Front. Earth Sci. 2014, 8(4): 512–523
DOI 10.1007/s11707-014-0426-y

RESEARCH ARTICLE

Comparison of multinomial logistic regression and logistic regression: which is more efficient in allocating land use?

Yingzhi LIN (✉)[1], Xiangzheng DENG[2,3], Xing LI[1], Enjun MA[1]

1 School of Mathematics and Physics, China University of Geosciences, Wuhan 430074, China
2 Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
3 Center for Chinese Agricultural Policy, Chinese Academy of Sciences, Beijing 100101, China

Abstract Spatially explicit simulation of land use change is the basis for estimating the effects of land use and cover change on energy fluxes, ecology and the environment. At the pixel level, logistic regression is one of the most common approaches used in spatially explicit land use allocation models to determine the relationship between land use and its causal factors in driving land use change, and thereby to evaluate land use suitability. However, these models have a drawback in that they do not determine/allocate land use based on the direct relationship between land use change and its driving factors. Consequently, a multinomial logistic regression method was introduced to address this flaw, and thereby, judge the suitability of a type of land use in any given pixel in a case study area of the Jiangxi Province, China. A comparison of the two regression methods indicated that the proportion of correctly allocated pixels using multinomial logistic regression was 92.98%, which was 8.47% higher than that obtained using logistic regression. Paired *t*-test results also showed that pixels were more clearly distinguished by multinomial logistic regression than by logistic regression. In conclusion, multinomial logistic regression is a more efficient and accurate method for the spatial allocation of land use changes. The application of this method in future land use change studies may improve the accuracy of predicting the effects of land use and cover change on energy fluxes, ecology, and environment.

Keywords multinomial logistic regression, land use change, logistic regression, land use suitability, land use allocation

Received April 3, 2013; accepted October 22, 2013

E-mail: linyz.ccap@igsnrr.ac.cn

1 Introduction

Global land use and cover change (LUCC) have exerted significant impacts on surface energy fluxes, ecology, and the environment at both local and global scales. To estimate the effects of future LUCC on energy fluxes and projected ecological and environmental changes, an accurate and spatially explicit simulation of land use change is vital (Kalnay and Cai, 2003; Pielke, 2005; Chen et al., 2006; Jiang et al., 2011; Meiyappan and Jain, 2012). In addition, modeling the spatial dynamics of regional land use assists in clarifying the dynamics of current land use and in projecting future land use trajectories for the purpose of making informed management decisions (Cao and Ye, 2013). This has already led to the development of a wide range of land use change models.

CLUE-S, one of the most widely used models, was specifically developed for generating a spatially explicit simulation of land use change based on an empirical analysis of location suitability, combined with a dynamic simulation of competition and interaction between the spatial and temporal dynamics of land use systems (Verburg et al., 2002). Currently, models such as CLUE-SII, Dyna-CLUE, and other variants of CLUE-S play leading roles in modeling the spatial dynamics of regional land use (Chen and Verburg, 2000; Duan et al., 2004; Verburg and Overmars, 2009; Lin et al., 2011).

Logistic regression (LR) is the key component of the aforementioned models. It is used to evaluate the statistical relationships between land use and its driving factors (Geoghegan et al., 2001; Serneels and Lambin, 2001). In these models, LR is used to indicate the probability of a certain grid cell to be dedicated to a certain land use type given a set of driving factors. The primary assumption of LR in these models is that the spatial pattern of land use is correlated with relevant driving factors (Braimoh and Onishi, 2007; Gellrich et al., 2007; Pueyo and Beguería, 2007; van Doorn and Bakker, 2007). The common form of LR is

$$p_{i,u} = \frac{\exp(\beta_{0,u} + \beta_{1,u}x_{1,i} + \cdots + \beta_{k,u}x_{k,i} + \cdots + \beta_{n,u}x_{n,i})}{1 + \exp(\beta_{0,u} + \beta_{1,u}x_{1,i} + \cdots + \beta_{k,u}x_{k,i} + \cdots + \beta_{n,u}x_{n,i})}, \tag{1}$$

where, $p_{i,u}$ is the probability or suitability of land use type u appearing in pixel i, $x_{k,i}$ is the variable value of the k^{th} driving factor in pixel i. LR helps to determine which type of land use is most likely to appear in a certain pixel, which can be described more succinctly as land use suitability. The process of identifying the probability of land use conversion in spatially explicit simulation models of land use change is known as land use allocation. A pixel with a high suitability for a specific type of land use has a higher probability of being converted to this type of land use from its original land use type.

Various attempts have been made to refine the LR method. For example, Duan et al. (2004) introduced neighborhood enrichment factors and average neighborhood enrichment factors in LR to improve simulation accuracy. Wu et al. (2010) introduced an autologistic regression method by incorporating components describing the spatial autocorrelation into LR, which was then used to improve the spatial analysis module of the CLUE-S model. In addition, in order to improve the accuracy of evaluating the relationships between land use and its driving factors, Verburg et al. (2002), Alabi (2011), and Sohl et al. (2012) used stepwise logistic regression to select significant driving factors from a large set of location characteristics. Nevertheless, these improvements do not address the key underlying shortcoming of LR in evaluating land use suitability. The key shortcoming of LR is that it neglects the crucial role of the pixel representing the original land use type in any evaluation of land use change. In other words, the probability of a certain grid cell to be dedicated specifically to a land use type is not applicable in land use allocation or in land use change simulation.

There are key distinctions between the probabilities of a certain grid cell to be devoted to a land use type, and a land use change given a set of driving factors. In the models discussed previously, the role of the original land use type of pixel is evaluated by land use conversion elasticity or similar matrices, which are performed subjectively and indistinguishably for all pixels (Verburg and Veldkamp, 2004). Undoubtedly, the subjectivity and homogeneity of this type of evaluation reduces the accuracy of land use allocation. This issue can be addressed easily by calculating the probability of a certain grid cell to be devoted to a land use change given a set of driving factors. However, the variables associated with land use changes are always polytomous-nominal, and are not compatible with LR. This necessitates the use of different methodologies to describe the relationships between these variables and their driving factors.

The main goal of this study is to introduce an alternative methodology, multinomial logistic regression (MNL), to describe the relationships between land use change and its driving factors, and to empirically demonstrate that this methodology is more efficient than LR in land use suitability evaluations. The theory behind MNL is explained in section 2. The case study area, data, and variables used in the comparative analysis of MNL and LR are introduced in section 3. The results are provided in section 4, and conclusions are presented and discussed in the final section.

