4. MODEL DEVELOPMENT

4.1 Structure of model development

There are three segments of structure in this study. The first and second part, theoretical exercises and framework development, show how this research was organized. The last part, conceptual map of modeling, shows the flow of modeling contribution of this study.

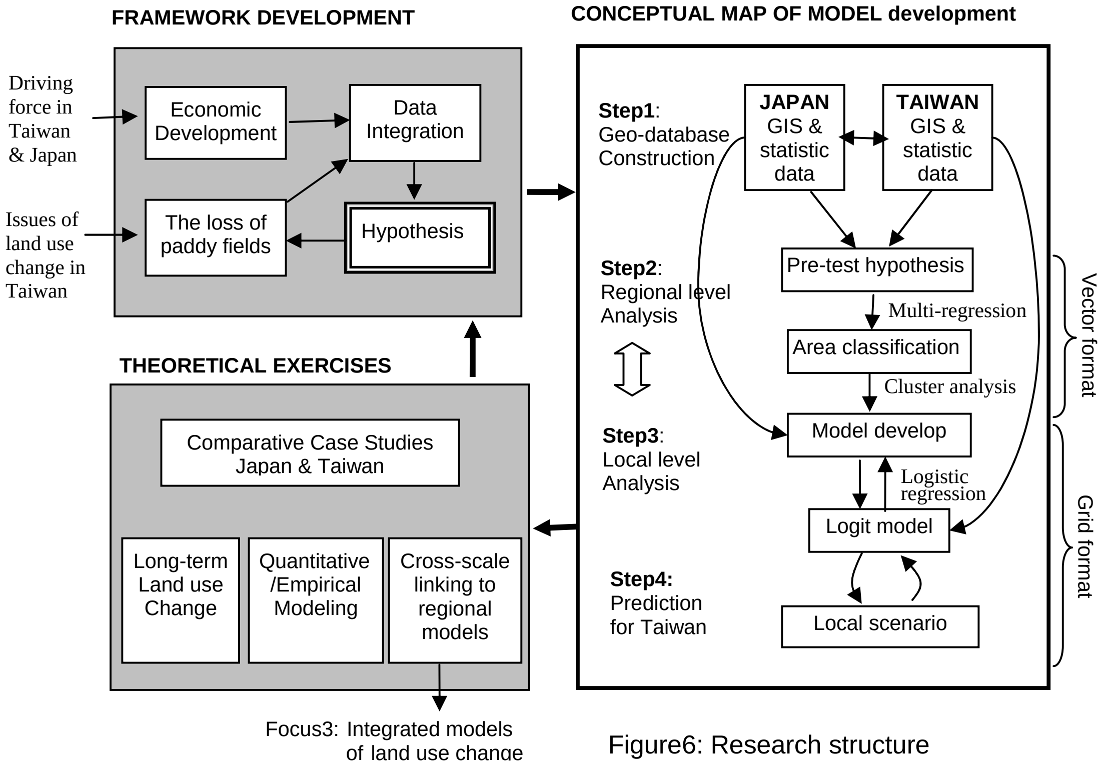

Figure6: Research structure

4.2 Geo-database construction

GIS is a powerful tool for storing, integrating, and visualizing spatial data for land use/cover change studies. Two GIS geo-databases, one for Japan and the other for Taiwan, were contracted for this study. Each database consist of digital land use data, national statistics on socioeconomic variables, county statistics on socioeconomic variables, DEM data, slope data and digital maps showing zoning, land suitability, roads, industrial parks, and other facilities.

