# MODELING SPATIAL LAND USE PATTERN USING AUTOLOGISTIC REGRESSION

Y. N. Zeng[a] *, G. P. Wu[a], F. B. Zhan[b] , H. H. Zhang[a]

[a] School of Info-Physics and Geomatics Engineering, Central South University, Changsha 410083, China – ynzeng@mail.csu.edu.cn
[b] Texas Center for Geographic Information Science, Texas State University, San Marcos, TX 78666, USA – zhan@txstate.edu

**WG II/1**

**KEY WORDS:** Land use pattern, Spatial autocorrelation, Autologistic regression, Modelling, GIS

**ABSTRACT:**

The significance of land cover as an environmental variable has made land use change an important subject in global environmental change and sustainable development. Modeling land use change has attracted considerable attention. Currently, empirical estimation models using statistical techniques are one of mostly used spatial models to simulate land use pattern and its changes. Empirical estimation methods can model the relationships between land use changes and the drivers. However, existing logistical regression models often ignore the spatial autocorrelation among land use data, which affect the goodness of fitting and accuracy of fitting of land use modeling. In this study we incorporate components describing the spatial autocorrelation into existing logistical regression and form an autologstic regression. Taking the Yongding County, Hunan province, China as study area, we simulate spatial pattern of different land use types using autologistic regression and compare with the existing logistical regression method. The results indicate that autologistic regression has better goodness of fitting and higher accuracy of fitting than the existing logistical regression method. Autologistic regression can improve the modeling result in some degree reasonably.

## 1. INTRODUCTION

The significance of land cover as an environmental variable has made land use change an important subject in global environmental change and sustainable development (Li, 1996; Vitousek, et al., 1997; Veldkamp, et al., 1997). Modeling land use change can formalize knowledge about land use change and understand the determinants of land use change. Therefore, land use modeling has attracted considerable attention (Gobim, et al., 2002; Lambin, 1997; Serneels, et al., 2001; Veldkamp, et al. 1996, 1997; Verburg, et al., 2002; Wu, F., et al., 1997). Currently, there are three kinds of spatial land use change models: empirical estimation models, dynamic simulation models and rule-base simulation models (He, et al., 2007). Most dynamic simulation cannot incorporate enough socioeconomic variables. As a rule-base simulation method, CA models can simulate spatial pattern but cannot interpret spatio-temporal processes of land use change and is more complicated to construct. However, empirical estimation methods using statistical techniques can model the relationships between land use changes and the drivers. As a result, the knowledge of the processes driving the change of spatial patterns can be obtained through the interpretation of the statistical models. As an empirical estimation method, logistical regression has been used deforestation analysis, agriculture, urban growth and farmland modeling. In many cases, logistical regression models fit spatial processes and land use change outcome reasonably well. However, there are many issues in modeling land use distribution using logistical regression models. Existing logistical regression models often ignore the spatial autocorrelation among land use data, which affect the goodness of fitting and accuracy of fitting of land use modeling. It is therefore essential to account for spatial correlation in land use change models(.Augustin, et al. 1996; Laurent, et al., 1993; Legendre, et al., 1998; Pontius, et al. 2001; Wu, et al., 1997).

In this study an approach to land use distribution modeling using autologistic regression is discussed. We incorporate components describing the spatial autocorrelation into existing logistical regression and form an autologstic regression model. Taking the Yongding County, Hunan province, China as study area, we simulate spatial distribution of different land use types using autologistic regression model and compare with the existing logistical regression method.

## 2. STUDY AREA

The Yongding County was chosen as the study area. The study area covers approximately 2174 km$^2$ located between 110°04´- 110°55´E and 28°52´- 29°25´N. It is a center for economics, culture, and transportation in northwestern part of Hunan province, China. As the economic development polices has implemented in this area, Yongding County is increasing in population and built-up areas. The areal expansion is through encroachment into the adjacent agricultural and rural regions. Rapid urbanization and accelerated urban sprawl converted the natural landscape to man-made landscape. Landscape pattern changes have significant impacts on environment and human life and are of great interest for diverse purposes such as urban planning, water and land resource management, etc. As a result, modeling land use pattern is significant to study

* Correspond author

regional environment as well as to control and plan future land use development.