2 Multinomial logistic regression

The MNL model is a regression model that generalizes the LR model by allowing for more than two discrete and unordered response variables. The multinomial logit was first introduced by Luce (1959). McFadden (1974) provided a general procedure for formulating the MNL model. Currently, the MNL model is widely used in biometrics, econometrics, psychometrics, sociometrics, and many other fields, where it uses a set of explanatory variables to predict the probabilities of different possible outcomes of categorically distributed responses (Dakin et al., 2006; Millington et al., 2007; Briz and Ward, 2009; Choi et al., 2011; dell'Olio et al., 2011). When considering all of the possible land use changes together as a set of possibilities and each land use change as a possibility within the set, MNL is applicable in land use suitability evaluations. Suppose that $Y_{i,u}$, the land use change of pixel i accompanied by original land use type u, is a nominal variable with U categories $C_{u,1}$, …, $C_{u,v}$, … , $C_{u,U}$ ($C_{u,v}=\{v\}$, denoting the occurrence of conversion from land use type u to land use type v), which are observed together with P explanatory variables $X_{1,i}$, … , $X_{p,i}$. Let

$$P_{i,u,v} = P(Y_{i,u} \in C_{u,v} | X_{1,i}, \cdots X_{p,i}), v = 1,...,U. \tag{2}$$

Eq. (2) denotes the probability that $Y_{i,u}$ belongs to $C_{u,v}$ (i.e., the occurrence probability of conversion from land use type u to land use type v in pixel i), under the condition (explanatory variables) $X_{1,i}$, … , $X_{p,i}$ of pixel i. Therefore, an observation $X_{1,i}$, … , $X_{p,i}$ fulfills the logistic assumption if

$$\log\left(\frac{P_{i,u,v}}{P_{i,u,u}}\right) = \beta_{0,u,v} + \beta_{1,u,v}X_{1,i} + \cdots + \beta_{p,u,v}X_{p,i,v} = 1,2,...,U. \tag{3}$$

Naturally, a general form of the MNL model is obtained

$$P_{i,u,v}=\begin{cases}\dfrac{1}{1+\sum\limits_{k\in\{1,\cdots,U\},k\neq u}\exp(\beta_{0,u,k}+\beta_{1,u,k}X_{1,k}+\cdots+\beta_{p,u,k}X_{p,i})}, & \text{for } v=u;\\ \dfrac{\exp(\beta_{0,u,v}+\beta_{1,u,v}X_{1,i}+\cdots+\beta_{p,u,v}X_{p,i})}{1+\sum\limits_{k\in\{1,\cdots,U\},k\neq u}\exp(\beta_{0,u,k}+\beta_{1,u,k}X_{1,i}+\cdots+\beta_{p,u,k}X_{p,i})}, & \text{for } v\neq u.\end{cases} \quad (4)$$

A semantic definition of $P_{i,u,v}$ is the probability of a land use change from land use type u to land use type v occurring in pixel i. This definition determines that the observations are divided into U groups by their original states of land use, and the parameters of Eq. (4) are estimated by group. Therefore, the role of original land use in determining land use change is taken into account in the MNL model. In this study, $P_{i,u,v}$ can be defined as the suitability of land use type v in pixel i. The parameters of MNL can be estimated using the Maximum Likelihood Estimation approach (Hosmer and Lemeshow, 2000).

3 Data and definitions

We used high quality remote sensing digital satellite images of land use as data for estimating the parameters of the MNL and LR models. The land use database was developed by the Chinese Academy of Sciences (CAS) using original data from Landsat TM/ETM images with a spatial resolution of 30 by 30 meters. These were aggregated by CAS into one kilometer by one kilometer picture elements ('pixels'), which are the observations used in this study. The database includes observations for three time periods: a) the mid-1990s, including Landsat TM/ETM scenes from 1995 and 1996 (henceforth, 1995); b) the late 1990s, including Landsat TM/ETM scenes from 1999 and 2000 (henceforth, 2000); and c) the mid-2000s, including Landsat TM/ETM scenes from 2005 and 2006 (henceforth, 2005). For each time period, more than 500 TM/ETM scenes were used to map all of China. The interpretations of the TM/ETM images and land use/cover classifications were then validated carefully against extensive field surveys (Liu et al., 2003; Deng et al., 2008). The data were then classified into one of six land use/cover classes, namely, cultivated land, forest, grassland, water area, built-up area, and unused land. In other words, every pixel on the map of China was classified into one of these categories.

3.1 The case study area

The Jiangxi Province (24°7′–29°9′N in latitude, 114°02′–118°28′E in longitude) has experienced dynamic changes in land use arising from human activity. Because our data were from three discrete sets of years, namely, 1995, 2000, and 2005, they were used to track these changes in land use. Over the ten year period from 1995 to 2005, 3.64×10^4 km^2 of land, which represents 21.8% of the province's total area, experienced significant land use change.

Economic development and environmental restoration were the two main drivers of land use and land cover changes in the Jiangxi Province. From 1995 to 2005, the GDP per capita in this province increased from 2,896 CNY to 4,851 CNY. Rapid economic development paralleled improvements in infrastructure and clear cutting for agriculture (Zhan et al., 2010). This process accelerated the expansion of both built-up area and cultivated land. On the other hand, some national and provincial environmental restoration programs, such as the Grain for Green Program, the Reclaimed Farmland to Lake Program, and the Fast-growing and High-yielding Program, have been responsible for converting significant tracts of cultivated land to forests and water area (Wang et al., 2012). By 2005, the forest coverage in the Jiangxi Province increased to 62.4 percent, well above the national average (<20 percent). Over 80 percent of the newly forested areas were originally classified as cultivated land.

The classification of land use into six types generates twelve dependent variables for MNL and LR, which contributes significantly to the complexity of the efficiency comparisons. Accordingly, we decided to focus on specific types of land use change. Cultivated land was involved in both economic development and environmental restoration in the period 1995–2005. Specifically, 1.47×10^4 km^2 of non-cultivated land was converted to cultivated land while 1.53×10^4 km^2 of cultivated land was converted to non-cultivated land. In other words, the latter type of conversion accounted for more than 40% of all land use changes in the province between 1995 and 2005. Therefore, we decided to focus only on this type of conversion. We used all 166,932 observations when estimating parameters for LR, of which a subset of 45,716 observations (cultivated land pixels in 1995) were selected for estimating the parameters for MNL, and 44,561 observations (cultivated land pixels in 2000) were selected for evaluating land use suitability in 2005 and for comparative analysis (Fig. 1).