Table3: Digital land use data in Japan and Taiwan

Japan			Taiwan		
Year	Scale	Data source	year	Scale	Data source
1976	100m mesh	Field survey	1982	1:5000	Aerial photo
1987	100m mesh	Field survey	1988	1:5000	Aerial photo
1992	100m mesh	Field survey	1994	1:5000	Field survey
1999	100m mesh	Field survey			

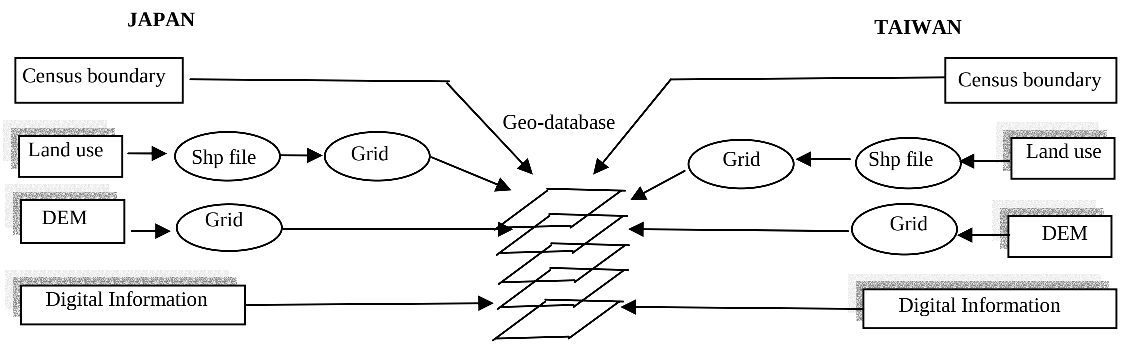

Figure8: geo database construction

4.3 Regional-level analysis

Regional-revel analysis consists of two parts. First, multiple regression analysis will be used to verify that losses of paddy filed follow the same pattern in Japan and Taiwan except for a time log in the past few decades. The second part is to use cluster analysis to identify sets of counties in Japan and Taiwan for local-level analysis.

- **Multiple regression analysis**

Run multiple regression analysis to verify that losses of paddy field can be explained by the same explanatory variables for 1980-90 county data from Japan and 1990-2000 county data from Taiwan.

A multiple regression model is defined by $y = a + b_1x_1 + b_2x_2 + . . . +b_nx_n$,

Dependent variable(Y): the rate of paddy fields losses during selected period
Independent variables (Xi):

% population growth,
% urban population,
per capita GDP,
change in % workers in each industrial sector,
% agriculture in GDP.
% change in land price

- **Cluster analysis**

Cluster Analysis is a multivariate analysis technique that seeks to organize information about variables so that relatively homogeneous groups, or "clusters," can be formed. This study will run cluster analysis to group counties in Japan and in Taiwan by paddy field losses and socioeconomic data, so that two counties, one from each study area, can be selected for further analysis. Variables used in the Cluster analysis were considered with three segmentations of factors, socio-economic driving factors. Land use policy and planning factors, and physical factors.

Table 4: variables of cluster analysis

Factors	Variables	Information type
Socio-economic driving factors	Population density	Person/km2
	Percentage of population under 64 years old	Percentage(Total population)
	Farm-household ratio	Percentage
	Percentage of full-time farm households	Percentage
	Percentage of part-time farm households	Percentage
	Percentage of workers in secondary industry	Percentage(to total workers)
	Percentage of workers in tertiary industry	Percentage(to total workers)
	Percentage of employees in secondary industry	Percentage(to total employees)
	Percentage of employees in tertiary industry	Percentage(to total employees)
	The Growth of Gross field husbandry product	Ratio(average)
	Numbers of employees per 100 persons	Persons
	GDP in agriculture/ GDP	Ratio
Land use policy and planning factors	Number of cars/population	Cars/ person
	Ratio of agriculture zone	Ratio
	Ratio of urbanization zone	Ratio
	The growth of land price	Ratio(average)
Physical Factors	Share of 0-3 degree slope area	Share in total area
	Share of 3-8 degree slope area	Share in total area
	Share of > 15 degree slope area	Share in total area
	Share of 0-100 elevation area	Share in total area
	Share of >200 elevation area	Share in total area

4.4 Local Level analysis

The purpose of local analysis is twofold: (1) to develop a binary logit model, relating cells of loss (1) or no loss (0) to biophysical and other factors, and (2) to use the logit model developed in Japan to predict losses of paddy field in Taiwan in 2010-2015.

