基于 CART-CA 模型的用地变化多尺度模拟
——以长江三峡库区为例

刘明皓，袁　敏，冯　源

（重庆邮电大学计算机科学与技术学院，空间信息系统研究中心，重庆　400065）

摘要：尺度依赖被认为是普遍存在的空间现象，寻找最佳模拟尺度及其影响因素对优化用地变化模型、提高模拟精度具有重要意义。以长江三峡库区为例，构建基于 CART-CA 模型的用地动态变化模型，并检验多尺度下建模方法差异及地类均衡性对模拟精度的影响。采用 CART-CA、DYNA-CLUE 和 Logistic-CA 模型，分别基于 200，300，500，800，1 000 m 5 种分辨率数据和 7 类、5 类、3 类、2 类用地分类方案实施模拟。结果显示：在 7 类、5 类、3 类、2 类 4 种分类方案下，最佳模拟尺度依次为 300，200，800，200 m；相同分类方案下的最佳模拟尺度不因采用模型的不同而不同；随着分辨率的降低，地类均衡度总体呈上升趋势；随着地类均衡度的升高，模拟精度降低。研究表明，CART-CA 模型用于多用途土地利用格局变化的动态模拟时，模拟精度较高；用地动态变化模型在尺度上存在最佳分辨率；分类均衡性是产生用地动态模拟精度高低的重要原因，最佳模拟尺度会因分类均衡度的变化而导致不同。

关键词：地理学；用地动态变化模拟；元胞自动机；分类回归树；多尺度；尺度效应

中图分类号：[K903]；F301.24　　**文献标识码**：A　　**文章编号**：1674-2850(2019)01-0162-13

Multiresolution simulation of land use change based on CART-CA model: a case study of the Yangtze Three Gorges Reservoir area

LIU Minghao, YUAN Min, FENG Yuan

(*Research Center of Spatial Information System*, *College of Computer Science and Technology*, *Chongqing University of Posts and Telecommunications*, *Chongqing*　400065, *China*)

Abstract: Scale dependence is considered to be a ubiquitous spatial phenomenon. It is very important for optimizing the land use change model and improving the accuracy of simulation to find the best simulation resolution and its influencing factors. Taking the Yangtze Three Gorges Reservoir area as an example, the dynamic change model of land use based on CART-CA model was constructed. On this basis, multiresolution schemes of land use are designed to test the effects of different modeling methods and equilibrium of the land use types on the accuracy of simulation. Based on the five resolution data of 200, 300, 500, 800, 1 000 m, and the land use classification schemes of 7, 5, 3, 2 categories, CART-CA, DYNA-CLUE and Logistic-CA models are used to carry out simulation experiments respectively. Results show that the best simulation resolution in the study area is 300, 200, 800, 200 m, in turn, based on the above land classification schemes of 7, 5, 3, 2 categories, which means that the best simulation resolution under the same classification scheme will not be different from the different adopted models. And with the decrease of the resolution, the equilibrium degree of the land use types is generally on the rise, otherwise

基金项目：重庆市应用开发计划重点项目（cstc2014yykfB30003）；重庆市教育委员会科学技术研究项目（KJ1400420）；重庆市教育委员会人文社会科学研究项目（18SKGH033）

作者简介：刘明皓（1970—），男，教授，主要研究方向：地理计算智能、城乡用地动态模拟和国土资源管理等. E-mail: liumh@cqupt.edu.cn

the simulation precision decreases with the increase of the equilibrium degree of the land use types. The research shows that CART-CA model has superior simulation precision which can be used for dynamic simulation of multi-purpose land use pattern change. There is an optimal resolution for the land use dynamic change model based on multiresolution. Classification equilibrium is an important reason for the accuracy of land use dynamic simulation. Therefore, the optimal simulation scale varies with the classification equilibrium.

Key words: geography; land use dynamic change simulation; cellular automata; classification and regression tree; multiresolution; scale effect

0 引言

元胞自动机（cellular automata，CA）模型因兼具数量模型与空间模型的优点，以及简洁、灵活、直观、易与地理信息系统（geographic information system，GIS）结合等特性而被广泛运用[1]。CA 模型用于用地变化模拟，其核心是提取元胞的用地转换规则。近年来，一些机器学习和人工智能算法等被用来与 CA 结合用于用地转换规则提取。尽管机器学习算法与 CA 的结合取得了较好的模拟效果，但也各自存在一定的缺陷。群体智能算法有较强的参数自适应和优化能力，但算法的时间复杂度高[2]。神经网络算法具有很强的非线性拟合能力，但属于黑箱操作，规则隐性，对模拟过程无法解释，且易过拟合[3]。决策树算法虽需要对树进行剪枝，但规则显性，解释性和模拟效果都较好[4~5]。C4.5 和分类回归树（classification and regression tree，CART）算法都是基于 ID3 算法进行改进的，而 CART 算法的应用更加广泛。CART 算法的基本原理是通过对数据集进行连续二分类来修正决策树的结构，该算法具有对数据的容错能力，可处理复杂的非线性关系数据和对特征向量进行重要性度量。