3.2 Definitions of dependent variables

To compare the relative efficiencies of MNL and LR in assessing the suitability of land uses, we evaluated two kinds of pixel-specific dependent variables of interest. The dependent variable for MNL was defined using data from two distinct sets of years, 1995 and 2000:

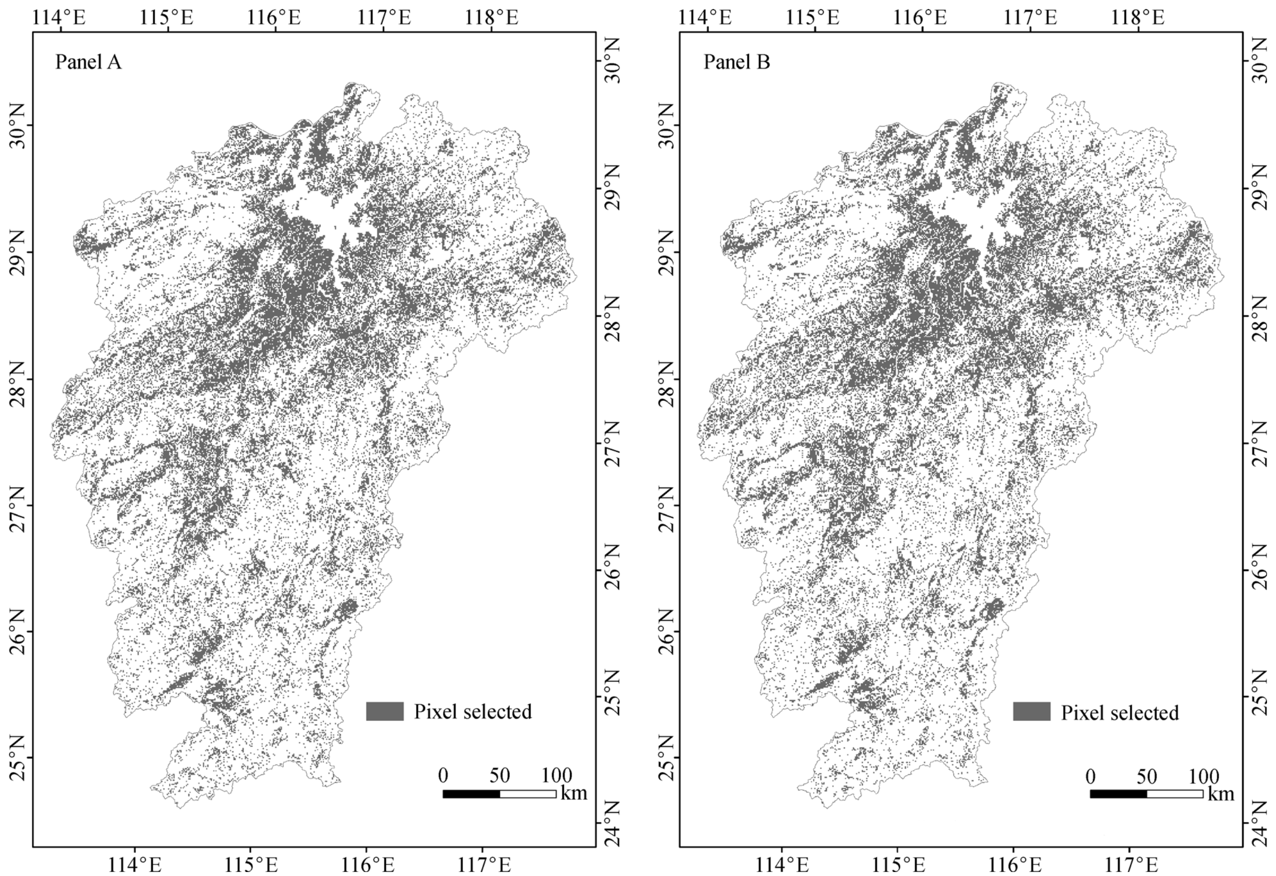

Fig. 1 Pixels selected for parameter estimation in multinomial logistic regression (Panel A), and for land use suitability evaluation and comparative analysis (Panel B).

$$DCLand_{1995-2000} = u, \text{ if the pixel is converted from cultivated land pixel to } u^{\text{th}} \text{ kind of land pixel from 1995 to 2000.} \tag{5}$$

u = 1, 2, 3, 4, 5, 6 refers to cultivated land, forest, grassland, water area, built-up area, and unused land, respectively, and $DCLand_{1995-2000}$ is the polytomous nominal variable representing the conversion from cultivated land to land use type u. Each of the 45,716 selected cultivated land pixels in 1995 has one of six designations: maintained cultivated land, converted from cultivated land to forest, converted from cultivated land to grassland, converted from cultivated land to water area, converted from cultivated land to built-up area, and converted from cultivated land to unused land.

The dependent variables for LR are defined as follows:

$$Land_{2000,u} = \begin{cases} 1, \text{ if the pixel is } u^{\text{th}} \text{ kind of land pixel in 2000;} \\ 0, \text{ if the pixel is not } u^{\text{th}} \text{ kind of land pixel in 2000.} \end{cases} \tag{6}$$

$Land_{2000,u}$ is the binary variable representing the existence of land use type u in the respective pixel. To generate the dependent variables for LR, all of the pixels mapping the Jiangxi Province in 1995 were considered.

3.3 Data for explanatory variables

Seventeen independent variables were created to account for the factors that drive land use changes. Four broad categories of variables have been used in other empirical studies in this area. Ostwald and Chen (2006), Walsh et al, (2008), Bahadur (2011), and others included a number of geographic and climatic variables. Zhong et al. (2011) and Nahuelhual et al. (2012) included measures of distance from the pixels to different features (such as, distance to the nearest city and distance to roads). Schaldach and Alcamo (2006) and Claessens et al. (2009) revealed that soil properties were primary determinants of land use changes. Other authors, such as Rozelle et al. (1997), Williams (2007) and Han et al. (2008) included demographic and economic variables.

In order to ensure the consistency of our analysis with the rest of the literature, we tried to collect all available information on all of the four sets of variables: geographic and climatic factors, measures of distance, soil properties, and demographic and economic factors. However, highly qualified demographic and economic data were unavailable. Given that demographic and economic variables are highly correlated with the proportions of both built-up area and cultivated land, we included the variables *bufferfarmland* and *bufferbuiltup* to identify a pixel that was surrounded by cultivated land or a built-up area, respectively. In addition, we created a variable, *bufferforest*, to identify a pixel that was surrounded by forest. The purpose of including this variable was to hold constant the impact on afforestation that could arise if a cultivated land pixel was situated near forested pixels. We categorized these three variables as buffer variables. Consequently, a total of four sets of independent variables were included in MNL and LR, geographic and climatic variables, measures of distance, soil properties, and buffer variables.