## 3. METHODS

### 3.1 Data sources

Land use map of the study area for 2005 was obtained from Land Resources Agencies in Yongding County. Population data was derived from statistic year book of Yongding County in 2005. DEM data with 90m resolution was obtained from NGA. Three dependent variables for this study, arable land, woodland, and built-up land were extracted from the map within geographical information system. Eight spatially explicit independent variables hypothesized to affect the dependent variables were also developed within a GIS. These include Euclidean distances from town cores, waterbodies, and major roads, population density, elevation, slope, aspect, and curvature. All variables were mapped at a resolution of 120m and produced within ArcGIS9.0 software. A total of 3500 points were randomly selected, out of which 2500 points were used for modeling and the remaining 1000 points for validation.

### 3.2 AutoLogstic regression model

An AutoLogstic regression model was developed by combining logistic regression model with autocorrelation effects. The model form is:

$$p\left(y_{ij} = 1 \middle| x_j, w_j\right) = \frac{e^{\alpha+\beta x_j + rw}}{1+e^{\alpha+\beta x_j + rw}} \tag{1}$$

and

$$\log itP(Y_{ij} = 1 | x_j, w_j) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta x_j + rw \tag{2}$$

Where $P(Yij=1|xj, w)$ is the expected value of the dependent variable $Yij$ (so that $Yij$ = 1 if cell $j$ belongs to land use type $i$ and $Yij$ = 0, otherwise)
$x_j$ is the vector of covariates (Table 1)
$w$ is autocorrelation weight of land use type $i$
$\alpha$ , $\beta$ and $r$ are the model parameters

The autocorrelation weight of land use type $i$ produced within Geoda software using following equation:

$$w = \begin{cases} 1/D \\ 0 \end{cases} \tag{3}$$

Where D is the distance one spatial point to another
While D less than given threshold, w =1/D
and w=0, otherwise.

While coefficient $r$ equates to zero, the equation (1) and (2) should be classic logistic regression models. In this study we use classic logistic regression and autologistic regression models to simulate land use pattern. Similar logistic and autologstic regressions were performed for two land use classes with different explanatory variables and autocorrelation weight of each land use type with SPSS13.0 software.

| Independent Variable | Meaning | Nature of variable |
|---|---|---|
| $X_1$ | Distances to the nearest town | continuous |
| $X_2$ | Distances to the nearest river | continuous |
| $X_3$ | Distances to the nearest major road | continuous |
| $X_4$ | Population density | continuous |
| $X_5$ | elevation | continuous |
| $X_6$ | slope | continuous |
| $X_7$ | aspect | continuous |
| $X_8$ | curvature | continuous |

Table 1. List of independent variable in regression model

### 3.3 Model validation

Model validation was carried out in three fold. The first method uses the Relative Operating Characteristic (ROC). The ROC compares binary data over the whole range of predicted probabilities. It aggregates into a single index of agreement, the ability of the model to predict the probability of arable and woodland distribution at various locations on the landscape. That is, the ROC is a measure of the ability of the model to correctly specify location. The second method uses the crosstab table to estimate simulation accuracy. The third method uses the Moran I indicator to compare the results of different models.

