2015 The 5th International Workshop on Computer Science and Engineering
doi: 10.18178/wcse.2015.04.130

Designing Basic Relational Database for Analysis Land-Use Change

Orawit Thinnukool[1], Noodchanath Kongchouy[2]

[1]Department of Modern Management and Information Technology, College of Arts, Media and Technology, Chiang Mai University,
Chiang Mai, Thailand, 50200
[2]Department of Mathematics and Statistics, Faculty of Science, Prince of Songkla University, Hat Yai Campus, Songkhla, Thailand. 90112
1Corresponding Author: orawit.t@cmu.ac.th

Keywords: Land-use analysis, land-use change, relational database, prediction

Abstract. The processing and analysis of spatial data especially land-use are becoming increasingly dependent on the methodology used in managing the data rather than only using only Geographic Information System (GIS) software. Our approach in handling spatial data for recorded land-use data using basic land-use concept saves cost and is effective for developing land-use analysis. The aim of the study is to design and explain how land use data can be retrieved and managed by a free. The example was used to analyze land-use change, with freely available tools. These tools can handle the spatial data very well. The computation was based on the actual number of observations via Google Earth with free version on pilot area. The results demonstrate the use of normal statistics analysis for prediction in the analysis of land use change is easy to understand.

Introduction

Land-use investigation has been changes the world when the Remote Sensing (RS) technology developed. Land-use change can explain the phenomenon of the activities on a land. The land-use data contain spatial information such as environmental, social, economics, geographical etc., which are of high value for planning or developing the world nowadays. The data recorded more than million records into database of the government official, such huge database are difficult for information retrievals and need the use of only the GIS software. Moreover, these data are important for the government making it difficult to get it from their office.

Researchers around the world developed a methodology and land-use database for studying land-use change with different positions. Recently, new database management has been developed to handle relational information. In the past database idea was been designed in three stages such as conceptual, logical andphysical [1]. Database has been used for land-use investigation in geographical information system (GIS) field. Several researchers used database to study land-use and managed land-use database.

Suk et al. developed urban GIS database for integrated land-use and property development in Seoul, South Korea. The database contains type of properties such as building, land-use type and land quality[2].

This database has useful information for land-use planning. Jose implemented land-use data from landsat thematic mapper (TM) as an operational relational database, which allows querying for detailed land cover/use information. RDBMS Microsoft Access was used to implement the database structure and structured query language (SQL) was used to analyze elementary terrain information[3]. Liu et al. developed land-use cover in China which greatly supported the nation land-use/land-cover (LUCC) change. The LUCC was developed by the definition of land-use categories and recorded into database in commercial program [4].

In addition, Zhou developed an object-relational prototype of a GIS-based disaster database that reported information on disastrous events, from perspectives of both emergency management and spatial database development. The development of database has been approach by programs such as web-based GIS systems, ArcGIS server and etc. [5]. Moreover, Maryati et al. managed GIS database for environmental management in Indonesia, where ArcGIS was used to handle the task and also used ArcCatalog software for classifying land-use type [6]. All studies can implement setups for achieving useful database but the solutions, recordings and transforming land-use data are quite complicated and the management depends on only the commercial software or other free software.

Although studying land-use database also requires the use of software, huge database of land-use nowadays has been developed by simple method to manage the data with the help of freely available software. Data sources for analysis of land-use change use Remote Sensing data (RS), which it comes from satellite image. But the cost of satellite image is high if you want to use in investigating land-use change. Google Earth provides freely information for investigating land-use if we want to estimate the changes and show basic statistics presentation. However, Thinnukool et al. developed program command for detecting land-use by using R program, which can detect land-use data from shape file. The question is how to manage land-use data or investigate land-use change without costing a dime if we have no information (data) and no software (GIS) [7].

First of all, the points that we need to consider are how to record land-use data, how to get the information from Google and how to analyze the changes that have occurred. The aim of this research is to design the basic land-use data analysis with relational database for analysis which starts with the questioning how we can record land-use data when we have no information and tool. Firstly, we can use eye estimation method to classify land-use type such as developed and undeveloped areas [8].