尺度依赖（scale dependence）被认为是普遍存在的空间现象[6~8]，尺度问题也是土地变化科学中的关键问题[9]。空间尺度一般分为幅度（空间范围）和粒度（最小可辨识单元的特征长度、面积或体积），本研究中的空间尺度指的是粒度，即分辨率问题。国内外在空间尺度方面的研究已取得了较为丰硕的成果，主要集中在尺度转换的基本理论、基本方法及尺度效应等方面。而在土地利用变化研究中的尺度效应主要关注数据获取、土地利用驱动因子选择和用地政策三方面[10]。在用地变化领域研究尺度效应的意义在于：1）恰当地选择尺度转换方法，有针对性地减少对研究目标有重大影响的要素信息损失；2）通过构建用地变化和驱动因子数据精度变化的尺度系谱，掌握尺度变化规律，甄别尺度域和尺度阈值，确定用地变化分析的最适尺度，建立相应的尺度转换关系和模型，提高分析和模拟精度，减少“尺度谬论”现象的发生[11~16]。

在数据获取与尺度的关系方面，陶舒等[17]通过定量研究发现，各种地表覆盖类型的损失精度对空间尺度的变化十分敏感。其中，平均面积较小的斑块在尺度变化过程中更容易被面积较大的斑块吞并，而且地表覆盖精度的尺度敏感性也与遥感数据获取如分类方法密切相关。在土地利用驱动因子的选择方面，用地变化与驱动因子相关分析的研究表明，驱动因子本身及其作用力大小呈现出显著的尺度依赖性规律，而且回归模型的解释能力会随尺度改变而发生变化[18~19]。尽管分类的均衡性对模型模拟精度的影响不容忽视[20]，但分类均衡性如何度量，其与分辨率之间是什么样的关系，建模方法差异是否存在尺度依赖，最佳模拟尺度是否受地类均衡度变化的影响等问题有待深入探讨。

本文以长江三峡库区为例，具体研究内容如下：1）将 CART 算法与 CA 结合，构建用地动态变化模型，并探讨驱动因子选择对模拟精度的影响；2）检验多尺度下，用地动态模型的尺度效应；3）检验多尺度下，地类均衡性对模拟精度的影响。

1　模型构建与技术路线

1.1　模型构建

提取用地转换规则是CA模型的核心任务，也是本文研究的重点内容之一。基于CART-CA的用地动态变化模型由数据处理模块、综合概率计算模块和动态迭代模块构成。一段时期内，用地类型是否发生改变，与地块的适宜性、地类的近邻性、规划的约束性和随机因子等因素密切相关[21~22]。本研究中，综合概率计算模块由地类适宜性评价、地类邻域效应评价和随机变量三部分构成；在综合概率计算基础上，动态迭代模块最终完成地类转换，即对某一栅格而言，当满足一定条件时，该栅格转换为某地类。

1）地类适宜性评价

地类适宜性评价模块作为CART-CA模型的重要组成部分，包括以下三个基本步骤：①将驱动因子数据作为特征数据，将基期年和末期年用地变化检测数据作为标签数据组合成数据集，采用分层随机抽样方式抽取一定比例的数据作为训练数据集；②用CART模型进行训练，生成分类回归树，完成模型训练；③测试集数据导入树模型中，获得元胞 ij 在 t 时刻从当前类型转变为第 l 类的概率 $P_c(ij,t,l)$.

2）邻域效应评价

根据地理学第一定理，CA模型中元胞与元胞之间的作用存在相互性，同时也具有局部性。某一元胞下一时刻的状态受其邻域元胞状态的影响。邻域效应可由式（1）表示：

$$\Omega_{ij}^{t}=\frac{\sum_{n\times n}\mathrm{con}(S_{ij}^{t}=\mathrm{Landuse}l)}{n\times n-1}, \tag{1}$$

其中，Ω_{ij}^{t} 为 t 时刻元胞 ij 的 $n\times n$ 邻域作用值，n>2；con(·)为条件函数；S_{ij}^{t} 为元胞当前的状态；Landuse$l$ 为第 l 类土地利用类别。

3）随机变量

将随机变量引入到CA模型中，可以使模拟更加接近真实情况。用地覆被变化是一个极为复杂的过程，涉及很多因素的影响。常规的驱动因子数据往往难以反映土地利用变化过程中随机因素的影响。本文在CART-CA模型中引入了随机变量。该随机变量可由式（2）表示：

$$\mathrm{RA}=1+(-\ln\gamma)^{\alpha}, \tag{2}$$

其中，γ 为一个取值范围在0～1之间的随机数（含0和1）；α 为一个控制随机变量大小的参数，取值范围为1～10之间的整数（含1和10）。

4）约束性因子

引入一些限制性条件来约束某些元胞向其他类型转变，如某些规划严格禁止向其他地类转换等[如式（3）所示]。由于本文没有获取到约束性因子数据，故在CART-CA模型中，对于约束性因子不予考虑。

