# Chapter 2

# Methodology

This section describes the methods used, including an overview of the management of data, statistical methods for data analysis, geometric correction, management of land-use code and land-use cover, digitization, complex regions, and statistical analysis.

## 2.1 GIS data structure

The Thailand Department of Land Development has recorded millions of records for land-use data from regular surveys and Remote Sensing (RS). The data are stored in several types, which need to be combined to obtain thematic maps. Such maps can be created directly from these files using commercial software such as MapInfo and ArcGIS. Alternatively, the files can be restructured into relational database tables and used with general purpose programs such as R. For MapInfo or ArcGIS use, the files for Phuket in 2009 are as shown in Figure 1.

Prince of Songkla University Pattani Campus


Figure 1. Files to illustrate thematic map from Thailand Department of Land Development.

To illustrate these files we focus on a subset comprising just one small island south of the Phuket mainland ("Coral Island") that contains 12 land-use plots. Note that each record in this MID file corresponds to a plot. The MID file has 8 fields. AREA is the total area in the polygon in square meters. PERIMETER is the total distance around the border of the polygon. LU_CODE refers to type of land-use. DES_TH and DES_EN are descriptions in Thai and English. RAI, ACRES and HECTARES specify these units, as shown in Figure 2. Note that the file does not contain a primary key and is thus not a relational database table, but it can be converted to such a table simply by inserting an index field.

| AREA | PERIMETER | LU_CODE | DES_EN | DES_TH | RAI | ACRES | HECTARES |
|---|---|---|---|---|---|---|---|
| 87607.007 | 1763.417 | A405 | มะพร้าว | Coconut | 55 | 21.648 | 8.761 |
| 235285.827 | 2348.238 | A405 | มะพร้าว | Coconut | 147 | 58.14 | 23.529 |
| 242126.513 | 2934.426 | U201/A405 | หมู่บ้าน/มะพร้าว | Village/Coconut | 151 | 59.831 | 24.213 |
| 39322.12 | 744.409 | A302 | ยางพารา | Para rubber | 25 | 9.717 | 3.932 |
| 17945.538 | 665.325 | A302 | ยางพารา | Para rubber | 11 | 4.434 | 1.795 |
| 66632.936 | 1101.126 | A302 | ยางพารา | Para rubber | 42 | 16.465 | 6.663 |
| 10080.674 | 461.133 | A302 | ยางพารา | Para rubber | 6 | 2.491 | 1.008 |
| 15269.779 | 492.801 | A302 | ยางพารา | Para rubber | 10 | 3.773 | 1.527 |
| 86619.535 | 1630.524 | A302 | ยางพารา | Para rubber | 54 | 21.404 | 8.662 |
| 3091006.074 | 13333.328 | F101 | ป่าดิบสมบูรณ์ | Dense evergreen forest | 1932 | 763.801 | 309.101 |
| 32929.351 | 1016.317 | A405 | มะพร้าว | Coconut | 21 | 8.137 | 3.293 |
| 657247.445 | 6928.178 | A302 | ยางพารา | Para rubber | 411 | 162.409 | 65.725 |

Figure 2. Example of data structure of original form from MID file

The corresponding MIF (MapInfo Data Format) file contains values of the coordinates X, Y and other information. Again the records are not indexed, and their physical order needs to be the same as the plots in the MID file.  Figure 3 illustrates this file structure for the Coral Island data. It contains 780 records. The first 15 records contain the names of the fields in the MID file, and the remaining 765 records comprise 12 groups corresponding to the plots in the MID file. The first record in each group is of the form "Region n", when n is 1 unless the plot contains "subplots" in which case n is the number of subplots. Subplots are of two kinds: "holes" and "islands".  For example, group 10 in the MIF file contains four such holes, labeled 10a, 10b, 10c and 10d, so n is increased to 5.  For each plot corresponding to a hole the data are duplicated: the same (X, Y) coordinates appear in plots 6, 8, 4 and 7, respectively. Note also that these coordinates are listed in clockwise order around the boundary of the plot, except that the duplicated points inside a plot are listed in the

Prince of Songkla University Pattani Campus

reverse order to ensure that the total area is correct, because the formula for the area of a polygon gives a result with the opposite sign when the order of the boundary points is reversed. "Island" subplots are clusters of plots not contained in any plot; Coral Island does not contain any subplots with this property. Each group in the MIF file also contains three trailing records, which specify how the vectors are graphed in the map and the position of the centroid of each plot.

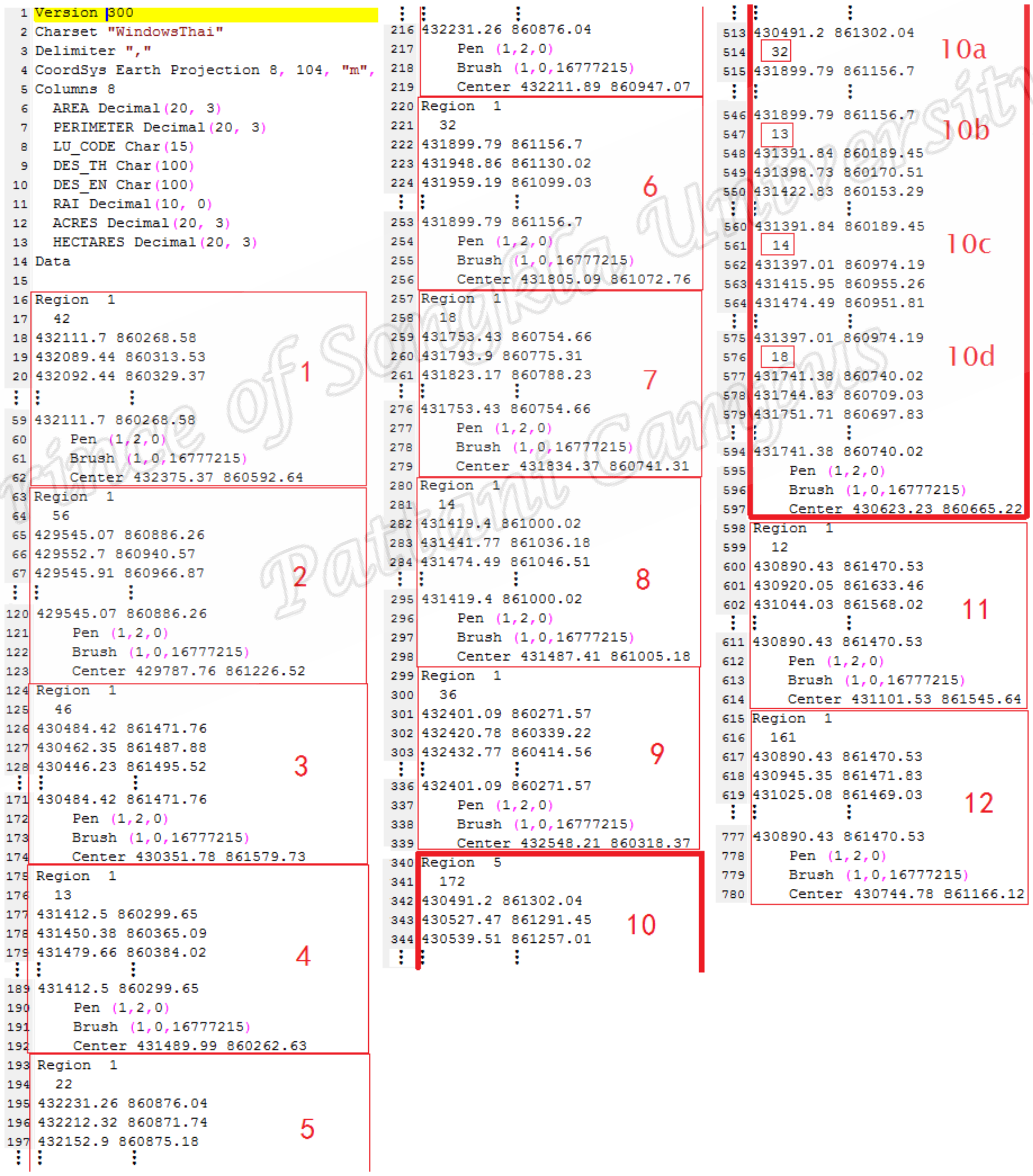



Figure 3. Data structure of MIF file in Coral Island (south of Phuket)

