011. Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software* 26(12): 1647-1659.

Martinez, M.E., Marshall, J.R., Sechrest, L. 1998. Invited Commentary: Factor Analysis and the Search for Objectivity. *American Journal of Epidemiology* 148(1): 17-19.

Moon, W.M. 1990. Integration of geophysical and geological data using evidential belief function. *IEEE Transactions on Geoscience and Remote Sensing* 28: 711-720.

Paul, C., Ghosal, T.B. 2009. An inventory of major landslides in Sikkim-Darjeeling Himalaya. *Special Publication* No. 94. Geological Survey of India, Kolkata.

Pradhan, B., Lee, S. 2010. Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. *Landslides* 7(1): 13-30.

Quiroz, J.C., Lintzer, M. 2013. The 2013 Resource Governance Index, Technical Report. *The Revenue Watch Institute*

Rowbotham, D.N., Dudycha, D. 1998. GIS modelling of slope stability in Phewa Tal watershed, Nepal. *Geomorphology* 26(1-3): 151-170.

Rummel, R.J. 1988. Applied Factor Analysis. *Northwestern University Press* 1- 617.

Saaty, T.L., Vargas, L.G. 2001. Models, methods, concepts & applications of the analytic hierarchy process. *International series in operations research & management science*. Springer, New York, 1- 333.

Streiner, D.L., Cairney, J. 2007. What’s under the ROC? An Introduction to Receiver Operating Characteristics Curves. *The Canadian Journal of Psychiatry* 52(2): 121-128.

Suzen, M.L., Doyuran, V. 2004. Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. *Engineering Geology* 71(3-4): 303-321.

van Westen, C.J., Castellanos, E., Kuriakose, S.L. 2008. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering Geology* 102(3-4): 112-131.
Varnes, D.J. 1978. Landslides: Analysis and Control. In: R.L. Schuster and R.L. Krizek (Editors), Slope Movement Types and Processes, Transportation Research Board, *National Academy of Sciences, Washington, DC*, pp. 11-33.

Westen, C.v., Ghosh, S., Jaiswal, P., Martha, T., Kuriakose, S. 2013. From landslide inventories to landslide risk assessment; an attempt to support methodological development in India. *Landslide Science and Practice*. Springer.

Yalcin, A., Reis, S., Aydinoglu, A.C., Yomralioglu, T. 2011. A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. *Catena* 85(3): 274-287.

Yong, A.G., Pearce, S. 2013. A Beginner’s Guide to Factor Analysis: Focusing on Exploratory Factor Analysis *Tutorials in Quantitative Methods for Psychology* 9(2): 79-94.

### Tables

**Table 1.** Factor matrix with communality. RR: Realative relief; LULC: Land use land cover.

| Attributes | Factor | 1    | 2    | 3    | 4    | 5    | 6    | Communality |
|------------|--------|------|------|------|------|------|------|-------------|
| Lithology  |        | .546 | .345 | -.119| -.049| -.011| -.112| 0.446       |
| RR         |        | -.410| .423 | -.159| .063 | -.097| .024 | 0.387       |
| Slope      |        | -.607| .196 | .251 | -.090| .102 | -.033| 0.49        |
| LULC       |        | .089 | .055 | .195 | .295 | -.133| .050 | 0.156       |
| Lineament  |        | .017 | .076 | -.257| .103 | .209 | .124 | 0.142       |
| Road       |        | .446 | .218 | .316 | -.050| .093 | .089 | 0.365       |
| Drainage   |        | .033 | .023 | -.050| -.230| -.161| .144 | 0.104       |
| Variance   |        | 23.06| 15.31| 14.82| 14.27| 12.86| 11.28|             |