因此，在CART-CA模型的总概率计算模块部分，即在 t 时刻，土地利用类型从当前类型转换为第 l 类的概率为

$$P(ij,t,l)=P_c(ij,t,l)\times\Omega_{ij}^{t}\times\mathrm{con}(S_{ij}^{t})\times\mathrm{RA}=P_c(ij,t,l)\times\frac{\sum_{n\times n}\mathrm{con}(S_{ij}^{t}=\mathrm{Landuse}l)}{n\times n-1}\times\mathrm{con}(S_{ij}^{t})\times[1+(-\ln\gamma)^{\alpha}], \tag{3}$$

其中，$P(ij,t,l)$ 表示 i 行 j 列的元胞在 t 时刻为第 l 类的综合概率；$\mathrm{con}(S_{ij}^{t})$ 为规划约束性因子。

动态迭代模块中，假定研究区域的土地利用类型有 m 种，即 l=(1,2,⋯,m)，对于每个元胞得到 l 种地类的概率。本文选取最大的地类概率 $P_{\max(ij,t)}$表示元胞 ij 在 t 时刻由当前地类转变为下一地类，可以表

示为

$$P_{\max}(ij,t)=\max(P(ij,t,l))=\max(P_{\mathrm{c}}(ij,t,l)\times\Omega_{ij}^{t}\times[1+(-\ln\gamma)^{\alpha}])\,. \tag{4}$$

因此，元胞发生土地利用转变的规则为

$$\begin{cases}P(ij,t,h)=P_{\max}(ij,t), & ①\\ P(ij,t,h)\neq P_{\max}(ij,t), & ②\end{cases} \tag{5}$$

其中，规则①表示元胞 *ij* 在 *t* 时刻由当前地类转换成第 *h* 种地类；规则②表示当前地类不发生转变。

1.2 技术路线

本文研究工作的重点可分为以下三部分：1）驱动因子选择与数据预处理；2）模型构建及检验，即综合概率计算；3）迭代计算与预测。技术路线图如图 1 所示。

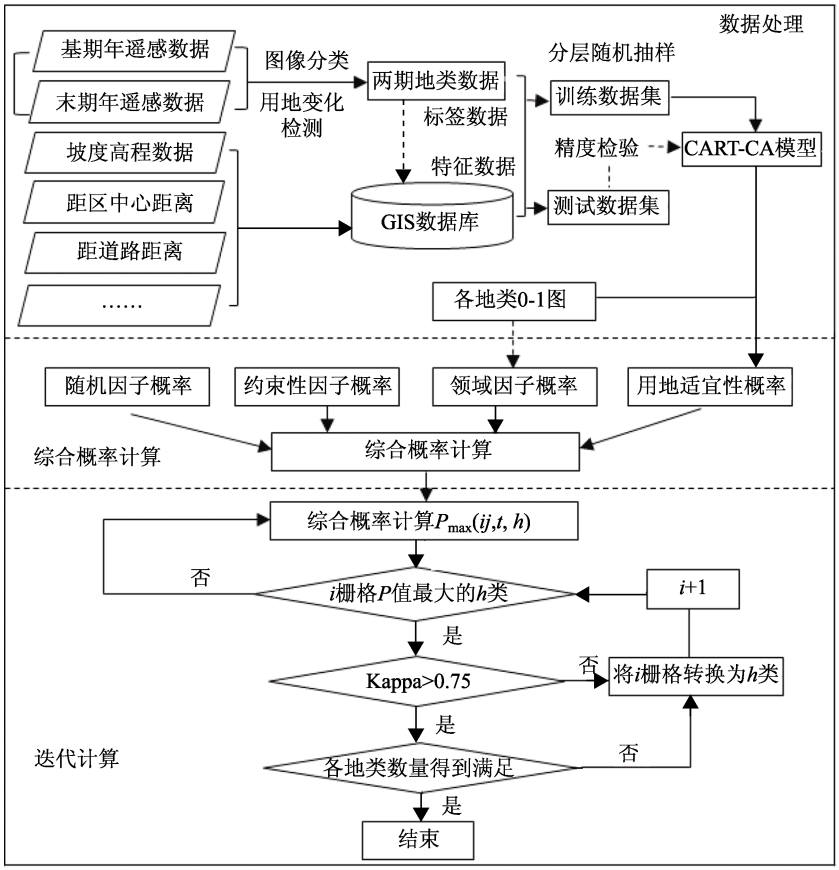

图 1 CART-CA 建模与研究技术路线图

Fig. 1 CART-CA modeling and research technical roadmap

2 研究区域与数据处理

2.1 研究区域概况

三峡库区是指三峡水利工程在设计的正常蓄水方案（175 m）下因淹没直接受影响的地区[23]，共涉及湖北省的 4 个县和重庆市的 23 个县（区），跨越大巴山及鄂西山地，位于 E105°4′～111°39′、N28°31′～31°44′之间，总面积约 7.9 万 km^2，总人口约 3 000 万人。三峡库区因其独特的自然条件使得该区域的土地利用、生态环境特征及社会经济发展状况与经济发达的平原地区有明显差异。当前我国土地利用/用地

