Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)
Available online at www.inia.es/sjar
Spanish Journal of Agricultural Research 2009 7(3), 687-698
ISSN: 1695-971-X

# Simulation of farmer decision on land use conversions using decision tree method in Jiangsu Province, China

T. Y. Zhong[1], X. Y. Zhang[2], and X. J. Huang[1]*
[1] *School of Geographic and Oceanographic Sciences, Nanjing University, Nanjing 210093, China.*
[2] *International Institute for Earth System Science, Nanjing University, Nanjing 210093, China.*

## Abstract

Understanding farmer decisions on land use conversions (LUC) in rural areas has significant importance to understand and predict the patterns of land use changes in China. Many methods have been developed to search for the influencing factors on land use changes at farm household level. However, these methods have difficulty in evaluating the intertwined influences between factors and achieving farmer decision rules on LUC. Taking three regions located in Jiangsu Province as the study areas, the present work proposed a data mining method, classification and regression tree (CART), to simulate farmer decisions on LUC at farm household level. The accuracy of the simulated LUC by CART was above 85.00%, indicating that the proposed method could be used to simulate farmer decisions. The simulation results also showed that farmer decision rules on LUC presented regional characteristics. In Jiangsu Province, 20 rules were inferred for LUC using 10 factors which were related to the household resources, land market, and so on. These factors were ranked in the decreased importance on LUC as labor transfer, land market, location, resources of household, and characteristics of household. In Rudong County, the which were related factors with LUC included land market, labor transfer, and household resource, while in Changshu County, the additional factor of location was involved.

**Additional key words**: classification and regression tree, determinants of rural land conversions, farm household level, regional differences.

## Resumen

### Simulación con métodos arborísticos de la decisión de los agricultores sobre la reconversión del uso de parcelas en la provincia de Jiangsu, China

Comprender las decisiones de los agricultores sobre la reconversión del uso de parcelas (LUC) en áreas rurales tiene una importancia significativa para comprender y predecir los patrones en los cambios en el uso de parcelas en China. Se han desarrollado muchos métodos para buscar los factores que influyen en los cambios del uso de parcelas a nivel agrícola familiar. Sin embargo, estos métodos presentan dificultades para evaluar las influencias entrecruzadas entre los factores y las reglas conseguidas para la decisión de los agricultores sobre la LUC. Tomando como áreas de estudio tres regiones de la provincia de Jiangsu, este trabajo propone un método de minería de datos, clasificación y árboles de regresión (CART) para simular las decisiones de los agricultores sobre el LUC a nivel familiar. Se consiguió una precisión >85% del LUC simulado con el CART, indicando que este método es capaz de simular las decisiones de los agricultores. Los resultados de la simulación también mostraron que las reglas de decisión de los agricultores sobre el LUC presentaban características regionales. Se dedujeron 20 reglas para el LUC en la provincia de Jiangsu utilizando 10 factores relacionados con los recursos familiares, el mercado del suelo, etc. Estos factores fueron clasificados en importancia decreciente en el LUC como transferencia de mano de obra, mercado del suelo, recursos familiares y características de las familias. En el condado de Rudong, los factores relacionados con el LUC fueron el mercado del suelo, la transferencia de mano de obra y los recursos familiares, mientras que en el condado de Changshu aparece la localización como factor adicional.

**Palabras clave adicionales:** árbol de clasificación y regresión, determinantes de reconversión de parcelas rurales, diferencias regionales, nivel agrícola familiar.

* Corresponding author: hxj369@nju.edu.cn; xjhuang.2008@163.com
Received: 19-10-08. Accepted: 23-06-09.

## Introduction

Since the economic reforms and an open door policy in 1978, China has been experienced rapid urbanization and industrialization, especially in the coastal regions, such as Yangtze River Delta region (Xu, 2004) and Pearl River Delta region (Weng, 2002; Li and Yeh, 2004). The accelerated industrialization and urbanization have greatly affected land use changes through increasing private enterprises and industry plants (Wu *et al.*, 2004). Although there have been numerous studies analyzing land use conversions (LUC) in eastern and coastal China, most of which focused on urban fringe areas (Weng, 2002; Tan *et al.*, 2005). In fact, many industrial plants have been boosted in rural areas (Xu, 2004), which leaded to nonfarm jobs in rural and small town enterprises growing rapidly. The transferred farm labor forces would influence the development of enterprises, and then influence LUC. Therefore, it is important to monitor land use changes in rural areas.

Jiangsu Province, located in the Yangtze River Delta, is one of the most developed provinces in China. In the recent three decades, heavily land use changes have been occurred, which focused many studies to investigate the land use changes and their driving forces in this area. For example, Ho and Lin (2004) believed that the cause to convert farm land to nonagricultural use were rural-urban migration, rapid economic growth, and increased investments in roads; Streets *et al.* (1995) used remote sensing data and social and economic data revealed the rapid growth of urban centers, commensurate declines in water surface area, and changing patterns of agriculture from 1976 to 1984 in southern Jiangsu Province; Xiao *et al.* (2007) studied the land use change patterns and their influence on population resource environment development system, and they concluded that the land use changed most heavily in the southern areas, and central and northern areas were followed in Jiangsu Province. These studies gave the insight to LUC and their driving forces. However most of them did at the macro level, not at farm household level. Farmers are the operators of LUC, thus it is significant important to explore driving forces of land use and land cover changes at farm household levels (Alisson *et al.*, 2005).

A wide range of factors affects farmer decisions on LUC. These factors include social, biophysical, economic and institutional components (Moran, 1981; Walker and Homma, 1996). To perform such complex processes, the remotely sensed data and household social surveys have often been used. Remotely sensed imageries can provide a continuous spatial coverage of land cover changes, but the data processing and field validating requirements necessary to obtain a higher order land cover classification are considerable (Evans *et al.*, 2001). While it is not easy to identify more concrete land use types from remotely sensed imageries because of the limited spectral and spatial resolution (Lillesand and Kiefer, 2000). Moreover, remote sensing techniques leave completely unaddressed the forces underlying farmer decision behind these trends (Pichón, 1997). Household social survey requires large teams of researchers due to travel time requirements and the complexity of the survey instruments, but it can provide indepth information at the level where land-use decisions are made. In this study, the analysis on driver forces for LUC was conducted based on household social survey data.

