Sample Position Affect Landslide Susceptibility Models in Hotspot Area of Nam Ma Basin, Lai Chau, Viet Nam

EGU2020-18278

Van-Trung Chu [1*], Shou-Hao Chiang [1,2], and Tang-Huang Lin[1]
1. Center for Space and Remote Sensing Research, National Central University, Taoyuan 32001, Taiwan
2. Department of Civil Engineering, National Central University, Taoyuan 32001, Taiwan
*Corresponding Author, +886-966765425, E-mail: chuvantrung@tuaf.edu.vn

Introduction

Data sampling is a necessary procedure to train a landslide susceptibility model. Different sampling strategies have been considered in previous studies, but few of them have focused on its effect on modeling performance. The aim of this study to analyze the effect of landslide sampling strategies on landslide susceptibility modeling. To conduct the analysis, the cutting-edge landslide sampling technique and sample partition approach were applied to produce three different subsets of training and testing data: the highest point within landslide polygon (SUB1), the centroid of landslide polygon (SUB2), and the highest point within the seed cell area of the landslide polygon (SUB3). In addition, the optimal strategy for non-landslide samples was also proposed. In this study, there are 18 landslide conditioning factor were considered, in which the land surface disturbance index (LSDI) and high above the nearest drainage (HAND) were newly introduced in the landslide susceptibility modeling. In order to have a more objective judgment about the main issue mentioned above, instead of using only one model, we applied three different models namely Random forest (RF), Logistic regression (LR) and Decision tree (DT) to perform three kinds of scenarios by difference subsets of landslides.

Objective

This study aims to explore the effect of landslide sampling strategies on landslide susceptibility modeling in the downstream area of Nam Ma watershed.