**Table 2.** Weights on Geo-environmental attributes by EWM, AHP and FRM.
| Variables   | Sub-variables             | EWM   | AHP   | FRM   |
|------------|---------------------------|-------|-------|-------|
| **LITHOLOGY** |                           |       |       |       |
| Gneiss     |                           | 0.065351 | 0.014 | 0.723 |
| Schist     |                           | 0.130702 | 0.041 | 0.705 |
| Phyllite & Quartzite |           | 0.196053 | 0.023 | 0.459 |
| Alluvium   |                           | 0.261404 | 0.224 | 0.000 |
| Older debris |                         | 0.326755 | 0.135 | 0.340 |
| Older alluvium fill |           | 0.392106 | 0.077 | 0.670 |
| Younger loose debris |           | 0.457457 | 0.346 | 3.310 |
| **RELATIVE RELIEF** |                       |       |       |       |
| Low        |                           | 0.129204 | 0.008 | 2.105 |
| Medium     |                           | 0.258408 | 0.031 | 2.716 |
| High       |                           | 0.387612 | 0.107 | 0.944 |
| **LINEAMENT** (Buffer distance in meter) |             |       |       |       |
| 501-1000   |                           | 0.012987 | 0.004 | 1.607 |
| 301-500    |                           | 0.025974 | 0.005 | 0.972 |
| 201-300    |                           | 0.038961 | 0.008 | 0.784 |
| 151-200    |                           | 0.051948 | 0.011 | 1.551 |
| 101-150    |                           | 0.064935 | 0.017 | 0.863 |
| 76-100     |                           | 0.077922 | 0.025 | 0.701 |
| 51-75      |                           | 0.090909 | 0.036 | 0.739 |
| 26-50      |                           | 0.103897 | 0.051 | 0.382 |
| 0-25       |                           | 0.116883 | 0.070 | 0.806 |
| **ROAD** (Buffer distance in meter) |            |       |       |       |
| 501-1000   |                           | 0.034299 | 0.013 | 0.065 |
| 301-500    |                           | 0.068598 | 0.017 | 0.177 |
| 201-300    |                           | 0.102897 | 0.025 | 0.547 |
| 151-200    |                           | 0.137196 | 0.037 | 0.473 |
| 101-150    |                           | 0.171495 | 0.056 | 1.065 |
| 76-100     |                           | 0.205794 | 0.082 | 1.513 |
| 51-75      |                           | 0.240093 | 0.120 | 1.850 |
| 26-50      |                           | 0.274392 | 0.169 | 5.167 |
| 0-25       |                           | 0.308691 | 0.233 | 5.422 |
| **DRAINAGE** (Buffer distance in meter) |          |       |       |       |
| 501-1000   |                           | 0.00888 | 0.001 | 0.000 |
| 301-500    |                           | 0.01776 | 0.001 | 0.336 |
| 201-300    |                           | 0.02664 | 0.002 | 0.624 |
| 151-200    |                           | 0.03552 | 0.003 | 0.797 |
| 101-150    |                           | 0.04444 | 0.004 | 1.791 |
| 76-100     |                           | 0.05328 | 0.006 | 1.088 |
| 51-75      |                           | 0.06216 | 0.009 | 0.851 |
Table 3. Area coverage and number of landslides on different susceptibility class of the models.

| Susceptibility class | EWMTC | EWMPC | AHP | FRM |
|----------------------|-------|-------|-----|-----|
|                      | Area (km²) | Lanslide (number) | Area (km²) | Lanslide (number) | Area (km²) | Lanslide (number) | Area (km²) | Lanslide (number) |
| Very low             | 26.12 | 0     | 29.63 | 0     | 21.87 | 3     | 61.85 | 4 |
| Low                  | 36.97 | 6     | 37.89 | 4     | 32.93 | 6     | 46.11 | 9 |
| Medium               | 39.04 | 13    | 33.07 | 9     | 55.79 | 8     | 21.02 | 5 |
| High                 | 31.54 | 15    | 31.97 | 16    | 28.46 | 16    | 14.62 | 11 |
| Very high            | 15.13 | 21    | 16.24 | 26    | 9.75  | 22    | 5.2  | 26 |

Table 4. Confusion matrix for ROC.

| Observed | Predicted |
|----------|-----------|
|          | Stable    | Unstable |
| Stable   | True Negative | False Positive (Type-I error) |
| Unstable | False Negative (Type-II error) | True Positive |
Figure 1

Location and extent of studied area as depicted on hillshade map generated from DEM showing the location of landslides.
Figure 2

A. Landslide probability map obtained after the EWM model considering total communality. B. Landslide susceptibility map after the EWM model considering partial communality. Inset shows a cumulative standard deviation plot starting with the highest standard deviation with only C1V1 and then gradually adding one after other CnVn with highest to lowest standard deviations.
Figure 3A was omitted by the authors in this version of the paper. A. Landslide susceptibility map after the AHP model. B. Landslide susceptibility map after the FRM model.
Figure 4

Plot showing area percentage and landslide concentration of classified susceptibility maps of different models.
Figure 5

ROC curve for all models.
Figure 6

Comparison of the Misclassification ratio of all models.
Figure 7

Cost curves for all models. A-B is the maximum cost difference.