覆被变化（land-use and land-cover change，LUCC）的研究区域大多侧重于经济发达或发展快速的城市拓展地区，因此关注生态环境相对脆弱、经济发展水平较低的三峡库区具有十分重要的现实意义[24]。研究区域的区位图如图 2 所示。

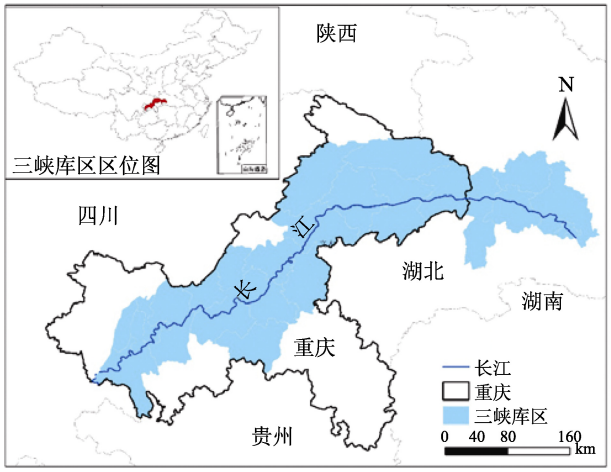

图 2　长江三峡库区位置示意图

Fig. 2　Location map of the Yangtze Three Gorges Reservoir area

2.2　数据来源与数据处理

本研究采用的用地变化数据来自全球 30 米分辨率地表覆盖数据（GlobeLand30）（http://www.globallandcover.com/home/background.aspx）。GlobeLand30 原始数据包括耕地、灌木地、林地、草地、水体、湿地、建设用地、冻土地带、裸地及永久雪地 10 种类型。在本文的研究区域中，用地数据包括 7 种地类，分别为耕地、林地、草地、灌木地、水体、湿地、建设用地。对数据进行分辨率统一处理，用三峡库区矢量图裁剪获取到 2000 年和 2010 年两期用地数据，地类分布图如图 3 所示。

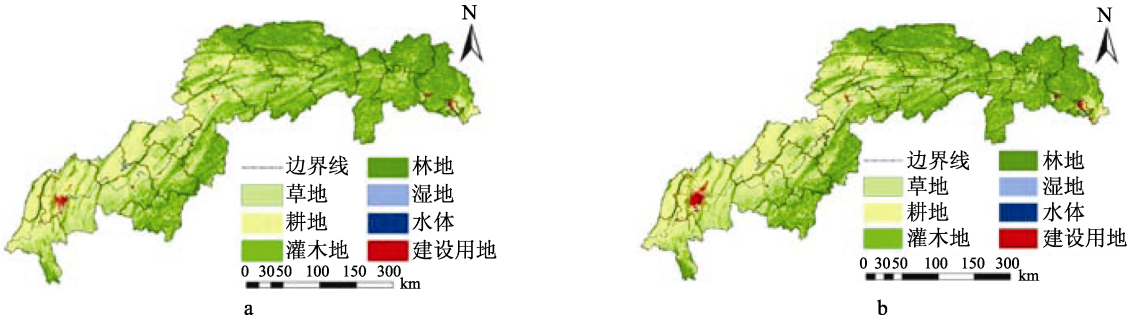

图 3　2000 年（a）和 2010 年（b）长江三峡库区土地利用类型分布图

Fig. 3　Distribution map of land use types in Yangtze Three Gorges Reservoir area in 2000 (a) and 2010 (b)

驱动因子数据选择高程、坡度、距河流距离（距主要河流距离、距其他河流距离）、距市场距离（距区和乡镇中心距离）、距道路距离（距铁路距离、距一级道路距离、距二级道路距离、距其他道路距离）。高程数据及衍生的坡度数据来自美国地质勘探局（United States Geological Survey，USGS）（https://hydrosheds.cr.usgs.gov/dataavail.php），分辨率为 90 m. 河流、道路和市场矢量数据来自自由街景（http://www.vdstech.com/osm-data.aspx）。上述数据经过投影变换、坐标系变换、裁剪等处理，得到空间数据集。

坡度采用 ArcGIS 中的坡度计算器计算，然后参考耕地坡度分级标准对研究区域进行坡度的重分类。坡度分为≤2°、2°～6°、6°～15°、15°～25°、25°～35°、35°～45°、45°～60°、60°～90°共 8 个坡度等级（上含下不含）。距离河流、道路和市场的距离采用 ArcGIS 的距离分析工具进行计算，生成相应的专题地图。其中，距乡镇中心距离首先在 http://www.86ditu.com/map/500000.html 网站上查找三峡区域乡、镇、街道中心经纬度坐标，形成 Excel 表格，接着通过添加 X、Y 数据，与 ArcGIS 关联，导出添加形成 SHP 点数据；最后借助欧氏距离分析工具生成相应的专题地图。三峡库区共涉及乡镇 604 个，其中街道 142 个，乡 124 个，镇 338 个。各距离驱动因子采用自然断点法分成 10 类，各类驱动因子空间数据示意图如图 4 所示。