Figure 1. Study area and distribution of landslide inventory

Methodology

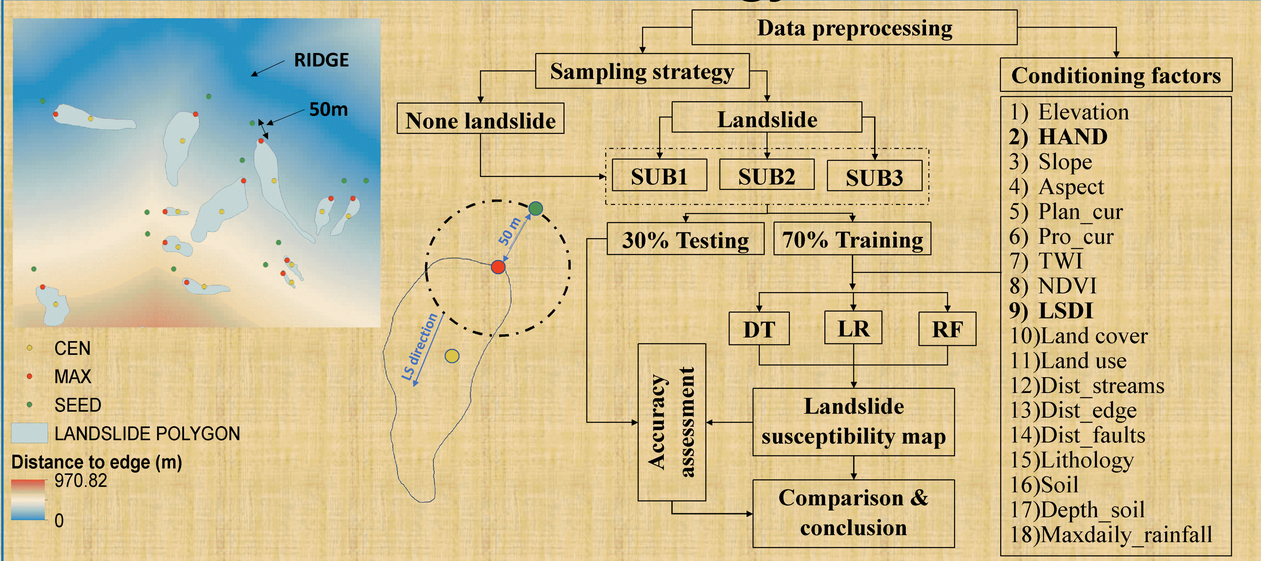
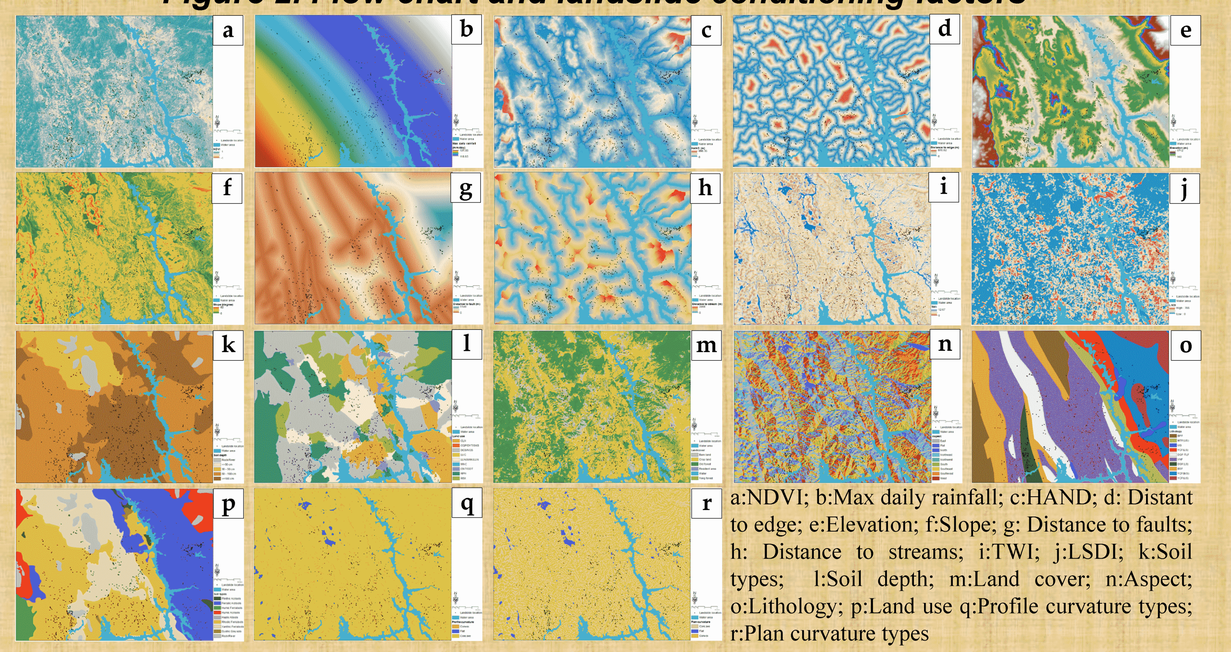


Figure 2. Flow chart and landslide conditioning factors



a:NDVI; b:Max daily rainfall; c:HAND; d: Distant to edge; e:Elevation; f:Slope; g: Distance to faults; h: Distance to streams; i:TWI; j:LSDI; k:Soil types; l:Soil depth; m:Land cover; n:Aspect; o:Lithology; p:Land use q:Profile curvature types; r:Plan curvature types

Figure 3. Maps of landslide conditioning factors:

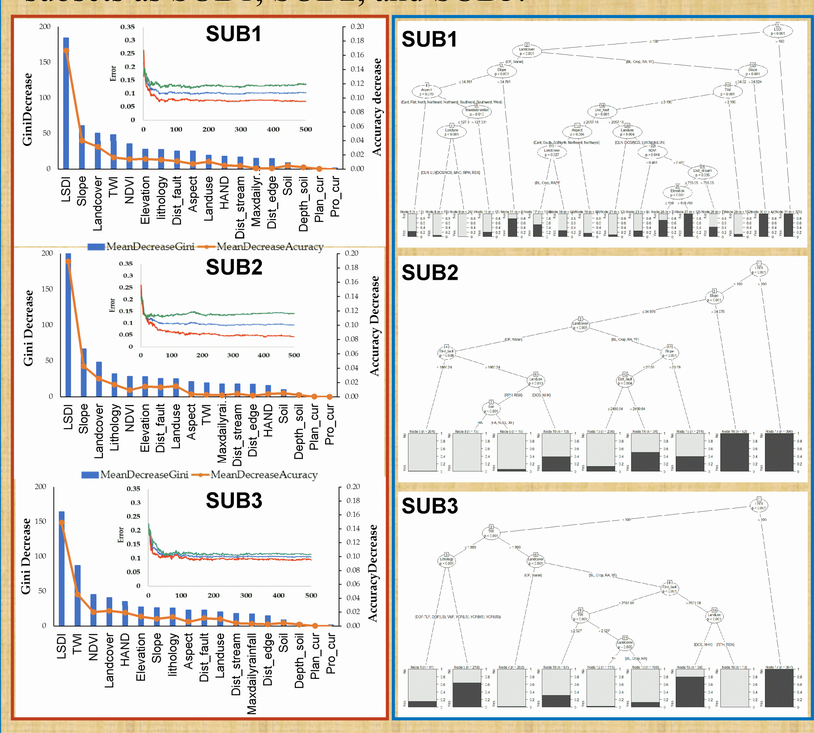
References

1. Süzen, M. L. and V. J. E. g. Doyuran (2004). "Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey." **71**(3-4): 303-321.
2. Lai, J.-S., Chiang, S.-H., & Tsai, F. (2019). Exploring Influence of Sampling Strategies on Event-Based Landslide Susceptibility Modeling. ISPRS International Journal of Geo-Information, 8(9), 397.
3. Kennedy, R. E., Yang, Z., & Cohen, W. B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. Remote Sensing of Environment, 114(12), 2897-2910

Results

- **Training model**

The RF, DT, and LR model generated in RStudio software using the same three training subsets as SUB1, SUB2, and SUB3.



Variable	Coefficients		
	SUB1	SUB2	SUB3
Slope	0.099**	0.15**	-0.006
Pro_cur	0.166	0.177	0.191
Dist_fault	0.0004**	0.0001**	0.0001**
Dist_stream	-0.0006	0.001	-0.001**
LSDI	0.025**	0.024**	0.023**
HAND	0.004**	0.001	0.008**
Dist_edge	-0.0002	0.0001	0.0001
Plan_cur	-0.012	0.052	0.017
NDVI	-0.936	-0.241	-1.086
Maxdailyrainfall	0.011	0.054	0.054
TWI	-0.11	0.234**	-0.54**
Aspect	0.961**	0.703*	1.047**
Lithology	-0.147	-0.125	-0.011
Landuse	-0.119	-0.04	0.179
Soil	0.499	0.602*	0.098
Depth soil	1.087*	0.592	1.266**
Landcover	0.06	0.032	0.08
Elevation	-0.003*	-0.003**	-0.003**
Constant	-6.705	-13.759	-8.918

* P-value < 0.05, ** P-value < 0.001

Figure 4. Model training performance: RF: OOB curve and variables important rank (left); DT: Tree plots (middle), and LR: Coefficients (right)

