Chapter 1

Introduction

The aims of this study were to develop an improved method for storing the rich and valuable land-use database comprising of the land-use plots recorded at regular surveys by the Thai Department of Land countrywide since 1967, and to use this method to explore land-use change on Phuket Island from 1967 to 2009. The method involves creating a digitized grid of geographical coordinates covering the whole study area and storing the land-use codes and plot identifiers as fields in database tables indexed by the grid coordinates, and then using logistic regression adjusted for spatial correlation, using a variance inflation factor defined on land-use categories, to predict land-use change in a specific category.

1.1 Rationale of study

Land-use has changed rapidly in Thailand since 1980. The Thai Department of Lands has been regularly surveying and recording land-use in hundreds of small plots from every province since 1967. These data contain a lot of valuable information about Thai history and culture development that is not available elsewhere, and tends to get forgotten and lost over the time due to data storage difficulties. Methods for extracting, managing and analyzing such land-use data are essential, but have not been fully developed. Such information is also valuable to planners and developers, to make it easier for government officials, university researchers and other interested parties to analyze land-use change.

Moreover, although land-use data exist in databases based on regular surveys of thousands of plots, plot shapes change over time, so analysis is difficult because the data are structured as polygonal "shape" files that change shape and diversify at each new survey. When land-use changes over time, polygons are replaced. Some disappear, others appear, and existing ones change their shape. The changes make it very difficult to analyze the more important determinants in the land-use itself, such as the pace and pattern of urbanization.

1.2 Thesis content

This thesis presents a method for measuring land-use by using digitization, which is based on data in Phuket Island from 1967 to 2009.

First, remote sensing data (RS) is our population for this research. This section reviews world accepted RS data. Next, we will explain data structure, such as block pixels and polygon, which type is suits presentation for land-use change. After that, land-use and land-cover (LUCC) will be explained. Land-use change and land-use modeling reviews shows the results of urbanization and perdition values from binary logistics. In addition, the methodology part contains data sources and the process has been explained (transformation from original polygon data to digital data using a grid and highlighting the problems of polygonal dada structure). The correction of coordinate shifts of plot boundaries is assumed by bilinear transformation, which this section describes.

The formats for the presentation of land-use are vector and raster format; the shape file data has been transformed to raster format by the grid-digitized method.
The populate grid point section describes how to construct a digital grid.

The problem of complex regions from shape file data has been solved by our program, which is illustrated in the complex region section.

In addition, statistics analysis is the methodology to analyze land-use with binary logistic regression. The result of land-use data from grid-digitized are illustrated land-use occurred to thematic map. This part show thematic maps of land-use in Phuket Island in 4 periods from 1967 to 2009.

The next section is discussion for digitization, land-use change in Phuket Island and urbanization of tourism area. The results show simple statistical display which it easier to understand like commercial software. Last section, concluding remarks and synthesis are summarized for this thesis, which it show urban growth occurs more than what was expected in Phuket Island.

1.3 Literature Review

Remote sensing data

The question of land-use change has attracted interest from a wide variety of researchers who focus on modeling the spatial and temporal patterns of land conversion and understanding the causes and consequences of these changes.

Land-use change has been investigated in many countries by geographers. In the past, researchers have argued about how to best monitor land-use change: by survey or remote sensing system (Hill, 1984). Hill also discussed how to study land-use change. Landsat data is appropriate for this purpose. When the remote sensing data provided high accuracy, land-use change was investigated. After land reform was initiated in 1987 by Remote Sensing (RS), land-use change was investigated by using high technology.

The majority of the researchers used remote sensing data to investigate land-use changes. RS data contains a lot of information that needs to be extracted, such as images, land properties, land valuation, and geography (Yang and Qiao, 2010; Strand *et al*., 2002; Weng, 2001; Manonmani and Suganya, 2010; Gret-Regamey *et al*., 2008). All of them also used RS data for investigating land-use change. GIS software computes the RS data which is used to develop land-use data and improve land-use planning. In 2010, researchers He-Bing and Su-Xia developed an application for land-use planning management information based on ArcGIS software which also uses RS data for analyzing land-use change. Yang and Qiao (2010) and Usha *et al.* (2012) also detected land-use change with image processing based on RS data.

In addition, another group of research scientists studied ecological systems and used GIS technology in environmental surveys (Gret-Regamet *et al.*, 2008). They also used RS data and link to show the ecosystem service.

