GEO-SPATIAL MODEL VALIDATION IN THE CONTEXT OF LOW-LAND CONVERSION DUE TO HOUSEHOLDS EMERGING IN THE COLOMBO SUB-URBAN

G.P.T.S Hemakumara*
Department of Geography, Faculty of Humanities and Social Sciences, University of Ruhuna, Matara, Sri Lanka
Ruslan Rainis
Geography Section, School of Humanities, University Sains, Malaysia

Abstract

This study examines how individual households have been emerging in the low lying areas during the period 2005-2012 in the core study area of Colombo Metropolitan Region (CMR) and also the process through which they have gradually established themselves as either stable or unstable households. Mass manipulation of geo-spatial factors in innumerable land plots has inevitably led to increasing negative environmental effects in the region. Hence, an attempt is made to build a geo-spatial model that can be used as a guide and index to help understand how the unending process of individual households emerging in the CMR. The primary focus of the study is comparison of main model data with validation model; hence, validation procedure of this study has been explained and compared with the main model. The typical individual household plot has been chosen as the unit of analysis. Information from 408 households was collected from the core study area to build the main model and perform validation and then the data was tested with a spatial logistic regression model. The main model indicated an accuracy of about 92.2% together with high significance levels for 8 variables out of the total 19 variables namely Household income (HI), Ground water surface level (GWS), Public participatory practice (PPP),Permanent plants growing in plot (PPG), Rain water remaining in the plot (RWR), Skilled jobs (SKJ), Technical skills and adaptation (TSA) and Low lying related plants availability (LLP). Predicted probability value of each housing plot mapped with GIS can be seen with the spatial distribution displayed clearly. Accuracy of the model validation process is 85.09% that indicates compatibly well. Based on predicted probability value of each land plot, both models have been run together with field data from the geo-spatial information system

Keywords: GIS, Logistic Model, Urban Planning, Land use Planning, Model Validation.

* Corresponding Author: GPTS Hemakumara; E-mail- shemakumara@hotmail.com

INTRODUCTION

The uncontrolled household emergence pattern due to the high demand for dwellings is one reason for the undesirable land conversion in urban city edges and sprawls (Carrion-Flores & Irwin, 2004; Correia, 1993). The actions of the householders who are dispersed over a wide area are very difficult to monitor until they become a huge issue. Eventually, this has threatened the entire sustainability prospects of large areas, which include eco-systems and large tracts of other environmentally sensitive lands (Chu et_al., 2010). While some of the problems are due to land filling by dumping earth and scrap materials onto the low lying areas (Barr, 2007; Oskamp et al., 1991), Other reasons such as inadequate garbage disposal, domestic water pollution, poor human waste disposal, bad drainage and flood control also impact negatively on the sustainability of cities. These are matters that require careful monitoring and control.

The Colombo Metropolitan region can be cited as an example that faces the problem of low lying land conversion issues, set in a tropical monsoonal climate. It is environmentally very sensitive during the monsoon periods and even during the inter-monsoonal rains (Divigalpitiya et_al., 2007). Therefore, bad practices occurring in environmentally sensitive areas have been identified and these mostly have to do with micro level land filling, encroachment, waste dumping etc. Therefore, it is necessary to monitor the household behavior and scatter pattern of households in low lying areas in their various aspects to carry out comprehensive modeling to understand the problem. This would help to pinpoint the various issues and devise appropriate solutions to overcome them in order to achieve a sustainable urban environment.

Clearly, the Geographic Information System in combination with Remote Sensing is one of the best tools to monitor above issues. However, according to (Staal et al., 2000) social aspects also exert much influence on the issues; but household linking in the field is one of the components lacking in GIS science. Even if many sub-models have been built in the field of land use with GIS, they can be used only to monitor the changes in the process (Liverman, 1998). Therefore, in order to gather additional data it is also necessary to survey the houses by conducting interviews and using questionnaires because socio-economic and behavior aspects are the main influential factors involved in the changing of low lying areas brought about by the emergence of households.

The main objective of this study is to focus on the main model’s data validation after running it with data from the particular conversion areas.

Spatial Distribution of Households in Study Area

The brown color plots in Figure 1 represent the spatial units (housing plots) of the model validation process. According to the questionnaire survey conducted in 2012, the 114 houses in the validation model process have been classified into two groups as non-stable (76) and stable (38), and these have been taken as the dependent variables of the model validation process. In addition, all housing plots in the figure that are marked in red indicate the 46 houses already converted before the year 2005. The 294 yellow color plots indicate all the non-stable houses that have existed since 2005 and these constitute the main database.

In addition to the above housing plots, the study area also contains some abandoned houses and low land plots numbering 35 that are mapped into the GIS. These plots too have the potential for future conversion. Table 1 indicates how the status of the houses has been changing over time within the low lying areas.

