Journal of Engineering and Applied Sciences 11 (7): 1676-1681, 2016
ISSN: 1816-949X

Land Use Change and Development Modeling Using Logistic Regression

[1]Cornelia B. Appianing, [1]Noodchanath Kongchouy, [2]Orawit Thinnukool
[1]Department of Mathematics and Statistics, Faculty of Science,
Prince of Songkla University, 90110 Songkhla, Thailand
[2]Departmemt of Modern Management and Information Technology,
College of Arts, Media and Technology, Chiang Mai University, 50200 Chiang Mai, Thailand

Abstract: This study analyzed land use change near Hat Yai international airport, Thailand, from 1991-2009 and modeled land development in order to detect the rate at which each land use type and location was prone to development. Land use data for 1991, 2000 and 2009 were obtained from Thailand Department of Lands. The data structure was improved by digitization where polygonal boundaries of land were converted to gridded sub-plots (cells). Logistic regression technique was utilized in modeling development that emerged. A classification accuracy result was used to measure the accuracy of the model. The result of the study showed development in the North-East increased from 21.8-66.3% and in the South-West increased from 61.3-81.4%. This shows 45% increase in North-East and 20% in South-West. Also, the accuracy of the model in predicting development from 1991-2000 was 87.2 and 81.7% from 2000-2009 indicating a good model. The study concluded there had been immense change in land use in the airport especially the North-East location since 1991-2009. Development had been on the increase throughout the period yet rubber plantation and paddy field and other agricultural land still reigned since they have been a major source of income to the community.

Key words: Land use change, digitization, lgistic regression, income, development

INTRODUCTION

Land Use Change (LUC) requires maintaining and adjusting the natural habitat into a built up area as a result of human interference or natural occurrence. The population of humans over the past years have been on a rise coupled with socio-economic activities thereby giving in to pressure on use of land. Conversion of the natural forest into urban or developed areas has intensified over the past years hence destabilizing the ecosystem and propelling multiple infrastructure and environmental challenges (Haack and Rafter, 2006). Study on the conversion of land use is part of the global land plan sponsored by International Geosphere-Biosphere Programme and International Human Dimensions Programme on Global Environmental Change. This has given much interest to researchers (Yu *et al*., 2011; Raine *et al*., 1994; Olson, 2004; Veldkamp and Lambin, 2001; Hu and Lo, 2007) to investigate the changes that have emerged and the forces that drive these changes. Such analyses provide insight to what has taken place in a locality and help the government in implementing suitable land policy in case of urban sprawl.

Several models have been developed by researchers in various field of study to analyze land use change, predict urban growth in future and single out the areas at risk. Spatially explicit modeling have been popular for some decades due to growth in computing skills, enhanced opportunity in acquiring spatial data and the desire for creative planning mechanisms to aid valid conclusions (Matthews *et al*., 1999; Castella *et al*., 2007). Methods such as empirical statistical models, optimization models, stochastic models and dynamic process based models have been used to analyze and predict land use change globally (Dessel *et al*., 2011). Such models offer the possibility to test the sensitivity of land use patterns to changes in selected variables. Most land use change models use inductive approach by specifying the model according to the statistical associations that exist between land use change and the independent variables (Overmars and Verburg, 2005). Logistic regression is one of the empirical-statistical methods and an example of discrete outcome modeling technique (Nong and Du, 2011) which has been applied in several research areas. The use of logistic regression is better compared with discriminant analysis because it analyzes data on all scales being it categorical or continuous (Chauhan *et al*.,

Corresponding Author: Noodchanath Kongchouy, Department of Mathematics and Statistics, Faculty of Science,
Prince of Songkla University, 90110 Songkhla, Thailand