## 4. RESULTS AND DISCUSSION

The results of autoLogistic regression and classic Logistic regression for different land use are summarized in Table 2 and 3.

| Variable | arable land | | woodland | |
|---|---|---|---|---|
| | Beta | Exp(B) | Beta | Exp(B) |
| constant | 0.374 | 1.454 | -1.836 | 0.160 |
| $X_1$ | – | – | – | – |
| $X_2$ | – | – | – | – |
| $X_3$ | – | – | – | – |
| $X_4$ | – | – | 0.001 | 1.001 |
| $X_5$ | -0.002 | 0.998 | 0.003 | 1.003 |
| $X_6$ | -0.059 | 0.942 | 0.066 | 1.068 |
| $X_7$ | -0.001 | 0.999 | 0.001 | 1.001 |
| $X_8$ | 0.063 | 1.065 | -0.071 | 0.932 |
| ROC | 0.851 | | 0.913 | |

Table 2. The results of logistic regression

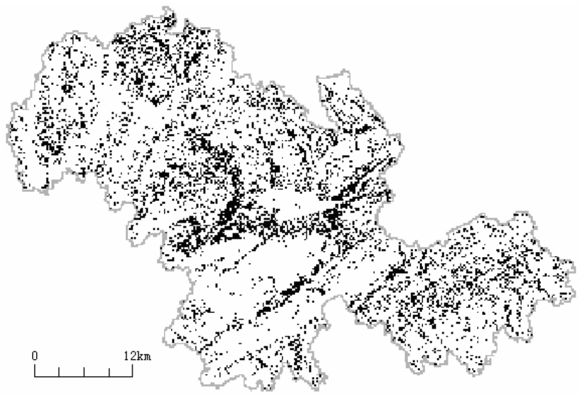
| Variable | arable land | | woodland | |
|---|---|---|---|---|
| | Beta | Exp（B） | Beta | Exp（B） |
| constant | 0.373 | 1.452 | -1.830 | 0.160 |
| $X_1$ | – | – | – | – |
| $X_2$ | – | – | – | – |

| | | | | |
|---|---|---|---|---|
| $X_3$ | – | – | – | – |
| $X_4$ | – | – | 0.001 | 1.001 |
| $X_5$ | -0.002 | 0.998 | 0.003 | 1.003 |
| $X_6$ | -0.059 | 0.942 | 0.066 | 1.068 |
| $X_7$ | -0.001 | 0.999 | 0.001 | 1.001 |
| $X_8$ | 0.062 | 1.064 | -0.071 | 0.931 |
| AutoValue | 0.001 | 1.001 | 0.002 | 1.002 |
| ROC | 0.893 | | 0.940 | |

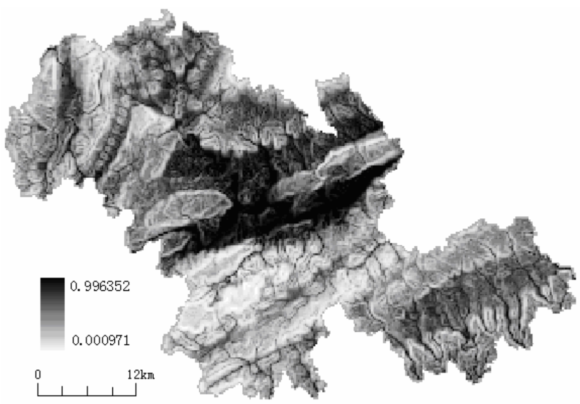
Table 3. The results of autologistic regression

### 4.1 Arable land

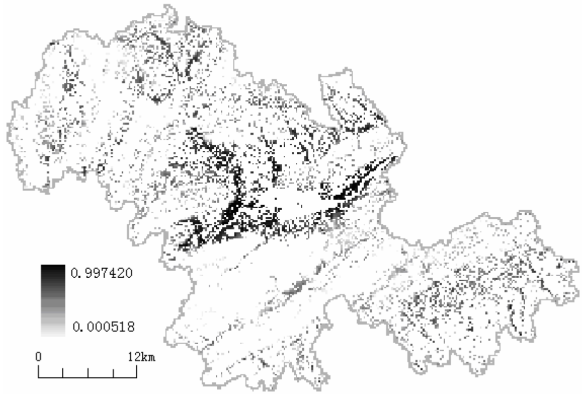
The elevation, slope, aspect and curvature are three variables that are thought to have a significant role in explaining the presence of arable land both in table 1 and 2. This indicates that both classic logistic regression and AutoLogistic regression models can explain relationship between the spatial pattern of arable land and drivers. Otherwise, AutoLogistic regression model have better goodness of fitting and higher accuracy of fitting. The Relative Operating Characteristic (ROC) of arable land is 0.85 and 0.895 for classic logistic regression and AutoLogistic regression models, respectively. The AutoLogistic regression models have higher modeling accuracy than classic logistic regression model. The overall accuracy of modeling is improved from 73% to 90.4%. The modeling result based on AutoLogistic regression models have closer Moran I indicate than that of classic logistic regression model. Therefore, the AutoLogistic regression model is reasonable to some extent for simulate the spatial patter of arable land (figure 1).