Stable and Non Stable Houses in Low Lying Areas

The following information provided by the government relating to the definition of stable house was used when some householders were uncertain about the status of their house's condition. The pre-questionnaire survey has been used only for demarcate the stable houses and non-stable houses in the study area.

a.) Structural materials

The Department of Census and Statistics of Sri Lanka (DCSSL) conducts a national population and housing survey every decade covering the whole country. They record information in respect of the structural materials used in the construction of the walls, roof and floor of every house they visit. Accordingly, the house is classified as being a permanent or non-permanent structure. This has been used as a guide to distinguish between stable and non-stable houses in the low lying areas.

b.) Infrastructure availability:

Structural materials of construction are not the only indicators used to determine the status of permanency of a household. In fact, the availability of infrastructure to an individual household has greater potential to decide its status as stable or non-stable in the low lying lands. On the other hand, only houses that have already been constructed of good materials will be able to incorporate many of the infrastructure facilities that will enable the household to become more stable. The following infrastructural amenities are some of the desirable ones that help to transform a non-stable house into a stable house as noted from the answers given namely Electricity supply , Pipe borne water supply, Accessibility (good roads), Sanitary facilities: Toilet, Sanitary facilities: Bathing and Fence marking the boundaries of the plot. Conversion of individual housing structures from non-stable to stable depend on above mentioned housing elements; however, it is not attained at the inception as in the case of a properly planned and designed housing project. The stable condition can only be achieved after incorporating at least some of the above elements into the household at different stages over a period of time.

Table 1: Summary of Status of Houses in the Core Study Area

No.	Status	Year	Number of Houses / Plots
1	Converted houses	Before 2005	46
2	Non-converted houses	As at 2005	294
3	Newly built houses	Between 2005 and 2012	114
4	Abandoned houses	Unknown	35

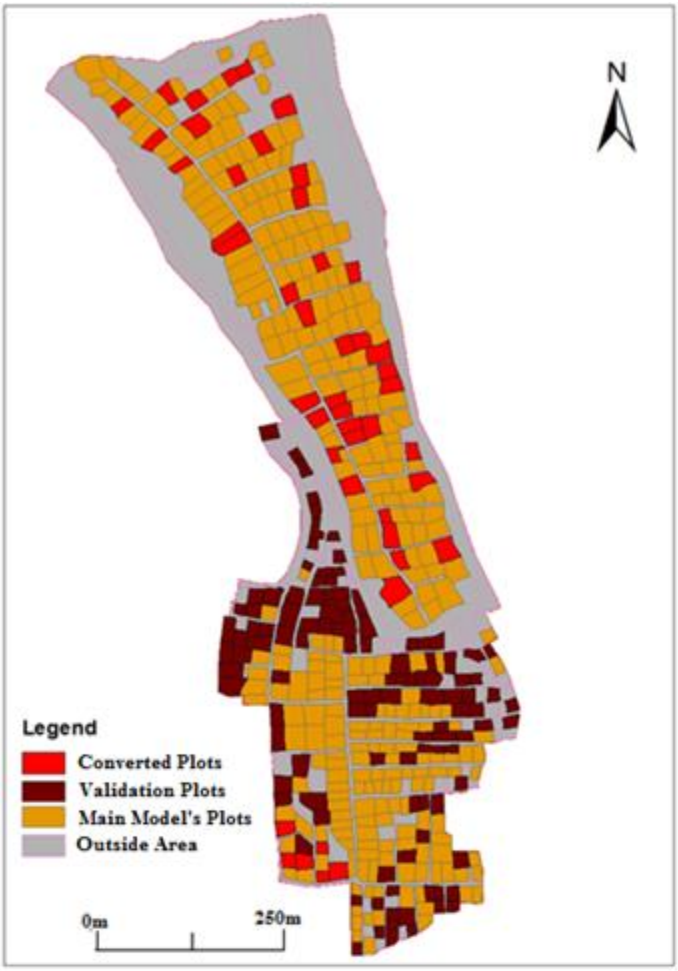

Figure 1: Location of Land Plots in the Validation Data Area

There are 114 houses that have been built up after the year 2005 in the core study area. All 114 houses were non-stable at the beginning and when the questionnaire survey was conducted in 2012, 76 of these houses were stable and the other 38 houses were non-stable. These 114 houses have been taken for the model validation process and are shown in brown color in Figure 1. At the questionnaire survey, It has been collected all household information related to below 19 variables of the main model as its validation model's variables mention table 4 below.

Literature Review

The empirical statistical model and applications are important in land use conversion modelling and should prove rather more helpful than a highly theoretical approach (Harman & McMinn, 2010) but their usefulness will be short term or long term based on ground experience of driving forces (Lambin et al., 2000; Lambin, Geist, & Lepers, 2003). As to its ability to perform in short term prediction, Verburg et al. (2004) have claimed that this model is able to forecast the situation twenty plus years ahead when used with simulation model and linked with sets of temporal and spatial data. The other advantage of empirical statistical modelling is the possibility of integration of sets of socio-economic variables for estimation and prediction of land use in relation to any specific issue (Irwin & Geoghegan, 2001; Irwin & Bockstael, 2004).