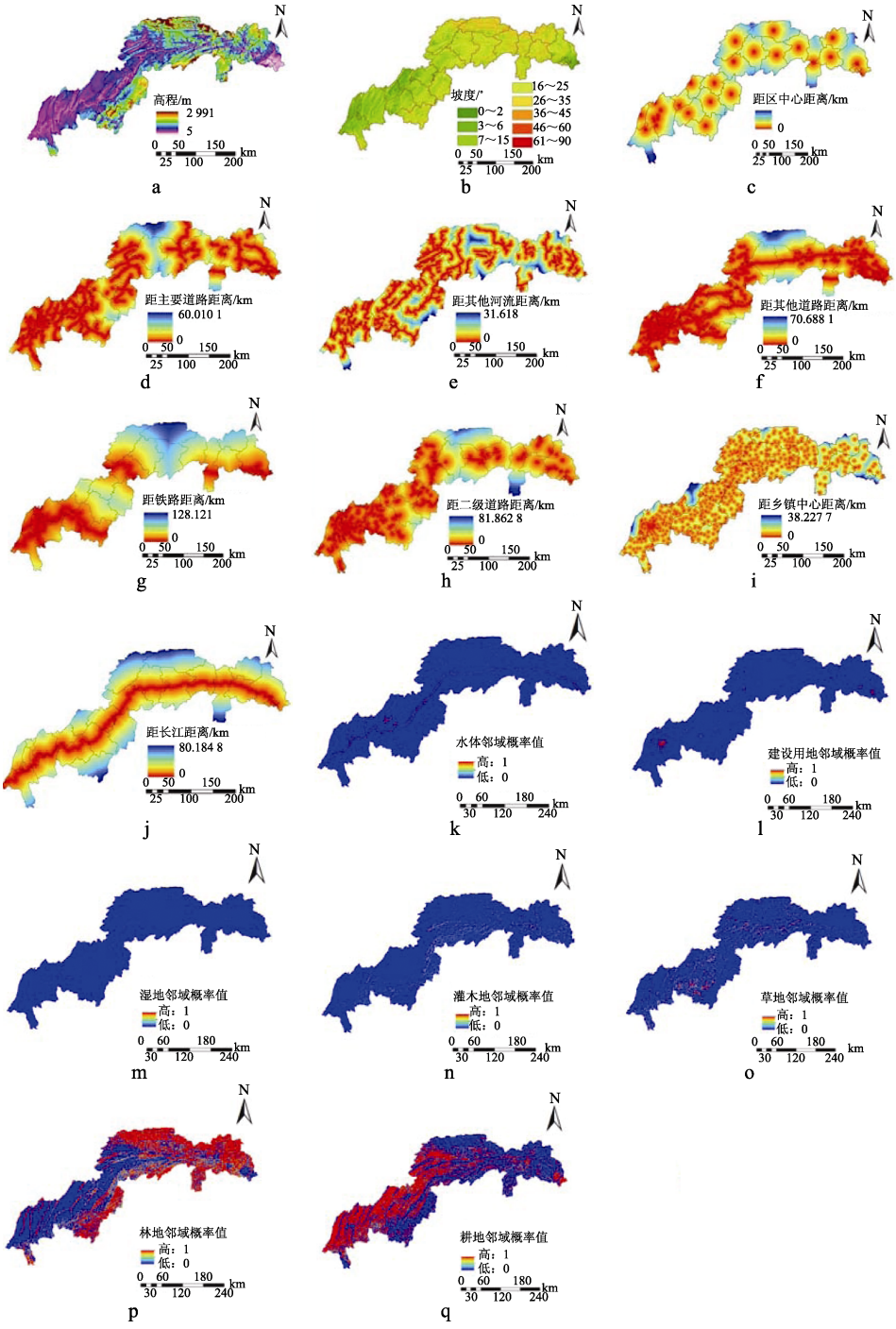

图 4 长江三峡库区各驱动因子空间分布示意图

Fig. 4 Spatial distribution map of each driving factor in the Yangtze Three Gorges Reservoir area

a—高程图；b—坡度图；c—距区中心距离图；d—距主要道路距离图；e—距其他河流距离图；f—距其他道路距离图；g—距铁路距离图；h—距二级道路距离图；i—距乡镇中心距离图；j—距长江距离图；k—水体邻域图；l—建设用地邻域图；m—湿地邻域图；n—灌木地邻域图；o—草地邻域图；p—林地邻域图；q—耕地邻域图

a-Elevation map; b-Slope map; b-Distance map from the county center; d-Distance map from main roads; e-Distance map from other rivers; f-Distance map from other roads; g-Distance map from railway; h-Distance map from secondary road; i-Distance map from township center; j-Distance from the Yangtze river; k-Water body neighborhood map; l-Construction land neighborhood map; m-Wetland neighborhood map; n-Shrub forest land neighborhood map; o-Grassland neighborhood map; p-Forest land neighborhood map; q-Cultivated land neighborhood map