(a)




(b)

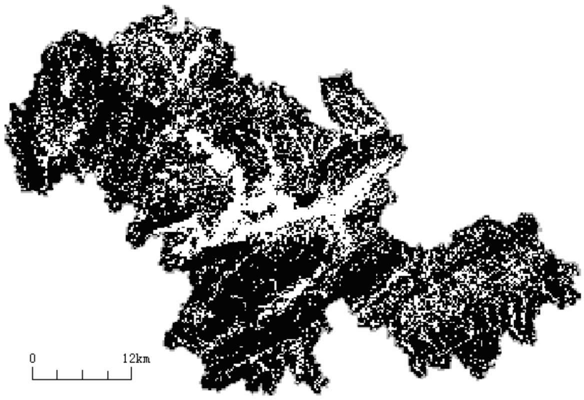



(c)

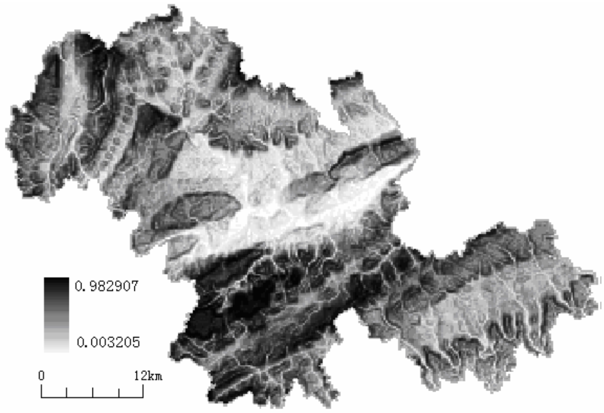
Figure 1. Arable land patterns (a), Logistic regression modelling result (b), and Autologistic regression modelling result

### 4.2 Woodland

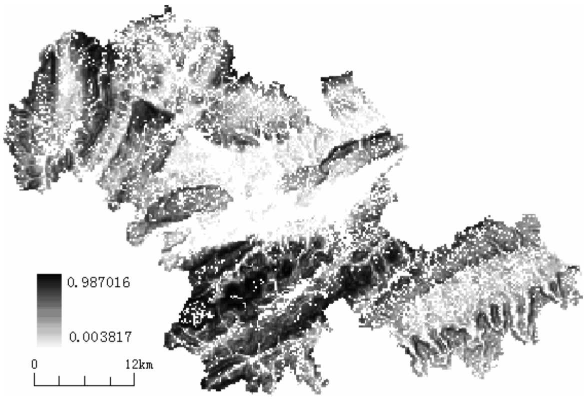
The physical variables and anthropogenic influence have a significant role in explaining the presence of woodland both in table 1 and 2. The elevation, slope, aspect, curvature, and population density can explain relationship between the spatial pattern of woodland and drivers both in classic logistic regression and AutoLogistic regression models. Otherwise, AutoLogistic regression model also have better goodness of fitting and higher accuracy of fitting. The ROC of woodland is 0.913 and 0.940 for classic logistic regression and AutoLogistic regression models, respectively. The AutoLogistic regression models have higher modeling accuracy than classic logistic regression model. The overall accuracy of modeling is improved from 70.3% to 87%. The modeling result based on AutoLogistic regression models have closer Moran I indicate than that of classic logistic regression model. Therefore, the AutoLogistic regression model is also reasonable to some extent for simulate the spatial patter of woodland (figure 2).