Regression analysis (single / multiple) is one of the best tools to apply with empirical statistical modelling with dependent variables because of its intuitive and explainable nature (Xie et al., 2005). Accordingly, they explain how the land use change model performs generally in the statistical domain as follows (Lambin et al., 2003): Land use = f (pressures, opportunities, policies, vulnerability, and social organization); further, its sectoral variables can be expanded into sub-variables.Logistic model is widely applied to modelling for spatial planning and disaster management such as landslide risk reduction (Ayalew & Yamagishi, 2005; Cheng & Masser, 2003; Dong, Tung, Chen, Liao, & Pan, 2011). There are two kinds of logistical models, multinomial logistic and binary logistic. In most empirical research the binary logistic model is applied. Multinomial logistic regression is very complex and its interpretation and understanding is more difficult than it is with binary logistic models. However, the selection of multinomial or binary logistic model for any research will depend on the kinds of dependent variables as explained in Table 2. Binary logistic modeling can be performed with the bivariate logistic which has only one independent variable with dependent dummy variable. However, most useful applied logistic modeling is done with multivariate logistic regression that has several independent variables, either categorical or continuous (Pradhan, 2010).

Applications of logistic regression models can be used to measure variations of probability of the independent variables for data sets of dichotomous variables such as 0 and 1 (or true and false) as dependent variable (Ayalew & Yamagishi, 2005; Chu, Lin, Huang, Hsu, & Chen, 2010). Probability value range indicates the relationship between presence and absence of a situation according to research hypothesis. Applications of logistic regression are widely used in the remote sensing based cell value classification in the spatial context (Xie, Huang, Claramunt, & Chandramouli, 2005).

Validation Process

The process of developing a model is commonly called calibration, whereas verifying a calibrated model is commonly called validation. Validation of a model involves comparison of different variations of the model by quadrants of the data set, sectors of the data set or other large areas of interest within the data set or outside of the data set (Hills & Trucano, 1999; Janssen & de Vries, 1998; Trucano, Swiler, Igusa, Oberkampf, & Pilch, 2006). Making use of an outside data set or splitting a data set into smaller units are also very good techniques for ensuring the "goodness of fit of the data" in addition to offering several advantages to model verification and validation (Halkidi, Batistakis, & Vazirgiannis, 2001). If possible, the model should be validated from a different database from the one that was used to build the model (Bewick, Cheek, & Ball, 2005).

DATA AND METHODS

Eight independent variables out of 19 variables of main model are used in the validation process. These variables are identical with the main model-significant variables namely household income (HI), ground water surface level (GWS), public participatory practice (PPP), permanent plants growing in plot (PPG), rain water remaining in the plot (RWR), skilled jobs (SKJ), technical skills and adaptation (TSA) and low lying related plants availability (LLP). Multivariate binary logistic model was applied via SPSS for the statistical analysis. Constant value of main model and beta coefficient value of each independent variable have also been assigned to the Validation process before running it. According to the above categories, the step by step procedure followed in testing validation of overall model has been schematically presented in Figure 2.