- **Landslide susceptibility maps**

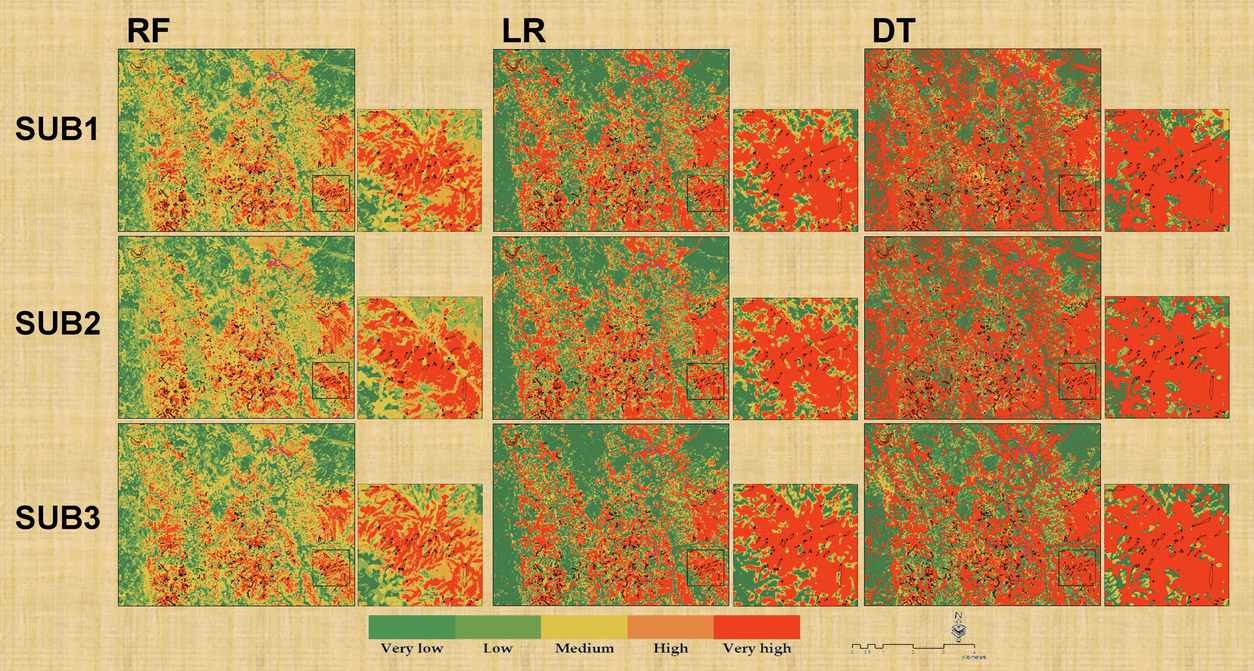


Figure 5. Landslide susceptibility maps (enlarge area on the right site)

- **Validation results**

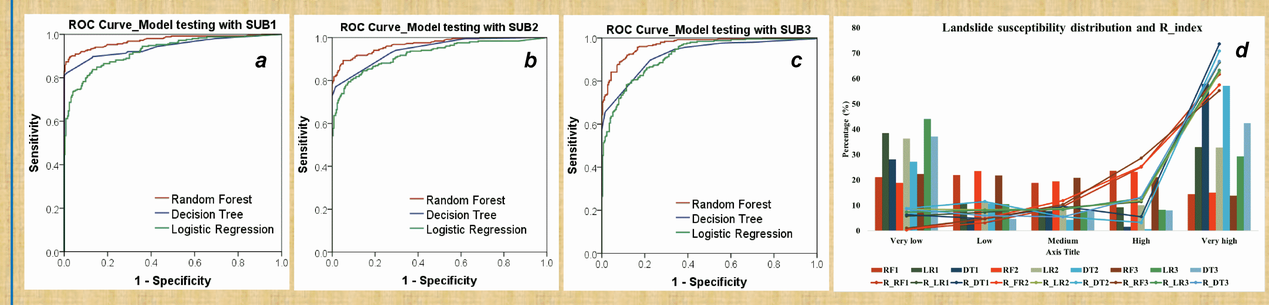


Figure 6. Accuracy of model validation and ROC_{AUC} plots (a,b,c), and the chart of landslide susceptibility distribution classes and relative density landslide index (d)

Table. Model validation

Index	SUB1			SUB2			SUB3		
	RF	LR	DT	RF	LR	DT	RF	LR	DT
Overall accuracy	0.921	0.848	0.881	0.911	0.846	0.864	0.907	0.822	0.836
Kappa	0.842	0.696	0.763	0.822	0.692	0.727	0.814	0.644	0.672
Recall (PA)	0.913	0.830	0.893	0.862	0.826	0.727	0.901	0.826	0.897
Precision (UA)	0.928	0.861	0.873	0.956	0.860	1.000	0.910	0.820	0.799
F1-score	0.920	0.845	0.883	0.906	0.843	0.842	0.907	0.8223	0.845
AUC	0.972	0.929	0.943	0.969	0.923	0.945	0.970	0.919	0.924
Std. Error	0.007	0.011	0.011	0.007	0.012	0.009	0.006	0.011	0.012

Conclusions

- Overall, our work were successfully employed the effect of different sample possession using three multivariate models. The results are an a abundance of consistency performance as the SUB1 gave the highest accuracy, followed by SUB2, and the lowest is SUB3. However, in terms of PA and UA indices, the SUB2 seems to be made underestimate (RF and DT). According to the result, LSDI, slope, TWI, landcover, and NDVI are the most important factors. Remarkably, LSDI contribution indicating its capability in landslide susceptibility modeling. In terms of model application, the Random forest is much recommended based on its contribution to our case study.