In the past decades, substantial advances have been made by developing a wide range of analytic tools to identify and simulate the driver forces for land use changes. The descriptive statistic methods (Odihi, 2003), linear regression techniques (Coxhead *et al.*, 2002), Probit model (Alix-Garcia *et al.,* 2005; Fisher and Shively, 2005), Logit model (Thapa Keshari *et al.*, 1996; Adesina *et al.*, 2000; Herath and Takeya, 2003; Tasser *et al.*, 2007), Tobit model (Godoy *et al.*, 1997), and so on, were often used in land use changes. These studies gain insight in the factors that influence land use decisions, and provide information about decision-making processes and human behaviors. However, in general, they have difficulty in evaluating the intertwined influences between these factors, especially intertwined effects caused by more than tow factors. They also have difficulty in explaining the distinguishing reaction of decision makers with different features defined by index value, because these analysis techniques are based on marginal analysis principle (Zhong, 2007). For example, young farmers and old farmers have weak motivation to change land use, and middle-aged farmers maybe have strong propulsion to convert land uses, however these

Abbreviations used: CART (classification and regression tree method), DTA (decision tree analysis), GDP (gross domestic product), LUC (land use conversions), LUCD (land use conversion decision).

methods mentioned above have difficulty to handle such situation.

The influencing factors could be used to simulate LUC, based on the method of decision tree analysis (DTA). The basic concept of a decision tree is to split a complex decision into several simpler decisions, which may lead to a solution that is easier to interpret. This method is especially to resolve complicated things involved of many intertwined influencing factors. DTA is increasingly being used in thematic mapping from remotely sensed data and habitat modeling in ecology (Zhou *et al*., 2003; Zhang, 2006; Zhang *et al.*, 2008). However, to our knowledge, the potential of DTA to understand the mechanism of LUC has received little attention.

Taking Jiangsu Province as the study area, the broad goals of this paper are to explore the intertwined influences of socio-economic, demographic, biophysical, and geographical variables on land use conversion decision (LUCD) by farmers, and to find LUCD rules for predicating the LUC situations using decision tree analysis.

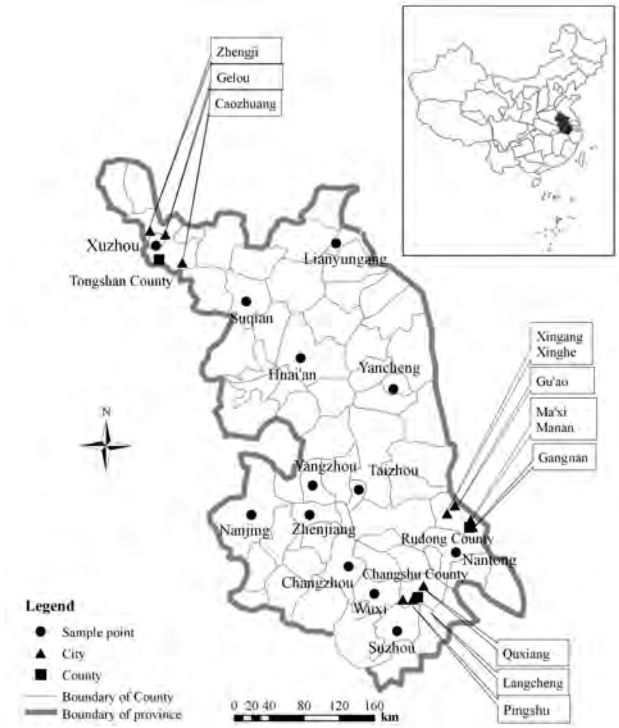


**Figure 1.** Spatial distribution of the investigated villages

## Study area and method

### Study area

The study area, Jiangsu Province ranges from 116°18′ E to 121°57′ E, and from 30°45′ to 35°20′ N, with an area of 106,700 km$^2$ (Fig. 1). Jiangsu is the lowest province in China, and most of the area is below 50 m in altitude. Low hilly region area accounts for about 14.30%, and low plain and water surface account for 85.71% of the study area. The northern subtropical monsoon climate dominates Jiangsu province year-round, with mean daily temperature of 13-16ºC, mean annual rainfall of 1060 mm, and mean annual non-frost period of 299 days, all of which are beneficial for agricultural production.

Jiangsu is one of the most developed provinces in China with great disparities among local regions within the province. According to the Statistical Yearbook of Jiangsu Province in 2007, the highest gross domestic product (GDP) per capita at county level reached to 141,064 yuan[1] (the base unit of the renminbi, people's currency in the mainland of the People's Republic of China), which gained 26.8 times of the lowest GDP per capita. The arable land accounted 45.10% of its total area. Orchard, grass, forest and the other farm lands accounted 2.90%, 0.02%, 3.10 and 12.40% respectively. The construction land and unused land accounted 16.90% and 17.60% respectively. From 1978 to 2006, farm land, particularly arable land, has been shrinking due to both urban sprawl and land requirements of villages, rural industries, and infrastructure. The arable land decreased by 16.9%, while urban settlements increased by 45.3%, rural settlements by 2.7%, industrial land by 73.8%, and communications and transport by 13.2%.

Jiangsu Province has been divided into three-fold divisions, namely developed southern Jiangsu (Sunan), moderate developed mid-Jiangsu (Suzhong) and poor northern Jiangsu (Subei) respectively. In our study, Changshu, Rudong and Tongshan County were selected to respectively represent the south, middle and north area of Jiangsu Province. According to the statistic yearbook of Jiangsu Province in 2007 (Table 1), Changshu County had the lowest total population, with the highest population density among the three counties. Changshu, with only 7.44% of the total population engaged in farming, provides most of the labor forces to produce commercial gains. While, in Rudong and Tongshan

[1] 1 yuan equals about €0.11.

