

# 基于 logistic 回归模型的农牧交错区 土地利用变化驱动力分析

## ——以内蒙古翁牛特旗为例

谢花林<sup>1,2</sup>, 李 波<sup>2,\*</sup>

(1 江西财经大学资源与环境管理学院, 江西 南昌 330032;

2 北京师范大学资源学院地表过程与资源生态国家重点实验室, 北京 100875)

**摘要:** 本文以农牧交错带的典型区域——内蒙古翁牛特旗为例, 考虑土地利用变化过程的空间变量, 建立了不同土地利用变化过程的 logistic 回归模型。结果表明: 模型中转为耕地的主要解释变量是到农村居民点的距离和农业气候区; 转为草地的主要解释变量是到农村居民点的距离、土壤表层有机质含量和到乡镇中心的距离; 转为林地的主要解释变量是到农村居民点的距离和海拔; 空间异质性和土地利用变化过程的时间变量共同影响着使用 logistic 回归模型来解释土地利用变化驱动力的能力; 通过对草地 logistic 回归模型的检验, 得出空间统计模型能较好地揭示不同土地利用变化过程的主要驱动力及其作用机理。

**关键词:** logistic 回归分析; 土地利用变化; 农牧交错区

文章编号: 1000-0585(2008)02-0294-11

## 1 引言

土地利用和土地覆盖变化 (LUCC) 研究由于其在全球气候变化、食物安全、土壤退化和生物多样性的重要作用, 受到国内外众多学者的关注<sup>[1~7]</sup>。我们通常采用不同的模拟方法来研究土地利用和土地覆盖变化。其中, 土地利用变化的空间显式模型是定量描述变化过程和检验我们理解这些过程的重要技术。土地利用变化模拟的目的是预测变化的空间格局。因此, 它需要解决两个问题: 一是土地利用变化将在什么地方发生? 二是土地覆盖变化的速率? 这两个问题被称为“点”和“数”问题<sup>[8,9]</sup>。土地利用变化的空间预测仅仅需要理解变化的可能原因。要想反映将来土地利用变化的速率则是一项相当艰巨的任务, 因为它需要很好地理解变化的潜在驱动力<sup>[9]</sup>。有时从空间和时间上观测变化的驱动力是细微的, 并且宏观经济变革和政策变化通常很难预料。

许多研究通过选用代表控制格局与过程相互作用的自然和人文景观变量来建立多元模型, 这些研究结果较好地模拟了土地利用变化的可能格局<sup>[10,11]</sup>。这些空间统计模型通过使用理论上驱动变化速率的外在变量来显式地辨识土地覆盖变化的可能原因, 大家普遍使

收稿日期: 2006-12-15; 修订日期: 2007-07-11

基金项目: 国家自然科学基金重点项目 (40435014) 和教育部人文社会科学项目 (05JD790117) 联合资助。

作者简介: 谢花林 (1979-), 男, 江西莲花县人, 博士。主要从事土地可持续利用和生态系统管理研究。

E-mail: xiehl\_2000@163.com。

\* 通讯作者: 李波, E-mail: libo@ires.cn

用的多元线性回归方法就是这一目的。运用统计模型模拟土地利用变化的难点在于它不能处理土地利用变化过程中的空间变量。大多数地区在土地利用/土地覆盖类型变化的驱动力(生物、自然和社会经济)或制度(政策)上有高度的地理变异性。这种空间异质性导致土地利用变化原因和过程的变异性,这样对于不同土地利用变化过程模型就有不同的参数。这就提出了需要对空间实体或区域进行最佳定义的问题,因为需要对一个给定的土地利用变化统计模型进行检验。一方面在大区域验证的单个模拟模型由于土地利用变化过程多样性的空间整合,可能会削弱这个模型的预测能力;另一方面,把一个区域分成许多的同质性实体,并且对不同的实体厘定不同的模型,这样却和寻求模拟行为的一般性不一致<sup>[12]</sup>。本研究通过对研究区分割成土地利用变化过程同质的不同区域,这一问题的解决将在本文中得以体现。

本研究的主要目的是关注土地利用变化过程的理解程度,这可以通过土地利用变化空间关系的描述模型来获得。土地利用变化的空间统计学模型能否预测变化的空间格局呢?一些学者认为它基于代表这一系统配置格局的观测能否推断系统动态过程的信息<sup>[13]</sup>。例如,土地利用变化某些类别倾向于分割景观(如小农场的扩展,小尺度的伐木等),另外一些土地利用变化却增加景观的同质性(如机械化耕作,大区域的农场经营等)。一些学者研究表明空间格局和农作系统的一些重要特征具有很好的相关性。Geoghegan 等<sup>[15]</sup>讨论了两种模拟方法去辨识埋藏在直接与社会科学和新主体相关的空间图像信息,并且用它来获悉与这些主体相关的概念和理论。换句话说,这些大量的文献表明从土地利用变化格局中,人们能够更好地理解土地利用变化是怎样发生以及为什么是这样发生的。而这些格局可以通过遥感来观测,并且能够发现观测到的土地覆盖变化和可能原因之间的空间相关性。

本研究的目标一是探讨如何建立一个空间上的 logistic 回归模型去发现农牧交错带土地利用变化不同过程的可能原因,这一模型考虑了土地利用变化过程空间变异性;二是通过 logistic 回归的原理,探讨这样的一个空间上的统计分析识别能在多大程度上理解土地利用变化的驱动力。