As is clear from Figure 3, the MIF file is not a relational database table because it is not indexed and it also contains a mixture of data and design information. To create thematic maps using a general purpose program, it is necessary to restructure the MIF file to be a relational database table. This new data structure is shown in Figure 4.

Prince of Songkla University
Pattani Campus

| plotID | pointID | x | y | |
|---|---|---|---|---|
| 1 | 1 | 432111.70 | 860268.58 | |
| 1 | 2 | 432-89.44 | 860313.53 | |
| 1 | 3 | 432092.44 | 860329.37 | |
| . | . | . | . | 1 |
| . | . | . | . | |
| 1 | 42 | 432111.7 | 860268.58 | |
| 2 | 1 | 429545.07 | 860886.26 | |
| 2 | 2 | 429552.7 | 860940.57 | |
| 2 | 3 | 429545.91 | 860966.87 | |
| . | . | . | . | 2 |
| . | . | . | . | |
| 2 | 56 | 429545.07 | 860886.26 | |
| 3 | 1 | 430484.42 | 861471.76 | |
| 3 | 2 | 430462.35 | 861487.88 | |
| 3 | 3 | 430446.23 | 861495.52 | |
| . | . | . | . | 3 |
| . | . | . | . | |
| 3 | 46 | 430484.42 | 861471.76 | |
| 4 | 1 | 431412.5 | 860299.65 | |
| 4 | 2 | 431450.38 | 860365.09 | |
| 4 | 3 | 431479.66 | 860384.02 | |
| . | . | . | . | 4 |
| . | . | . | . | |
| 4 | 13 | 431412.5 | 860299.65 | |
| 5 | 1 | 432231.26 | 860876.04 | |
| 5 | 2 | 432212.32 | 860871.74 | |
| 5 | 3 | 432152.9 | 860875.18 | |
| . | . | . | . | 5 |
| . | . | . | . | |
| 5 | 22 | 432231.26 | 860876.04 | |
| 6 | 1 | 431899.79 | 861156.7 | |
| 6 | 2 | 431948.86 | 861130.02 | |
| 6 | 3 | 431959.19 | 861099.03 | |
| . | . | . | . | 6 |
| . | . | . | . | |
| 6 | 32 | 431899.79 | 861156.7 | |
| 7 | 1 | 431753.43 | 860754.66 | |
| 7 | 2 | 431793.9 | 860775.31 | |
| 7 | 3 | 431823.17 | 860788.23 | |
| . | . | . | . | 7 |
| . | . | . | . | |
| 7 | 18 | 431753.43 | 860754.66 | |
| 8 | 1 | 431419.4 | 861000.02 | |
| 8 | 2 | 431441.77 | 861036.18 | |
| 8 | 3 | 431474.49 | 861046.51 | |
| . | . | . | . | 8 |
| . | . | . | . | |
| 8 | 14 | 431419.4 | 861000.02 | |
| 9 | 1 | 432401.09 | 860271.57 | |
| 9 | 2 | 432420.78 | 860339.22 | |
| 9 | 3 | 432432.77 | 860414.56 | |
| . | . | . | . | 9 |
| . | . | . | . | |
| 9 | 36 | 432401.09 | 860271.57 | |

| | | | | |
|---|---|---|---|---|
| 10 | 1 | 430491.2 | 861302.04 | |
| 10 | 2 | 430527.47 | 861291.45 | |
| 10 | 3 | 430539.51 | 861257.01 | |
| . | . | . | . | 10 |
| . | . | . | . | |
| 10 | 172 | 430491.2 | 861302.04 | |
| 10 | 173 | NA | NA | |
| 10 | 174 | 431899.79 | 861156.7 | |
| 10 | 175 | 431858.47 | 861165.32 | |
| 10 | 176 | 431815.42 | 861172.21 | |
| . | . | . | . | 10a |
| . | . | . | . | |
| 10 | 205 | 431899.79 | 861156.7 | |
| 10 | 206 | NA | NA | |
| 10 | 207 | 431391.84 | 860189.45 | |
| 10 | 208 | 431398.73 | 860170.51 | |
| 10 | 209 | 431422.83 | 860153.29 | |
| . | . | . | . | 10b |
| . | . | . | . | |
| 10 | 219 | 431391.84 | 860189.45 | |
| 10 | 220 | NA | NA | |
| 10 | 221 | 431397.01 | 860974.19 | |
| 10 | 222 | 431415.95 | 860955.26 | |
| 10 | 223 | 431474.49 | 860951.81 | |
| . | . | . | . | 10c |
| . | . | . | . | |
| 10 | 234 | 431397.01 | 860974.19 | |
| 10 | 235 | NA | NA | |
| 10 | 236 | 431741.38 | 860740.02 | |
| 10 | 237 | 431744.83 | 860709.03 | |
| 10 | 238 | 431751.71 | 860697.83 | |
| . | . | . | . | 10d |
| . | . | . | . | |
| 10 | 253 | 431741.38 | 860740.02 | |
| 11 | 1 | 430890.43 | 861470.53 | |
| 11 | 2 | 430920.05 | 861633.46 | |
| 11 | 3 | 431044.03 | 861568.02 | |
| . | . | . | . | 11 |
| . | . | . | . | |
| 11 | 12 | 430890.43 | 861470.53 | |
| 12 | 1 | 430890.43 | 861470.53 | |
| 12 | 2 | 430945.35 | 861471.83 | |
| 12 | 3 | 431025.08 | 861469.03 | |
| . | . | . | . | 12 |
| . | . | . | . | |
| 12 | 161 | 430890.43 | 861470.53 | |

Prince of Songkla University Pattani Campus

Figure 4. Data in Figure 3 after restructuring & indexing as a relational database table

Figure 5 shows a map created from the data in Figure 4 using Excel. This map may be created by starting with the set of (X, Y) coordinates for plot 10 with cells containing

NAs omitted, and then adding separate points for the other 15 sets (including the sets for 10a, 10b, 10c and 10d, again omitting the cells containing NAs).