**Table 1.** Social and economic conditions in Changshu, Rudong and Tongshan County

| Characteristics | Changshu | Rudong | Tongshan |
|---|---|---|---|
| Total population (x $10^4$) | 105.13 | 107.41 | 120.16 |
| Population density (person km$^{-2}$) | 964 | 618 | 643 |
| Percent of population engaged in agriculture (%) | 7.44 | 21.99 | 47.12 |
| GDP (x $10^6$ yuan)[a] | 809.28 | 175.02 | 168.66 |
| GDP per capita (yuan) | 76979 | 16295 | 14036 |
| Ratio of GDP from agriculture to GDP (%) | 2.16 | 17.20 | 14.96 |
| Average wage of fully employed staff and workers (yuan) | 25411 | 18436 | 15604 |
| Annual net income per capita of rural household (yuan) | 9293 | 5420 | 5591 |

[a] yuan is the base unit of the renminbi, people's currency in the mainland of the People's Republic of China; 1 yuan equals about €0.11.

County, the percentages of population engaged in farming was respectively 21.99% and 47.12%. All of the GDP, GDP per capita, average wage of fully employed staff and worker and annual net income per capita of rural households in Changshu were the highest among the three counties, and they were listed at the lowest in Tongshan except annual net income per capita of rural households. While, Rudong had the highest percent of GDP from agriculture, Tongshan was ranked in the medium, and Changshu was listed at the last.

The crop land accounting for its agricultural land in Rudong, Tongshan, and Changshu County was 75%, 56% and 53% respectively, while its multiple crop index was 1.58, 1.56 and 1.38 respectively in 2006. The amount of fertilizer put into arable land was 5800, 4400 and 9000 Mg ha$^{-1}$ respectively in the above three counties, while its crop production was 58900, 63900, and 55800 Mg ha$^{-1}$ respectively. This indicated that the cultivated land use efficiency in Tongshan was lowest among the three counties.

## Field surveys

One of the reasons for choosing Changshu, Rudong and Tongshan County to study is that we enjoyed particularly good access to villages in these counties. In Changshu, Tongshan, and Rudong County, three, three, and four towns were selected based on the distance from their capitals to these towns. Considering the village numbers in every town and the town areas, one or two villages were randomly selected to conduct farm household surveys. Totally, 12 villages were selected. The detailed information about the selected villages is listed in Table 2 and the locations of the villages are described in Figure 1.

In every village, about 20 to 40 farm households were randomly selected and interviewed in July 2006. The data relating to farm households were obtained through structured interviews, informal discussions with village elders and local government leaders. In total, we obtained 343 questionnaires. However, some of them were incomplete or one questionnaire answered by different respondents, which were thought invalid. Finally, 329 copies were determined as valid. There were 89, 104, and 136 valid questionnaires in Changshu, Tongshan, and Rudong County respectively.

The following information was included in the questionnaires: agricultural production, resource of household, land uses, LUC in the recent 10 years, labor transfer, credit and saving, land market, locations, and so on. LUC in this study referred to the conversions determined by farmers themselves, not by the government. It should be noted here that some LUC, such as arable land to construction land, was illegal, but it in fact occurred. In total, 50 households converted their land uses in the three counties. Among them, 16 households converted arable lands to vegetable plots and orchard lands, 17 converted arable lands to fishery and aquaculture lands, and 17 converted arable lands to construction lands. Fisheries are very popular here because Jiangsu province is one of the famous "a land of fish and rice" due to its special climatic and land characteristics.

## Theoretical framework and independent variables

The decision on land use at a given parcel is made by a profit-maximizing "operator" of the land, and the "operator" may be "a single person, household, or group of people in the case of common property ownership" (Nelson and Geoghegan, 2002). The underlying motivation for operator to covert one use to another use is assumed to be maximization of expected return over an

**Table 2.** The surveyed villages and investigated samples

| County | Town | Village | Distance to the county city (km) | | | Number of interviewed household |
|---|---|---|---|---|---|---|
| | | | Minimum | Maximum | Mode | |
| Changshu | Meili | Quxiang | 20 | 20 | 20 | 28 |
| | Shajiabang | Langcheng | 17 | 24 | 20 | 29 |
| | Xinzhuang | Pingshu | 15 | 15 | 15 | 32 |
| Tongshan | Shanji | Caozhuang | 40 | 50 | 50 | 34 |
| | Hanwang | Gelou | 4 | 4 | 4 | 36 |
| | Zhengji | Zhengji | 25 | 30 | 25 | 34 |
| Rudong | Juegang | Gangnan | 2 | 3.5 | 2 | 29 |
| | Yangkou | Gu'ao | 30 | 35 | 35 | 22 |
| | Matang | Ma'nan | 12.5 | 17.5 | 15 | 26 |
| | | Ma'xi | 15 | 16 | 16 | 14 |
| | Chahe | Xingang | 23 | 30 | 25 | 26 |
| | | Xinghe | 23 | 28 | 25 | 19 |

infinite time horizon (Irwin and Geoghegan, 2001). The return of land uses is determined by the costs and benefits, so the operator makes the decision on LUC through comparing costs and benefits of alternative land uses (Gellrich *et al.,* 2007). Based on this theoretical supposition and given that there are no cost of converting from one land use to another land use, a model describing the operator's decision of land use can be developed. The parcel *k* will be converted from state *i* to state *j* in time *t* if

$$\begin{aligned} &V_{jt} \succ V_{it} \\ &V_{jt} = W_{jt} - C_{jt} \\ &V_{it} = W_{it} - C_{it} \end{aligned} \quad [1]$$

Where $v_{jt}$ is defined as the net present value of the future stream of net returns to parcel *k* in state *j* at time *t*, $v_{it}$ is the net present value of the future stream of returns to parcel *k* in state *j* at time *t*, $w_{jt}$ is the future stream of benefits to parcel *k* in state *j* at time *t*, $c_{jt}$ is the future stream of cost under operator control all for land use *j* at parcel *k*, $w_{it}$ is the future stream of benefits to parcel *k* in state *i* at time *t*, and $c_{it}$ is the future stream of cost under operator control all for land use *i* at parcel *k*.

The underlying response function is expressed through:

$$Y^* = \begin{cases} 1 \ if \ Y \succ 0 \\ 0 \ if \ otherwise \end{cases} \quad [2]$$

$$Y = V_{jt} - V_{it} = f(X)$$

Where $Y^*$ is a dummy variable, *X* is a vector of covariates including benefit-related variables and cost-related variables.