## 2 研究区概况

本研究区为农牧交错带的典型区域——内蒙古自治区赤峰市翁牛特旗,位于  $117^{\circ}49'48'' \sim 120^{\circ}45'58''\text{E}$ ,  $42^{\circ}27'37'' \sim 43^{\circ}25'31''\text{N}$ , 面积  $11882\text{km}^2$ 。该地区位于大兴安岭隆起带东侧向辽河沉降带的过渡地段,大致可分为西部中山台地、中部低山丘陵和东部沙地坳甸三个类型区。平均海拔  $600 \sim 800\text{m}$ , 最高为  $2000\text{m}$ 。年平均气温  $4.4 \sim 7.7^{\circ}\text{C}$ , 年降雨量  $310 \sim 470\text{mm}$ ,  $\geq 10^{\circ}\text{C}$  活动积温大多在  $2200 \sim 3250^{\circ}\text{C}$  之间,西部高寒山区少于  $1500^{\circ}\text{C}$ , 年日照时数为  $2650 \sim 3090\text{h}$ , 属于温带半干旱大陆性季风气候。

研究区具有以下四个典型特点: ①地形地貌的多层次性: 由内蒙古高原向东北平原的过渡地带, 由燕山山脉向松辽平原的过渡地带, 由浑善达克沙地向科尔沁沙地的过渡地带; ②典型的生态脆弱区: 风蚀沙化严重(沙地沙化面积占总土地面积的  $46\%$ )、水土流失严重(水土流失面积占总土地面积的  $48\%$ ); ③典型的干旱区: 年降水少, 变率大, 十年九旱, 特别是春旱较为严重; ④典型的农牧区: 2003 年农林牧渔总产值中农业产值占  $41\%$ , 牧业产值占  $54\%$ 。

### 3 数据获取和研究方法

#### 3 1 数据来源

本研究所使用的数据集包括空间显现的土地利用数据、自然和社会经济数据。基于图件和统计数据而创建的这些数据集主要是为了发现土地利用变化的驱动因素，它们能较好地描述土地利用变化的过程。本研究在两个不同时期（1985~ 1995 年和 1995~ 2000 年）上，建立 GIS 数据库（见表 1）。其中因为 1995~ 2000 年这个时期林地变化的样本点太少，不太符合 logistic 回归分析法的要求，所以 GIS 数据库中因变量中林地变化仅考虑 1985~ 1995 年这一阶段。本文在因变量选择中社会经济数据仅取人口密度和农业人口密度两个指标，主要是考虑本文研究对象社会经济数据的尺度效应以及资料的可获取性。

利用 1985 年、1995 年和 2000 年 3 期 Landsat TM 数字遥感影像作为基本信息源。数据处理是以翁牛特旗 1: 10 万地形图为参照，并对 3 期 TM 图像选取控制点，在 MGE 的 image Analysis 模块下，对 TM 图像进行几何校正和镶嵌处理，并实现县界与图像套合。结合野外实测资料，建立解译标志，然后以最大似然法进行监督分类，根据野外考察、GPS 数据以及相关资料，修改混分和错分类结果，最后参考有关辅助图件，并借助专家目视判读对分类图像进行精度检验，研究区三期土地利用分类结果精度均较高，分别达到 75.3%、78.9% 和 83.5%。

交通分布图、河流分布图和农村居民点分布图通过手工数字化 1: 25 万地形图获取，并且通过最近的卫星影像更新数据。人口密度分布图来自于 1980 年、1995 年和 2000 年翁牛特旗的人口统计数据。土壤表层有机质含量分布图通过手工数字化翁牛特旗 1: 50 万土壤养分分布图（1986 年）得到。

所有的数据都采用 GIS 的栅格数据类型，并且在 500m 的分辨率上进行重采样。这一分辨率能最大程度反映因变量的本质。虽然在更粗的分辨率上也许更适合自变量的特征，但这可能会改变细尺度上模拟土地覆盖变化的空间格局。

#### 3 2 因变量

20 年的大部分变化是植被覆盖的减少。土地覆盖变化三种过程很明显：（1）耕地面积的增加；（2）林地面积的起伏变化；（3）草地面积的减少。它们的空间分布模拟分为两个阶段：1985~ 1995 年和 1995~ 2000 年。

表 1 GIS 数据库

Tab 1 GIS database

变量	类型	单位
因变量		
耕地的变化（1985~ 1995 年）	二分类	0~ 1
耕地的变化（1995~ 2000 年）	二分类	0~ 1
林地的变化（1985~ 1995 年）	二分类	0~ 1
草地的变化（1985~ 1995 年）	二分类	0~ 1
草地的变化（1995~ 2000 年）	二分类	0~ 1
自变量		
到最近农村居民点的距离	连续型	km
到最近道路的距离	连续型	km
到最近河流的距离	连续型	km
到最近乡镇中心的距离	连续型	km
海拔	连续型	m
坡度	多分类	1~ 3
坡向	多分类	1~ 5
农业气候区	多分类	1~ 3
土壤表层有机质含量	连续型	%
总人口密度变化（1985~ 1995 年）	连续型	Cap/km <sup>2</sup>
总人口密度变化（1995~ 2000 年）	连续型	Cap/km <sup>2</sup>
农村人口密度变化（1985~ 1995 年）	连续型	Cap/km <sup>2</sup>
农村人口密度变化（1995~ 2000 年）	连续型	Cap/km <sup>2</sup>
农业从业人员占农村人口百分比变化（1985~ 1995 年）	连续型	%
农业从业人员占农村人口百分比变化（1995~ 2000 年）	连续型	%