To generate the independent variables for our analysis, we used data and information from various sources. The elevation (*elevation*, m), terrain slope (*slope*, degree), and aspect (*aspect*, degree) variables, which characterize the nature of the terrain of each pixel, were generated from China's digital elevation model data set within the basic CAS data base. The data for measuring the mean rainfall (*precipitation*, mm) and averaged temperature (*temperature*, ºC) of two periods, 1995 to 1999 and 2000 to 2004, were obtained from the CAS data center after initially being collected and organized by the Meteorological Observation Bureau of China from more than 600 national climatic and meteorological data centers. For our study, we interpolated the relevant observations from the Jiangxi Province into surface data using an approach called the Thin Plate Smoothing Spline Method. These two variables in the period from 1995 to 1999 were used for estimating the parameters of MNL and LR (Table 1). Their equivalents in the period from 2000 to 2004 were used

Table 1 Descriptive statistics of the variables at the pixel level involved in logistic regression and multinomial logistic regression

Variable	Logistic regression*			Multinomial logistic regression**		
	Median	Mean	Std. Dev.	Median	Mean	Std. Dev.
Dependent variables						
$DCLand_{1995–2000}$		—		1	1.32	0.79
$Land_{2000,1}$	0	0.27	0.44		—	
$Land_{2000,2}$	1	0.62	0.48		—	
$Land_{2000,3}$	0	0.04	0.20		—	
$Land_{2000,4}$	0	0.04	0.20		—	
$Land_{2000,5}$	0	0.02	0.13		—	
$Land_{2000,6}$	0	0.01	0.07		—	
Independent variables						
Elevation/m	200.00	251.64	229.43	88.00	129.67	133.71
Slope/(°)	1.67	2.73	2.99	0.54	1.31	1.90
Aspect/(°)	93.51	93.75	52.95	103.76	100.43	53.42
Precipitation/mm	1846.70	1841.41	152.62	1922.39	1878.90	146.77
Temperature/°C	18.18	17.94	1.23	18.12	17.81	1.24
Dtoroad/km	1.00	0.68	0.82	0.00	0.50	0.66
Dtocity/km	159.00	173.27	82.05	130.00	146.68	81.44
Dtowater/km	0.00	0.34	18.31	0.00	0.02	4.68
pH	4.60	4.69	0.76	4.60	4.71	0.67
Loam/%	24.00	26.00	4.68	24.00	26.44	5.02
Organics/%	3.14	3.08	0.90	3.14	3.01	0.67
Nitrogen/%	0.16	0.15	0.03	0.16	0.15	0.03
Phosphorous/%	0.04	0.04	0.01	0.04	0.04	0.01
Potassium/%	1.79	1.78	0.28	1.79	1.77	0.21
Bufferfarmland/%	16.53	27.26	28.88	62.81	60.38	25.46
Bufferbuiltup/%	0.00	1.66	5.66	0.00	3.31	5.81
Bufferforest/%	73.55	62.23	34.92	23.97	30.46	28.05

Notes: *166,932 observations were used when estimating parameters for LR; ** 45,716 observations were used when estimating parameters for MNL.

Table 2 Descriptive statistics of the variables at the pixel level prepared for the evaluation of land use suitability estimating from 44,561 observations

Variable	Median	Mean	Std. Dev.
Elevation/m	84.00	125.85	127.92
Slope/(°)	0.54	1.29	1.88
Aspect/(°)	105.86	101.65	53.19
Precipitation/mm	1,609.04	1,610.64	57.57
Temperature/°C	18.38	18.05	1.15
Dtoroad/km	0.00	0.49	0.64
Dtocity/km	129.00	146.11	81.11
Dtowater/km	0.00	0.02	4.71
pH	4.60	4.71	0.67
Loam/%	24.00	26.43	5.02
Organics/%	3.14	3.01	0.66
Nitrogen/%	0.16	0.15	0.03
Phosphorus/%	0.04	0.04	0.01
Potassium/%	1.79	1.77	0.21
Bufferfarmland/%	62.81	60.51	25.27
Bufferbuiltup/%	0.00	3.36	5.85
Bufferforest/%	23.97	30.07	27.79

for evaluating land use suitability in 2005 (Table 2).

We also created several measures of distance (in kilometers), defined separately for each pixel in our sample (Table 1 and Table 2). The variable, distance to road (*dtoroad*), measures the distance from each grid cell to the nearest road. Distance to city (*dtocity*) measures the distance (by the shortest road route) from each pixel to certain cities. We generated a variable, distance to the nearest water area (*dtowater*), to predict the availability of irrigation.

Data on soil properties, obtained from the CAS data center, were originally collected by a special nationwide research and documentation project (the Second Round of China's National Soil Survey) that was organized by the State Council and run by a consortium of universities, research institutes, and soil extension centers. We used the data to specify five variables: soil pH value (*pH*), proportion of loam in soil (*loam*, %), organic matter content in top soil (*organics*, %), nitrogen content of soil (*nitrogen*, %), phosphorus content of soil (*phosphorus*, %), and potassium content of soil (*potassium*, %). Using a conventional Kriging algorithm, we interpolated the soil information into surface data to generate more disaggregated information on soil properties over space for each pixel (Table 1 and Table 2).

The buffer variables were developed by measuring the percentages of the areas of cultivated land, built-up area, and forest within a 121 km^2 square centered on the pixel of interest. As with *precipitation* and *temperature*, buffer variables from two periods were used, namely, 1995 to 1999 and 2000 to 2004, to estimate regression parameters and to predict land use suitability, respectively. Summary statistics for the independent variables prepared for MNL and LR are indicated in Table 1, while those for the variables used in evaluating land use suitability in 2005 are indicated in Table 2.

Strong collinearities can lead to the erroneous estimation of parameter values, which further affects the prediction results. The Variance Inflation Factor (VIF) was chosen in this study to identify multicollinearity among the seventeen independent variables. According to the threshold suggested by Chatterjee and Price (1991), VIF = 10 is large enough to indicate the problem of collinearity. The multicollinearity test results showed that the VIFs of *organics*, *nitrogen,* and *pH* are larger than 10, which indicated that there was multicollinearity existing in the originally selected seventeen variables (Table 3, column 2). In addition, we found that the variable *dtowater* was omitted because of collinearity when estimating the parameters in LR and MNL. When we rejected the variables *organics* and *dtowater*, the multicollinearity test results showed that the VIF of each remaining variable is less than 10 (Table 3, column 3). Therefore, we used the remaining fifteen variables to estimate the parameters in LR and MNL in this study.