Land use will be converted on parcel *k*, if $Y^*$ takes the value 1; otherwise, it will not be converted. According to the theoretical analysis of farmer decisions on land use choice, and the former studies in LUC (Moran, 1981; Walker and Homma, 1996; Zhong, 2007), factors influencing farmer decisions were selected. They represented the characteristics of households, resources of households, labor transfer, land market, credit and saving, and location respectively. The demographic structure indicative of the characteristic of households exerts significant effects on the prominence of land uses (Perz, 2001), and demographic structure changes have the most significant effects on land use and land cover change (Evans *et al*., 2001; Stephen *et al*., 2006; VanWey *et al*., 2007). Resources of household included the area of arable land, the area of total land including arable land, fishery, forest land and orchard land, and the number of labor force aged between 16 to 65 years old, which impose strong influence on land uses (Dogliotti *et al.,* 2005). The labor transfer characteristics involved five factors: number of wage employees, number of self-employees, number of off-farm employment (the sum of the number of wage and self employees), income from off-farm activities, and per capita income from off-farm activities. The opportunity for off-farm employment and the income from off-farm activities have been proven to strongly influence land use changes (Godoy *et al*., 1997; Perz, 2001; Shriar, 2002; Shively and Pagiola, 2004). The land market is also believed an important factor to influence LUC because market will determine what is economically optimal (Drozd and Johnson, 2004). However, farm land sale market is forbidden in the mainland of China, where only agricultural land leasing market is permitted by law. Therefore, the area for rent-in and rent-out arable land, and the rent-in fishery land are selected here. The relation between the credit and saving, and land conversions is often discussed, on which there exists contrary views (Angelsen, 1999). Location is considered to be one of the most important determinants of land use and land cover change (Verburg *et al*., 2004). In this study, the distance from the household to its county city and the local counties were selected to represent the location characteristics.

In total, 19 independent variables were selected in this study. The detailed information is described in Table 3. The correlation coefficients between independent variables above 0.5 will bring multicollinearity problem (Wang and Guo, 2001), and the independent variables should remove redundancy before performing the simulation processes (Zhong *et al*., 2008a,b). However, considering the ability of the decision tree model can process nonlinear problems, the correlation analysis is not conducted in this study.

## Decision tree

This study attempts to simulate farmer decisions to pursuit the maximization of expected return through LUC at farm household level, based on a data mining approach, decision tree model. The basic concept of a decision tree is to split the complex decisions using most independent variables into several simpler decisions using conditional methods, which may lead to a solution that is easier to interpret. In a decision tree

**Table 3.** Independent variables used in this study

| Variable | Description | Mean | Standard deviation |
|---|---|---|---|
| **Characteristics of household** | | | |
| AGE | The age of the household head | 53.29 | 10.78 |
| GEN | Gender of the household head: M = male, F = female | 96% male | |
| EDU | Educational attainment of the household head (Years) | 6.89 | 3.25 |
| **Resource of household** | | | |
| ARA | Area of the contracted arable land by households (ha) | 3.67 | 2.17 |
| TAL | Area of the contracted land by households (ha) | 4.10 | 2.47 |
| NLA | Labor forces aged between 16 and 65 | 3.03 | 0.99 |
| **Labor transfer** | | | |
| NEE | Number of wage employees | 1.32 | 1.02 |
| NSE | Number of self-employees | 0.31 | 0.64 |
| TNE | Number of labor forces with off-farm employment | 1.67 | 1.05 |
| INE | Total income from off-farm activities (yuan) | 16526.10 | 22390.06 |
| AINA | Per capita income from off-farm activities (yuan) | 5392.68 | 6970.25 |
| **Land market** | | | |
| AHA | Area of the rent-in arable land by the households (ha) | 0.09 | 0.61 |
| ALA | Area of the rent-out arable land by the households (ha) | 0.25 | 0.92 |
| AHW | Area of the rent-in fishpond by household (ha) | 0.60 | 3.10 |
| ARL | Area of the rented land (ha): positive value for rent-in land, negative value for rent-out land | ---- | |
| **Credit and saving** | | | |
| CRE | credit or not: Credit = 1, and no credit = 0 | 22% credited | |
| SAV | The amount of saving by household ($10^4$ yuan) | 2.50 | 2.51 |
| **Location** | | | |
| DIS | The distance from the site of household to its county city (km) | 20.87 | 12.79 |
| COU | County name: CS = Changshu, RD = Rudong, TS= Tongshan | ---- | |

approach, the social-economic variables are the predictor variables whereas the LUC are referred to as the target variable. Several decision tree models have been developed to solve the environmental problems, such as ID3, CART, C4.5, See 5.0, Orange (Tooke *et al*., 2009). Among them, the model of CART (classification and regression tree), proposed by Breiman *et al*. (1984), has been widely used because of their non-parametric nature, simplicity, flexibility, and computational efficiency (Friedl and Brodley, 1997).

In CART, a tree-structured decision space is estimated by recursively splitting the data at each node on the basis of a statistical test that increases the homogeneity of the training data in the resulting descendant nodes (McLachlan, 1992). The decision tree grows by means of the successive subdivision until a stage is reached when there is no significant increase in homogeneity with further nodes. At this stage, the node will not subdivide further and automatically becomes a terminal node. Training accuracy is used to evaluate the simulation results, which equals to the simulated LUC of the samples inferred by simulation rules dividing by their actual land use conversations. In this study, the CART analysis was conducted with the software S-PLUS 2000.

To perform the accuracy assessment on the process of CART, the training accuracy is used. Firstly, the rules for simulating farmer decision on LUC are inferred based on the social and economic factors and farmer decisions on the sample questionnaires, and then farmer decisions are simulated using the social and economic factors on the sample questionnaires. The training accuracy is calculated by the simulated farmer decisions and the investigated farmer decisions. The sum of training accuracy and misclassification error equals to 1.

# Results

## Farmer decisions on land use conversions in Jiangsu Province

Using the 329 household questionnaires in Jiangsu Province, the CART model was used to simulate farmer decisions on LUC. The accuracy of the CART training process in rightly attributing LUC was 93.31%, indicating that the simulation results were acceptable. The simulated tree included 20 nodes and 10 independent variables, which were COU, ARL, AHW, AINA, AGE, TAL, DIS, TNE, INE, and NSE respectively (Figure 2). There were 20 branches to predict farmer decisions, among which 13 branches were for not convert land uses, and the other seven were for converting land uses.