3 3 自变量

到最近农村居民点的距离、到最近道路的距离、到最近河流的距离、到最近乡镇中心的距离等自变量, 是利用 ArcGIS9. 0 的空间分析模块计算到相关数据要素的直线距离得到的。坡度和坡向等自变量是利用 ArcGIS 9. 0 的空间分析模块, 从 1: 25 万的 DEM 中派生出来的, 并进行了重分类。其中坡向分为五类: 平坡、北坡、东坡、南坡和西坡, 分别赋予 1~ 5 值。坡度分为三类: < 5°、5°~ 15°以及> 15°, 分别赋予 1~ 3 值。

农业气候分区是利用 ≥10°活动积温、生长季降水量、生长季日照时数和无霜期四个气候要素指标分别赋予适当的权重系数来划分的。翁牛特旗农业气候区分为三类: 农业气候区 I 即西部温凉半干旱牧林气候区; 农业气候区 II 即中部温暖半干旱半农半牧气候区; 农业气候区 III 即东部温热半干旱沙地牧业气候区 (图 1), 分别赋予 1~ 3 值。

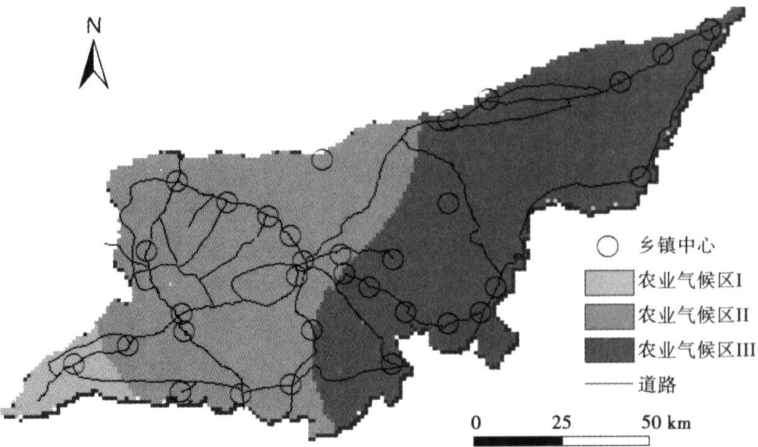


图 1 翁牛特旗农业生态区的划分  
Fig 1 Agro- regions zoning in Ongniud County

3 4 多元 logistic 回归模型

本研究运用的方法是多元 logistic 回归。线性回归模型在定量分析的实际研究中也许是最流行的统计分析方法了, 然而在许多情况下, 线性回归会受到限制, 特别当因变量是一个分类变量而不是一个连续变量时, 线性回归就不适用, logistic 回归模型能很好地解决这一问题<sup>[16]</sup>。多元 logistic 回归技术方法基于数据的抽样, 能为每个自变量产生回归系数。这些系数通过一定的权重运算法则被解释为生成特定土地利用类别的变化概率。多元 logistic 回归已经被成功地运用到野生动植物栖息地研究<sup>[17, 18]</sup>, 森林火灾的预测<sup>[19]</sup>以及森林采伐分析<sup>[20]</sup>等方面。

多元 logistic 回归能确定解释变量  $X_n$  在预测分类应变量  $Y$  发生概率的作用和强度。假定  $X$  是反应变量,  $P$  是模型的响应概率, 相应的回归模型如下:

$$\ln\left(\frac{p_1}{1 - p_1}\right) = \alpha + \sum_{k=1}^k \beta_k X_{ki} \tag{1}$$

式 (1) 中,  $p_1 = P(y_i = 1 | x_{1i}, x_{2i}, \dots, x_{ki})$  为在给定系列自变量  $x_{1i}, x_{2i}, \dots, x_{ki}$  的值时事件的发生概率,  $\alpha$  为截距,  $\beta$  为斜率。

发生事件的概率是一个由解释变量  $X_i$  构成的非线性函数, 表达式如下:

$$p = \frac{\exp(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni})}{1 + \exp(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni})} \tag{2}$$

发生比率 (odds ratio) 用来对各种自变量 (如连续变量、二分变量、分类变量) 的 logistic 回归系数进行解释<sup>[17]</sup>。在 logistic 回归中应用发生比率来理解自变量对事件概率

的作用是好的方法, 因为发生比率在测量关联时具有一些很好的性质<sup>[21]</sup>。发生比率用参数估计值的指数来计算<sup>[22]</sup>:

$$\text{odd}(p) = \exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n) \quad (3)$$

本研究中, 多元 logistic 回归是用 SAS 统计软件的 logistic 函数来操作完成的。logistic 回归模型预测能力通过得到最大似然估计的表格来评价, 它包括回归系数、回归系数估计的标准差、回归系数估计的 Wald $\chi^2$  统计量和回归系数估计的显著性水平。正的回归系数值表示解释变量每增加一个单位值时发生比会相应增加。相反, 当回归系数为负值时说明增加一个单位值时发生比会相应减少。Wald $\chi^2$  统计量表示在模型中每个解释变量的相对权重, 用来评价每个解释变量对事件预测的贡献力。

模型估计完成以后, 我们需要评价模型如何有效地描述反应变量及模型配准观测数据的程度。用来进行拟合优度的检验指标有皮尔逊  $\chi^2$ 、偏差 D 和 Homsmer-Lemeshow 指标 (HL) 等。当自变量数量增加时, 尤其是连续自变量纳入模型之后, 皮尔逊  $\chi^2$ 、偏差 D 不再适用于估价拟合优度。在应用包括连续自变量的 logistic 回归模型时, HL 是广为接受的拟合优度指标<sup>[14]</sup>。因此, 本文用 HL 指标来进行土地利用变化的 logistic 回归模型拟合优度检验。当 HL 指标统计显著表示模型拟合不好; 相反, 当 HL 指标统计不显著表示模型拟合好。HL 指标是一种类似于皮尔逊  $\chi^2$  统计量的指标, 其公式如下:

$$HL = \sum_{g=1}^G \frac{(y_g - n_g \hat{p}_g)^2}{n_g \hat{p}_g (1 - \hat{p}_g)} \quad (4)$$