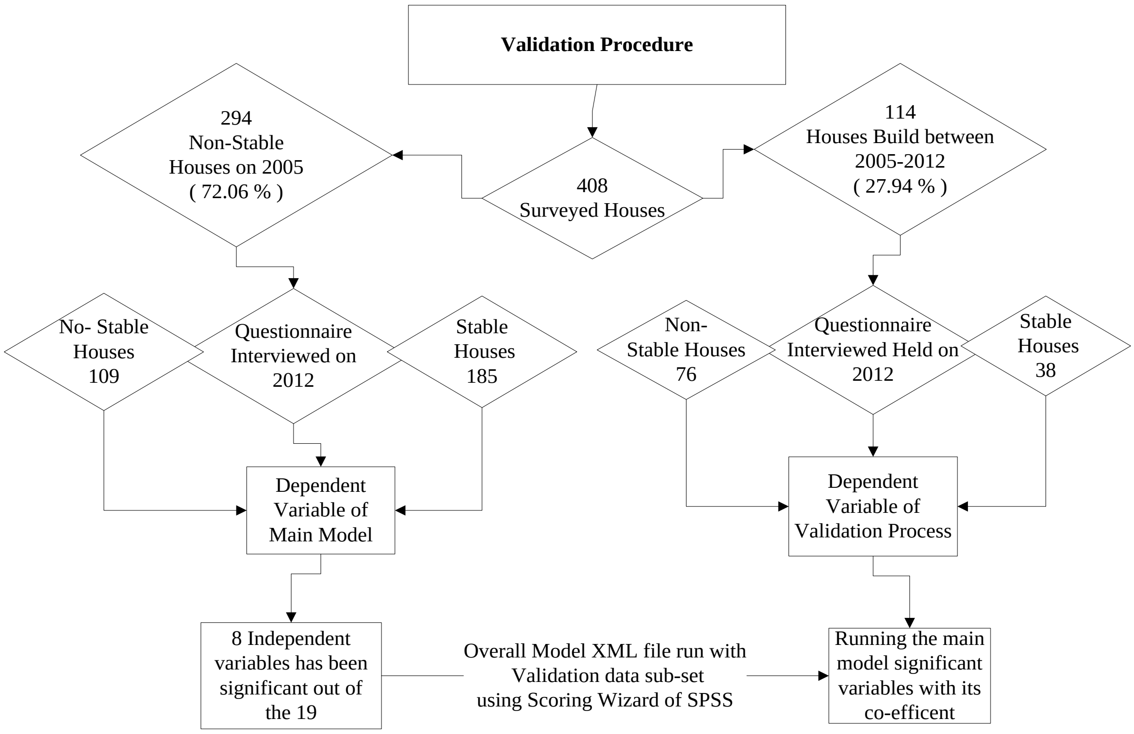

Figure 2: Step by Step Procedure of Validation Model Method

The Main Model

Validation model run based on the main model beta values, because validation procedure confirms the main model consistency. It was found that there were 46 households that had been converted and were stable in 2005. These houses were not taken for further analysis, because the conversion had already been effected.

According to the questionnaire interviews, there were 294 non-stable houses in the year 2005, but later on many of these houses had gradually changed from low land plots to home gardens. By mid-2012, it was found that there were 185 stable houses and 109 non-stable houses. Therefore, these 294 households that had existed in 2005 have been taken as the main database, accounting for 72.06% of the entire database. The geo-spatial model was built based on spatial logistic development, as it is mentioned in the literature that the logistic model has the capacity to better elaborate on spatial problems (Kwan, 2000; Xie, et al,. 2005). Hence, this research has been run on the geo-statistical model that comprises 19 variables covering socio-economic, behavioral and geo-spatial aspects. The model can be expressed as a Multivariate Logistic Regression Equation as follows:

$$Y = a + b_1 x_1 + b_2 x_2 + \dots\dots\dots + b_n x_n$$

$$Y = \log_e\left(\frac{p}{1-p}\right) = \log it(P)$$

$$P = Exp(Y)/1 + Exp(Y)$$

$$= B_0 + B_1 HS + B_2 HI + B_3 HHEL + B_4 SE + B_5 IMF + B_6 SKJ + B_7 VA + B_8 HHA$$

$$+ B_9 MU + B_{10} TS + B_{11} LT + B_{12} SDD + B_{13} PG + B_{14} PPP + B_{15} LLP + B_{16} GWS$$

$$+ B_{17} FL + B_{18} DSO + B_{19} RWR \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

Where p is the probability that the dependent variable (Y) is 1, p (1-p) is the so called Odds or likelihood ratio, B_0 is the intercept, and $B_1, B_2 \dots, B_{19}$ are coefficients, which measure the contribution of the independent factors listed below to the variations in Y.

HS	=	Household Size
HI	=	Household Income
HHEL	=	Head of Household Education Level
SE	=	Settlement year
IMF	=	Immigration From
SKJ	=	Skilled Jobs
VA	=	Vehicle Availability
HHA	=	Household Assets
MU	=	Material Used
TS	=	Technical Skills and adaptation
LT	=	Living Time per week in house
SDD	=	Smoke, Drink and Drug habits of inhabitants
PG	=	Permanent plants Growing in a plot
PPP	=	Public Participatory Practice
LLP	=	Low Land Plants availability in a plot
GWS	=	Ground Water Surface level in a plot
FL	=	Flood Level in a plot
DSO	=	Distance to Sensitive Objects from the plot
RWR	=	Rain Water Remaining days in a plot

In order to interpret the meaning of Eq. (1) correctly, it is necessary to express the coefficients as a power of the natural log (*e*), which represents the Odds ratio. In the data analysis section, Main Model explains how all variables pertaining to individual households have been contributing to the dependent variables relating to Stable houses and Non-stable houses of the study area.

Beta Value of Main Model for Model Validation

Validation model can be run using only beta value of above mention main model, according to Table 2. The beta values of the significant variables of the main model have been applied to the validation model's formula (Lee & Koval, 1997). When considering the results of the main model, 8 variables have significant levels out of 19 variables. Odds ratio as well as the Wald value of each significant variable has very good level. Therefore, it has been attempted to repeat these results to confirm consistency so as to be able to make recommendations to use the common model by running the validation model as explained of this study.