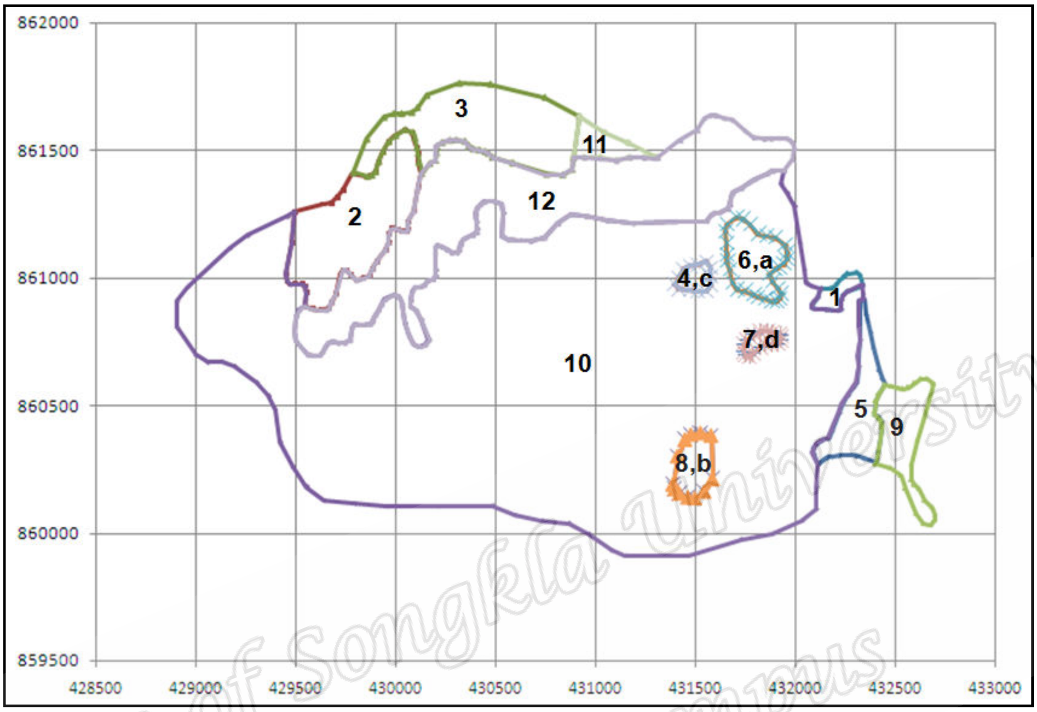



Figure 5. Map created from data in Figure 4 using Excel

Prince of Songkla University Pattani Campus

**2.2 Coordinate shifts (geometric correction)**

Image Geometry Correction is the process of digitally manipulating image data so that the image's projection precisely matches a specific projection surface or shape.

Data from RS needs to be checked for geometric correction because of distortion created by incorrect positioning. The standards for measuring GIS data have changed in recent decades. These coordinate shifts are quite substantial and complicate the accurate measurement of land-use change. Assuming that coordinates available from Google Earth maps are correct and that these locations have not changed substantially over recent decades, it is desirable to convert all land-use coordinates to agree with the corresponding Google Earth coordinates. The method we use for this conversion is based on a bilinear transformation of the form

$$u = a_1 + b_1x + c_1y + d_1xy \qquad (1)$$

$$v = a_2 + b_2x + c_2y + d_2xy \qquad (2)$$

In these equators ($x$, $y$) are coordinates in the files that need to be corrected and ($u$, $v$) are their corresponding corrected values that agree with Google coordinates.

The parameters ($a_1$, $b_1$, $c_1$, $d_1$, $a_2$, $b_2$, $c_2$, $d_2$) in equations (1) and (2) are determined by using the data for the coordinate shifts (dx, dy) at the four locations mapped in Figure 6. These equations are expressed in matrix form as

$$\boldsymbol{g} = \boldsymbol{F}\,\boldsymbol{h} \qquad (3)$$

In this formulation $\boldsymbol{g}$ is the column vector ($u_1$, $v_1$, $u_2$, $v_2$, $u_3$, $v_3$, $u_4$, $v_4$), $\boldsymbol{h}$ is the column vector ($a_1$, $b_1$, $c_1$, $d_1$, $a_2$, $b_2$, $c_2$, $d_2$) and $\boldsymbol{F}$ is an 8 × 8 matrix, as follows.

$$\begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \\ u_4 \\ v_4 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & x_1y & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_1 & y_1 & x_1y \\ 1 & x_2 & y_2 & x_2y & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_2 & y_2 & x_2y \\ 1 & x_3 & y_3 & x_3y & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_3 & y_3 & x_3y \\ 1 & x_4 & y_4 & x_4y & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_4 & y_4 & x_4y \end{bmatrix} \times \begin{bmatrix} a_1 \\ b_1 \\ c_2 \\ d_2 \\ a_3 \\ b_3 \\ c_4 \\ d_4 \end{bmatrix}.$$

Prince of Songkla University Pattani Campus

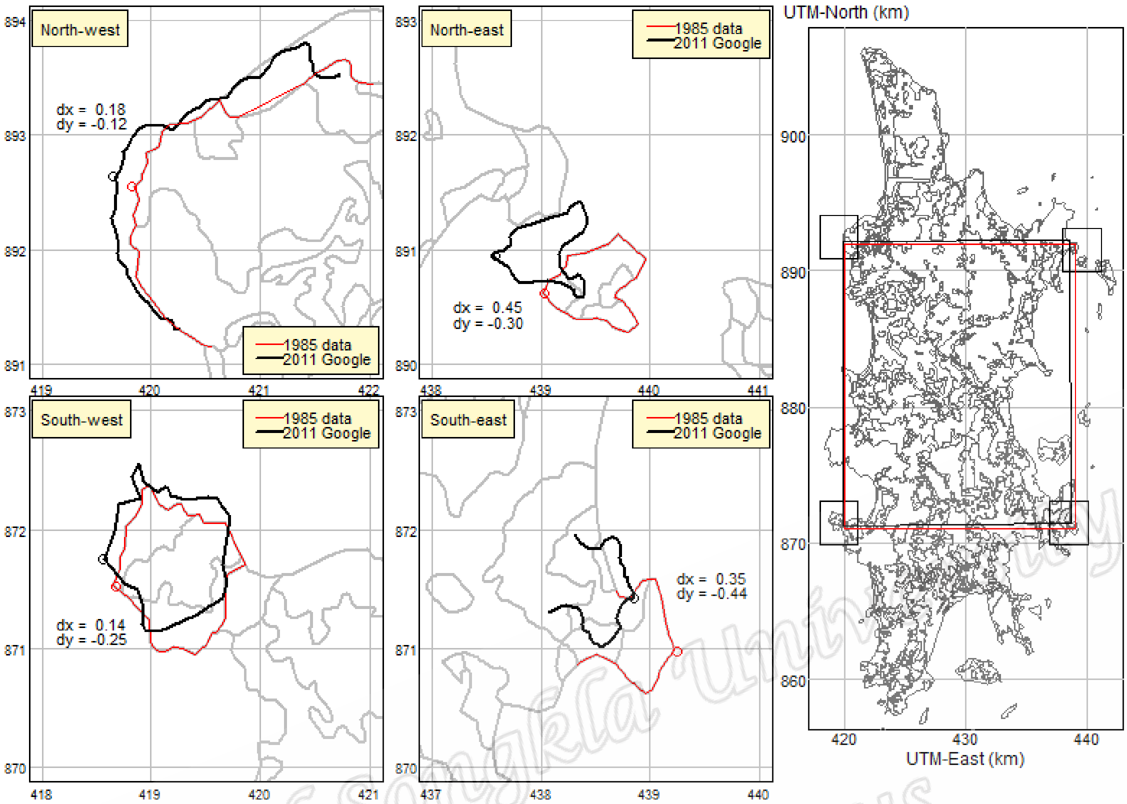



Figure 6. Geometric correction (in kilometers) of UTM coordinates positions based on land-use records recorded by the Thailand Department of Land Development in 1985 to those in Google Earth in 2011

Table 1 shows (x,y) coordinates of the location of the rectangle (coloured red) in the right panels, with corresponding (u,v) coordinates.