式 (4) 中,  $G$  代表分组数, 且  $G \leq 10$ ;  $n_g$  为第  $g$  组中的案例数;  $y_g$  为第  $g$  组事件的观测数量;  $\hat{p}_g$  为第  $g$  组的预测事件概率;  $n_g \hat{p}_g$  为事件的预测数。

为了拟合模型, 本文选用了逐步模型选择法和概念模型法相结合的方法。在统计模型中, 我们先选用概念模型中的解释变量, 然后用逐步回归法选用主要的解释变量, 最后基于饱和模型分析哪些变量对解释土地利用变化有明显贡献。

### 3.5 抽样过程

为了使用 logistic 回归模型, 本文选用了分层随机抽样方法选择了均匀分布整个研究区的  $n$  个观测点。选择对观测点进行随机抽样是为了避免数据的空间自相关性。对于每一个抽样观测点, 记录其因变量和一系列的自变量值。对于每个模型, 随机抽样了 2000 个观测点, 确保因变量的 0 和 1 观测值有相等的数量。不相等的抽样比例不会影响解释变量在 logistic 回归模型中的系数估计, 但是会影响模型的常数项<sup>[16]</sup>。当用模型来模拟时, 常数项通过  $(\ln p_1 - \ln p_2)$  其中  $p_1$  和  $p_2$  分别是因变量 0 和 1 的观测频数<sup>[23, 24]</sup>。

## 4 结果分析

### 4.1 初步统计诊断

各自变量之间共线性关系很小, 即相关程度小。决定自变量相对其他自变量相关的系数  $R^2$  在 0.02~0.56 之间, 均低于临界值 0.8<sup>[25]</sup>。这样所有的自变量都可以被纳入到 logistic 回归模型中。

### 4.2 转化为耕地的 logistic 回归模型

在耕地空间模型中, 农业气候区用两个虚拟变量分别代表农业气候区 I 和农业气候区 II, 农业气候区 II 作为它们的参照对象。坡向用 4 个虚拟变量分别代表北坡、东坡、南坡和西坡, 平地作为它们的参照对象。坡度用一个二分类变量表示, 较适宜作为耕地的坡度级

别 I 和 II ( $\leq 15^\circ$ ) 用代码 1 表示, 不太适宜作为耕地的坡度级别 III ( $> 15^\circ$ ) 用代码 0 表示。

在第一阶段 (1985~ 1995 年) 和第二阶段 (1995~ 2000 年), 其他土地利用类型转为耕地的 logistic 回归模型都有很好的拟合度, HL 指标分别为 7.32 和 8.41, p 值分别为 0.312 和 0.242, 统计检验都不显著, 即两个模型很好地拟合了数据 (表 2a, b)。根据 Wald $\chi^2$  统计量, 第一阶段 (1985~ 1995 年) 较为重要的解释变量是到最近农村居民点的距离、到最近河流的距离、农业气候区、土壤表层有机质含量和坡向。而在第二阶段 (1995~ 2000 年) 较为重要的解释变量是到最近农村居民点的距离、坡度、土壤表层有机质含量和农村人口密度变化。

在第一阶段 (1985~ 1995 年) 和第二阶段 (1995~ 2000 年), 最重要的解释变量都为到最近农村居民点的距离。模型中该解释变量负的回归系数, 表明其他土地利用类型转为耕地的概率随着到农村居民点距离的减小而增大。对于离农村最近农村居民点每减小 1km, 其他土地利用类型转变为耕地的概率将增大 1.6 倍 (即 1km 的发生率是  $e^{-0.474 \times 1} = 0.62$ ; 见表 2a)。在第一个阶段 (1985~ 1995 年), 第二个

表 2 转化为耕地: (a) 1985~ 1995 年,  $p = 0.312$ ;  
(b) 1995~ 2000 年,  $p = 0.312$   
Tab 2 Expansion of arable land: (a) between 1985 and 1995,  $p = 0.42$ ; (b) between 1995 and 2000,  $p = 0.636$