Table 2: Results of Main Model; Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 8[a]	HI	.168	.048	12.282	1	.000	1.183	1.077	1.300
	SKJ(1)	1.093	.542	4.074	1	.044	2.984	1.032	8.628
	PPG	.376	.115	10.667	1	.001	1.457	1.162	1.825
	RWR	-.207	.098	4.466	1	.035	.813	.671	.985
	GWS	.139	.029	22.545	1	.000	1.149	1.085	1.217
	PPP			9.414	2	.009			
	PPP(1)	1.387	.662	4.387	1	.036	4.003	1.093	14.659
	PPP(2)	2.092	.713	8.615	1	.003	8.099	2.004	32.736
	LLP			6.070	2	.048			
	LLP(1)	.292	.928	.099	1	.753	1.339	.217	8.258
	LLP(2)	1.523	.776	3.855	1	.050	4.586	1.003	20.972
	TSA			7.175	2	.028			
	TSA(1)	1.603	.666	5.793	1	.016	4.966	1.347	18.312
	TSA(2)	1.693	.776	4.762	1	.029	5.438	1.188	24.891
	Constant	-7.977	2.074	14.797	1	.000	.000		

a. Variable(s) entered in step 8: SKJ.

Validation Database

An additional 114 houses were built between the years 2005 and 2012 and of these 76 were non-stable houses and 38 were stable houses as at mid-2012. These have been taken for the model validation process, accounting for about 27.94% of the entire database. Therefore, it is seen that approximately 25% of the data has been taken for model validation (Xie et al., 2005).

Encoding of the Validation Process

The following tables explain the common statistical parameters for the validation process. Table 3 indicates the *n* value (number of observations) of the database, which is 114, being equal to the number of houses established between 2005 and 2012. There are no missing data for any of the variables that have been tested in the models.

Table 3: Dependent Variable Encoding for Validation Process

Variable	N	Internal Value
Non-Stable House	76	0
Stable House	38	1

In this study, vacant lands and abandoned housing plots have not been considered for the model because there was no responding person to represent such spatial units. However, a value of 0 is assigned to the non-stable housing plots and a value of 1 assigned to the stable housing plots. In the SPSS database the values have been assigned as 0 for non-stable housing plots and as 1 for stable housing plots as indicated in Table 3 showing dependent variable encoding.

Descriptive Analysis of the Validation Data Process

Validation process considers the 8 variables that have a definite relationship with the main model because usually the constant value and co-efficient beta value of the original model are assigned to the validation process. Therefore, the validation model must be comprised of the original variables of the main model. Table 4 shows the descriptive statistics for scalar data of the validation process according to their units. However, dummy data have also been displayed with their coding to enable one to get an overall idea about the outcome of the validation process

Table 4: Descriptive Analysis of Validation Data Process

No.	Description of Variable labels	Units	Mean	Standard Deviation	Min	Max
1	Household income(HI)	SL Rupees ,000	18.16	7.05	6.00	41.00
2	Ground water surface (GWS) level	cm	14.17	12.14	-15.00*	60.00
3	Public participatory practice (PPP)		Variable with three categories 1 = Low, 2 = Moderate, 3 = High			
4	Permanent plants growing in plot (PPG)	Number of plants	2.15	2.46	0	13
5	Rain water remaining in the plot (RWR)	Days	10.90	2.95	2	16
6	Skilled jobs (SKJ)		Dummy Coded 0 = No, 1 = Yes			
7	Technical skills and adaptation (TSA)		Variable with three categories 1 = Low, 2 = Moderate, 3 = High			
8	Low lying related plants availability (LLP)		Variable with three categories 1 = Low, 2 = Moderate, 3 = High			

**Some housing plots have surface water instead of ground water because of very low elevation. Those houses having surface water have been assigned a minus value under the ground water surface category*.

Categorical Variable Settings for Validation Process

In the validation model process, there are four categorical variables out of a total of 8 variables as indicated in Table 5. Public participatory practice, technical skills and adaptation, and low lying related plants availability comprise three categories marked as low, moderate and high. The skilled jobs variable is a dummy category marked as 1 and 0. The other 4 variables are scalar data that are not required for reference setting of the logistic model (Hosmer Jr & Lemeshow, 2004).

Table 5: Categorical Variables Coding for Validation Process

		Frequency	Parameter coding	
			(1)	(2)
PPP	1	29	0.000	0.000
	2	45	1.000	0.000
	3	40	0.000	1.000
TSA	1	50	0.000	0.000
	2	42	1.000	0.000
	3	22	0.000	1.000
LLP	1	31	1.000	0.000
	2	67	0.000	1.000
	3	16	0.000	0.000
SKJ	No	71	0.000	
	Yes	43	1.000	

Following methods were following to set the reference into stepwise logistic model;

- Householders from non-stable houses engage in fewer public participatory practices than householders from stable houses. Then ‘less public participatory practice’ is the reference variable.
- Less technical skills and adaptation are more applicable to non-stable householders than stable householders. Then ‘less technical skills and adaptation’ is the reference variable.
- Availability of low lying related plants is higher in non-stable housing plots than in stable housing plots. Then ‘high low lying related plants’ is the reference variable.
- Fewer numbers of skilled workers are living in non-stable houses than stable houses. Then ‘no skills’ is the reference variable.