CART provides the relative importance of independent variables on LUCD. As in our study, the other nine variables did not appear in the simulated decision tree. However, CART can not quantitatively assess the influence of these variables involved in the simulated tree on LUCD. To understand the quantitative importance of independent variables, the 239 training samples were used to evaluate the 10 variables of COU, ARL, AHW, AINA, AGE, TAL, DIS, TNE, INE, and NSE. The prediction error and the number of terminal nodes of the constructed tree model when using different variable groups are listed in Table 4.

The numbers of terminal nodes for the 14 tree models ranged from 17 to 21, indicating these tree models have similar complicated structures. Variables related to labor transfers including AINA, TNE, INE, and NSE were most important on LUCD (omitting these variables increased misclassification error to 10.64%), followed by the land market factors (misclassification error increased to 9.73%), location characteristics (misclassification error increased to 9.42%), household resources (misclassification error increased to 8.51%) and the characteristics of households (misclassification error increased to 7.90%). If comparing the importance

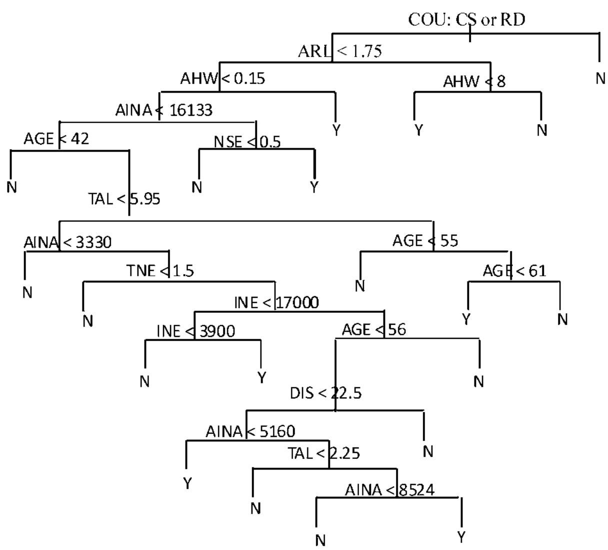


**Figure 2.** Decision tree for simulating farmer decisions on land use conversions in Jiangsu Province

**Table 4.** Mis-predication errors on land use conversion decision using different independent variables in Jiangsu Province

| Variables [a] | Misclassification error rate (%) | Number of terminal nodes |
|---|---|---|
| All | 6.69 | 20 |
| Missing AGE | 7.90 | 17 |
| Missing TAL | 8.51 | 19 |
| Missing ARL | 9.42 | 20 |
| Missing AHW | 9.42 | 21 |
| Missing ARL and AHW | 9.73 | 21 |
| Missing AINA | 7.90 | 19 |
| Missing TNE | 8.21 | 17 |
| Missing INE | 8.21 | 18 |
| Missing NSE | 9.12 | 19 |
| Missing AINA, TNE, INE, NSE | 10.64 | 17 |
| Missing COU | 7.60 | 19 |
| Missing DIS | 7.29 | 18 |
| Missing COU and DIS | 9.42 | 20 |

[a] Abbreviations: see Table 3.

of single dependent variables on LUCD, it was concluded that ARL, AHW and NSE had the strongest influence on LUCD, followed by TAL, TNE and INE, the factors of AGE, AINA, COU and DIS had the relative lower effect on LUCD.

Since there were social and economic differences among Rudong, Changshu, and Tongshan, the following analyses were conducted respectively in the three counties. In this study, only two households converted their land uses during the past 10 years in Tongshan County; and the other 48 households were located in Changshu and Rudong County. Therefore, the following analyses were conducted in Rudong and Changshu.

## Farmer decisions on land use conversions in Changshu County

The 136 questionnaires were used to simulate farmer LUCDs in Changshu County. Among these investigated households, 24 converted land uses. The training accuracy by CART was 89.71%, indicating the simulated results were acceptable. The simulation tree for LUCD included 12 nodes and seven independent variables (Figure 3). The factors of DIS, AINA, TAL, ARA, AGE, and NSE were involved in the simulation results. The branches for predicting LUCD in Changshu were 12, among which seven were for not converting land uses, and the other five were for converting land uses.

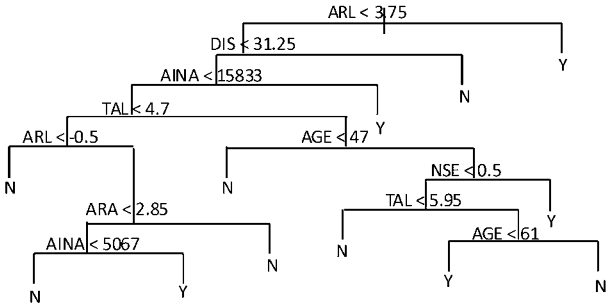


**Figure 3.** Decision tree for simulating farmer decisions on land use conversions in Changshu County

The mis-predication errors and the number of terminal nodes of the constructed tree model when using different variable groups in Changshu are listed in Table 5. The number of terminal nodes for the eight tree models ranged from eight to 13, indicating these tree models had similar complicated structures. Variables related to labor transfer including AINA and NSE, and household resource involved of NLA, TAL and ARA were most important on LUCD (omitting these variables increased misclassification error to 14.71%), followed by labor transfer factors (misclassification error increased to 12.24%) and location (misclassification error increased to 12.50%). If comparing the importance of single dependent variable on LUCD, it was concluded that NSE, DIS, TAL, and AINA had the strongest influence on LUCD, followed by APA and NLA and NLA.

**Table 5.** Mis-predication errors on LUCD using different independent variables in Changshu County

| Variables [a] | Misclassification error rate (%) | Number of terminal nodes |
|---|---|---|
| All | 8.82 | 12 |
| Missing ARL | 8.51 | 13 |
| Missing DIS | 12.50 | 8 |
| Missing NLA | 8.51 | 19 |
| Missing TAL | 12.50 | 11 |
| Missing ARA | 9.56 | 12 |
| Missing NLA,TAL, and ARA | 14.71 | 13 |
| Missing AINA | 10.29 | 13 |
| Missing NSE | 13.24 | 12 |
| Missing AINA and NSE | 14.71 | 13 |

[a] Abbreviations: see Table 3.