变量	参数估计( $\beta$ )	标准误差(SE)	Wald $\chi^2$ 统计量	$Pr > \chi^2$	发生比率
(a) 1985~ 1995 年, HL= 7.32, $p = 0.312$					
常数	3.359	0.327	105.358	0.0001	-
到最近农村居民点的距离	- 0.474	0.035	187.918	0.0001	0.622
到最近道路的距离	- 0.045	0.016	7.598	0.006	0.956
到最近河流的距离	- 0.088	0.011	68.586	0.0001	0.916
到最近乡镇中心的距离	0.002	0.002	3.875	0.035	1.002
海拔	0.376	0.363	4.073	0.03	1.457
坡向 II (北坡)	0.84	0.159	27.961	0.0001	2.317
坡向 III (东坡)	0.657	0.153	18.491	0.0001	1.928
坡向 IV (南坡)	0.603	0.188	10.3	0.001	1.828
坡向 V (西坡)	0.973	0.196	24.725	0.0001	2.647
农业气候区 I	1.272	0.502	6.43	0.011	3.569
农业气候区 II	- 1.134	0.146	60.528	0.0001	0.322
土壤表层有机质含量	- 0.725	0.108	44.744	0.0001	0.484
总人口密度变化 (1985~ 1995)	0.016	0.004	17.737	0.0001	1.016
农村人口密度变化 (1985~ 1995)	- 0.037	0.006	36.738	0.0001	0.963
农业从业人员占农村人口百分比变化 (1985~ 1995)	0.001	0.006	1.06	0.0806	1.001
(b) 1995~ 2000 年, HL= 9.51, $p = 0.242$					
常数	- 0.529	0.509	1.077	0.08	-
到最近农村居民点的距离	- 0.553	0.054	103.911	0.0001	0.575
到最近道路的距离	- 0.069	0.24	8.421	0.004	0.933
到最近乡镇中心的距离	0.014	0.003	19.172	0.0001	1.014
海拔	- 0.516	0.415	1.543	0.014	0.933
坡向 II (北坡)	- 0.33	0.287	1.325	0.016	0.719
坡向 III (东坡)	0.528	0.238	4.937	0.026	1.695
坡向 IV (南坡)	0.709	0.23	9.48	0.002	2.032
坡向 V (西坡)	0.618	0.257	5.777	0.016	1.855
坡度级 I ( $< 15^\circ$ )	2.102	0.272	59.879	0.0001	8.186
土壤表层有机质含量	- 0.927	0.162	32.668	0.0001	0.396
农村人口密度变化 (1995~ 2000)	0.056	0.016	12.287	0.0001	1.057
农业从业人员占农村人口百分比变化 (1995~ 2000)	0.013	0.006	5.226	0.022	1.013

重要的解释变量为到最近河流的距离, 并且转化概率随着道河流距离的增加而减小, 这表明耕地更有可能出现在离河流较近的地方发生。而在第二阶段 (1995~ 2000 年) 到最近河流的距离这个解释变量却被排除在 logistic 回归模型之外, 其原因可能是离河流近的其

它土地利用类型在第一阶段已转化成耕地, 而离河流距离较远的其他土地利用类型却不太适合用做耕地。

在第一阶段 (1985~ 1995 年), 第三个重要的解释变量是农业气候区, 这表明其他土地利用类型转变为耕地的概率更有可能出现在适合农作物生长的农业气候区。农业气候区 II 这一解释变量转化为耕地发生比率是 0. 322, 表明在农业气候区 II 中其他土地利用类型转为耕地的概率是在农业气候区 II 的 3 倍。在第二阶段 (1995~ 2000 年), 第二个重要的解释变量是坡度。坡度级别 I 这一解释变量转化为耕地发生比率是 8 186, 表明在坡度级别 I 中其他土地利用类型转为耕地的概率是在参照坡度级别的 8 倍。

在第一阶段 (1985~ 1995 年), 随着海拔的增加, 耕地转移率也增加。而在第二阶段 (1995~ 2000 年) 却相反, 随着海拔的增加, 耕地转移率减小。在第一阶段 (1985~ 1995 年) 和第二阶段 (1995~ 2000 年), 随着土壤有机质含量的增加, 转化为耕地的概率却减少, 这表明土壤养分不是当地农民把其他土地利用类型转为耕地来种植农作物的先决条件, 土壤养分的增加可以通过增施肥料等途径达到。

### 4 3 转化为林地的 logistic 回归模型

在林地空间模型中, 坡向用 4 个虚拟变量分别代表平地、北坡、东坡和南坡, 西坡作为它们的参照对象。农业气候区用一个二分类变量表示, 农业气候区 I 和 II 作为农业气候区 III 的参照对象。

在林地空间模型第一阶段 (1985~ 1995 年), HL 指标为 13 173, p 值为 0. 096, 统计检验不太显著, 其他土地利用类型转化为林地的 logistic 回归模型有较好的拟合度, 即这一模型能勉强拟合数据(表 3)。

在第一阶段 (1985~ 1995 年), 较为重要的解释变量是到最近农村居民点的距离、海拔、农业气候区。其中最重要的解释变量是到最近农村居民点的距离, 模型中该解释变量负的回归系数, 表明其他土地利用类型转为林地的概率随着

表 3 转化为林地 (1985– 1995 年), p= 0. 096  
Tab 3 Woodland modification between 1985 and 1995, p= 0. 096

变量	参数估计(β)	标准误差(SE)	Waldχ <sup>2</sup> 统计量	Pr> χ <sup>2</sup>	发生比率
常数	4. 161	0. 548	57. 577	0. 0001	–
到最近农村居民点的距离	– 0. 398	0. 054	54. 913	0. 0001	0. 672
海拔	– 0. 003	0. 001	22. 229	0. 0001	0. 997
坡向 I (平地)	– 1. 266	0. 321	15. 578	0. 0001	0. 282
坡向 II (北坡)	– 0. 461	0. 266	3. 004	0. 083	0. 630
坡向 III (东坡)	– 0. 128	0. 266	3. 232	0. 063	0. 88
坡向 IV (南坡)	– 0. 838	0. 308	7. 408	0. 006	0. 432
农业气候区 III	0. 866	0. 21	17. 04	0. 0001	0. 421
土壤表层有机质含量	– 1. 269	0. 192	43. 876	0. 0001	0. 281
农村人口密度变化 (1985~ 1995 年)	0. 007	0. 008	5. 744	0. 038	1. 007
农业从业人员占农村人口百分比变化 (1985~ 1995 年)	0. 024	0. 01	5. 544	0. 19	1. 024