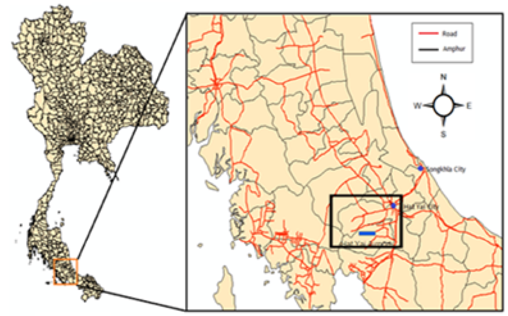
2010). It is robust to the violation of multinormality assumption (Johnson, 1998; Ohlmacher and Davis, 2003) and has fewer theoretical assumptions than discriminant analysis (Ayalew and Yamagishi, 2005). It also combines the independent variables to estimate the probability that a particular event will occur. Logistic regression has been applied in land use and urban growth study (Lin *et al*., 2011; Rutherford *et al*., 2007; Cheng and Masser, 2003). This study tends to analyze land use change near Hat Yai international airport, Songkhla, Thailandfrom 1991-2009 and model development using logistic regression in order to detect the rate at which each land use type and location was prone to development.

MATERIALS AND METHODS

Study area: Hat Yai is the fourth largest city in Thailand and is situated at latitude 7°1'N and longitude 100°28'E a distance 946 km South of Bangkok. The total population of the area is over 800,000. Hat Yai's economy relies on agriculture and tourism having enormous income from the timber, fishing industries, rubber and palm oil plantations. Hat Yai International airport is dominant in southern Thailand. It is situated at Klong Hoi Kong district close to Hat Yai district. Klong Hoi Kong used to be a sub district under Hat yai district until 1992 where it was separated and made a district on its own. Since, the airport was established much development has taken place near the area which needs to be given proper attention in order to prevent sprawl. Figure 1 shows the Hat Yai international airport.

Data and land use classification: Land use data for 1991, 2000 and 2009 was obtained from Thailand Department of Lands. The data structure was improved by digitization where polygonal boundaries of land were converted to gridded sub-plots (cells) which were 100×100 m in area. Digitization helped to replace the polygons that varied in shape and size by a regular and unchanging grid of points (Thinnukool *et al*., 2014). There were 1,199,200 cells and after removing the cells with water and areas far from the airport it remained 5,685 cells for the analysis.

Land use was classified into four types for this study. These are undeveloped land, paddy field and other agricultural land, rubber plantation and developed (urban) land. The reason behind classifying land use into four types and precisely separating paddy field and other agricultural land and rubber plantation was due to the fact that Hat Yai district and Southern Thailand as a whole depend on these two plantations as their major sources of income hence separated to see their trend of change as



Hat Yai airport, Songkhla province

Fig. 1: Map of Hat Yai airport

Table 1: Land use types used for the study

LU type	Description	Categories
UD	Undeveloped land	Orchard, evergreen forest, rangeland, marsh, beach
PF+	Paddy field and other agricultural land	Rice plantation, pasture, horticulture, field crop perennial and aquatic plants
RP	Para rubber	Rubber plantation
Dev	Developed (urban) land	City, allocated land, institutional area, utility, industrial land, recreationalarea, Hotel concentration, roads

urban growth occurred. It was also to measure the rate of development over the years. Table 1 gives the detail of how the classification was done. The locations considered for this study are North-East (NE) and South-West (SW) of the airport.

Preliminary data analysis and graph presentation: The classified land use data was assigned numbers 1-4 (i.e., 1: UD, 2: PF+, 3: RP and 4: Dev). Colors such as pale green orange, yellow and red were used to show distinct view of each land use type. Land use types were displayed in thematic maps for the three periods, 1991, 2000 and 2009.

A bar chart was also generated to support the map. It showed the exact increase and decrease in area of land use. Bubble plot matrices for land use were also created to show the percentage change of each type of land for the two periods (i.e., 1991-2000 and 2000-2009). This was carried out by computing a cross-tabulation for each period. Cross-tabulation for each period was divided by its total and multiplied by 100%. The percentage change was then rounded up. Hence, a change below 1% was not seen in the plot though the areas were made visible.