(a)




(b)




(c)

Figure 2. Woodland patterns (a), Logistic regression modelling result (b), and Autologistic regression modelling result

## 5. CONCLUSION

Both classic logistic regression and AutoLogistic regression models can explain relationship between the land use spatial pattern and its drivers. The elevation, slope, aspect and curvature are three variables that are thought to have a significant role in explaining the presence of arable land. The elevation, slope, aspect, curvature, and population density can explain relationship between the spatial pattern of woodland and drivers. Otherwise, classic logistic regression model have no ability to describe the autocorrelation in land use variables.

AutoLogistic regression models incorporate components describing the spatial autocorrelation into classic logistic regression model and result in better goodness of fitting and higher accuracy of fitting in land use modeling.

## REFERENCES

Augustin, N., Mugglestone, M. and Buckland, S., 1996. An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, 33, pp. 339-347.

Gobim, A, Campling, P, Feyen, J. 2002. Logistic modeling to derive agricultural land determinants: a case study from southeastern Nigeria. *Agriculture, Ecosystem and Environment*, 89, pp. 213-228.

He, Z., Lo, C., 2007. Modeling urban growth in Atlanta using logistic regression. Computers, Environment and Urban Systems, 31(6), 667-688

Laurent, A., Golmard, J., Mallet, A., 1993. An autologistic model for the genetic analysis of familial binary data. *Am.J.Hum.Genet*, 53, pp. 894-907.

Legendre. P., Legendre, L., 1998. Numerical Ecology. Developments in environmental modelling.20. Amsterdam: Elsevier.

Lambin, E,Baulies, X.,Bockstael, N. et al. 1999. Land-use and land-cover change(LUCC) implementation strategy. IGBP Report No.48 and HDP Report No.10.Stockholm: IGBP.

Lambin, E., 1997. Modeling and monitoring land-cover change processes in tropical regions. *Progress in Physical Geography*, 21(5), pp.375-393.

Li, X., 1996. A review of the international researches on land use/land cover changes. *Acta Geographica Sinica*, 51(5), pp.553-558.

Pontius. G., Cornell, J., Hall, C. 2001. Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica. *Agric. Ecosyst. Environ*, 85, pp. 191-203.

Serneels, S., Lambin, E., 2001. Proximate causes of land use change in Narok district Kenya: a spatial statistical model. *Agriculture, Ecosystem and Environment*, 85, pp. 65-82.

Veldkamp, A., Lambin, E., 2001. Predicting land-use change. *Agriculture, Ecosystems and Environment*, 85, pp. 1-6.

Veldkamp, A., Fresco, L., 1997. Exploring land use scenarios, an alternative approach based on actual land use. *Agriculture Systems*, 55, pp. 1-17.

Veldkamp, A., Fresco, L., 1996. CLUE: a conceptual model to study the conversion of land use and its effects. *Ecological Modelling*, 85, pp. 253-270.

Verburg, A., Soepboer, P., Veldkamp, W., et al.2002. Modeling the spatial dynamics of regional land use: the CLUE-S Model. *Environmental Management*, 30(3), pp. 391-405.

Vitousek, P.M., Mooney, H.A., Lubchenco, J., Melillo, J.M., 1997. Human domination of Earth's ecosystems. *Science*, 277, pp. 494–499.

Wu, F., Yeh, A.G., 1997. Changing spatial distribution and determinants of land development in Chinese cities in the transition from a centrally Planned Economy to a Socialist Market Economy: a case study of Guangzhou. *Urban Studies*, 34 (11), pp. 1851-1879.

Wu, H., Huffer, W., 1997. Modelling the distribution of plant species using the autologistic regression model. *Environmental and Ecological Statistics*, 4, pp. 49-64.

## ACKNOWLEDGMENT

The research is supported by the National Science Foundation of China (NO. 40771198).