到最近农村居民点距离的减小而增大。对于离农村最近农村居民点每减小 1km, 其他土地利用类型转变为林地的概率将增加 1. 5 倍 (即 1km 的发生率是  $e^{-0.398 \times 1} = 0.67$ ; 见表 3)。另外一个较为重要的解释变量是海拔, 结果表明其他土地利用类型转为林地的概率随着海拔的增加而减小。农业气候区、也是林地空间模型的重要解释变量。在第一阶段 (1985~ 1995 年) 林地空间模型中, 农业气候区 II 这一解释变量转化为林地发生比率是 0. 421, 表明在农业气候区 II 转为林地的概率是在参照对象下的 2. 4 倍。

### 4 4 转为草地的 logistic 回归模型

在草地空间模型中, 在第一阶段 (1985~ 1995 年) 和第二阶段 (1995~ 2000 年), 农

业气候区、坡度两个分类变量在逐步回归法中统计不显著,被排除在模型之外。坡向用 4 个虚拟变量分别代表北坡、东坡、南坡和西坡,平地作为它们的参照对象。

在第一阶段(1985~1995 年)和第二阶段(1995~2000 年),其他土地利用类型转为草地的 logistic 回归模型都有很好的拟合度,HL 指标分别为 8 143 和 6 069, p 值分别为 0.42 和 0.636,统计检验都不显著,即两个模型很好的拟合了数据(表 4a, b)。这两个回归模型纳入了相同的解释变量,根据 Wald $\chi^2$  统计量,第一阶段(1985~1995 年)较为重要的解释变量是到最近农村居民点的距离、土壤表层有机质含量、到最近乡镇中心的距离、总人口密度的变化。而在第二阶段(1995~2000 年)较为重要的解释变量却为农业从业人员占农村人口百分比的变化、到最近农村居民点的距离和到最近河流的距离。

在第一阶段(1985~1995 年),最重要的解释变量为到最近农村居民点的距离,模型中该解释变量负的回归系数,表明其他土地利用类型转为草地的概率随着到最近农村居民点距离的减小而增大。对于离农村最近农村居民点每减小

1km,其他土地利用类型转变为草地的概率将增大 2.1 倍(即 1km 的发生率是  $e^{-0.476 \times 1} = 0.62$ ;见表 4a)。在第一阶段(1985~1995),另外一个较为重要的解释变量是总人口密度的变化,模型中该解释变量负的回归系数,这表明其他土地利用类型转变为草地的概率随着到总人口密度的增加而减小。对于总人口密度每平方公里增加 1 人,其他土地利用类型转变为草地的概率将增大 1 倍(即 1km 的发生率是  $e^{-0.043 \times 1} = 0.96$ ;见表 4a)。

表 4 转化为草地: (a) 1985~ 1995, p= 0.42;  
(b) 1995~ 2000, p= 0.636  
Tab 4 Expansion of grassland: (a) between 1985 and 1995,  
p= 0.42; (b) between 1995 and 2000, p= 0.636

变量	参数估计( $\beta$ )	标准误差(SE)	Wald $\chi^2$ 统计量	Pr> $\chi^2$	发生比率
(a) 1985~ 1995 年, HL= 8.134, p= 0.42					
常数	- 0.367	0.49	2.562	0.045	-
到最近农村居民点的距离	- 0.476	0.065	53.746	0.0001	0.621
到最近河流的距离	0.023	0.019	4.449	0.023	1.023
到最近乡镇中心的距离	0.016	0.004	14.252	0.0001	1.016
海拔	0.001	0.001	6.539	0.011	1.001
坡向 II (北坡)	0.295	0.276	4.14	0.028	1.343
坡向 III (东坡)	0.647	0.263	6.039	0.014	1.909
坡向 IV (南坡)	1.048	0.332	9.972	0.002	2.853
坡向 V (西坡)	0.178	0.339	4.275	0.021	1.195
土壤表层有机质含量	- 0.868	0.202	18.545	0.0001	0.42
总人口密度变化(1995~2000)	- 0.043	0.012	12.537	0.0001	0.958
农村人口密度变化(1995~2000 年)	0.045	0.014	10.695	0.001	1.046
农业从业人员占农村人口百分比变化(1995~2000 年)	0.034	0.011	10.299	0.001	1.034
(b) 1995~ 2000 年, HL= 6.096, p= 0.636					
常数	1.919	1.949	3.969	0.045	-
到最近农村居民点的距离	- 0.788	0.245	10.339	0.001	0.455
到最近河流的距离	- 0.216	0.09	5.776	0.016	0.805
到最近乡镇中心的距离	0.027	0.012	4.855	0.028	1.027
海拔	- 0.001	0.002	0.704	0.042	0.999
坡向 II (北坡)	1.184	1.076	1.212	0.027	3.269
坡向 III (东坡)	0.614	0.991	0.384	0.053	1.849
坡向 IV (南坡)	0.557	1.369	0.166	0.068	1.746
坡向 V (西坡)	2.186	1.397	2.448	0.011	8.897
土壤表层有机质含量	- 0.847	0.598	2.007	0.007	0.429
总人口密度变化(1995~2000)	- 0.037	0.041	0.842	0.035	0.963
农村人口密度变化(1995~2000 年)	- 0.287	0.172	2.806	0.094	0.75
农业从业人员占农村人口百分比变化(1995~2000 年)	- 0.11	0.032	12.273	0.0001	0.895