### Farmer decisions on land use conversions in Rudong County

Using the 89 questionnaires in Rudong County, CART model was used to simulate farmer decisions on LUC at the farm household level. Among these investigated households, 24 converted land uses, and 65 did not convert their land uses. The accuracy of the CART training process in rightly attributing land use changes was 85.39%. The simulated tree included seven nodes and four independent variables, which were ARL, AINA, ARA, and INE respectively (Figure 4). There were three rules for changing land use, and the other four for keeping the same land use format.

Among the four independent variables involved in the simulated tree in Rudong County, ARA was considered the most important influencing factor on LUCD (omitting this variable increased misclassification error to 21.35%, Table 6). This demonstrated that the arable land directly affected farmer decisions on LUCD in Rudong County. The labor transfer variables of AINA and INE were also included in the simulated decision tree, indicating that the number of employment and the income by employment had strong influence on LUCD. If INE and AINA were not involved in the simulated tree model, the misclassification error increased to 22.47%. The factor ARL also showed effect on farmer LUCD.

## Discussion

### Comparison of the simulated results by CART, Logit and Tobit model in Jiangsu Province

Based on the same questionnaires used in this study, Zhong *et al.* (2008a,b) simulated the influencing factors on LUC in Jiangsu Province using a Binary Logit Model and Tobit Model respectively. The simulation results showed that GEN, TAL, NSE, AHW, DIS and

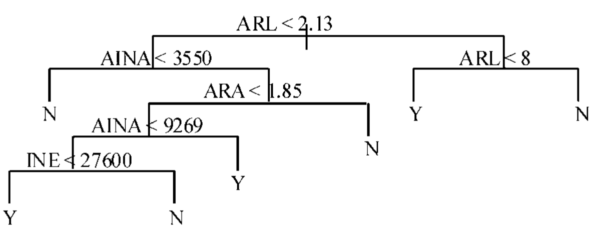


**Figure 4.** Decision tree for simulating farmer decision rules on land use conversions in Rudong County

**Table 6.** Mis-predication errors on LUCD using different independent variables in Rudong County

| Variables [a] | Misclassification error rate (%) | Number of terminal nodes |
|---|---|---|
| All | 14.61 | 7 |
| Missing ARL | 15.73 | 10 |
| Missing AINA | 17.98 | 8 |
| Missing ARA | 21.35 | 9 |
| Missing INE | 15.73 | 9 |
| Missing INE and AINA | 22.47 | 10 |

[a] Abbreviations: see Table 3.

COU had strong influence on farmer LUCDs by Logit model, while an additional factor SAV was involved in the simulation results by Tobit. In this study, the five factors of TAL, NSE, AHW, DIS and COU also showed strong influence on farmer decisions, which was almost agreed with the work by Zhong *et al*. (2008a,b).

The COU in the three studies presented strong influence on farmer decisions on LUC, indicating that farmer LUCDs had regional characteristics. The COU was the root node in the simulated tree, and most of the braches to predict farmer decisions were for Changshu and Rudong County, because only two households converted their land uses in Tongshan County. The simulation branches on the two household decisions were cut during the simulation process, considering the balance between the cost and the efficiency (McLachlan, 1992). The reason for regional difference might be the social and economic differences. In Changshu County, there are many industries which provide many opportunities for employment; in Rudong County, the home workshops for texture are popular there; and in Tongshan County, there are not many industries and home workshops.

In the two work by Zhong *et al*. (2008a,b), DIS showed negative influence on LUC. The households leave more far to its county city, the land would not be converted uses. This was mainly because the distance to the county city nearly represented the distance to the land market (Zhong *et al*., 2008b). In this study, DIS had effect on farmer LUCDs under some conditions of COU, ARL, AWH, AINA, AGE, TAL, and INE, and it did not show the always negative or positive influence on LUCDs (Figure 2). Under the conditions mentioned above, the households would not convert land uses when DIS greater than 22.5 km; while the conditions combing different conditions of TAL and NIAN when DIS less than 22.5, the households would have different decisions on LUC. This is the merit of the DTA method. It could fully consider the intertwined interactions among the factors in the simulating process.

The area of the contracted land by households showed positive influence on farmer decisions in the work by Zhong *et al*. (2008a,b). The more area of the contracted land by household, the more possibility of the LUC occurs. In this study, the influence of TAL on LUC would also determined by the other conditions of COU, ARL, AWH, AINA, AGE, DIS, INE and TNE. The influence of land market on LUCDs showed the area of the rent-in fishpond had positive influences (Zhong *et al*., 2008a,b). In this study, under some conditions of COU and ARL, the households would change their land uses when AWH was greater than 0.15 ha; while the condition of AWH smaller than 0.15 ha and other conditions would determine farmer decisions on LUC. The number of self-employees of the households showed positive influence on farmer LUCDs in Zhong *et al*. (2008a,b). In this work, similar results to Zhong *et al*. (2008a,b) were obtained under some conditions of COU, ARL, and AHW. When NSE was < 1, land use would not be converted; otherwise, it would.

### Comparison of factors and rules for LUCD between local counties

It was obvious that farmer decisions on LUC were more complicated in Jiangsu Province than that in local counties (Figures 2, 3, and 4). Totally, 20 branches for simulating LUC using 10 independent variables were obtained in Jiangsu Province by CART. There were 12 rules with 7 variables in Chaungshu County, and 7 rules with 4 variables in Rudong County to simulate farmer LUCDs. The influencing factors on farmer decisions in Jiangsu Province included labor transfer, land market, location, resource and characteristics of household, while the characteristics of household were not involved in Changshu and Rudong County.

The simulated decision trees for farmer LUCDs in Changshu County had more complicated structure than that in Rudong County (Figures 3 and 4). In Rudong, the contracted arable land area by households, the total income from employment, the average income from off-farm activities, and the rent-in land by households were the main factors to influence farmer decisions on LUC. In Changshu County, the three factors of resources of households, the number of self-employees, per capita income from off-farm activities, the area of rent-in land, and distance to its county city showed strong influence on farmer LUCDs. Comparison with the two simulation results in Changsu and Rudong County, the additional factor of location was involved in the simulated tree in Changshu. This demonstrated that farmer decisions on LUC were influenced by more complicated conditions in Changshu than that in Rudong County.