The data were managed and recorded by the use of a common tool other than GIS commercial software. It is expected that basic land-use databases can retrieve the information for analysis, planning, and can also display the basic statistics of an area.

Materials and Methods

The first step was to get the land-use data from Google Earth. The land-use data were retrieved from Google to display the area that we wanted to record. The study area is close to Chiang Mai University (Fig. 1) The picture was taken in 2013, bird eye view capture at 1.48 km. Eye estimation method has been used to classify land-use type [8]. Although, this method does not provide accuracy in position of a specific area, is commonly used. The advantage of this method is because it’s fast to get data (saves time), and the shortcoming is that the precision is low.

Fig.1 Pilot area

This step we used the satellite image from Google Earth in Fig. 2, with the area been 10x10 meter (square). We used these symbols to represent the land use type (D indicates developed area and U undeveloped area). For example, first row and fourth column indicates D, whereas the area on the map located by the building at bottom right indicates U because it consists of trees. In the next step, positions of each area (square) which is ordered as X and Y are located, and transferred into the table.

Table 1: Records the positions of land-use types, which corresponds to Fig. 2(example, 2 rows 2 columns).

Position	X	Y	Type
X_1Y_1	494.760	207.775	D
X_2Y_1	494.770	207.775	D
X_3Y_1	494.780	207.775	D
.	.	.	.
.	.	.	.

$X_{10}Y_1$	494.850	207.775	D
X_1Y_2	494.760	207.765	D
X_2Y_2	494.770	207.765	D
X_3Y_2	494.780	207.765	D
.	.	.	.
.	.	.	.
$X_{10}Y_2$	494.850	207.765	U

Image Transformation

Image transformation of land-use map in Google Earth is capture via google earth software (free version)

Imagery Date:2/9/2013 47Q494848.35m E2077713.94 mN elev 100m eye alt 500m

Fig. 2 Google map has been divided into squares, 10x10 meters (1 grid) illustrating the land-use types. Note that example of the area of study has 10 rows and 7 column, which is ordered by coordinate x, y with the red lines indicating the built up areas.

Data Management

Creating database one needs to have data table to link the data which are related in order to have accuracy in results. Basically, database concepts are designed to handle data structure information. It starts with recording the positions of points and considering land-use type into each box (raster format), which is known in remote sensing as “supervise classification and unsupervised classification”. Google Earth provides high resolution map, which land-use type easily is indicated in the map by zooming in and out. Moreover, 3D function can show clearly images on the map. After downloading the map and recording the positions and land-use types into the table, we use the information from Fig. 2 to change and insert Primary key (PK), which is illustrated in table 2 right panel. Land-use data for each year can shows historical imagery by using Google Earth software (free version) to see the changes on the Google map. Note that, we can record land-use data in each year to investigate land-use change.

Basically, the left and right panels for table 3 are related by PointID. These two tables can be extended to more tables such as finding the density of population, temperature or type of building and environmental conditions etc. (example in right panel of Table 3), depending on information for the study.

Conceptual model illustrated in Figure 3demonstrates land-use dataset mainly contains PlotID, and X Y positions, which shows the relationship between these two attributes land-use position and land-use type. Land-use type contains PlotID, land-use type (Developed or Undeveloped), and

Land-use description (explains the characteristic of land-use). Other attributes such as ownership of land, population, and price of land are factors that can be included in finding the changes in land use.

Table 2: The left table shows land-use positions and land-use types, but with the table on the right the position has been renamed as PointID.