4 Results

The estimation results of LR presented in Table 4 demonstrate that the suitability of cultivated land was significantly correlated with most of the geographic and

Table 3 Results of the multicollinearity test

Variable	Variance Inflation Factor (VIF)	
	Originally prepared variables	Finally selected variables
Organics	15.42	
Nitrogen	15.20	5.45
pH	10.17	8.75
Potassium	7.31	6.06
Phosphorous	6.11	6.02
Bufferforest	4.23	4.20
Bufferfarmland	3.69	3.67
Dtocity	3.16	3.15
Elevation	2.80	2.75
Precipitation	2.41	2.40
Slope	2.17	2.17
Loam	1.81	1.62
Temperature	1.80	1.76
Dtoroad	1.25	1.25
Bufferbuiltup	1.21	1.21
Aspect	1.08	1.08
Stowater	1.00	
Mean VIF	4.75	3.44

climatic variables (*elevation*, *slope*, *aspect,* and *temperature*), measures of distance (*dtoroad* and *dtocity*), soil properties (*loam* and *phosphorous*), and buffer variables (*bufferfarmland*, *bufferbuiltup*, and *bufferforest*). It is consistent with common sense that the cultivated land was more likely to be distributed in areas with low elevation, southern aspect, high temperature, short distance to road, and high phosphorus content of soil (Table 4, column 2, rows 3, 5, 7, 8, and 13). But it was inexplicable that cultivated land was more inclined to appear in the areas with steeper terrain slope, farther away from city, and less loam in soil (Table 4, column 2, rows 4, 9, and 11). Such inexplicabilities can also be found in the estimates of LR models with dependent variables of land use suitability of forest, grassland, water area, built-up area, and unused land (Table 4, columns 3 to 7).

The estimation results of MNL presented in Table 5 showed that there was little inexplicability when explaining the relationships between land use suitability and its driving factors. The only confusion was that the estimated parameter showed that cultivated land was more likely to be converted to forest area in the flat areas (Table 5, column 2, and row 4). Therefore, MNL was more accurate in explaining the relationships between land use suitability and its driving factors.

To compare the efficiencies of MNL and LR in evaluating land use suitability, we determined the number of correctly allocated pixels. Such a pixel was defined as one with a numerically higher suitability for its actual land use type (than other types) in 2005. For example, with respect to cultivated land, we considered any cultivated land pixel exhibiting a suitability that was highest for cultivated land as a correctly allocated pixel. The similar principle was used for assigning the accuracies of allocations for all other land use types. For instance, if the suitability for built-up area (0.25) was larger than that of any other land use type except for cultivated land, including forest (0.18), grassland (0.11), water area (0.10), and unused land (0.04) for a built-up area pixel, it was considered to be a correctly allocated pixel. The accuracy of land use allocation was the proportion of pixels being allocated correctly, which could also be measured by the number of pixels being allocated correctly in this study.

The statistical analysis indicated that there were 37,813 correctly allocated pixels (out of the total of 44,561) using either MNL or LR (Fig. 2). An additional 3,620 pixels were correctly allocated using MNL but not LR, while the opposite was true for 386 other pixels. Consequently, a total of 41,433 pixels were correctly allocated using MNL, as opposed to only 38,199 pixels using LR. For MNL, the proportion of correctly allocated cultivated land pixels was 97.53% (= 27,531/28,227), which was 13.91% higher than that for LR (Table 6). Meanwhile, the proportions of correctly allocated water area and built-up area pixels were increased by 8.14% and 3.11%, respectively, when using MNL instead of LR.

Table 4 Logistic regression results of land use suitability

	Dependent variable					
	$Land_{2000,1}$	$Land_{2000,2}$	$Land_{2000,3}$	$Land_{2000,4}$	$Land_{2000,5}$	$Land_{2000,6}$
Elevation	-5.79×10^{-4} (7.37)***	-3.09×10^{-4} (4.88)***	-1.29×10^{-3} (13.81)***	-8.32×10^{-3} (21.14)***	-7.06×10^{-4} (2.13)**	-4.68×10^{-4} (0.43)
Slope	0.02 (5.48)***	0.01 (3.42)***	0.08 (11.10)***	−0.06 (3.58)***	0.02 (1.29)	−0.53 (3.64)***
Aspect	1.02×10^{-3} (6.58)***	3.84×10^{-4} (2.37)**	-4.56×10^{-3} (15.79)**	1.17×10^{-3} (5.45)***	6.65×10^{-4} (1.51)	2.32×10^{-3} (3.34)***
Precipitation	-0.81×10^{-4} (1.01)	-3.51×10^{-4} (2.07)**	-4.69×10^{-3} (14.08)***	-2.94×10^{-3} (6.62)***	-1.99×10^{-4} (0.38)	5.50×10^{-3} (3.17)**
Temperature	0.02 (1.87)*	0.04 (3.79)***	0.04 (2.00)**	0.10 (4.50)***	0.03 (0.94)	−0.28 (3.76)***
Dtoroad	−0.16 (12.72)***	−0.03 (2.31)**	−0.54 (27.15)***	−0.14 (9.01)***	−0.14 (3.64)***	0.14 (6.40)***
Dtocity	1.17×10^{-3} (6.89)***	2.37×10^{-4} (1.49)	7.86×10^{-3} (26.08)***	-1.36×10^{-3} (3.68)***	1.99×10^{-4} (0.45)	-0.02×10^{-4} (9.10)***
pH	3.58×10^{-3} (0.09)	0.06 (1.55)	0.68 (18.35)***	0.29 (6.10)***	0.06 (0.45)	0.35 (3.61)***
Loam	−0.01 (4.84)***	-3.29×10^{-3} (1.39)	−0.07 (18.41)***	−0.01 (2.87)***	-1.75×10^{-4} (0.03)	0.02 (1.94)*
Nitrogen	0.58 (0.78)	0.84 (1.07)	7.16 (12.10)***	1.19 (1.27)	1.57 (0.67)	3.76 (1.11)
Phosphorous	8.19 (2.19)**	−7.18 (2.03)**	23.76 (5.84)***	−36.68 (5.85)***	−3.82 (0.28)	59.47 (3.71)***
Potassium	−0.11 (1.32)	0.06 (0.70)	−1.52 (14.90)***	−0.51 (2.56)***	0.19 (0.66)	−1.63 (3.07)***
Bufferfarmland	0.07 (116.70)***	2.29×10^{-3} (3.68)***	−0.06 (78.94)***	−0.06 (76.36)***	0.02 (11.93)***	−0.06 (23.70)***
Bufferbuiltup	−0.01 (6.36)***	3.89×10^{-3} (1.92)*	−0.10 (21.74)*	−0.04 (20.71)***	0.14 (52.71)***	−0.06 (5.38)***
Bufferforest	3.89×10^{-3} (7.31)***	0.08×10^{-3} (134.45)***	0.08 (118.57)***	−0.06 (70.97)***	−0.01 (5.58)***	−0.07 (13.84)***
Intercept	−4.00 (16.49)***	−3.16 (10.96)***	−7.53 (14.67)***	6.15 (10.99)***	−6.30 (7.92)***	−6.15 (3.22)***
Pseudo R^2	0.44	0.54	0.38	0.52	0.39	0.59

Note: $Land_{2000,u}$ is the binary variable representing the existence of land use type u in the respective pixel. *, **, *** indicate significance at 0.1, 0.05, and 0.01 levels respectively.