### Further studies in simulating farmer decisions on land use conversions at farm household level

Although we tried to select enough influencing factors on LUC at farm household levels, there might be other factors to influence farmer decisions. For example, the government policy has strong influence on farmer decisions on LUC (Lichtenberg and Ding, 2008). This factor was not involved in our study because we have not enough data. In the further studies, the government policies would be involved in the simulation process.

Furthermore, it would be more useful to break down land-use changes to several categories surveys and identify the major changes from the surveys and investigate why these major changes happen. However, in our study area, there were only 16, 17, and 17 households converted their arable lands to vegetable plots and orchard lands, fishery and aquaculture lands, and construction lands. The samples were not enough when classifying the LUC.

Moreover, there are also some problems needed to further study. For instance, could the selection of villages be a potential problem for the findings about location which is important in one county but not in another? How the imbalance on the numbers of household that have changed and have not changed land use could affect the results? How many villages should be enough to study the problem of farmer decisions on LUC?

## Conclusions

In this study, farmer decisions on LUC at farm household level were simulated by a decision tree model in Jiangsu Province. This model obtained reasonable accuracy for predicating farmer LUCDs of 93.31%, 89.71% and 85.39% in Jiangsu Province, Changshu County and Rudong County respectively. The simulation results also showed that farmer decisions presented regional characteristics. The decision tree structure in Jiangsu Province was more complicated than that in Changshu County, which was more complicated than that in Rudong County. There were 20 branches with 10 indepent variables, 12 branches with 7 variables and 7 branches with 4 variables in the simulated trees to predict farmer decisions on LUC in Jiangsu Province, Changshu County and Rudong County respectively.

Moreover, the decision tree model provides some insight into the social and economic variables that are most responsible for driving LUC. The factors influencing farmer LUCDs in Jiangsu Province were ranked as decreased importance as labor transfer, land market, location, resource of household, characteristics of household, according to the importance of the effect on LUCD. In Changshu County, the related factors on LUCD ranked as the decreased importance as NSE, DIS, TAL, AINA, ARA, NLA, and ARL; while in Rudong County, the 4 involved factors ranked ARA, AINA, ARL, and INE.

## Acknowledgments

This contribution was financed by National Natural Science Foundation of China (No. 40801063 and No. 70373029), and Nanjing University Talent Development Foundation.

## References

ADESINA A.A, MBILA D., NKAMLEU G.B., ENDAMANA D., 2000. Econometric analysis of the determinants of adoption of alley farming by farmers in the forest zone of southwest Cameroon. Agr Ecosyst Environ 80(3), 255-265. doi: 10.1016/S0167-8809(00)00152-3.

ANGELSEN A., 1999. Agricultural expansion and deforestation: modeling the impact of population, market forces and property rights. J Dev Econ 58, 185-218. doi: 10.1016/S0304-3878(98)00108-4.

ALISSON B., RICHARD B., WILLIAM P., 2005. Farm household lifecycles and land use in the Ecuadorian Amazon. Popul Environ 27, 1-27. doi: 10.1007/s11111-005-0013-y.

ALIX-GARCIA J., DE JANVRY A., SADOULET E., 2005. A tale of two communities: explaining deforestation in Mexico. World Dev 33, 219-235. doi: 10.1016/j.wiekddev.2004.07.010.

BREIMAN L., FRIEDMAN J.H., OLSHEN R.A., STONE C.J., 1984. Classification and regression trees. Wadsworth International Group, Belmont, California, pp. 1-451.

COXHEAD I., SHIVELY G., SHUAI X., 2002. Development policies, resource constraints, and agricultural expansion on the Philippine land frontier. Environ Dev Econ 7(3), 341-363. doi:10.1017/S1355770X02000219.

DOGLIOTTI S., VAN ITTLERSUM M.K., ROSSING W.A.H., 2005. A method for exploring sustainable development options at farm scale: a case study for vegetable farms in South Uruguay. Agr Syst 86, 29-51. doi: 10.1016/j.agsy.2004.08.002.

DROZD D.J., JOHNSON B.B., 2004. Dynamics of rural land market experiencing farmland conversion to acreages: the case of Saunders County, Nebraska. Land Econ 80, 294-311. doi:10.3368/le.80.2.294.

EVANS T.P., MANIRE A., CASTRO F.D., BRONDIZIO E., MCCRACKEN S., 2001. A dynamic model of household decision-making and parcel level landcover change in the eastern Amazon. Ecol Model 143, 95-113. doi: 10.1016/S0304-3800(01)00357-X.

FISHER M., SHIVELY G., 2005. Can income programs reduce tropical forest pressure? Income shocks and forest use in Malawi. World Dev 33, 1115-1128. doi:10.1016/j.worlddev.2005.04.008.

FRIEDL M.A., BRODLEY C.E., 1997. Decision tree classification of land cover from remotely sensed data. Remote Sens Environ 61, 399-409. doi: 10.1016/S0034-4257(97)00049-7.

GELLRICH M., BAUR P., KOCH B., ZIMMERMANN K.E., 2007. Agricultural land abandonment and natural forest regrowth in the Swiss mountains: A spatially explicit economic analysis. Agr Ecosyst Environ 118, 93-108. doi: 10.1007/s10666-006-9062-6.

GODOY R., O'NEILL K., GROFF S., 1997. Household determinants of deforestation by Amerindians in Honduras. World Dev 25, 977-987. doi:10.1016/S0305-750X(97)00007-7.

HERATH P.H.M.U., TAKEYA H.Y., 2003. Factors determining intercropping by rubber smallholders in Sri Lanka: a logit analysis. Agr Econ 29, 159-168. doi:10.1016/ S0169-5150(03)00045-8.

HO S.P.S., LIN G.C.S., 2004. Converting land to nonagricultural use in China's coastal provinces-Evidence from Jiangsu. Modern China 30, 81-112. doi: 10.1177/0097700403259131.

IRWIN E.G., GEOGHEGAN J., 2001. Theory, data, methods: developing spatially explicit economic models of land-use change. Agr Ecosyst Environ 85, 7–23. doi: 10.1016/S0167-8809(01)00200-6.

LI X., YEH A.G.O., 2004. Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS. Landscape Urban Plan 69, 335-354. doi:10.1016/j.landurbplan.2003.10.033.