Data structure (polygonal, block and pixel)

Remote Sensing data is beneficial for the study of land-use change. Analysis of land-use change is done by the representation of land-use type. As described in Thinnukool *et al*., (2014) that used and discussed polygonal, block and pixel method for analyzing land-use change. Klajnsek and Zalik (2005), Bach et al (2006), and Mizutani (2009) used GIS data to analyze polygonal shaped land-use data. They also focused on shape change and used polygon events and status to understand land-use change. The results from different inputs of land-use representation showed differences in accuracy all depended on software.

Nowadays, researchers use commercial software, such as AreGIS, AreView, MapInfo, Intergraph, IDRISI, SAGA GIS, and others. Important land-use changes are displayed on a thematic map and this requires the appropriate software. However, the most efficient software for detecting land-use change is also the most expensive. Nowadays, geographers have been able to develop GIS programs such as MapWindows GIS and Quantum GIS for detecting land-use change, but the limitation is that some techniques are not available as free software. The majority of RS data in the past depended on polygonal with aerial photography to study land-use change. Though polygonal data structures can provide thematic maps for displaying patterns for a given year, the data are difficult to analyze because the polygons change. Hun et al (2011), Stehman and Wickham (2011), Frazier and Wang (2011), Bach *et al*.

(2006) and Guo *et al*., (2011) described the use of pixels, blocks and polygons to construct accurate maps. One group of researchers, such as Whiteside *et al.,* (2011) confirmed that pixel-based construction can accurately show land-use maps from RS data. Based on pixels, the grid-digitization method provides a data structure that could be used directly for statistical analysis of land-use change.

Land-use cover and land-use code (LUCC)

Land-use is defined as human activity carried out on land which has natural resources. (Irwin and Geoghegan, 2001; Manonmani and Suganya, 2010; Madureira *et al*., 2007 and Rebelo, 2009). Land-use classification refers to a representation of an area that explains what is carried out on that particular land. It would be good to consider the land-use categories; notably, LUCC used for the land-use data in Phuket Island from 1967 to 2009 has changed. LUCC has been classified into past data and present data. In the past, land-use and land-cover data were sampled from digital land-use and land-cover files obtained from the United State Geological Survey (USGS).

Some of the land-use categories in the US didn't correspond to that of United Nations Educational, Scientific and Cultural Organization (UNESCO). According to the Los Angeles country planning commission, the land was classified into 10 groups starting from 000 to 999 of code (The Los Angeles County Planning Commission, 1941). In this thesis, the Thai Department of Land Development has classified the LUCC into 3 levels, but some LUCC in Thailand has been changed in 1985. This resulted in LUCC corresponding to real world land-use. In Thailand, LUCC had been modified in support of the new LUCC in sub-class. The new LUCC were re-organized to consider the similarity between old and new LUCC of Thailand land-use based on the Thailand Department of Land Development (Thinnukool *et al.*, 2014b).

Land-use change

Starting from 1987, remote sensing has been used to investigate land-use change (Hill, 1984). Several researchers have studied land-use changes which have affected the quality of human life. The disorder resulting from globalization has used up a lot of resources, especially the land. An example is the ecological system in Shenzhen, China (Li *et al.*, 2008), which has been adversely effected. The researchers listed below also studied land-use change: Bamesh (1989) focused on land-use in the Chiang Mai area using secondary data from aerial photographs. Nine groups of land-use have been classified and the results showed that the urban area has increased to other land-use categories and the maximum agriculture area has been converted to urban land over a period of 12 years. In western Thailand, Raine (1994), studied land-use change in the Chanthaburi province, which is a coastal zone. Deforestation was estimated to have decreased from 1975 to 1989.The pattern of change from forest area to urban area was also studied by Jim and Liu (2001). They researched the association between land-use and forests and their relation to the culture, history, biodiversity and pattern of change in Guangzhou, China. Moreover, they found the land-use scale as an important characteristic that affected the amount of land conserved for forests.

Hascic and Wu (2006) studied land-use change of the forest, and reported that the quality of drinking water in catchment areas had been affected adversely.

Martin and Assenov (2008) focused on Southeast Asia, especially the popular tourist areas on Phuket Island, such as Kata-Karon and Patong. They found that these areas have been encroached upon by urbanization in 2007.