在第二阶段 (1995~ 2000 年), 最重要的解释变量是农业从业人员占农村人口百分比的变化, 模型中该解释变量负的回归系数, 表明其他土地利用类型转为草地的概率随着农业从业人员占农村人口百分比的增大而减小。对于农业从业人员占农村人口百分比每增加 1 个百分点, 其他土地利用类型转变为草地的概率将减小 1.1 倍 (即 1km 的发生率是  $e^{-0.11 \times 1} = 0.89$ ; 见表 4b)。在第二阶段 (1995~ 2000 年) 对于坡向这一解释变量, 在北坡转化为草地的发生比率是 1.343, 表明在北坡转为草地的概率是在参照坡向 (平坡) 的 1.34 倍。

#### 4.5 模型验证

就草地而言, 尽管两个阶段的 logistic 回归模型的系数不一样, 但它们包含了相同的空间解释变量。本文用第一阶段草地的 logistic 回归模型的参数来检验在第二阶段可能转为草地的准确性。用公式 (2) 产生一个转为草地的概率分布图, 它是代表在假定第二阶段和第一阶段有同样的驱动过程下, 1995 年以后转为草地的风险区域 (图 2)。通过与实际在第二阶段其他土地利用类型

转化为草地的像元数和转化为草地的预测概率相对照, 结果表明在已转为草地像元数中, 预测概率小于 50% 的像元占总数还不到 15%, 而 85% 以上的像元的预测概率是大于 50%。第二个高峰是 70~ 85% 的预测概率, 实际上有 15~ 20% 的像元数已转化为草地。这些像元数主要分布在牧区, 并且离河流的距离较近, 主要是人工草地或改良草场。虽然第二阶段土地利用变化的各解释变量同第一阶段不同, 但从模拟的结果来看, 第一阶段的模型能较好地用来进行第二阶段的模拟。

### 5 结论与讨论

(1) 研究区土地覆盖变化各个阶段有着不同的重要驱动因素。就草地变化而言, 在第一阶段主要的驱动因素是到最近农村居民点的距离, 而在第二阶段主要的驱动因素是农业从业人员占农村人口百分比的变化。通过对草地 logistic 回归模型的检验, 得出空间上的统计模型能较好地揭示不同土地利用变化过程的主要驱动力<sup>[25]</sup>。

(2) 通过对研究区土地利用类型的变化 logistic 回归分析表明, 空间异质性和土地利用变化过程的时间变量共同影响着使用 logistic 回归模型来进行推断的能力。

(3) 本研究中因变量中社会经济数据仅取人口密度和农业人口密度进行分析显得不

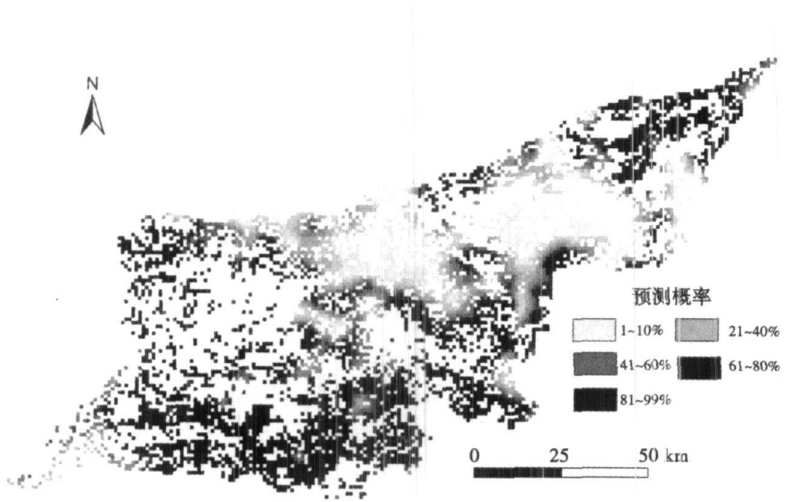


图 2 1995 年以后转为草地的概率分布图

(其中 1995 年已为草地的区域被掩码)

Fig 2 Probability map for conversion to grassland after 1995

(Locations of already grassland in 1995 are masked)

足, 其实 LUCC 与人口、技术、富裕程度、政治结构和价值与观念直接相关, 因此, 从这方面选取一些可以量化的指标应是下一步研究的重点内容。

(4) 本研究建立的空间模型尽管能够预测在不久的将来土地覆盖转换在什么地方发生, 能够揭示导致这些变化的一些驱动因素, 但还是不能预测土地覆盖转换在什么时候能发生。因此为了提高土地覆盖变化的预测概率, 我们需要建立考虑驱动力的动态模型。这些驱动力应包括新政策的制定、农产品的价格变化等动态变化因素, 这样代表变化可能性原因的驱动力 (如到市场的交通费用) 在空间模型中就可以显现。例如, 修建一条新的道路可能带来的影响就可以在与土地覆盖变化预测概率相关联的地图中显示出来。

(5) 土地利用和土地覆盖变化研究是一个复杂的系统工程, 本文只是利用 logistic 回归模型的原理探讨变化的可能驱动因素, 其理论和实践还有待进一步深入研究。