While the proportions of correctly allocated pixels corresponding to cultivated land, water area, and built-up area were significantly higher when using MNL than LR; the proportions of forest pixels and grassland pixels correctly allocated with MNL are actually lower than those with LR. The number of correctly allocated forest pixels was 11,835 with LR, which was 1.12% (131 pixels) higher than that obtained using MNL. Similarly, the proportion of correctly allocated grassland pixels using MNL was 15.97% lower than with LR.

In consideration of the argument that judgment criteria may have influenced the analysis of evaluation efficiencies, we conducted a paired t-test (Lam and Longnecker, 1983; Hsu and Lachenbruch, 2008) to test the significance of the difference in the suitability of a pixel for different land uses. The significance level of such a t-test demonstrates whether a pixel representing a specific land use type is clearly distinguishable from others. The paired t-test results showed that the suitability of all types of land uses evaluated by MNL were significantly different ($p < 0.01$) from one another (Table 7). On the other hand, the difference in the suitability of a pixel for grassland and water areas, as evaluated by LR, was not significant ($t = 0.10$). This implies that the grassland and water area pixels could not be statistically distinguished when using LR.

5 Discussion and conclusions

It was evident that MNL was more efficient than LR in evaluating land use suitability. This regression method was

Table 5 Multinomial logistic regression results of land use suitability

	Dependent variable: $DCLand_{1995-2000}$ (= 1 is the comparison group)				
	(= 2)	(= 3)	(= 4)	(= 5)	(= 6)
Elevation	-6.42×10^{-5} (0.40)	1.85×10^{-3} $(3.83)^{***}$	-6.06×10^{-3} $(5.67)^{***}$	2.54×10^{-4} (0.42)	4.55×10^{-3} (0.61)
Slope	−0.03 $(3.69)^{***}$	−0.03 (0.94)	−0.07 $(1.48)^{***}$	4.47 (0.14)	−0.71 (1.09)
Aspect	-2.17×10^{-3} $(7.24)^{***}$	-2.88×10^{-3} $(3.36)^{**}$	2.22×10^{-3} $(3.20)^{**}$	1.24×10^{-3} $(1.86)^{*}$	2.33×10^{-3} (0.57)
Precipitation	-1.15×10^{-4} (0.74)	9.59×10^{-5} (0.18)	-7.95×10^{-4} $(1.66)^{*}$	-9.95×10^{-4} $(2.47)^{**}$	1.83×10^{-3} (0.47)
Temperature	−0.08 $(5.03)^{***}$	0.21 $(3.94)^{***}$	0.12 $(3.32)^{***}$	0.02 (0.73)	−0.56 $(3.37)^{***}$
Dtoroad	−0.01 (0.58)	−0.10 (1.34)	0.36 $(8.25)^{***}$	−0.03 (0.51)	0.16 (0.63)
Dtocity	4.87×10^{-3} (1.49)	2.61×10^{-4} (0.25)	4.53×10^{-3} $(4.95)^{***}$	-1.77×10^{-3} $(2.27)^{**}$	-8.24×10^{-3} (1.15)
pH	−0.04 (0.40)	0.45 $(2.15)^{**}$	−0.03 (0.15)	0.09 (0.39)	−0.97 (1.57)
Loam	-6.50×10^{-3} (1.45)	−0.05 $(4.24)^{***}$	0.01 (1.34)	4.04×10^{-4} (0.05)	0.01 (0.28)
Nitrogen	−1.20 (0.64)	10.69 $(3.08)^{***}$	1.07 (0.30)	1.88 (0.45)	−16.67 (1.13)
Phosphorous	1.52 (0.18)	6.15 (0.28)	13.96 (0.61)	0.90 (0.04)	−22.95 (0.29)
Potassium	0.16 (0.81)	−1.29 $(2.64)^{***}$	−0.44 (0.77)	−0.31 (0.61)	4.60 (1.51)
Bufferfarmland	-6.21×10^{-3} $(2.96)^{***}$	−0.09 $(35.06)^{***}$	−0.06 $(28.80)^{***}$	-2.26×10^{-3} (0.64)	−0.09 $(8.55)^{***}$
Bufferbuiltup	−0.03 $(4.88)^{***}$	−0.11 $(8.96)^{***}$	−0.05 $(7.32)^{***}$	0.10 $(22.40)^{***}$	−0.17 $(2.20)^{**}$
Bufferforest	0.04 $(22.26)^{***}$	−0.08 $(33.81)^{***}$	−0.05 $(19.99)^{***}$	−0.01 $(2.52)^{**}$	−0.14 $(3.68)^{***}$
Intercept	−1.14 $(2.18)^{**}$	−0.75 (0.48)	1.29 (1.12)	−2.70 $(2.36)^{**}$	6.43 $(0.85)^{***}$
Pseudo R^2			0.23		

Note: $DCLand_{1995-2000}$ is the polytomous nominal variable representing the conversion from cultivated land to other land use types. *, **, *** indicate significance at 0.1, 0.05, and 0.01 levels respectively.

employed by theoretically considering that the actual process of land use change was determined not only by its driving factors but also by its original land use type. The MNL empirically described the direct statistical relationships between the driving factors and subsequent land use change, thereby demonstrating superior accuracy over LR in modeling land use change. The case study in the Jiangxi Province indicated that 41,433 of the 44,561 pixels under scrutiny were correctly allocated using MNL, an improvement of 8.47% over the LR method. Based on the paired t-test results, MNL was also more efficient than LR in evaluating land use suitability.