3 实验

3.1 模型实现与检验

本研究所使用的编程环境为 Python 3.6.0，土地适宜性评价部分的概率通过调用 sklearn.tree 库下的 DecisionTreeClassifier 包计算得到。实验数据集以三峡库区驱动因子数据为特征数据，以 2000 年至 2010 年土地利用类型变化数据为标签数据；采用分层随机抽样方法抽取数据集的 70%作为训练数据集，将所有数据作为测试数据集。过程可简要概括为：1）导入数据并分层抽样得到训练数据集；2）使用 DecisionTreeClassifier()方法对训练数据进行拟合生成分类回归树（训练规则）；3）利用生成的分类回归树对测试集进行训练，即导入测试集于步骤 2）中的拟合结果中，得到用地适宜性概率。这里将测试集中各地类数据作为测试数据输入到生成的分类回归树模型中，得到地类转换的概率。

就某一栅格而言，用地分为几种类型，就有几种用地转换的可能性，即有几种转换概率。动态迭代计算过程中，首先将各地类转换概率与模型中的其他概率作乘积运算，得到每个栅格 7 种地类的概率；接着将概率最大的地类作为该栅格的地类。动态迭代完毕后，得到模拟结果。通过与 2010 年实际用地结果的对比，得到混淆矩阵，计算出总体精度和总体 Kappa 指数分别为 88.21%和 80.14%（如表 1 所示）. 通过对比研究，在同等情况下，DYNA-CLUE 和 Logistic-CA 模型的总体 Kappa 指数分别为 79.03%和 78.26%（如表 2 所示）.

表 1　长江三峡库区 2010 年实际用地和模拟用地分布混淆矩阵

Tab. 1　Confusion matrix of actual and simulated land use distributions in Yangtze Three Gorges Reservoir area in 2010

地类	草地	耕地	灌木地	林地	湿地	水体	建设用地
草地	47 713	2 093	67	2 052	3	60	38
耕地	357	327 444	42	2 231	2	100	137
灌木地	223	630	21 915	1 832	0	14	4
林地	717	3463	306	353 161	0	21	16
湿地	0	10	0	0	588	48	3
水体	42	668	3	44	7	9 401	31
建设用地	6	324	1	6	0	29	5 299
总体精度/%				88.21			
总体 Kappa 指数/%				80.14			

注：混淆矩阵中数据单位为个（栅格），横、纵项分别表示模拟和实际的地类数量，研究区域共 781 151 个栅格，分辨率为 300 m

表 2　不同分辨率下三种模型的模拟结果对比（%）

Tab. 2　Comparison of simulation results of three models at different resolutions (%)

模型	精度类型	分辨率/m				
		200	300	500	800	1 000
CART-CA	总体精度	87.89	88.21	85.76	82.46	81.21
	总体 Kappa 指数	79.98	80.14	79.06	77.57	77.03
Logistic-CA	总体精度	83.46	84.17	82.76	80.34	79.24
	总体 Kappa 指数	77.65	78.26	77.16	75.79	75.03
DYNA-CLUE	总体精度	85.02	85.89	83.37	81.68	79.31
	总体 Kappa 指数	78.38	79.03	77.83	76.65	75.42

上述实验结果表明，CART-CA 模型精度较高，可用于多用途土地利用格局变化的动态模拟。长江三峡库区 2010 年实际用地分布图和模拟用地分布图如图 5 所示。

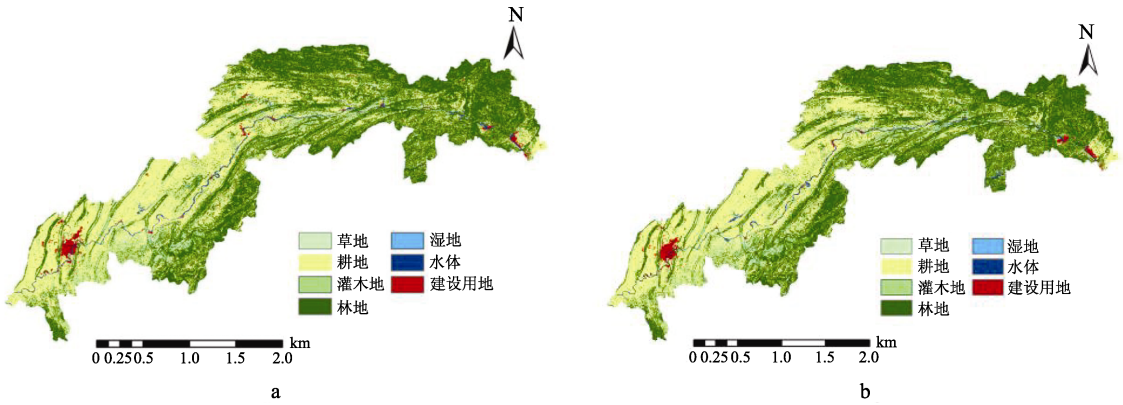

图 5 长江三峡库区 2010 年实际用地分布（a）与模拟用地分布（b）

Fig. 5 Actual land use distribution (a) and simulated land use distribution (b) in Yangtze Three Gorges Reservoir area in 2010