Position	X	Y	Type	PointID	X	Y	Type
X_1Y_1	494.760	207.775	D	1	494.760	207.775	D
X_2Y_1	494.770	207.775	D	2	494.770	207.775	D
X_3Y_1	494.780	207.775	D	3	494.780	207.775	D
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
$X_{10}Y_1$	494.850	207.775	D	10	494.850	207.775	D
X_1Y_2	494.760	207.765	D	11	494.760	207.765	D
X_2Y_2	494.770	207.765	D	12	494.770	207.765	D
X_3Y_2	494.780	207.765	D	13	494.780	207.765	D
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
$X_{10}Y_2$	494.850	207.765	U	20	494.850	207.765	U
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
				70	494.850	207.705	U

Table 3: Left table shows land-use position (X and Y) and PointID (PK), and the right one shows PointID and land-use classification (Type)

PointID	X	Y
1	494.760	207.775
2	494.770	207.775
3	494.780	207.775
.	.	.
.	.	.
.	.	.
10	494.850	207.775
11	494.760	207.765
12	494.770	207.765
13	494.780	207.765
.	.	.
.	.	.
.	.	.
20	494.850	207.765
.	.	.
.	.	.
.	.	.
70	494.850	207.705

PointID	Type	PointID	Variable	Variable
1	D	1		
2	D	2		
3	D	3		
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
10	D	10		
11	D	11		
12	D	12		
13	D	13		
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
20	U	20		
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
70	U	70		

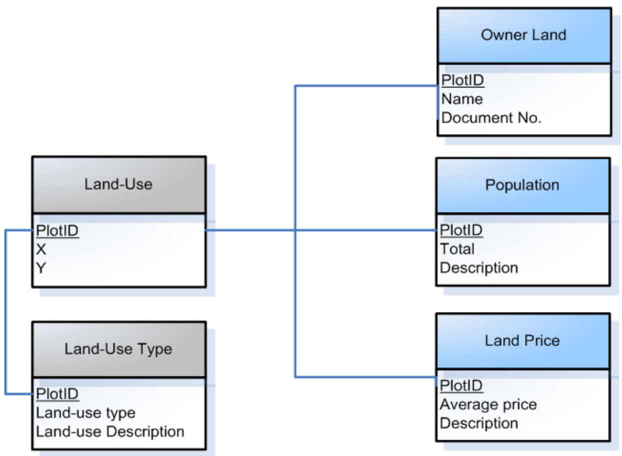

Fig.3: Conceptual model

Next step we record land-use data into a database by R program download at http://cran.r-project.org/bin/windows/base/. R program is a freely available software, which is once of programming language and environmental software for statistical computing and graphics. R program can analyze data in statistical term especially spatial data. Land-use data has been recorded into notepad, which it contains attribute of land-use data such as PlotID, x, y, type2000, type2005, type2010, type2013 and development2013 which are shown in figure 4 below:

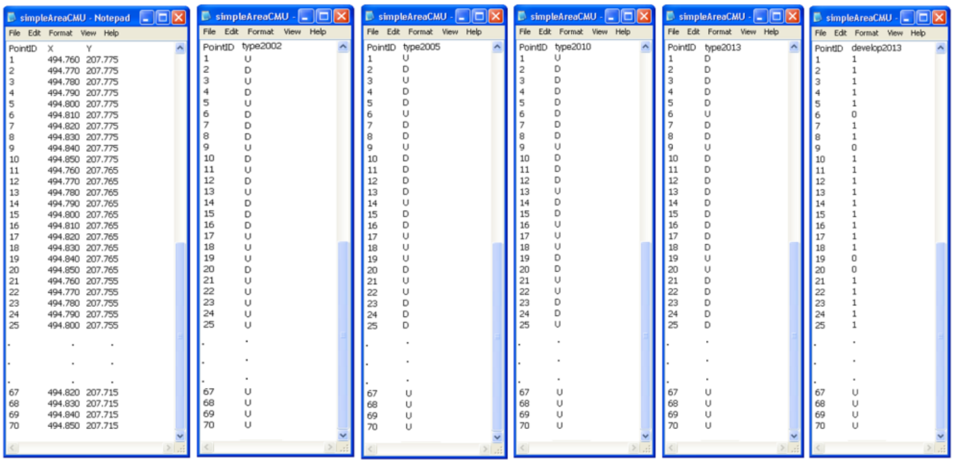

Fig. 4 Notepad files contains land-use data that was recorded from Google Earth free version. Next step, when R program has been install into the computer, the file in Fig. 5 are then computed by the program.Command line in R program starts to read the notepad file (.txt) in to R program and import file name “simpleAreaCMU.txt” into p1 (variable).

```
> options(scipen=4)# display numbers properly
> setwd("h:/")
> read.table("simpleAreaCMU.txt",h=T,as.is=T) -> p1
> p1
   PointID      X       Y type2002 type2005 type2010 type2013 develop2013
1        1 494.76 207.775        U        U        U        D           1
2        2 494.77 207.775        D        D        D        D           1
3        3 494.78 207.775        U        U        D        D           1
4        4 494.79 207.775        D        D        D        D           1
5        5 494.80 207.775        U        D        D        D           1
6        6 494.81 207.775        D        U        D        U           0
7        7 494.82 207.775        D        D        D        D           1
8        8 494.83 207.775        D        D        D        D           1
9        9 494.84 207.775        U        U        U        U           0
10      10 494.85 207.775        D        D        D        D           1
.        .      .       .        .        .        .        .           .
.        .      .       .        .        .        .        .           .
.        .      .       .        .        .        .        .           .
67      67 494.82 207.715        U        U        D        U           0
68      68 494.83 207.715        U        U        U        U           0
69      69 494.84 207.715        U        U        U        U           0
70      70 494.85 207.715        U        U        U        U           0
> |
```

Fig. 5 The land-use data in the R program is contained in p1 (variable). Note that, attribute develop2013 show land-use changes to developed area (indicate 1) and no changes (indicate 0), which this attribute investigate land-use change between each year and 2013.

Analysis of Land-use

This section explains how to analyze land-use data which was recorded in previous section in a table form. The data were recorded for a period of seven year, 2002, 2003, 2005, 2010, 2011, 2012, and 2013. Next step the land-use data from all the years will be managed by Excel to count the number of land-use type. After which statistical model will used to predict the probability of land-use change in the years and this model is known as logistic regression. Logistic regression was used because the outcome variable of the model is binary.

Classifications of the outcome variable in this article are developed land or undeveloped land. The model creates a linear relationship between independent variable (X) and the dependent variable (Y). The logistic regression formula is indicated below;

$$P = (Y = 1/X_i) = \frac{1}{1+e^{\alpha+\sum_{i=1}^{n}\beta_i X_i}} \tag{1}$$

P is the probability of occurrence of developed land. Y is the dependent variable which is developed land in each year. X_i depicts the independent variable. Normally using logistic regression need to classify observed and predicted values of model. Using binomial logistic regression need to check percentages of correction. However, this paper only shows the conceptual framework for develop and management land-use data.The logistic regression is an easy command in R, Fig. 4 presents the command to analyze land-use change.

```
# logistic regression models
options(contrasts=c("contr.sum","contr.poly"))
glm(data=p1,family=binomial,develop2013~factor(type2002)) -> mod
summary (mod)
```

Fig. 6: R (programming) command for analyzing land-use change with logisticregression

Results and Discussion

According to the command in Fig. 6 land-use change was analyzed, where the probability for a change in area of land was found (developed or undeveloped). The results from using statistical model such as logistic regression is shown using the example of land-use change from 2002-2013, which data for small sample of 70 points (70 hectares) were selected from Fig. 2 Result from fitting logistic regression model shows undeveloped area is very likely to change to developed area (p-value for no change 0.0075). Fig. 7 (small rectangle) explains the estimated value 0.1863, which tells the probability of land-use change in 2013 based on 2002, the occurrence of urban growth is 1.20 times the occurrence of no urban growth. Standard error was 0.2833, z-value was 0.658 and p-value was 0.51078.

```
> # logistic regression models
> options(contrasts=c("contr.sum","contr.poly"))
> glm(data=p1,family=binomial,develop2013~factor(type2002)) -> mod
> summary (mod)

Call:
glm(formula = develop2013 ~ factor(type2002), family = binomial,
    data = p1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5829  -1.4152   0.8203   0.9569   0.9569

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)         0.7300     0.2833   2.576  0.00999 **
factor(type2002)1   0.1863     0.2833   0.658  0.51078
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 90.008  on 69  degrees of freedom
Residual deviance: 89.565  on 68  degrees of freedom
AIC: 93.565

Number of Fisher Scoring iterations: 4

> |

> exp(0.1863)
[1] 1.204784
> |
```