Expression of Validation Process

The validation model explained how validation variables of individual households have been contributing to the dependent variables in respect of housing type of the study area. Especially, validation process has a definite relationship with the main model because usually the constant value and co-efficient beta value of the original model are assigned to the validation process. Hence, beta value has been written as B_{VX} . This model process can be expressed in terms of the multivariate logistic regression equation as follows:

$$Y = a + b_1 x_1 + b_2 x_2 + ... + b_n x_n$$

$$Y = \log_e \left(\frac{p}{1-p} \right) = \log it(P)$$

$$P = Exp(Y)/1 + Exp(Y) = B_{v0} + B_{v1}HI + B_{v2}GWS + B_{v3}PPP$$

$$+ B_{v4}PPG + B_{v5}RWR + B_{v6}SKJ + B_{v7}TSA + B_{v8}LLP \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (2)$$

Where p is the probability that the dependent variable (Y) is 1, p (1-p) is the so called odds or likelihood ratio, B_{V0} is the intercept of the overall model that has been taken as the constant value of model validation process, and $B_{v1}, B_{v2} ..., B_{v8}$ are coefficients of the overall model (significant variables) that measure the contribution of independent factors HI, GWS, PPP, PPG, RWR, SKJ, TSA, and LLP to the variations in Y.

In order to appropriately interpret the meaning of Eq. 2, it has to use the coefficients as a power of the natural log (*e*), which represents the odds ratio. In the data analysis section the validation process explained how all variables of individual households have been contributing to the dependent variables of housing types. This is shown in Table 6 that gives details of how the main model values are applied for the model validation.

Table 6: Coefficient values assigned for Validation Process

No.	Variable Name and Code	Beta Code for Validation Process	Beta Value for Validation Process
1	Household Income (HI)	B_{v1} = B_1 value of overall Model	0.168
2	Ground Water Surface Level (GWS)	B_{v2} = B_2 value of overall Model	0.139

3	Public Participatory Practice (PPP)	B_{v3} = B_3 value of overall Model	2.092
4	Permanent Plants Growing in Plot (PPG)	B_{v4} = B_4 value of overall Model	0.376
5	Rain Water Remaining in the Plot (RWR)	B_{v5} = B_5 value of overall Model	-0.207
6	Skilled Jobs: Dummy Coded (SKJ)	B_{v6} = B_6 value of overall Model	1.093
7	Technical Skills and Adaptation (TSA)	B_{v7} = B_7 value of overall Model	1.693
8	Low Lying Related Plants Availability (LLP)	B_{v8} = B_8 value of overall Model	1.523
	Constant Value of overall Model (B_{v0})		- 7.977

Therefore, Equation 3 shows the testing formula for the model validation process with its constant value and coefficient beta values.

$$Y = a + b_1x_1 + b_2x_2 + \dots\dots\dots + b_nx_n$$

$$Y = \log_e\left(\frac{p}{1-p}\right) = \log it(P)$$

$$P = Exp(Y)/1 + Exp(Y) = -7.977 + (0.168xHI) + (0.139xGWS) + (2.092xPPP)$$

$$+ (0.376xPPG) + (-0.207xRWR) + (1.093xSKJ) + (1.693x.TSA) + (1.523xLLP) \cdots\cdots\cdots\cdots (3)$$

Model Running Procedure

Validation Process has been run using SPSS 21 version software. XML file of the Main Model's significant variables has been saved. Thereafter, validation data set has been opened with SPSS and the scoring wizard run with XML file of overall model to produce the predicted probability value of model validation. Cut-off value 0.5 has been assigned for dividing into two classes. A spatial logistic probability table has been linked to the GIS to emphasize the spatial distribution of the relevant land plots.

ANALYSIS AND RESULTS

Validation Process and Accuracy

The classification tables can be used to validate the model, estimating it on the basis of validation data sample in 2012. The model estimation should be able to predict the dependent variables accurately according to the values indicated. The percentage of accurate predictions when compared against observed data reflects the validity of the estimated model. The accuracy of the prediction can be verified in different ways. Firstly, when the observed value of dependent variable is 0 (Y=0) for a particular case in the data set and the predicted value by model is also 0 (Y=0). Secondly, when the observed value of dependent variable is 1 (Y=1) for a particular case in the data set and predicted value by the model is also 1 (Y=1). But errors can occur occasionally.

The percentage of correct predictions out of the total number of cases is known as model accuracy. As an example, if the percentage of correct predictions out of the total number of cases is higher than 60%, then it can be accepted as a good validation process (Altman & Royston, 2000). As shown in table 7.A, Main model indicates accuracy about 92.2%. Referring to the table 7.B, the validation model indicates an accuracy of 85.09% which considerably high accuracy and good validation.