Interestingly, the improvements in evaluation efficiencies obtained using MNL were true only for cultivated land, water area, and built-up area. One reasonable conjecture is that the land use change process in the Jiangxi Province varied between the periods 1995–2000 and 2000–2005. The Grain for Green Program, which is one of the six largest ecological forest programs in China, and which has been implemented in 25 provinces, played a key role in prompting this variability (Wang et al., 2012). The stated aim of this program is to convert tracts of cultivated land, exhibiting serious soil erosion, desertification, or salinization, to grassland/forest in accordance with local climatic and geological conditions, and to restore natural vegetation in a planned, gradual manner. Terrain slope is one of the criteria used in determining if a pixel is involved in the Grain for Green Program (Feng et al.,

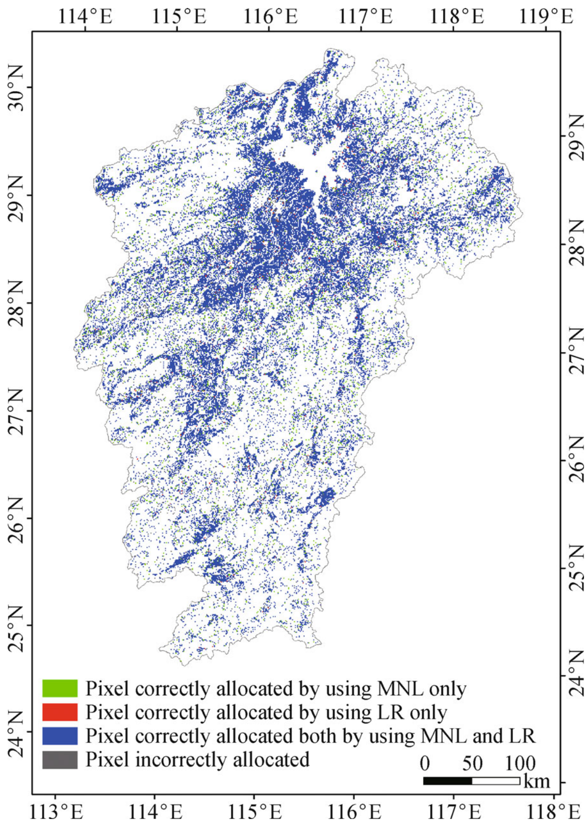

Fig. 2 Land use allocation results of multinomial logistic regression (MNL) and logistic regression (LR) in the Jiangxi Province in 2005.

2005). A cultivated land pixel with a steep slope is more likely to be a candidate for conversion to a forest or grassland pixel. Our statistics also showed that the distributions of the terrain slopes in pixels converted from cultivated land to either forest or grassland during the period 2000–2005 were more right-skewed (skewness values of 2.09 and 3.25, respectively) compared with their counterparts during the period 1995–2000 (skewness values of 2.03 and 2.20, respectively). Theoretically, variability in the land use change process will reduce the efficiencies of evaluating the suitability of both forest and grassland using MNL.

The process of using MNL to allocate land uses based on land use suitability was established in this study. However, to highlight the issues in evaluating the relationships between land use change and its driving factors, and the spatial dynamics of regional land use modeling, we focused our discussion, not on the land use suitability evaluation process, but on land use allocation. We surmised that the improvements in understanding the land use change process were more important than a mere analysis of land use change modeling. We did not drop the independent variables that were individually insignificant and inexplicable in MNL and LR in order to make the comparison analysis as fair as possible. The evaluation accuracy of land use suitability using MNL would likely have been improved if the independent variables had been selected more strictly.

Table 6 Number and proportion of selected pixels being allocated correctly in 2005

Land use type	Remote sensing	Pixels being allocated correctly			
		Logistic regression		Multinomial logistic regression	
		Number	Proportion/%	Number	Proportion/%
Cultivated land	28,227	24,169	85.62	27,531	97.53
Forest	12,151	11,835	97.40	11,704	96.32
Grassland	1,117	501	44.85	421	37.69
Water area	1,171	602	51.41	651	55.59
Built-up area	1,862	1,092	58.65	1,126	60.47
Unused land	33	0	0.00	0	0.00
Total		38,199	85.72	41,433	92.98

Table 7 Paired *t*-test results of land use suitability evaluated by multinomial logistic regression (MNL) and logistic regression (LR)

	(1)	(2)	(3)	(4)	(5)	(6)	MNL
(1)	—	413.89	865.64	898.11	924.46	993.63	(1)
(2)	129.09	—	151.89	135.92	132.59	170.00	(2)
(3)	404.25	161.02	—	–40.27	–35.04	56.92	(3)
(4)	418.51	160.14	0.10	—	–3.62	115.49	(4)
(5)	429.33	150.24	–14.93	–17.68	—	111.64	(5)
(6)	444.60	186.11	84.43	120.29	97.17	—	(6)
LR	(1)	(2)	(3)	(4)	(5)	(6)	

Note: (1) stands for cultivated land; (2) stands for forest; (3) stands for grassland; (4) stands for water area; (5) stands for built-up area; (6) stands for unused land.

Acknowledgements This research was financially supported by the National Basic Research of China (2010CB950900) and the National Natural Science Foundation of China (Grant Nos. 71225005 and 41071343). Two anonymous reviewers are sincerely acknowledged for their valuable comments which have significantly improved the manuscript.

References

Alabi M O (2011). Analytical approach to examining drivers of residential land use development in Lokoja, Nigeria. British Journal of Educational Research, 1(2): 144–152

Bahadur K C K (2011). Linking physical, economic and institutional constraints of land use change and forest conservation in the hills of Nepal. For Policy Econ, 13(8): 603–613

Braimoh A K, Onishi T (2007). Spatial determinants of urban land use change in Lagos, Nigeria. Land Use Policy, 24(2): 502–515

Briz T, Ward R W (2009). Consumer awareness of organic products in Spain: an application of multinomial logit models. Food Policy, 34 (3): 295–304

Cao K, Ye X Y (2013). Coarse-grained parallel genetic algorithm applied to a vector based land use allocation optimization problem: the case study of Tongzhou Newtown, Beijing, China. Stochastic Environ Res Risk Assess, 27(5): 1133–1142

Chatterjee S, Price B (1991). Regression Analysis by Example. New York: John Wiley & Sons, 85–120

Chen B, Chen G Q, Yang Z F (2006). Exergy-based resource accounting for China. Ecol Modell, 196(3–4): 313–328

Chen Y Q, Verburg P H (2000). Modeling land use change and its effects by GIS. Ecologic Sci, 19(3): 1–7

Choi S W, Sohngen B, Alig R (2011). An assessment of the influence of bioenergy and marketed land amenity values on land uses in the Midwestern US. Ecol Econ, 70(4): 713–720

Claessens L, Schoorl J M, Verburg P H, Geraedts L, Veldkamp A (2009). Modelling interactions and feedback mechanisms between land use change and landscape processes. Agric Ecosyst Environ, 129(1–3): 157–170

Dakin H A, Devlin N J, Odeyemi I A O (2006). “Yes”, “No” or “Yes, but”? Multinomial modelling of NICE decision-making. Health Policy, 77(3): 352–367

dell’Olio L, Ibeas A, Cecin P (2011). The quality of service desired by public transport users. Transp Policy, 18(1): 217–227