Table 1. Coordinate shifts in Phuket Island based on information in Figure 6

| Rectangle Corner | *x, y* | km | *u, v* |
|---|---|---|---|
| North east | 439.0, 892.0 | -0.45, 0.30 | 438.55, 892.30 |
| North west | 420.0,871.0 | -0.18, 0.12 | 420.18, 870.89 |
| South east | 439.0, 871.0 | -0.35, 0.44 | 438.65, 871.44 |
| South west | 420.0, 871.0 | -0.14,0.25 | 419.86, 871.25 |

Prince of Songkla University Pattani Campus

**2.3 Digitization concept**

Polygonal shapes are difficult to use for analysis of land-use change because both the plot shapes and their data change. The left panel of Figure 7 shows thematic maps of land-use on Coral Island for 2000 and 2009. The top left panel shows the coordinate shift needed to make the reported and corrected maps agree using the bilinear transformation described in the preceding section. The right panel shows a 100×100 meter square grid to convert the polygonal shape file to raster format.  The six points in the plot with plotID 397 in 2009 are (431.75, 861.15), (431.85, 861.15), (431.75, 861.05), (431.85, 861.05), (431.75, 861.95) and (431.85, 861.95) respectively.

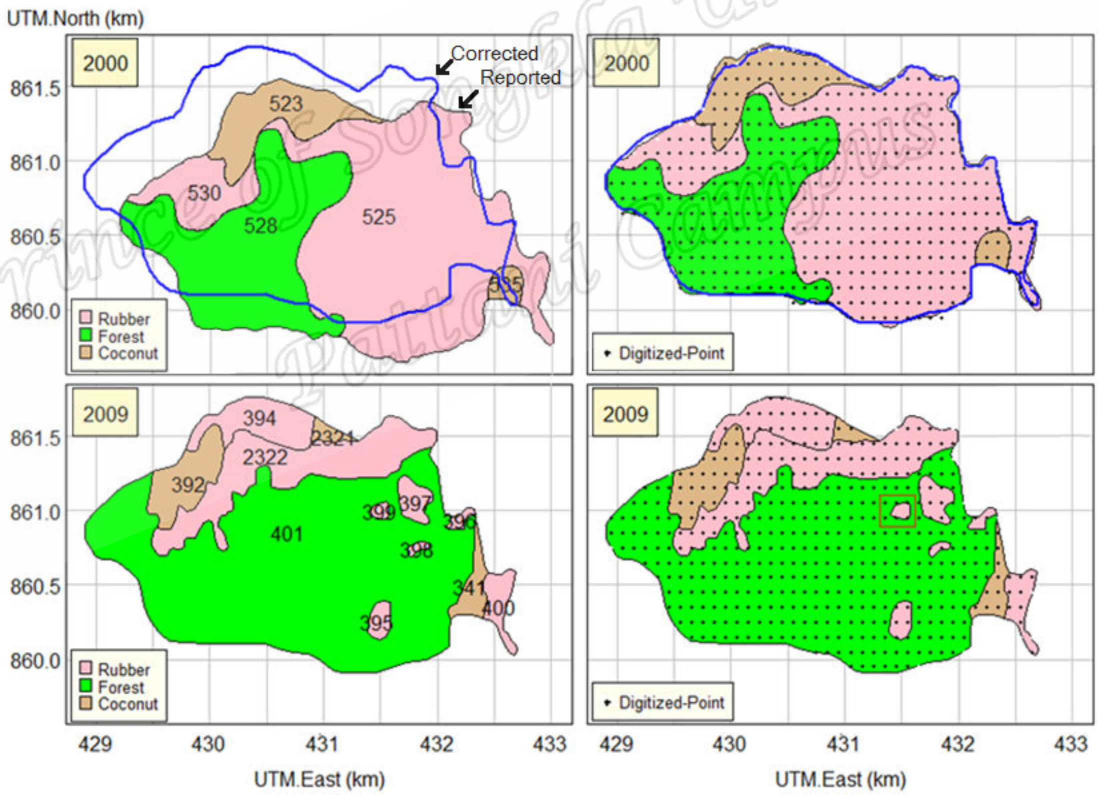



Figure 7. Thematic map of land-use in Coral Island, Phuket, in 2000 (upper panel) and 2009 (lower panel) with 100×100 meter grid used for digitization.

Prince of Songkla University Pattani Campus

The right panel of Figure 8 details a subset of the grid points within a 5×5 matrix, which illustrates the comparison between the polygonal and digitized data structures. These data structures for the polygon with plotID 397 in 2009 are shown in Table 2 below.

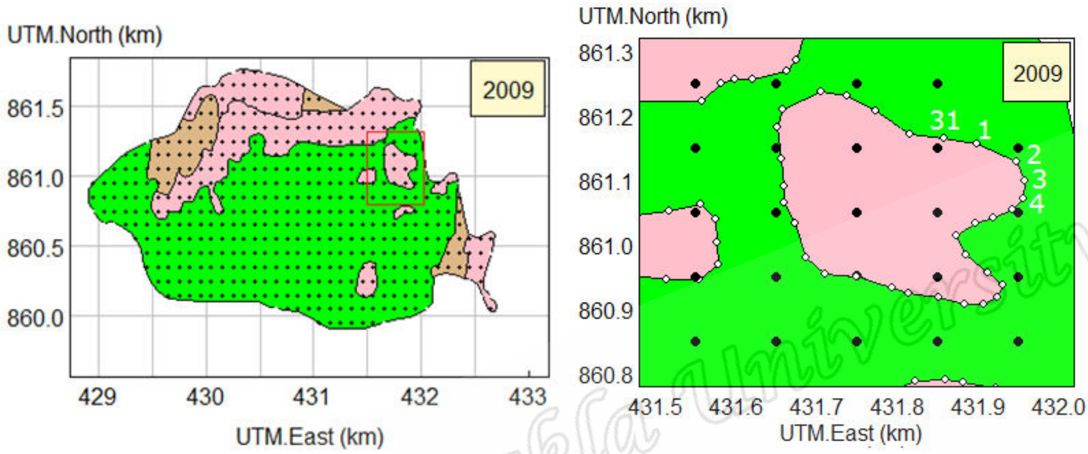



Figure 8. Grid points in raster format for Coral Island in 2009 with details for plotID 397

Table 2: Data structures for polygonal (left table) and digitalized (right table) data structures for plotID 397 in 2009

Polygon (vector)

| plotID | pointID | x | y |
|---|---|---|---|
| 397 | 1 | 431.90 | 861.16 |
| 397 | 2 | 431.95 | 861.13 |
| 397 | 3 | 431.96 | 861.10 |
| 397 | 4 | 431.96 | 861.07 |
| 397 | 5 | 431.94 | 861.05 |
| 397 | 6 | 431.92 | 861.04 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| 397 | 30 | 431.82 | 861.17 |
| 397 | 31 | 431.86 | 861.17 |
| 397 | 32 | 431.90 | 861.16 |

Point (raster)

| x | y | plotID |
|---|---|---|
| 431.75 | 861.15 | 397 |
| 431.85 | 861.15 | 397 |
| 431.75 | 861.05 | 397 |
| 431.85 | 861.05 | 397 |
| 431.75 | 860.95 | 397 |
| 431.85 | 860.95 | 397 |

Prince of Songkla University Pattani Campus

### 2.4 Digitization method

This section explains how to convert a polygonal shape file to a digitized data structure using a computer program. The computation involves assigning grid points to polygons. This pseudo code is

*for* each polygon p*i* in the specified region

label all grid points inside p*i* as *i*

*end*

This algorithm can be implemented in any language that accommodates *for…end* loops, provided the language has a function that determines which elements of a specified set of points contained within a specified polygon. The R program language has a library *sp* that contains a function *point.in.polygon*() documented below.