Table 7: Classification Table of Model Validation

Classification Table of Main Model: A			
Observed	Predicted		
	Non Stable (0)	Stable (1)	Correct %
Non Stable (0)	96	13	88.1
Stable (1)	10	175	94.6%
Overall %			92.2%

Classification Table of Validation Model: B			
Observed	Predicted		
	Non Stable (0)	Stable (1)	Correct %
Non Stable (0)	66	10	86.84%
Stable (1)	7	31	81.58%
Overall %			85.09%

Conversion of Housing Plots and Empirical Evidence

Figure 3 indicates the predicted probability classification in two classes with converted housing plots in magenta colour and non-converted housing plots in light blue colour. Conversion status (probability level) and the spatial distribution of the plots can be seen from the developed geo-spatial database. Further, the validation process results can be verified by using field evidence as illustrated in Figure 3 of this empirical study. Several randomly selected houses with predicted probability values were re-checked to see whether the predictions made were empirically correct (Ayalew & Yamagishi, 2005). As an example, in Figure 3 the housing plot reference Number 19 of the validation survey database indicates a high probability value of 0.9700 under the validation running process. Thereafter, this particular house was re-observed to verify the housing elements to determine whether it had really converted.

Figure 3: Empirical evidence of stable house on the study area

Comparison of Overall Model and Validation Process

Main model accuracy was 92.2% when its constant and co-efficient values were applied to the validation model's data set during the validation process. Accuracy of model validation process was 85.09%, which indicates that the estimated regression coefficients are compatible, and that the model fits with validation data. Also, the same variables can be seen to be important for future use. Hence, there is no reason to doubt the overall model's accuracy when applying its data for development work on future low land planning. Further, validation data has been tested with empirical evidence in the field as shown in Figure 4.

The figures below show the converted land plots and the non-converted land plots. The Main Model is shown in Figure 4.A and the Validation Model in Figure 4.B according to their probability values.

(A) **(A)**

Figure 4: Comparison of Model Output

Based on the findings of this study, it is suggested that any local government or other planning body involved in the preservation of low lying areas could use these results profitably in their work; they could apply this knowledge in their rehabilitation programs for low lying areas or for maintaining the sustainability of a region more effectively. Methods and values of probabilities of the variables can be used as key indicators of the rehabilitation process. One example of such a project that has been implemented in Sri Lanka is the rehabilitation of the coastal belt in various localities where extensive limestone quarrying has been carried out (Kumarasinghe et al., 2013). Even though the nature of the problems that need to be tackled in various rehabilitation projects may be different from the scenario of this study, the methods and concepts used in this study may bear much relevance to rehabilitation projects in other spheres and other localities. Hence, it is suggested that the probability of conversion of each housing plot be determined before rehabilitation. For example, non-stable houses can be removed from the low lying areas by means of low cost rehabilitation processes. Therefore, as shown in Fig. 4, the status of each household can be identified clearly with housing location, photos, conversion rates and all other information stored in the geo-spatial information system.

CONCLUSION

The Model explained how validation variables of individual households have been contributing to the dependent variables in respect of housing type of the study area. Especially, validation process has definite relationship with the main model because usually the constant value and co-efficient beta value of the original model are assigned to the validation process. In this model validation approach, it is explained how model validation is done using 27.94% of the overall data. Technically, validation data should split approximately 25% off the main database. The number of correct predictions out of the total cases is known as model accuracy. Basically, if the percentage of correct predictions out of the total cases is higher than 60 % can be accepted as good validation process. This study shows that the validation model has 85.09% accuracy level when it runs with the overall model's (main model's) B-coefficient data. Hence, this validation model can be recommended as a decent model that will serve the intention target because the overall model (main model) also has 92.2% accuracy. Even though it scored a little bit lower than the main model, the model validation process was able to achieve very good accuracy with its eight independent variables, which indicate that the estimated regression coefficients are compatible, while these same variables are important for future estimation and model fit with validation data.

References

Altman DG and Royston P. (2000) What do we mean by validating a prognostic model? Statistics in medicine 19: 453-473.

Ayalew, L., & Yamagishi, H. (2005). The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology, 65*(1), 15-31.

Barr, S. (2007). Factors influencing environmental attitudes and behaviors. *Environment and behavior, 39*(4), 435.

Bewick, Viv, Cheek, Liz, & Ball, Jonathan. (2005). Statistics review 14: Logistic regression. *Crit Care, 9*(1), 112-118.

Carrion-Flores, Carmen, & Irwin, Elena G. (2004). Determinants of residential land-use conversion and sprawl at the rural-urban fringe. *American Journal of Agricultural Economics, 86*(4), 889-904.

Cheng, Jianquan, & Masser, Ian. (2003). Urban growth pattern modeling: a case study of Wuhan city, PR China. Landscape and urban planning, 62(4), 199-217.