Ying *et al.*, (2009) studied urbanization in Beijing, they predicted urban growth would increase from 2020 to 2049 and urbanization would disturb other categories land.
Zhao *et al.*, (2011) studied the interference of natural resources, especially the forest area, by deforestation in northeast of China.
Bagan and Yamagata (2012) estimated urban growth in Tokyo 40 years ahead: urban or new built-up density should decrease in the metropolitan inner core as the city center experiences depopulation. Liao and Wei (2012) researched urban growth in Dongguan, China, and the outcome was that the urban area has increased by 181 % from 1988 to 2006.
Kurt (2013) studied the land-use change of Black sea coastal regions in Istanbul. Agricultural and forest areas changed to urban area and the particular urban area increased to 122% over the two decades from 1987-2007.
Masucci *et al.,* (2013) studied urban growth in London; the limitation of urban growth in London's street network affected the urban growth corresponding to the street. The results showed that the trend of urbanization has increased.
Recently, on the topic of the City in 2050, by Brown (2013), with regard to the city as it grows, from rural to urban, there is a strong need for planning and managing recourses such as infrastructure, transportation, and regional planning for urbanization.

Land-use change modeling

Studying models of land-use change has been developed to analyze change. A land-use change model is important to explore possible future development in the land-use system (Verburg *et al*., 2006). The several techniques of modeling have been approached based on the underlying theory of statistics (Verburg *et al*., 2004). Land-use change models include stochastic models, optimization models, simulation models, empirical models, and logistic regression model (Li and Yeh 2002; Verburg *et al*., 2002; Dai *et al*., 2005; Castella *et al*., 2007; Dendoncker *et al*., 2007).

Logistic regression is one of the statistical methods used to analyze land-use data. Several researchers also use logistic regression using binary outcomes are used for predicting events that will happen or not happen. Logistic regression parameterizes the relationship between binary outcome and dependent variables using the logit link function, which constrains values dependent viable to be between 0 and 1 (Lee and Pradhan, 2007). Examples of such dependent variable of interest are land-use change or not change, presence or absence of landslides or in health science, low birthweight or not low birthweight. This model can handle both continuous and categorical variables. Also, the variables need not have normal distributions as required in the case of discriminate analysis (Ohlmacher and Davis, 2003).

Logistic regression coefficients can be used to estimate odds ratios for each of the independent variables in the model (Kundu *et al*., 2013) The dependent variable, Y has a binary value of 1 and 0 for Yes and No, respectively.

Moreover, logistic regression has been applied in several fields such as health science, ecology and environmental sciences.

Ramani *et al*. (2011) presented an approach for the analysis and modeling of landslide data using rare events logistic regression and applied the approach to an area in Lianyungang, China. In the present situation, the dependent variable is a binary variable representing the presence or absence of landslides by using the statistical analysis software, SPSS. It is said that using logistic regression may cause some problems if the total area affected by landslides is much smaller than the total study area. To avoid these problems, the rare events logistic regression (OLR) was used to produce landslide-susceptibility mapping in this study. The rare events logistic regression was computed with the relogit module. The first correction deals with the selection of a representative sample. It is generally recommended in logistic regression to use equal landslide rather than no-landslide pixels.

Olaniyi *et al* (2012), who studied coastal land use change in Malaysia between 1990 and 2006, biophysical and socio-economic data were as independent variables. Results showed that agricultural practices were particularly the cause for coastal land-use change. Coastal lands were converted to agricultural uses as a result of increased accessibility, suitable slope and favorable climatic condition of the areas.

In addition, Ayalew and Yamagishi (2005) studied landslide susceptibility maps in the Kakuda-Yahiko Mountains of Central Japan. The use of logistic regression is to find the best fitting function to describe the relationship between the presence or absence of landslides (dependent variable) and a set of independent parameters such as slope angle and lithology. Here, an inventory map of 87 landslides was used to produce a dependent variable, which takes a value of 0 for the absence and 1 for the presence of slope failures. Lithology, bed rock-slope relationship, lineaments, slope gradient, aspect, elevation and road network were taken as independent parameters. Using a predicted map of probability, the study area was classified into five categories of landslide susceptibility: extremely low, very low, low, medium and high. The medium and high susceptibility zones make up 8.87% of the total study area and involve mid-altitude slopes in the eastern part of Kakuda Mountain and the central and southern parts of Yahiko Mountain.

Manzo *et al*. (2012) studied landslide susceptibility (Yes/No) in Sicily, Italy by using binary logistic regression model. 31 independent variables comprise both the continuous as well as categorical variables. The model was implemented in a GIS environment by using the ArcSDM (Arc Spatial Data Modeller) module, modified to develop spatial prediction through regional data sets. In this study, Receiver Operator Characteristic (ROC) analysis was used to assess the performance of binary logistic regression model for landslide prediction. The area under curve showed 0.952 corresponding to landslide prediction accuracy of 95.2 %. The result also showed a strong agreement between distribution of existing landslides and predicted landslide susceptibility zones.