Prince of Songkla University Pattani Campus

point.in.polygon {sp} R Documentation

**do point(s) fall in a given polygon?**

**Description**

verifies for one or more points whether they fall in a given polygon

**Usage**

```
point.in.polygon(point.x, point.y, pol.x, pol.y, mode.checked=FALSE)
```

**Arguments**

| | |
|---|---|
| point.x | numerical array of x-coordinates of points |
| point.y | numerical array of y-coordinates of points |
| pol.x | numerical array of x-coordinates of polygon |
| pol.y | numerical array of y-coordinates of polygon |
| mode.checked | default FALSE, used internally to save time when all the other argument are known to be of storage mode double |

**Value**

integer array; values are: 0: point is strictly exterior to pol; 1: point is strictly interior to pol; 2: point lies on the relative interior of an edge of pol; 3: point is a vertex of pol.

**References**

Uses the C function InPoly(). InPoly is Copyright (c) 1998 by Joseph O'Rourke. It may be freely redistributed in its entirety provided that this copyright notice is not removed.

**Examples**

```
# open polygon:
point.in.polygon(1:10,1:10,c(3,5,5,3),c(3,3,5,5))
# closed polygon:
point.in.polygon(1:10,rep(4,10),c(3,5,5,3,3),c(3,3,5,5,3))
```

[Package *sp* version 1.0-5 Index]

Figure 9. Documentation for function *point.in.polygon*() in *sp* library for R

The program commands in Figure 10 show how to remove holes in a complex region and assign the values of the plotIDs to grid points.

```
# digitize land-use for Phuket from analog form for 2009

area <- function(X) {                    # function to compute area of a polygon
 X <- rbind(X,X[1,])
 x <- X[,1]
 y <- X[,2]
 lx <- length(x)
 sum((x[2:lx]-x[1:lx-1])*(y[2:lx]+y[1:lx-1]))/2
}

#holesFound <- 1
holesFound <- 0

if (digitized==0 & holesFound==0) {

 read.table("luPhuket52.txt",h=T,as.is=T) -> p52data
 str(p52data)
 read.table("luPhuket52.xy",h=T,as.is=T) -> p52xy

 p52xy <- subset(p52xy,is.finite(p52xy$x))
 p52xy <- subset(p52xy,p52xy$pointID>0)
 p52xy <- subset(p52xy,p52xy$plotID>0)
 p52xy <- subset(p52xy,!(is.na(p52xy$x) & p52xy$y==1))
 p52xy$x <- ifelse(is.na(p52xy$y),NA,p52xy$x)

 p52xy$x <- p52xy$x/1000                 # convert meters to km
 p52xy$y <- p52xy$y/1000

 p52data$hole <- 0*p52data$plotID        # add a field to identify
                                         # plots with holes for 2009

# find plots with holes and delete the holes

 if (holesFound==0)
 rxya <- NA+p52xy[1,]

 for (j in p52data$plotID) {
  rxyj <- subset(p52xy,p52xy$plotID==j)
  rxyjNA <- subset(rxyj,is.na(rxyj$x))
  if (dim(rxyjNA)[1]>0) {
   ptID1 <- min(rxyjNA$pointID)
   rxyj <- subset(rxyj,rxyj$pointID<ptID1)
   p52data$hole[p52data$plotID==j] <- 1
  }
  rxya <- rbind(rxya,rxyj)
 }

 rxy0 <- rxya[-1,]
 write.table(rxy0,"phuket52OK.xy",row.names=F,quote=F)
 write.table(p52data,"phuket52OK.txt",row.names=F,quote=F)

}
 if (digitized==0 & holesFound==1) {

 # region 1 (north of Phuket )

  ymax <- 908; ymin <- 881; xmin <- 418; xmax <- 443

  ncols <- (xmax-xmin)*10                # specify grid
  nrows <- (ymax-ymin)*10

  pt.x <- xmin-0.05+0.1*c(1:ncols)
  pt.x <- rep(pt.x,nrows)

  pt.y <- ymax-0.05+0*c(1:ncols)
  for (i in c(1:(nrows-1))) {
   pt.y <- c(pt.y,ymax-0.05-i*0.1+0*c(1:ncols))
  }

  plotIDs <- 0*pt.x                      # populate grid with plotIDs

  read.table("phuket52OK.xy",h=T,as.is=T) -> Rxy
  read.table("phuket52OK.txt",h=T,as.is=T) -> Rdata

  rxy <- subset(Rxy,(Rxy$x>xmin & Rxy$x<xmax & Rxy$y>ymin & Rxy$y<ymax))
  set <- unique(rxy$plotID)
  rxy <- subset(Rxy,Rxy$plotID %in% set)
  rd <- subset(Rdata,Rdata$plotID %in% set)

  rd <- rd[order(rd$area,decreasing=F),]
  plotIDs1 <- rd$plotID

  library(sp)

  for (j in plotIDs1) {
   pol <- subset(rxy,rxy$plotID==j)
   pol.x <- pol$x
   pol.y <- pol$y
   point.in.polygon(pt.x,pt.y,pol.x,pol.y) -> grid
   plotIDs <- ifelse((grid==1 & plotIDs==0),j,plotIDs)
  }

  plots <- as.data.frame(plotIDs)
  names(plots) <- "plotID"
  plots$x <- pt.x
  plots$y <- pt.y

  merge(plots,rd,by.x="plotID",by.y="plotID",all.x=T)[,c(2,3,4)] -> plot52
  merge(plot52,plots,by.x=c("x","y"),by.y=c("x","y")) -> lu01

  write.table(lu01,"phuket5201dig.txt",row.names=F,quote=F)

 }
 #-----------------------------end of digitization
```

Figure 10 Program commands for computing digitized land-use

Prince of Songkla University Pattani Campus

The program contains two sections. The first section removes holes from polygons. This is needed because polygons corresponding to holes are duplicated and are thus redundant, as described in a preceding section. Once their holes have been removed, it is necessary to order the polygons by increasing area, because once the grid-points within a polygon containing a hole have been assigned with a plotID value, they cannot be overwritten. Note that the MID file contains the area of each polygon as one of its fields, but if this information is not available, the *area*() function can be used to compute the area of a polygon.  If the number of grid-points in the whole region is large the program will take much longer, so it is better to spit the whole region into smaller components containing whole polygons and digitize them separately.