3.2 实验方案与结果分析

在上述 CART-CA 模型构建与验证的基础上，本研究设计了如下几种不同的实验方案：1）采用驱动因子逐步删减法检验驱动因子选择对模拟精度的影响；2）设计不同分辨率的数据来检验用地动态模型的尺度效应；3）设计多尺度、多分类方案，检验分类均衡性与模拟精度的关系以及分辨率与分类均衡性之间的关系。

3.2.1 驱动因子选择与影响力测度

在用地变化动态模型的构建过程中，邻域因子起着重要的作用[25]。有研究表明，邻域窗口大小的选择对模拟精度有一定的影响[26]。本研究设置不同的邻域窗口大小（3×3、5×5、7×7、9×9 和 11×11）进行了 5 组实验，发现最优邻域窗口为 3×3.

在对分类回归树进行训练后，通过 Scikit-learn 包下的 feature_importances_方法可得到该 CART 模型下驱动因子影响力（重要性）的度量，如表 3 所示。可以看出，高程和坡度两个因素对用地类型变化的影响最大，占所有因子影响力的 25.47%和 15.14%；其次是距其他河流和距乡镇中心距离，分别占到 10.10%和 9.55%；距长江、距区中心和距铁路距离等驱动因子也有一定的影响。

表 3 不同驱动因子的影响力（重要性）测度（%）

Tab. 3 Measurement of the importance of different driving factors (%)

驱动因子	重要性
高程	25.47
坡度	15.14
距区中心距离	7.57
距一级道路距离	6.57
距二级道路距离	5.82
距三级道路距离	5.35
距铁路距离	6.43
距长江距离	8.00
距其他河流距离	10.10
距乡镇中心距离	9.55

不同土地利用类型变化与驱动因子之间的解释能力也可通过 Logistic Regression 模型得到，但 Logistic Regression 模型只能处理线性相关的变量，而土地利用变化与驱动因子之间一般是非线性的关系，故利用 Logistic Regression 模型来测度驱动因子对土地利用变化的影响力并不是最合适的方式。CART 模型属于自然的非线性模型，能够更好地用来分析评价驱动因子的影响力。

采用驱动因子逐步删减法（逐步减去影响小的驱动因子）计算模拟精度，可以看出，当驱动因子减少到 5 个时，总体精度和总体 Kappa 指数开始出现较大幅度的下降；而驱动因子在 5 个以上及减少到 3 个时，随着驱动因子数量的减少，总体精度和总体 Kappa 指数基本上保持不变（如表 4 所示）。比较而言，选择前 5 个重要驱动因子进行模拟可达到既降低维度、减少计算量，又减少精度损失的效果。据此，可

根据不同驱动因子的影响力测度结果对同类变量通过加权方式达到既保持模拟精度不变，又降低维度的目的。在定量获取驱动因子影响力后，采用同类变量加权降维的方法与采用特尔菲法[27]降低维度相比，方法更可靠，这也是CART-CA这类用地变化动态模型的建模优势所在。

表4　驱动因子个数依次减少时的模拟结果（%）

Tab. 4　Simulation results when the number of driving factors successively reduces (%)

精度类型	驱动因子个数									
	10	9	8	7	6	5	4	3	2	1
总体精度	88.21	88.18	88.15	88.11	88.00	87.71	87.28	87.16	87.15	87.10
总体Kappa指数	80.14	80.08	80.00	79.92	79.67	78.91	77.71	77.33	77.30	77.21

3.2.2　尺度效应分析

分别采用200，300，500，800，1 000 m 5种分辨率的用地数据进行5次模拟实验（如表2所示）。以CART-CA模型为例，当分辨率由200 m降为300 m时，模拟精度小幅提高。随着尺度的继续降低，模拟精度也较大幅度地降低；当分辨率由800 m降到1 000 m时，精度降低较缓慢。实验表明，模拟尺度不同，模拟精度也会随之改变，但并不是简单的线性关系。就本研究区域而言，用地变化的最佳模拟尺度为300 m. 尽管本文所选数据分辨率的尺度间距较大，但基本能够反映尺度变化与模拟结果之间的关系。由表2可以看出，最佳模拟尺度出现在300 m附近，可在200～500 m之间采用小粒度尺度间距进行模拟，以得出更加精确的最佳模拟尺度。