Statistical analysis of developed land: Binary logistic regression was used to analyze the data since the outcome at each grid point was binary, thus either the specific land-use of interest (developed, say) or not. Logistic regression as a conventional statistical method assumes the individual samples are not correlated but this

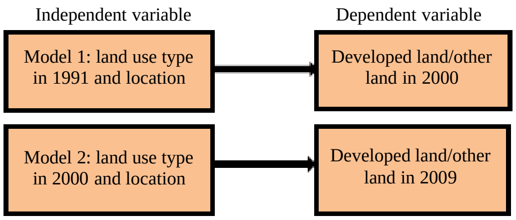


Fig. 2: Block diagram of study

doesn't work for the digitized data just 100×100 m apart. Hence variation inflation factors (Mathew *et al*., 2009; Rao and Scott, 1992) were used to account for spatial correlation within sub-plots (cells) and valid confidence intervals obtained using the inflation factors. The probability, p that a developed land (Dev) will occur given land use type and location was demonstrated in Eq. 1:

$$p = \frac{\exp(\alpha+\beta X)}{1+\exp(\alpha+\beta X)} \quad (1)$$

Where:

X = The independent variable
α and β = The model parameters

which are estimated using maximum likelihood method based on the values of the independent variable and the status of the dependent variable in the sample cells. Taking a logit transformation of the first equation we produce a model that is linear in the parameters:

$$1n\left(\frac{p}{1-p}\right) = \alpha+\beta X \quad (2)$$

Where:

1-p = The probability that other land will occur. It must be noted that "other land" comprised undeveloped land, rubber plantation, paddy field and other agricultural land.
α = A constant
β = The magnitude of location by land-use type 9 years earlier

Figure 2 demonstrates the independent and dependent variables for the logistic regression modelling of development in the study. All analyses were performed using R version 3.00 (Team , 2012).

RESULTS AND DIASCUSSION

Land use change: Figure 3 and 4 show the land use change in thematic map view and bar chart view respectively. Note that the colors used in both figures mean same. Pale green is Undeveloped land (UD) orange is Paddy field and other agricultural land (PF+), yellow is Rubber plantation (RP) and red is Developed land (Dev). The diagonal line in Fig. 3 indicates the location (North-East of the airport) while the black lines represent the roads around the airport. Both figures demonstrate UD and PF+ decreased in area after 1991, RP increased in 2000 and Dev increased after 1991 especially, near the airport. The diagonal plots with colours pale green, yellow orange and pink in Fig. 5 show no change (maintained

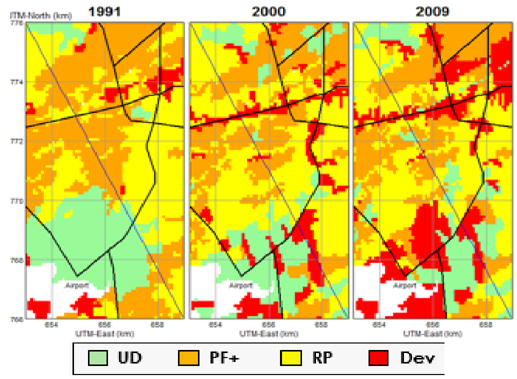


Fig. 3: Thematic map showing various land use in each year

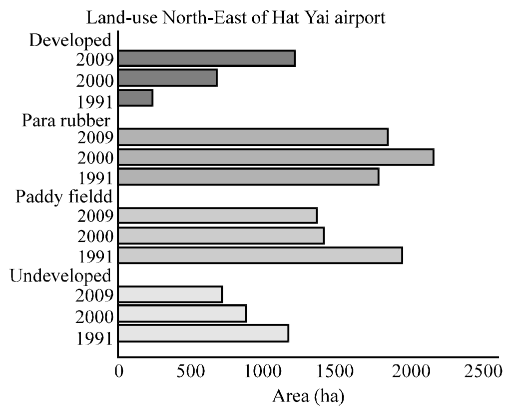


Fig. 4: Change in area of land use given in bar chart view

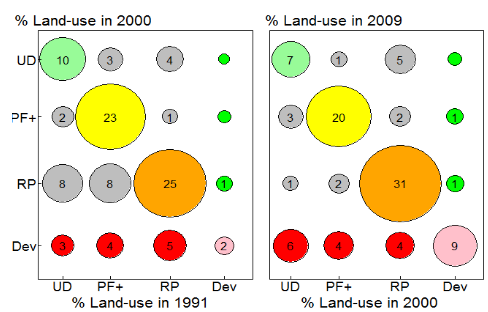


Fig. 5: Bubble plot matrix giving a clear view of the land use change in percentages