Note that the *point.in.polygon*() function needs to determine whether or not each specified (x, y) point is inside or outside a polygon specified by its set of boundary points. It does this by computing the number of times a vector originating from the point to a boundary of the region crosses the boundary of the polygon. If this number is odd, the point is inside the polygon; if the number is even, the point is outside the polygon.

Prince of Songkla University Pattani Campus

## 2.5 Varying the grid density

We used a 100×100 meter grid for digitization, so that each grid point corresponds to one hectare of land. The digitization method can fail to accurately identify small polygons that fall between grid points. However, this problem can be solved by varying the density of the grid where needed. Figure 11 shows an example.

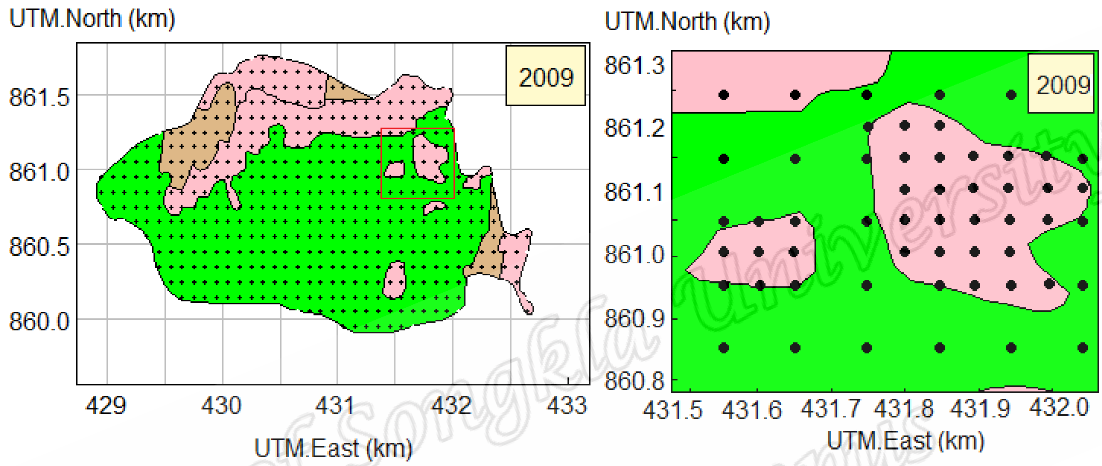



Figure 11. Varying grid for Coral Island containing extra grid-points (right panel)

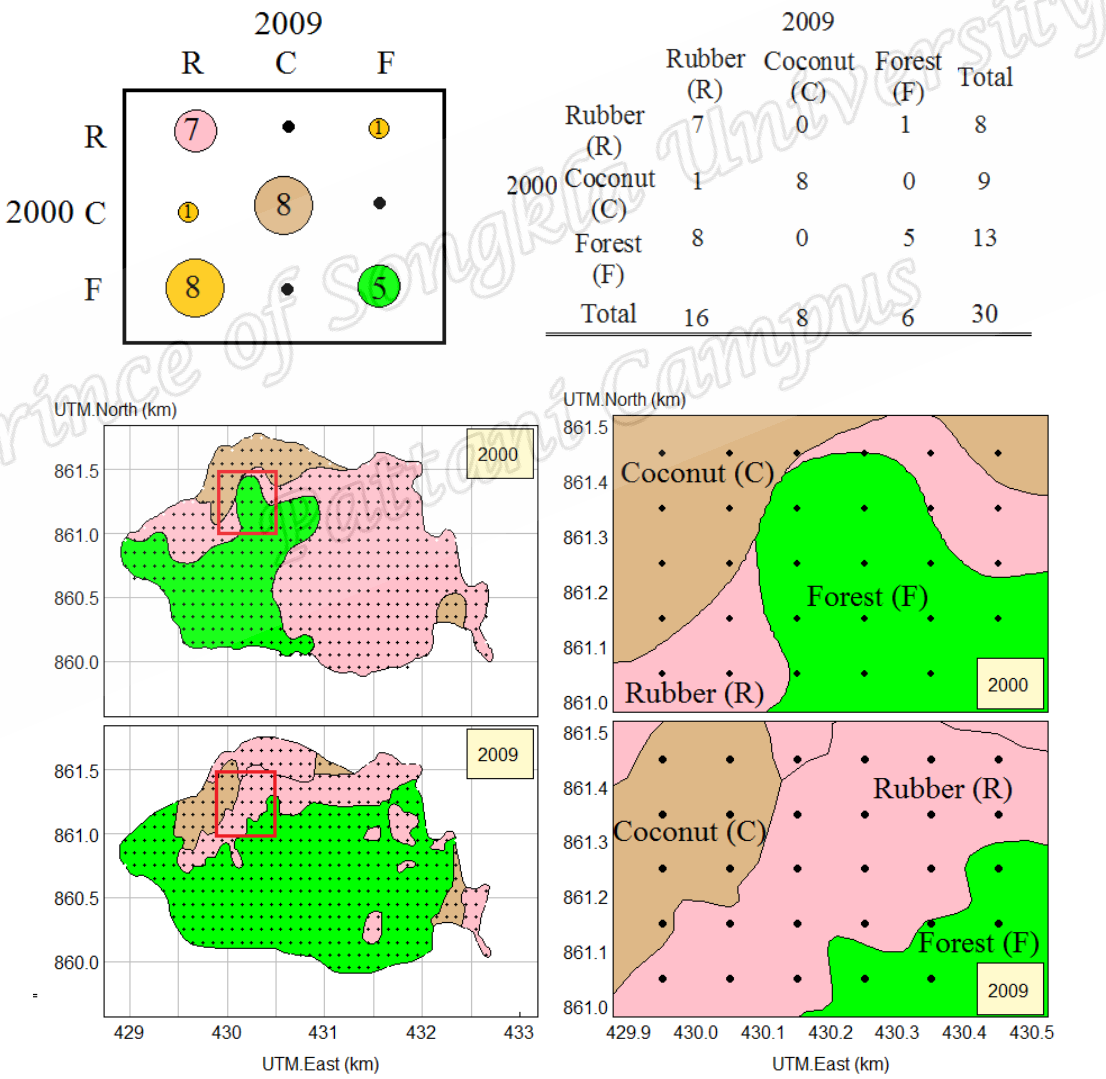
Table 3: Data structures for 6×5 point rectangle with additional grid-points to improve accuracy in small polygons (asterisked)