- **Binary logistic regression**

A binary logit model will be first developed for the selected county in Japan.

$$\text{Logit}(p) = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots$$
$$p = 1 / [1 + e^{\text{logit}(p)}]$$

p: probability of occurrence of the dependent variable
e: natural exponent
x_i: independent variable i,
a,b_i: regression coefficients.

1. Select the paddy fields form the 1976 and 1999 digital map, convert to raster format with a cell resolution of 500m, and drive a grid showing the loss of paddy filed.
2. Prepare a grid for each independent variable: the distance to built-up area cell, the distance to road, slope, elevation, and land suitability etc. (table 5)
3. Combine all grids to produce an input data for binary logistic regression.

Table5: the independent variables of Binary logistic regression

category	layer	Information type
Topography	elevation	Quantitative(m)- Base map
	Slope	Quantitative(%)- Derived map
	Aspect	Qualitative (8 classes)- Derived map
Land use/cover	Built-up area	500m mesh
	Forest	500m mesh
	Paddy field	500m mesh
	Fields	500m mesh
	Other	500m mesh
Policy & planning feature	Urban area	Qualitative(2 class)- Derived map
	Industrial park	Qualitative(2 class)- Base map
	Protected agricultural zone	Qualitative(2 class)- Base map
	Road network	Qualitative(2 class)- Base map
Spatial relationships	Distance to road	Quantitative(m)- Derived map
	Distance to urban area	Quantitative(m)- Derived map
	Distance to fields	Quantitative(m)- Derived map
	Distance to forest	Quantitative(m)- Derived map
	Distance to industrial park	Quantitative(m)- Derived map
	Distance to protected zone	Quantitative(m)- Derived map
	Distance to slope	Quantitative(m)- Derived map
	Distance to elevation	Quantitative(m)- Derived map

- **Logit model**

After running the binary logistic regression using Japan's data from 1976 to 1999, the result will be a logit model for predicting and locating losses of paddy field. This study will use the logit model with a paddy field map of 1994 to predict and locate which paddy field cells will be lost in 2015 for the selected Taiwan County.

4.5 Prediction with local scenarios

In this spatially based empirical model, local-level analysis can be altered to simulate different scenarios. As addressed in the LITERETURE REIVEW, land use change is often triggered by shock event or an important policy decision in a short time scale. The high speed rail way system in Taiwan is the most important infrastructure in the beginning of 21 century. The system goes through the West Taiwan Corridor, which the land use in plain area will be facing violent changes in the next couple of years. (Figure9)

To better and more accurately predict the loss of paddy filed in the next decade of Taiwan, this study will add a new transportation factor, such as the stations of Taiwan high speed rail way system in local-level analysis.

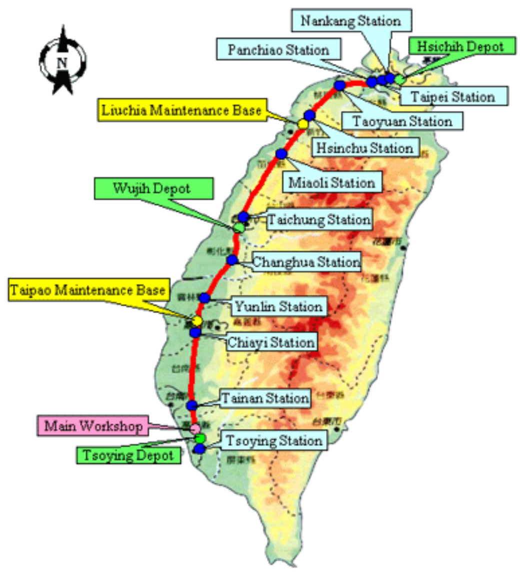

Figure9: Taiwan's high speed rail system