Chu, Hone-Jay, Lin, Yu-Pin, Huang, Chun-Wei, Hsu, Cheng-Yu, & Chen, Horng-Yng. (2010). Modelling the hydrologic effects of dynamic land-use change using a distributed hydrologic model and a spatial land-use allocation model. *Hydrological Processes, 24*(18), 2538-2554. doi: 10.1002/hyp.7667

Correia, T. (1993). Land abandonment: changes in the land use patterns around the Mediterranean basin. *Cahiers Options Méditerranéennes, 1*(2), 97-112.

Divigalpitiya, Prasanna, Ohgai, Akira, Tani, Takeru, Watanabe, Kojiro, & Gohnai, Yoshimizu. (2007). Modeling land conversion in the Colombo Metropolitan Area using cellular automata. *Journal of Asian Architecture and Building Engineering, 6*(2), 291-298.

Dong, J.J., Tung, Y.H., Chen, C.C., Liao, J.J., & Pan, Y.W. (2011). Logistic regression model for predicting the failure probability of a landslide dam. Engineering Geology, 117(1), 52-61.
Halkidi, Maria, Batistakis, Yannis, & Vazirgiannis, Michalis. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems, 17*(2-3), 107-145.
Harman, Mark, & McMinn, Phil. (2010). A theoretical and empirical study of search-based testing: Local, global, and hybrid search. Software Engineering, IEEE Transactions on, 36(2), 226-247.
Hills, Richard G, & Trucano, Timothy G. (1999). Statistical validation of engineering and scientific models: Background. *Sandia National Laboratories, SAND99-1256*, 36.
Hosmer Jr, David W, & Lemeshow, Stanley. (2004). *Applied logistic regression*: John Wiley & Sons.
Irwin, EG, & Geoghegan, J. (2001). Theory, data, methods: developing spatially explicit economic models of land use change. Agriculture, Ecosystems & Environment, 85(1-3), 7-24.
Irwin, Elena G, & Bockstael, Nancy E. (2004). Land use externalities, open space preservation, and urban sprawl. Regional science and urban economics, 34(6), 705-725.
Kumarasinghe A, Pradeep W, Peabotuwage P, et al. (2013) An attempt to reduce impacts of limestone quarries through biodiversity assessment and translocation: A case study at the Holcim Limestone Quarry Site in Puttalam, Sri Lanka. Asian Journal of Conservation Biology.
Janssen, Marco, & de Vries, Bert. (1998). Global modelling: managing uncertainty, complexity and incomplete information. *Validation of Simulation Models, SISWO, Amsterdam*.
Kwan, Mei-Po. (2000). Analysis of human spatial behavior in a GIS environment: Recent developments and future prospects. *Journal of Geographical Systems, 2*(1), 85-90.
Lambin, E. F., Rounsevell, M. D. A., & Geist, H. J. (2000). Are agricultural land-use models able to predict changes in land-use intensity? Agriculture, Ecosystems & Environment, 82(1-3), 321-331. doi: Doi: 10.1016/s0167-8809(00)00235-8
Lambin, E.F., Geist, H.J., & Lepers, E. (2003). Dynamics of land-use and land-cover change in tropical regions. Annual review of environment and resources, 28(1), 205-241.
Lee, Kang I, & Koval, John J. (1997). Determination of the best significance level in forward stepwise logistic regression. Communications in Statistics-Simulation and Computation, 26(2), 559-575.
Liverman, D.M. (1998). *People and pixels: Linking remote sensing and social science*: National Academies Press.
Oskamp, S, Harrington, MJ, Edwards, TC, Sherwood, DL, Okuda, SM, & Swanson, DC. (1991). Factors influencing household recycling behavior. *Environment and behavior, 23*(4), 494.
Pradhan, Biswajeet. (2010). Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate logistic regression model on three test areas in Malaysia. Advances in Space Research, 45(10), 1244-1256.
Staal, S, Wolff, T, Baltenweck, I, Romney, D, Waithaka, M, Njoroge, L, . . . Thorpe, W. (2000). *Combined household and GIS analysis of farmer strategies: an application to feeding practices on smallholder Kenyan dairy farms*.
Trucano, Timothy G, Swiler, Laura Painton, Igusa, Takera, Oberkampf, William L, & Pilch, Martin. (2006). Calibration, validation, and sensitivity analysis: What's what. *Reliability Engineering & System Safety, 91*(10), 1331-1357.
Verburg, PH, Schot, PP, Dijst, MJ, & Veldkamp, A. (2004). Land use change modelling: current practice and research priorities. GeoJournal, 61(4), 309-324.
Xie, Chenglin, Huang, Bo, Claramunt, Christophe, & Chandramouli, C. (2005). *Spatial logistic regression and gis to model rural-urban land conversion.* Paper presented at the Proceedings of PROCESSUS Second International Colloquium on the Behavioural Foundations of Integrated Land-use and Transportation Models: Frameworks, Models and Applications.