## 参考文献:

- [1] Turner B L, Skole D, Sanderson S, *et al.* Land Use and Land-Cover Change Science/ Research Plan 1995, IGPB Report No. 35 and HDP Report No. 7, 132
- [2] Geist H J, Lambin E F Proximate causes and underlying driving forces of tropical deforestation BioScience, 2002, 52(2): 143~ 150
- [3] Serneels S, Lambin E F. Proximate causes of land-use changes in Narok district, Kenya: A spatial statistical model Agriculture, Ecosystem and Environment, 2001, 85: 65~ 81
- [4] 王长科, 吕宪国, 蔡祖聪, 等. 土地利用方式对白浆土氧化甲烷的影响. 地理研究, 2006, 25(2): 335~ 341
- [5] 刘纪远, 张增祥, 庄大方, 等. 20 世纪 90 年代我国土地利用变化时空特征及其成因分析, 地理研究, 22(1), 2003, 1~ 12.
- [6] 许月卿. 土地利用对地下水位下降的影响——以河北平原为例. 地理研究, 2005, 24(2): 222~ 228
- [7] 摆万奇, 张永民, 阎间忠, 等. 大渡河上游地区土地利用动态模拟分析. 地理研究, 2005, 24(2): 206~ 212
- [8] Pontius Jr R G, Schneider L C Land-use change model validation by a ROC (relative operating characteristic) method Agric Ecosyst Environ., 2001, 85: 239~ 248
- [9] Riebsame W E, Meyer W B, Turner II B L Modelling land use and cover as part of global environmental change Clim. Change, 1994, 28: 45~ 64
- [10] Chomitz K M, Gray D A. Roads, land use, and deforestation: A spatial model applied to Belize World Bank Econ Rev, 1996, 10: 487~ 512
- [11] Mertens B, Lambin E F. A spatial model of land-cover change trajectories in a frontier region in southern Cameroon Ann. Assoc. Am Geogr., 2000, 90: 467~ 494
- [12] Mertens B, Lambin E F Spatial modelling of deforestation in southern Cameroon: Spatial isaggregation of diverse deforestation processes. Appl Geogr., 1997, 17: 143~ 162
- [13] Liverman D, Moran E F, Rindfuss R R, *et al.* People and pixels: Linking remote sensing and social science Washington, DC National Academy Press, 1998
- [14] Gilruth P T, Hutchinson C F. Assessing deforestation in the Guinea Highlands of West Africa using remote sensing Photogramm. Eng. Remote Sens., 1990, 56: 375~ 1382
- [15] Geoghegan J, Pritchard L, Ogneva-Himmelberger Y, *et al.* "Socializing the pixel" and "pixelizing the social" in land-use and land-cover change. In: Liverman D, Moran E F, Rindfuss R R, *et al.* (eds.). People and Pixels: Linking Remote Sensing and Social Science Washington, DC: National Academy Press, 1998. 51~ 69
- [16] 王济川, 郭志刚. Logistic 回归模型——方法与应用. 北京: 高等教育出版社, 2001
- [17] Pereira J M C, Itami R M. GIS-based habitat modeling using logistic multiple regression: A study of the Mt. Graham red squirrel Photogramm. Eng. Remote Sens., 1991, 57: 1475~ 1486
- [18] Narumalani S, Jensen J R, Althausen J D, *et al.* Aquatic macrophyte modelling using GIS and multiple logistic regression. Photogramm. Eng. Remote Sens., 1997, 63: 41~ 49

- [19] Vega Garcia C, Woodard P M, Titus S J, *et al.* A logit model for predicting the daily occurrence of human caused forest fires Int J Wildl Fire, 1995, 5: 101~ 111
- [20] Stokes M A, Davis C S, Koch G G. Categorical Data Analysis Using the SAS System SAS Institute, Inc., Cary, NC, 1995.
- [21] Feiberg Stephen The analysis of Crossclassified Categorical Data (2<sup>nd</sup> ed). Cambridge, MA: MIT Press, 1985.
- [22] Hosmer D W, Lemeshow S Applied Regression Analysis New York: Wiley, 1989.
- [23] Maddala G S. Introduction to Econometrics New York: Macmillan, 1988.
- [24] Menard S. Applied Logistic Regression Analysis. Quantitative Applications in the Social Sciences, No. 106. Sage, London, 1995
- [25] 邓祥征, 刘纪远, 战金艳, 等. 太仆寺旗土地利用变化时空格局的动态模拟. 地理研究, 2004, 23(2): 147~ 156

## Driving forces analysis of land-use pattern changes based on logistic regression model in the farming-pastoral zone: A case study of Ongiud Banner, Inner Mongolia

XIE Huailin<sup>1,2</sup>, LI Bo<sup>2</sup>

(1. College of Resource & Environment Management, Jiangxi University of Finance and Economics, Nanchang 330032, China; 2 State Key Laboratory of Earth Surface Processes and Resource Ecology, School of Resource Science, Beijing Normal University, Beijing 100875, China)

**Abstract:** Land use and land cover change (LUCC) research has been given close attention during the past decade, because of the pivotal role of LUCC in many aspects like global climatic change, food security, soil degradation and biodiversity. This study attempts to identify how much driving forces of land-use changes can be understood through logistic regression analysis

So, spatial statistical models of the proximate causes of different processes of the land-use change in the farming-pastoral zone (Inner Mongolia) were developed, taking into account the spatial variability of the land-use change processes. The descriptive spatial models developed here suggested some important factors driving the land-use changes that can be related to some well-established theoretical frameworks. The explanatory variables of the spatial model of arable land suggest a model, where conversion to arable land is controlled by the rural settlements and agro-climatic zones. Grassland conversion is controlled by the distance to the nearest rural settlement, organic matter in the soil and distance to the nearest river. Expansion of forest is also controlled by the rural settlements and altitude. Spatial heterogeneity as well as the variability in time of land-use change processes affect human ability to use regression models for wide ranging extrapolations. These spatial models can identify the main driving forces of different land-use changes through validation of the logistic regression model of the grass expansion.

**Key words:** logistic regression analysis; land-use change; farming-pastoral zone