| pointID | x | y | pointID | x | y | pointID | x | y |
|---|---|---|---|---|---|---|---|---|
| 2322 | 431.55 | 861.25 | *397 | 431.85 | 860.10 | *397 | 431.85 | 860.00 |
| 2322 | 431.65 | 861.25 | *397 | 431.90 | 860.10 | *397 | 431.90 | 860.00 |
| 401 | 431.75 | 861.25 | *397 | 431.95 | 860.10 | *397 | 431.95 | 860.00 |
| 401 | 431.85 | 861.25 | *397 | 432.00 | 860.10 | 399 | 431.55 | 859.95 |
| 401 | 431.95 | 861.25 | *397 | 432.05 | 860.10 | *399 | 431.60 | 859.95 |
| 401 | 432.05 | 861.25 | 399 | 431.55 | 860.05 | 399 | 431.65 | 859.95 |
| *397 | 431.75 | 861.20 | *399 | 431.60 | 860.05 | 401 | 431.75 | 859.95 |
| *397 | 431.80 | 861.20 | 399 | 431.65 | 860.05 | 397 | 431.85 | 859.95 |
| *397 | 431.85 | 861.20 | 401 | 431.75 | 860.05 | *397 | 431.90 | 859.95 |
| 401 | 431.55 | 860.15 | *397 | 431.80 | 860.05 | 397 | 431.95 | 859.95 |
| 401 | 431.65 | 860.15 | 397 | 431.85 | 860.05 | *397 | 432.00 | 859.95 |
| 401 | 431.75 | 860.15 | *397 | 431.90 | 860.05 | 401 | 432.05 | 859.95 |
| *397 | 431.80 | 860.15 | 397 | 431.95 | 860.05 | 401 | 431.55 | 859.85 |
| 397 | 431.85 | 860.15 | *397 | 432.00 | 860.05 | 401 | 431.65 | 859.85 |
| *397 | 431.90 | 860.15 | 401 | 432.05 | 860.05 | 401 | 431.75 | 859.85 |
| 397 | 431.95 | 860.15 | *399 | 431.55 | 860.00 | 401 | 431.85 | 859.85 |
| *397 | 432.00 | 860.15 | *399 | 431.60 | 860.00 | 401 | 431.95 | 859.85 |
| 397 | 432.05 | 860.15 | *399 | 431.65 | 860.00 | 401 | 432.05 | 859.85 |
| *397 | 431.80 | 860.10 | *397 | 431.80 | 860.00 | | | |

Prince of Songkla University Pattani Campus

**2.6 Graphical display**

We have seen how thematic maps are appropriate for displaying land-use data defined as polygons. While thematic maps can also be created from digitized data, if the grid is too coarse these maps will be less informative than corresponding maps based on polygons. Other relevant graphical displays include bubble plots, bar charts and plots of confidence intervals. Land-use change for two given years can be graphed as thematic map combined with a bubble plot as shown in Figure 12.




| | | 2009 | | | |
|---|---|---|---|---|---|
| | | Rubber (R) | Coconut (C) | Forest (F) | Total |
| 2000 | Rubber (R) | 7 | 0 | 1 | 8 |
| | Coconut (C) | 1 | 8 | 0 | 9 |
| | Forest (F) | 8 | 0 | 5 | 13 |
| | Total | 16 | 8 | 6 | 30 |

Figure 12. Geographical display showing land-use change in Coral Island from 2000 to 2009

Prince of Songkla University Pattani Campus

Graphing land-use change is complicated because many additional colours are needed to illustrate the data from a cross-tabulation that has more than a small number of land-use categories. However, analysis of such change can be simplified by focusing on a binary outcome, such as a specific land-use of interest, such as "developed" land.

### 2.7 Logistic regression

Logistic regression is an appropriate method for analyzing remote sensing (RS) data where the outcome is binary, such as urban land. This model parameterizes a linear relationship between a set of independent variables and the binary variable using a logit link function, which constrains values of the probability associated with the specific outcome to be between 0 and 1 (Lee and Pradhan, 2007).

This model can be used to handle as independent variables both continuous and categorical predictor variables. Also, these predictors need not have normal distributions as required in discriminant 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). For $p$ numeric predictor variables $X_1, X_2, ..., X_p$ the model is formulated as follows (Kleinbaum,1994).

$$\text{logit}\left( prob[Y_i = 1 \,|\, X_1 = x_{i1}, X_2 = x_{i2}, \dots, X_p = x_{ip}\right) = \alpha_i + \sum_{j=1}^{p} \beta_j x_{ij} \qquad (4)$$

In this equation $x_{i1}, x_{i2}, \dots, x_{ip}$ are the observed values of $X_1, X_2, ..., X_p$ for case $i$. For a single categorical (factor) predictor variable $\boldsymbol{W}$, say, this model takes the form

$$\text{logit}\left( prob[Y_i = 1 | \boldsymbol{W}_i = \boldsymbol{w}_i]\right) = \alpha_i + \sum_{j=2}^{p} \beta_j w_{ij} \ , \qquad (5)$$

where the $w_{ij},\ j = 1,2,..,p$ are indicator variables taking binary values 0 or 1. Note that $w_{i1}$ is omitted from the model because when the intercept term $\alpha_i$ is included one of the factor levels must be omitted to avoid over-parameterization. For example, if ***W*** is land-use with three categories corresponding to forest, agriculture or urban land-use, for case $i$, $w_{i1}$ is 1 for forest and 0 for other use, $w_{i2}$ is 1 for agriculture and 0 other use, and $w_{i3}$ is 1 for urban and 0 for other use.

For analyzing land-use change in the simple situation where no covariates are considered, the binary dependent variable $Y$ denotes the specified binary land-use for each grid-point in a specified year, and $X$ is the land-use type (not necessarily binary) at the same grid-point for a preceding year. We now show how the logistic regression model is applied to the data for Coral Island, where the binary outcome is Rubber Plantation or other land-use in year 2009 and the predictor is the land-use (Rubber Plantation, Coconut, or other Forest) in year 2000. First, we choose the small sample containing just 30 grid-points as indicated in Figure 12. Since both the outcome variable $Y$ and the predictor variable W is categorical, the data to be fitted by the logistic regression model may be listed simply as a contingency table of counts, as shown in the left side of Figure 13. Using a logistic regression program with lu2000 as the predictor and rubber2009/other2009 as the binary outcome, we get the result shown in the right side of Figure 13.

| lu2000 | Rubber2009 | Other2009 |
|---|---|---|
| Rubber.A302 | 7 | 1 |
| Coconut.A405 | 1 | 8 |
| Forest.F101 | 8 | 5 |

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.946      1.069   1.820  0.06872 .
lu43.A405     -4.025      1.506  -2.673  0.00752 **
lu43.F101     -1.476      1.212  -1.218  0.22315
---