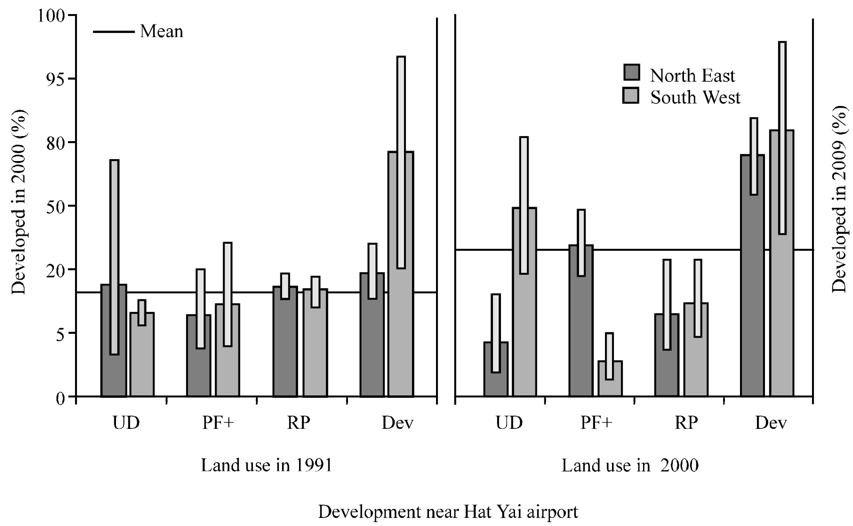


Fig. 6: Development in North-East and South-West locations of Hat Yai airport demonstrated in bar chart view

their use) only the off diagonal plots with red, gray and bright green colours indicate changes. Note that the colours used for the plots do not have same meaning as that of thematic map and bar chart but just used to show change in area from one type to the other. From 1991-2000, undeveloped land, paddy field and other agricultural land and rubber plantation became developed by 3, 4 and 5%, respectively. Rubber plantation gained 8% each from undeveloped land and paddy field and other agricultural land. Paddy field and other agricultural land lost 3% to undeveloped land. Also, from 2000-2009, land use for undeveloped land, paddy field and other agricultural land and rubber plantation changed to developed land by 6, 4 and 4%, respectively implying that much of the undeveloped land had lost to the developed land thereby causing an increase in development. Developed land which didn't undergo any changes from 1991-2000 is just 2% and that of 2000-2009 is 9% indicating development is on the increase as the years go by.

Modeling results: Results from the Logistic regression modelling in Fig. 6 demonstrates how each land use type coupled with the location is likely to turn developed in 2000 and 2009 based on percentages instead of the usual Odd Ratios (OR). Note that the independent variable is categorical. The blue vertical line segments denote a 95% confidence interval of development. The bars in yellow represents developments in the North-East (NE) and those in gray are developments in the South-West (SW). The average developments in both years are shown by the horizontal red lines. The expected (mean) development from 1991-2000 was 13 and 24% from 2000-2009 which were quite high. Confidence intervals raised above the mean indicates the land use types that were most likely to turn developed.

Development from 1991-2000 indicated that, Undeveloped land (UD), paddy field and other agricultural land (PF+) and Rubber Plantation (RP) in the NE changed by 17, 8 and 16% respectively to developed land while UD, PF + and RP in the SW changed to developed land by 10 , 11 and 14% , respectively. Developed land that stayed developed during this period (1999-2000) was vast (61%) in the SW and NE (22%). Though other land use body (UD, PF+ and RP) changed to developed land, it was at a lesser rate since the confidence intervals were either below or across the expected (mean) development. Also, development from 2000-2009 revealed that, UD, PF+ and RP in the NE changed to developed land by 6, 25 and 8%, correspondingly , while those in the SW changed by 48, 2 and 10% correspondingly. Thus, other land use body (UD, PF+ and RP) seldom turned developed in both locations except UD (48%) in the SW, since their confidence intervals were either below or across the mean. Developed land in both locations were most likely to maintain their use. Though, development has taken place in both locations, it is obvious that much development occurred in the North-East of the airport. For instance, the development in 2000 for the South-West (61.3%) was far greater than the North-East (21.8%) but in 2009, development in the north-east increased by 45% (i.e., from 21.8-66.3%) which was quite a lot of change, that of the south-west increased by only 20% (i.e., from 61.3-81.4%).