Deng X Z, Huang J K, Rozells S, Uchida E (2008). Growth, population and industrialization, and urban land expansion of China. J Urban Econ, 63(1): 96–115

Duan Z Q, Verburg P H, Zhang F R, Yu Z R (2004). Construction of a land-use change simulation model and its application in Haidian District, Beijing. Acta Geogr Sin, 59(6): 1037–1047

Feng Z M, Yang Y Z, Zhang Y Q, Zhang P T, Li Y Q (2005). Grain-for-green policy and its impacts on grain supply in West China. Land Use Policy, 22(4): 301–312

Gellrich M, Baur P, Koch B, Zimmermann N E (2007). Agricultural land abandonment and natural forest re-growth in the Swiss mountains: a spatially explicit economic analysis. Agric Ecosyst Environ, 118(1–4): 93–108

Geoghegan J, Villar S C, Klepeis P, Mendoza P M, Ogneva-Himmelberger Y, Chowdhury R R, Turner B L II, Vance C (2001). Modeling tropical deforestation in the southern Yucatán peninsular region: comparing survey and satellite data. Agric Ecosyst Environ, 85(1–3): 25–46

Han J G, Zhang Y J, Wang C J, Bai W M, Wang Y R, Han G D, Li L H (2008). Rangeland degradation and restoration management in China. Rangeland J, 30(2): 233–239

Hosmer D, Lemeshow S (2000). Applied Logistic Regression. New York: John Wiley & Sons, 31–46

Hsu H, Lachenbruch P A (2008). Paired *t* Test. Wiley Encyclopedia of Clinical Trials, 1–3

Jiang Y, Liu J, Cui Q, An X H, Wu C X (2011). Land use/land cover change and driving force analysis in Xishuangbanna Region in 1986–2008. Frontiers of Earth Science, 5(3): 288–293

Kalnay E, Cai M (2003). Impact of urbanization and land-use change on climate. Nature, 423(6939): 528–531

Lam F C, Longnecker M T (1983). A modified Wilcoxon rank sum test for paired data. Biometrika, 70(2): 510–513

Lin Y P, Chu H J, Wu C F, Verburg P H (2011). Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling: a case study. Int J Geogr Inf Sci, 25(1): 65–87

Liu J Y, Zhang Z X, Zhuang D F, Wang Y M, Zhou W C, Zhang S W, Li R D, Jiang N, Wu S X (2003). A study on the spatial-temporal dynamic changes of land-use and driving forces analyses of China in the 1990s. Geographical Research, 22(1): 1–12 (in Chinese)

Luce R D (1959). Individual Choice Behavior: A Theoretical Analysis. New York: John Wiley & Sons, 139–141

McFadden D (1974). Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. Frontiers in Econometrics. New York: Academic Press, 105–142

Meiyappan P, Jain A K (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. Frontiers of Earth Science, 6(2): 122–139

Millington J D A, Perry G L W, Romero-Calcerrada R (2007). Regression techniques for examining land use/cover change: a case study of a Mediterranean landscape. Ecosystems (N. Y.), 10(4): 562–578

Nahuelhual L, Carmona A, Lara A, Echeverría C, González M E (2012). Land-cover change to forest plantations: proximate causes and implications for the landscape in south-central Chile. Landsc Urban Plan, 107(1): 12–20

Ostwald M, Chen D L (2006). Land-use change: impacts of climate variations and policies among small-scale farmers in the Loess Plateau, China. Land Use Policy, 23(4): 361–371

Pielke R A Sr (2005). Atmospheric science. Land use and climate change. Science, 310(5754): 1625–1626

Pueyo Y, Beguería S (2007). Modelling the rate of secondary succession after farmland abandonment in a Mediterranean mountain area. Landsc Urban Plan, 83(4): 245–254

Rozelle S, Huang J K, Zhang L X (1997). Poverty, population and environmental degradation in China. Food Policy, 22(3): 229–251

Schaldach R, Alcamo J (2006). Coupled simulation of regional land use change and soil carbon sequestration: a case study for the state of Hesse in Germany. Environ Model Softw, 21(10): 1430–1446

Serneels S, Lambin E F (2001). Proximate causes of land use change in

Narok district Kenya: a spatial statistical model. Agric Ecosyst Environ, 85(1–3): 65–81

Sohl T L, Sleeter B M, Zhu Z L, Sayler K L, Bennett S, Bouchard M, Reker R, Hawbaker T, Wein A, Liu S G, Kanengieter R, Acevedo W (2012). A land-use and land-cover modeling strategy to support a national assessment of carbon stocks and fluxes. Appl Geogr, 34: 111–124

van Doorn A M, Bakker M M (2007). The destination of arable land in a marginal agricultural landscape in South Portugal: an exploration of land use change determinants. Landscape Ecol, 22(7): 1073–1087

Verburg P H, Overmars K P (2009). Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. Landscape Ecol, 24(9): 1167–1181

Verburg P H, Soepboer W, Veldkamp A, Limpiada R, Espaldon V, Mastura S S (2002). Modeling the spatial dynamics of regional land use: the CLUE-S model. Environ Manage, 30(3): 391–405

Verburg P H, Veldkamp A (2004). Projecting land use transitions at forest fringes in the Philippines at two spatial scales. Landscape Ecol, 19(1): 77–98

Walsh S J, Messina J P, Mena C F, Malanson G P, Page P H (2008). Complexity theory, spatial simulation models, and land use dynamics in the Northern Ecuadorian Amazon. Geoforum, 39(2): 867–878

Wang J, Chen Y Q, Shao X M, Zhang Y Y, Cao Y G (2012). Land-use changes and policy dimension driving forces in China: present, trend and future. Land Use Policy, 29(4): 737–749

Williams N S G (2007). Environmental, landscape and social predictors of native grassland loss in western Victoria, Australia. Biol Conserv, 137(2): 308–318

Wu G P, Zeng Y N, Feng X Z, Xiao P F, Wang K (2010). Dynamic simulation of land use change based on the improved CLUE-S model: a case study of Yongding County, Zhangjiajie. Geographical Research, 29(3): 460–470 (in Chinese)

Zhan J Y, Shi N N, He S J, Lin Y Z (2010). Factors and mechanism driving the land-use conversion in Jiangxi Province. J Geogr Sci, 20 (4): 525–539

Zhong T Y, Huang X J, Zhang X Y, Wang K (2011). Temporal and spatial variability of agricultural land loss in relation to policy and accessibility in a low hilly region of southeast China. Land Use Policy, 28(4): 762–769