Moreover, Hu and Lu (2007) researched on modeling urban growth in Atlanta using logistic regression. RS data in a raster was transformed into a grid form for predicting land-use change. Population density, distances to nearest urban clusters, activity centers and roads, and high/low density urban uses (all with odds ratios < 1); and (2) distance to the CBD, number of urban cells within a 7×7 cell window, bare land, crop/grass land, forest, and UTM northing coordinate (all with odds ratios > 1) were found to affect urban growth in different degrees as indicated by odd ratios.

Nong and Du (2011) studied urban growth pattern modeling using logistic regression and it was used to testify the logistic regression compared to actual 2007 land use map to measure the simulated and real change. For the dependent represents urban growth results, Y (outcome) has a binary value of 1 and 0 for Yes and No, respectively. Independent variables have six involve population density, total volume of final industrial products, agriculture forestry, distance to economic center, distance to the major road, and slope. They also used ArcView GIS 3.2 and analyze by SPSS, which found that changed of land and an accuracy of prediction urban growth was 60.5% of classification.

However, this research also ignored the spatial autocorrelation in geographical phenomenon also has influence and ROC curve is particularly important for evaluating how good the decision making is at discriminating between stable versus unstable areas) values was and this logistic model has the largest area under the curve 0.891 with standard error 0.001.

Eyoh *et al*. (2012) studied modeling and predicting future urban expansion of Lagos, Nigeria from remote sensing data using logistic regression that the prediction for 46 years was obtained for the period 1984-2030. Ten independent variables contain distance to water, distance to less dense urban, distance to dense urban, distance to major roads, distance to railway, distance to Lagos Island, distance to international airport, distance to seaport, distance to University of Lagos, and distance to Lagos State University. Kappa statistics which was computed from a performance matrix resulting from the comparison of the reference with the predicted data was 0.6998. However, this research didn't consider spatial autocorrelation. The results expanded urban growth in Lagos by 56.90% from 1984-2000 and by 64.04% from 1984-2005.

Chakir and Gallo (2013) also used RS data to predict land-use allocation in France from 1992 to 2003, which used logistic regression. Four groups (independent) were considered such as land use agriculture, forest, urban and other land-use. They used a pooled model to predicted urban growth and ignored spatial autocorrelation. The results shows that urban gained was 17.65%, forest gained was 5.33% and agriculture loosed was 3.48%.

Example in health science research, Bonellie (2012) applied logistic regression to investigate changes in birthweight for term singleton infants in Scotland between 1994 and 2003. The model was used to investigate changes over time in size of babies particularly in relation to social deprivation, age of the mother and smoking. The results showed that, smoking during pregnancy is shown to have a detrimental effect on the size of infants at birth.

In conclusion, the conception for using logistical regression for land-use change, we have to set independent variables (x) location, population, distance to urban, volume for forest, distance to river, etc. for example which they are causes to gain of urban growth probability (y). Note that an independent correlation mean on that is not influenced by values of the other predictor variables. In case of land-use change or land-use prediction by using a logistical regression approach, one must be cautious about spatial autocorrelations that often exist in spatially referenced data because they may violate the assumption of the model (Hu and Lo, 2007).

However, this assumption clearly does not hold for data defined at grid-points just 100 meters apart. In other studies with geographical data such spatial correlation can be handled by aggregating data into larger regions with acceptably small correlation between adjoining regions, or by using factor analysis and multivariate regression to adjust for spatial correlations. For our land-use data, we used a conventional method widely used in survey sampling, based on variation inflation factors (Rao and Scott, 1992). The conventional statistical analysis assumes that errors after fitting a model are independent. In fact this assumption fails for your data because observations at the grid-points in your samples are spatially correlated. This method computes effective sample sizes for each land-use plot based on their sample variances, from which standard errors are applied to fitted values from a logistic model to compute confidence intervals.

1.4 Objectives

The study of analyzing land-use change using digitization follows the objectives below:

First, develop a new data structure of the land-use data using digitization.

Second, investigate the land-use change of the Phuket Island from 1967 to 2009 using the digital grid method.

Third, predict land-use change in Phuket Island and tourist areas from 1967 to 2009.