    Null deviance:  11.825 on 2  degrees of freedom
Residual deviance:       0 on 0  degrees of freedom
```

Figure 13. Data for small sample of 30 grid-points (30 hectares) selected from Figure 12 and result from fitting logistic regression model

Prince of Songkla University
Pattani Campus

Note that the estimate for the predictor level Rubber.A302 is 0 because this parameter is absorbed by the intercept as indicated in Equation (5), and the coefficients and corresponding standard errors and p-values denote the differences from this omitted category. Thus the result shows that coconut was very likely to change to rubber (p-value for no change 0.0075), but there was no evidence that forest changed to rubber. These results are consistent with the maps in the right panels of Figure 12.

The left panel of Figure 14 depicts a bar chart of the data in the left side of Figure 13. It shows that 77.5% of land devoted to rubber plantation in 2000 remained rubber plantation in 2009, whereas 10% and 61.5% of the land that was coconut and forest, respectively, in 2000 became rubber plantation in 2009.

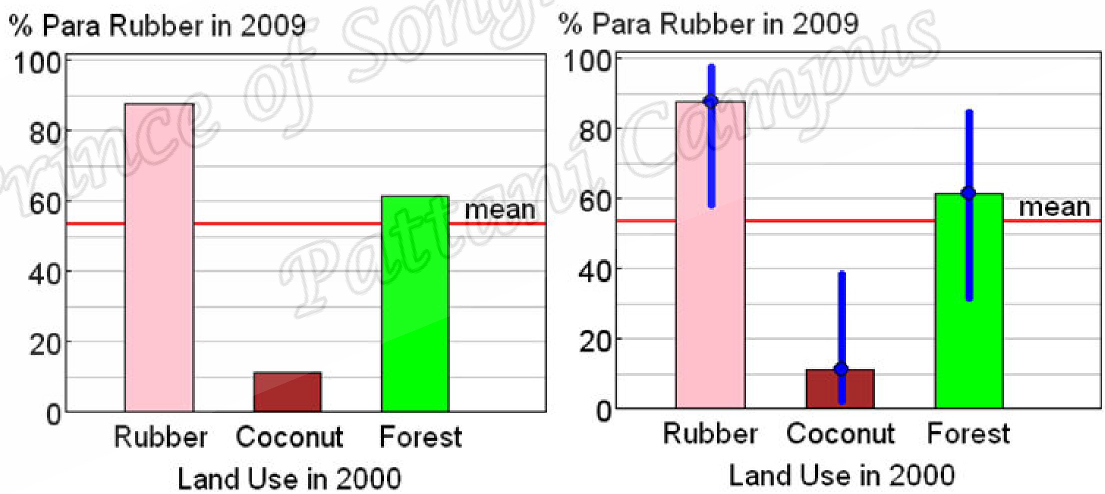


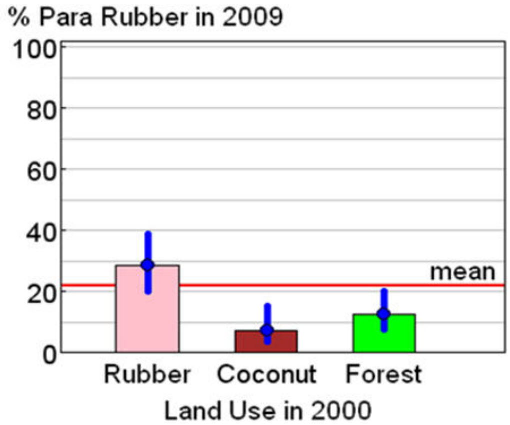

Figure 14. Bar chart of percentages of land that changed to rubber plantation in a sample of 30 hectares in Coral island from 2000 to 2009, with 95% confidence intervals for comparing them (right panel)

Prince of Songkla University Pattani Campus

It was not necessary to use the logistic regression method just to get these percentages because they are obtained simply from the table on the left of Figure 13. However, the logistic regression model is needed because it also provides standard errors for these percentages, based on the assumption that the 30 hectare area is a subset from a larger population from which the 30 hectares is a randomly selected sample. These standard errors can be used to construct 95% confidence intervals for comparing the population percentages, and these are superimposed on the bar chart on the right side of Figure 14. They show that the population percentages for the changes from coconut and forest to rubber differed. This conclusion is consistent with the p-value 0.0075 shown in the logistic regression result on the right side of Figure 13, which shows evidence that the percentage of coconut land that changed to rubber was less than the percentage of rubber that remained rubber.

When analyzing land-use change from an earlier survey year to a later year it is important to ensure that the samples are taken from the same area. However, with shifting coordinates and changes in coastlines it can happen that these areas are not exactly the same. For example, the land area of Coral Island in 2000 was reportedly 440 hectares and 434 hectares in 2009, and its total land area that did not change was 424 hectares. So analyzing land-use change, we consider only this common area. Figure 15 shows the data and results from fitting the same model to the common area of Coral Island as before for the smaller sample of just 30 hectares.

Prince of Songkla University
Pattani Campus

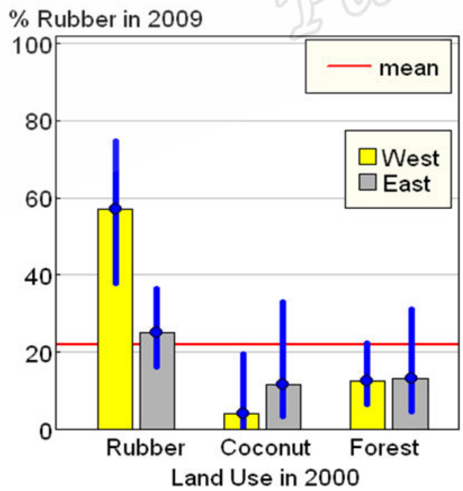



| lu2000 | Rubber2009 | Other2009 |
|---|---|---|
| Rubber.A302 | **75** | **189** |
| Coconut.A405 | **3** | **38** |
| Forest.F101 | **15** | **104** |

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.9243     0.1365  -6.773  0.00000 **
lu43.A405     -1.6147     0.6150  -2.625  0.00866 **
lu43.F101     -1.0121     0.3081  -3.285  0.00102 **
---
      Null deviance: 1.9389 on 2  degrees of freedom
Residual deviance:       0 on 0  degrees of freedom
```

Figure 15. Data for Coral Island (424 hectares) and fitted logistic regression model

Since location is an important predictor of land-use change, next we give a simple illustration of how the logistic regression model can be extended to include this addition predictor, by dividing Coral Island into two sections, one west of UTM 430.5 km East and the other east of this longitude. Taking the combination of land-use in 2000 and this east/west location as a single factor, Figure 16 shows the result of fitting the logistic model, which computed by program commands in appendix.




| lu2000.loc | Rubber2009 | Other2009 |
|---|---|---|
| Rubber.A302.West | **16** | **12** |
| Rubber.A302.East | **59** | **177** |
| Coconut.A405.West | **1** | **23** |
| Coconut.A405.East | **2** | **15** |
| Forest.F101.West | **12** | **84** |
| Forest.F101.East | **3** | **20** |

```
Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.0986     0.1503  -7.308  0.00000 ***
lu43.A302.West   1.3863     0.4104   3.378  0.00073 ***
lu43.A405.East  -0.9163     0.7676  -1.194  0.23261
lu43.A405.West  -2.0369     1.0325  -1.973  0.04852 *
lu43.F101.East  -0.7985     0.6371  -1.253  0.21010
lu43.F101.West  -0.8473     0.3433  -2.468  0.01358 *
---
    Null deviance: 31.662  on 5  degrees of freedom
Residual deviance:      0  on 0  degrees of freedom
```

Figure 16. Data for Coral Island (424 hectares), which separated to west and east and fitted logistic regression model

Prince of Songkla University Pattani Campus

**2.8 Adjusting for spatial correlation**

Conventional statistical analysis such as logistic regression assumes that data samples ($Y$) are independent. However, this assumption does not hold for data defined at grid-points (raster) just 100 meters apart. The data from neighboring plots are likely to be mutually correlated, violating the independence assumption, giving incorrect standard errors. Previous studies with geographical data such as those by McNeil and Chooprateep (2013) handled spatial correlation by aggregating data in larger regions with acceptably small correlation between overlapping regions, or used factor analysis to adjust for these correlations.The generalized estimating equations (GEE) method could be used by dividing the region into groups of plots (clusters) and estimating common fixed correlations between plots in the same group (Zeger and Liang 1986).

However, this method assumes a single common correlation, whereas it is likely that the spatial correlation in land-use data will vary with the type of land-use. For binary outcomes a simpler method is available, based on variation inflation factors, proposed by Rao and Scott (1992). This method avoids the problem in the GEE method by computing effective sample sizes for sets of land-use plots based on their sample variances, and then expanding the confidence intervals given by the logistic regression model by multiplying them by the square roots of these factors.

Prince of Songkla University
Pattani Campus

Prince of Songkla University
Pattani Campus