Table 2: Classification result for 1991-2000

	Predicted Y		
Observed	Other land (0)	Developed land (1)	Correct (%)
Other land (0)	4892	682.0	87.8
Developed land (1)	43	68.0	61.3
Overall percentage		87.2	

*Recall = 0.613

Table 3: Classification result for 2000-2009

	Predicted Y		
Observed	Other land (0)	Developed land (1)	Correct (%)
Other land (0)	4120	815.0	83.5
Developed land (1)	223	527.0	70.3
Overall percentage		81.7	

*Recall = 0.703

Accuracy assessment of model: The classification table is another method to evaluate the predictive accuracy of the logistic regression modelling. There are several ways to assess the accuracy of a logistic regression model and one of them is using the classification table popularly known as the confusion matrix. Classification accuracy results for binomial logistic regression consist of both observed and predicted values of the model. The model classified the data into two categories since the response is binary, thus developed land and other land with their respective values 1and 0. This grouping was done by setting up a cut off value of 0.5. The values above 0.5 were classified as developed land and those below as other land. Table 2 shows classification results for developed (urban) land from 1991-2000. The number of cases observed to be other land were 5,574 and 4,892 (87.8%) out of these 5,574 were correctly predicted and the number of cases observed to be developed land were 111 with 68 (61.3%) out of the 111 been correctly predicted. Hence, the overall percentage of accurate prediction of the model is 87.2%. Table 3 also gives the classification results for developed (urban) land from 2000 to 2009. The number of cases observed to be other land were 4,935 and 4,120 (83.5%) out of these 4,935 were correctly predicted and the number of cases observed to be developed land were 750 with 527 (70.3%) out of the 750 been correctly predicted. Hence, the overall percentage of accurate prediction of the model is 81.7%.

CONCLUSION

A search was carried out on the land use change in the area near Hat Yai airport. The results from the study revealed several changes took place near the airport. These changes were made visible by the thematic maps, bar chart and the bubble plot matrices. Though several changes had taken place, rubber plantation and paddy field and other agricultural land were still reigning. The reason is the inhabitants of Hat Yai still depend highly on these resources as their major source of income (Chuangchang and Tongkumchum, 2014). Furthermore, great development (urban land) has showed up in this area especially places extremely close to the airport as well as the major roads (black lines in Fig. 3. Increase in developed (urban) land near the airport is mostly expected (Swangjang and Iamaram, 2011; Cidell, 2004) due to population growth and building of settlements such as hotels and factories. Most factories are into importation and exportation of goods so does prefer their sites near the airport. For instance PTT Public Company Limited, a very vibrant company owned by the state and deals in production of oil and gas for both local and international use is situated near the airport.

Airports built in highly populated regions are easily besieged with enormous developments. Therefore, proper measures need to be utilized during land use planning for both local and international airports so as to prevent sprawl, since a country cannot do without Aviation as one of the major sources of transportation. Hat Yai airport is one of the best in Thailand which handles over 1, 500,000 passengers per year and it's just 9km from the main city making it easy to fuel great development in the area. Visitors or tourists often lodge in hotels near the airport when they arrive originating much development in the area especially in the North-East location since 1991. It is expected that future development may even be unbearable if proper measures are not taken by the government and land use policy makers. Also, the logistic regression model was able to predict development well. The accuracy in predicting development from 1991-2000 was 87.2% with recall of 61.3 and 81.7% accuracy from 2000-2009 with a recall of 70.3%. This shows a good prediction of the developed area and a reliable model despite the use of only one independent variable (land use type and location).

In conclusion, there has been immense change in land use in the airport especially the North-East location since 1991 to 2009. Development (urban growth) has been on the increase throughout the period yet rubber plantation and paddy field and other agricultural land still reigned. In our further study, we would be looking at the current change in land use and use in predicting future change. Also factors such as population density and economic conditions that contribute to the changes in land use will be incorporated in the model.

ACKNOWLEDGEMENTS

The researchers would like to thank Prof. Don McNeil and reviewers for comments and suggestions. The

Thailand Department of Lands also supported by giving us the GIS data and Prince of Songkla University (research assistant scholarship) for funding the research.