Fig. 7: R (programming) command for analyzing land-use change

First period, land-use change between 2002 and 2005 is not too much, which logistic regression predicts land-use change from undeveloped area to developed area is 0.205 times. Second period, growth of development is 0.596 times. Third period, land-use change from 2010 to 2013 is 1.20

times. Figure 8 illustrates bar chart for land-use change in four periods. To see the land-use change year by year, bubble plot illustrated land-use change three periods.

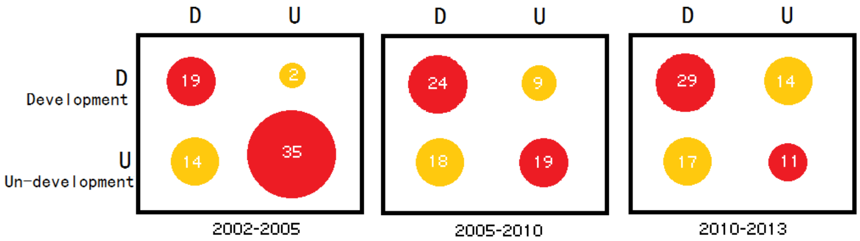

Fig. 8: Bubble plots show numbers of land-use change in hectare unit.

Conclusions

The example was used to analyze land-use change, with freely available tools such as Google Earth, and R program. These tools can handle the spatial data very well. Although, this research ignores factors that support land-use change which was stated from the beginning of the paper, the analysis was basically for investigation of land-use change, in which a researcher can try to learn without any GIS program. Researchers can develop programming command to show bar chat or any statistical display, which make it easy to read and understand. However, the spatial data need to be checked for an accuracy of thematic map. We expect that the freely available software will release the program with up to date information, which provides a good opportunity to management land-use or land-use planning.

Acknowledgements

The authors thank Prof. McNeil D. for his suggestions. Department of Modern Management and Information Technology, College of Arts, Media and Technology, Chiang Mai Universityis also gratefully acknowledged.

References

[1] Connolly, T.M. and Begg, C.E. Database System: A Practical Approach to Design, Implementation and Management. 4rd Edn., Addison Wesley, 2004.

[2] Suk O. H., Kim, H.C., &Seong, C.J. An urban GIS Database Model for Integrated land and Building Property Management.Michigan Academician, 32, pp. 421-437, 2000.

[3] Jose, A. M. A. cartographic and database approach for land cover/use mapping and generalization from remotely sensed data. International Journal of Remote Sensing, 21, pp.1825-1842, 2000.

[4] Liu, J., Liu, M., Deng, X., Zhuang, D., Zheang, Z. and Luo, D. The land use and land cover change database. Environmental Science, 11, pp.706-714, 2011.

[5] Zhou, Y., Liu, G. J., Fu, E. J, and Zhang, K. F. An object-relational prototype of GIS-based disaster database.Procedia Earth and Planetary Science, 1, pp.1060-1066, 2009.

[6] Sri Maryati, Hideki Shimada, Takashi Sasaoka, Akihiro Hamanaka, Kikuo Matsui, and Hideaki Nagawa.GIS Database Template for Environmental Management of Mining in Indonesia.Journal of Geographic Information System, 4, pp. 62-70, 2012.

[7] Thinnukool, O., Chuangchang, P., Khongkraphan, K.Analyzing land use change using grid-digitized method. Songklanakarin Journal of Science and Technology, 36(2), pp. 235-240,2014.

[8] Hui, L., Pei, L., Junshi, X., Huasheng, S., Kun, T., and Shuwei, D. Research on Technology for land-use Map renewal and Database Construction. Procedia Environmental Science, 11, pp.706-712, 2011.

Author Index