LICHTENBERG E., DING C., 2008. Assessing farmland protection policy in China. Land Use Policy 25, 59-68. doi:10.1016/j.landusepol.2006.01.005.

LILLESAND T.M., KIEFER R.W., 2000. Remote sensing and image interpretation. 4th edition. John Willey and Sons, NY, USA. 724 pp.

McLACHLAN G.J., 1992. Discriminate analysis and statistical pattern recognition. Wiley Interscience, NY, USA, 323 pp.

MORAN E.F., 1981. Developing the Amazon. Indiana University Press, Bloomington, IN.

NELSON G.C., GEOGHEGAN J., 2002. Deforestation and land use change: sparse data environments. Agr Econ 27, 201–216. doi:10.1016/S0169-5150(02)00080-4.

ODIHI J., 2003. Deforestation in afforestation priority zone in Sudano-Sahelian Nigeria. Appl Geogr 23(4), 227-259. doi: 10.1016/j.apgeog.2003.08.004.

PERZ S.G., 2001. Household demographic factors as life cycle determinants of land use in the Amazon. Popul Res Policy Rev 20, 159-186. doi: 10.1007/s10745-006-9039-8.

PICHÓN F.J., 1997. Settler households and land-use patterns in the Amazon frontier: Farm-level evidence from Ecuador. World Dev 25(1), 67-91. doi: 10.1016/S0305-750X(96)00091-5.

SHIVELY G.E., PAGIOLA S., 2004. Agricultural intensification, local labor markets, and deforestation in the Philippines. Environ Dev Econ 9, 241-266. doi: 10.1017/S1355770X03001177.

SHRIAR A.J., 2002. Food security, land use, and deforestation in northern Guatemala. Food Policy 27, 395-414. doi: 10.1016/S0306-9192(02)00071-4.

STREETS D., CHUNG C., SU H., 1995. Remote sensing of global change: growth in China's Jiangsu province. Int J Sustainable Dev World Ecol 2, 257-266.

STEPHEN P., ROBERT W., MARCELLUS C., 2006. Beyond population and environment: household demographic life cycles and land use allocation among small farms in the Amazon. Hum Ecol 34, 829-849. doi: 10.1007/s10745-006-9039-8.

TAN M.H., LI X.B., XIE H., LU C.H., 2005. Urban land expansion and arable land loss in China-a case study of Beijing-Tianjin-Hebei region. Land Use Policy 22, 187-196. doi:10.1016/j.landusepol.2004.03.003.

TASSER E., WALDE J., TAPPEINER U., TEUTSCH A., NOGGLER W., 2007. Land-use changes and natural reforestation in the Eastern Central Alps. Agr Ecosyst Environ 118, 115-129. doi:10.1016/j.agee.2006.05.004.

THAPA K.K, BILSBORROW R.E., MURPHY L., 1996. Deforestation, land use, and women's agricultural activities in the Ecuadorian Amazon. World Dev 24(8), 1317-1332. doi: 10.1016/0305-750X(96)00041-1.

TOOKE T.R., COOPS N.C., GOODWIN N.R., VOOGT J.A., 2009. Extracting urban vegetation characteristics using spectral mixture analysis and decision tree classifications. Remote Sens Environ 113, 398-407. doi: 10.1016/j.rse.2008.10.005.

VANWEY L.K., D'ANTONA A.O., BRONDÍZIO E.S., 2007. Household demographic change and landuse/land cover change in the Brazilian Amazon. Popul Environ 28, 163-185. doi: 10.1007/s11111-007-0040-y.

VERBURG P.H., OVERMARS K.P., WITTE N., 2004. Accessibility and land-use patterns at the forest fringe in the northeastern part of the Philippines. Geographical J 170, 238-255. doi: 10.1111/j.0016-7398.2004.00123.x.

WALKER R., HOMMA A.K., 1996. Land use and land cover dynamic in the Brazilian Amazon: an overview. Agr Econ 18, 67-80. doi:10.1016/0921-8009(96)00033-X.

WANG J.C., GUO Z.G., 2001. Regression - method and application. High Education Press House, Beijing, China. 251 pp. [In Chinese].

WENG Q.H., 2002. Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modeling. J Environ Manage 64, 273-284. doi:10.1006/jema.2001.0509.

WU L.X., SUN B., ZHOU S.L., HUANG S.E., ZHAO Q.G., 2004. A new fusion technique of remote sensing images for land use/cover. Pedosphere 14, 187-194. doi: 1002-0160.0.2004-02-008.

XIAO S.S., HUANG X.J., PENG B.Z., 2007. Coordinative development between land use change and regional population-environment-development system- A case study of Jiangsu Province. Chin Geogra Sci 17, 289-296. doi: 10.1007/s11769-007-0289-1.

XU W., 2004. The changing dynamics of land-use change in rural China: a case study of Yuhang, Zhejiang Province. Environ Plann A 36, 1595-1615. doi: 10.1068/a36185.

ZHANG X.Y., 2006. Detecting urban vegetation categories based on objected-oriented method from high resolutionary remotely sensed data in Nanjing City. Nanjing University.

ZHANG X.Y., LIN F.F., JIANG Y.G., WANG K., WONG M.T.F., 2008. Assessing soil Cu content and anthropogenic influences using decision tree analysis. Environ Pullut 156, 1260-1267. doi: 10.1016/j.envpol.2008.03.009.

ZHONG T.Y., 2007. Impact of labor mobility on land use and land use change: analysis at rural household level. Nanjing University. [In Chinese].

ZHONG T.Y., ZHANG X.Y., HUANG X.J., 2008a. Impact of labor transfer on agricultural land use conversion at farm household level based on Logit Model. Chin Geogra Sci 18, 300-307. doi: 10.1007/s11769-008-0300-5.

ZHONG T.Y., HUANG X.J., ZHANG X.Y., HU J., 2008b. Analysis of agricultural landuse conversion at household level based on Tobit model. Bull Soil Water Conserv 28, 760-766. doi: STTB.0.2008-05-035. [In Chinese].

ZHOU B., XU H.W., WANG R.C., 2003. Soil organic matter mapping based on classification tree modeling. Acta Pedologica Sinica 40, 801-808. doi: 0564-3929.0.2003-06-000 [In Chinese].