REFERENCES

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

Castella, J.C., S.P. Kam, D.D. Quang, P.H. Verburg and C.T. Hoanh, 2007. Combining top-down and bottom-up modelling approaches of land use-cover change to support public policies: Application to sustainable management of natural resources in Northern Vietnam. Land Use Policy, 24: 531-545.

Chauhan, S., M. Sharma and M.K. Arora, 2010. Landslide susceptibility zonation of the Chamoli region, Garhwal Himalayas, using logistic regression model. Landslides, 7: 411-423.

Cheng, J. and I. Masser, 2003. Urban growth pattern modeling: A case study of Wuhan city, PR China. Landsc. Urban Plann., 62: 199-217.

Chuangchang, P. and P. Tongkumchum, 2014. Modeling land development along highway 4 in Southern Thailand. Songklanakarin J. Sci. Technol., 36: 719-725.

Dessel, V.W., V.A. Rompaey and P. Szilassi, 2011. Sensitivity analysis of logistic regression parameterization for land use and land cover probability estimation. Int. J. Geog. Inf. Sci., 25: 489-508.

Haack, B.N. and A. Rafter, 2006. Urban growth analysis and modeling in the Kathmandu Valley, Nepal. Habitat Int., 30: 1056-1065.

Hu, Z. and C.P. Lo, 2007. Modeling urban growth in Atlanta using logistic regression. Comput. Environ. Urban Syst., 31: 667-688.

Johnson, D.E., 1998. Applied Multivariate Methods for Data Analysts. Duxbury Press, Pacific Grove, USA., ISBN: 9780534237967, Pages:567.

Lin, Y.P., H.J. Chu, C.F. Wu and P.H. Verburg, 2011. Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling-a case study. Int. J. Geog. Inf. Sci., 25: 65-87.

Mathew, J., V.K. Jha and G.S. Rawat, 2009. Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. Landslides, 6: 17-26.

Matthews, K.B., A.R. Sibbald and S. Craw, 1999. Implementation of a spatial decision support system for rural land use planning: Integrating geographic information system and environmental models with search and optimisation algorithms. Comput. Electron. Agric., 23: 9-26.

Nong, Y. and Q. Du, 2011. Urban growth pattern modeling using logistic regression. Geo Spatial Inf. Sci., 14: 62-67.

Ohlmacher, G.C. and J.C. Davis, 2003. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Eng. Geol., 69: 331-343.

Olson, J.M., 2004. Multi-scale analysis of land use and management change on the Eastern slopes of Mt. International Livestock Research Institute, Nairobi, Kenya.

Overmars, K.P. and P.H. Verburg, 2005. Analysis of land use drivers at the watershed and household level: Linking two paradigms at the Philippine forest fringe. Int. J. Geographical Inform. Sci., 19: 125-152.

Team, R.C., 2012. A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing, Vienna, Austria.

Raine, R.M., 1994. Current land use and changes in land use over time in the coastal zone of Chanthaburi Province, Thailand. Biol. Conserv., 67: 201-204.

Rao, J.N.K. and A.J. Scott, 1992. A simple method for the analysis of clustered binary data. Biometrics, 48: 577-585.

Rutherford, G.N., A. Guisan and N.E. Zimmermann, 2007. Evaluating sampling strategies and logistic regression methods for modelling complex land cover changes. J. Appl. Ecol., 44: 414-424.

Swangjang, K. and V. Iamaram, 2011. Change of land use patterns in the areas close to the airport development area and some implicating factors. Sustainability, 3: 1517-1530.

Thinnukool, O., P. Chuangchang and K. Kongkraphan, 2014. Analyzing land-use change using grid-digitized method. Songklanakarin J. Sci. Technol., 36: 235-240.

Veldkamp, A. and E.F. Lambin, 2001. Editorial: Predicting land-use change. Agric. Ecosyst. Environ., 85: 1-6.

Yu, W., S. Zang, C. Wu, W. Liu and X. Na, 2011. Analyzing and modeling land use land cover change (LUCC) in the Daqing City, China. Applied Geogr., 31:600-608.