通过与不同尺度下的DYNA-CLUE和Logistic-CA模型的比较发现，其最佳模拟尺度也是300 m（如表2所示）. 因此，就某一区域而言，不是分辨率越低，模拟精度越高，而会存在一个最佳分辨率使得模拟精度最高；远离这一分辨率，模拟精度会相应降低；且这一最佳分辨率不会因建模方法的不同而变化。但由表5的结果来看，最佳模拟尺度随着分类方案的不同而变化，当依次分7类、5类、3类、2类时，最佳模拟尺度依次为300，200，800，200 m，说明分类不均衡会引起最佳模拟尺度的变化。

表5　不同用地分类、不同分辨率下的均衡度及模拟精度评价结果

Tab. 5　Evaluation results of equilibrium degree and simulation accuracy under different land classifications and different resolutions

分辨率/m	均衡度（用地结构信息熵）				总体Kappa指数/%				总体精度/%			
	7类	5类	3类	2类	7类	5类	3类	2类	7类	5类	3类	2类
1 000	0.574 6	0.654 1	0.663 0	0.069 7	77.03	74.49	74.29	90.26	81.21	80.64	80.24	**95.28**
800	0.572 9	0.652 5	0.661 5	0.068 2	76.71	75.17	**75.83**	90.92	82.46	81.27	81.48	**95.83**
500	0.572 1	0.651 2	0.662 1	0.067 9	79.06	76.51	75.06	91.46	85.76	81.95	81.17	**962.4**
300	0.411 2	0.650 9	0.661 9	0.067 5	**80.14**	76.96	75.39	92.19	**88.21**	82.26	81.52	**96.79**
200	0.432 7	0.650 2	0.661 6	0.067 1	79.98	**77.57**	75.72	**92.86**	87.89	**82.79**	**81.87**	**97.27**

注：加粗部分表示不同分类方案下的最佳模拟尺度

3.2.3　多尺度与多分类方案的对比

在机器学习中，数据倾斜（分类的均衡性）通常被认为是引起学习精度的重要原因[28]。下面引入土地利用结构信息熵方法度量分类的不均衡性[29]，据此验证分类的均衡性对模拟精度的影响。

熵值可以反映系统的稳定性，熵值越小，系统有序度越高，稳定性越好。假设研究区域的总面积为S，该区域的地类共有n种，第i种地类的面积为S_i，其中i=1,2,⋯,n，那么有

$$S=\sum_{i=1}^{n}S_i\ . \tag{6}$$

各地类占总面积的百分比 P_i 可以表示为

$$P_i=\frac{S_i}{S}\ . \tag{7}$$

根据香农信息熵公式可以定义地类分布的信息熵为

$$H=-\sum_{i=1}^{n}(P_i\ln P_i)\ , \tag{8}$$

其中，H 为土地利用结构的信息熵（Nat）。信息熵的值直接描述区域土地利用结构的均衡度，当各地类面积相同时土地利用结构达到均衡态。土地利用结构的信息熵越低，反映出各类土地的均衡度越高，对于合理利用土地资源更为有利。一般来说，均衡度更能描述土地利用结构特征，其表达式为

$$J=H/H_{\max}=-\sum_{i=1}^{n}(P_i\ln P_i)/\ln n\ , \tag{9}$$

其中，J 为均衡度，即土地利用结构信息熵与最大熵的比值。当 J=1，也就是各类土地利用类型数量完全相等时，土地利用结构信息熵最大，土地利用结构处于最平衡的状态；当 J 接近于 0 时，表示土地利用结构极不均衡。

采用用地结构信息熵表达分类的均衡性，对研究区域不同分辨率的用地类型数据进行信息熵计算，结果如表 5 所示。

本文分别采用 200，300，500，800，1 000 m 分辨率对用地进行重分类，将地类分为 7 类、5 类、3 类、2 类 4 种分类方案，检验分类均衡性与模拟精度的关系以及分辨率与分类均衡性之间的关系。7 类用地，即草地、耕地、灌木地、林地、湿地、水体、建设用地；在 7 类用地的基础上，合并草地和灌木地为灌草地，合并湿地和水体为湿水地，即灌草地、耕地、林地、湿水地、建设用地共 5 类用地；在 7 类用地的基础上，合并耕地、湿地、水体为其他用地，合并林地、灌木地、草地为林灌草地，即林灌草地、建设用地、其他用地共 3 类用地；在 7 类用地的基础上，合并草地、耕地、灌木地、林地、湿地、水体为非建设用地，即建设用地、非建设用地共 2 类用地。在 5 种分辨率下对 4 种分类方案进行模拟实验，模拟精度评价结果如表 5 所示。

由表 5 可以发现，不同分辨率的总体 Kappa 指数都在 74%以上，同一分类体系中，随着分辨率的变化，总体 Kappa 指数的波动幅度不大，最大为 3.43%. 但同一分辨情况下，随着分类的变化，总体 Kappa 指数的波动幅度较大，最小波动幅度为 15.75%，最大波动幅度为 17.14%. 说明分类对精度产生的影响比尺度效应产生的影响更明显。分辨率与分类不均产生的叠加导致模拟精度变化最大为 18.57%.

总体 Kappa 指数与均衡度之间存在一定的关系，即均衡度升高，总体 Kappa 指数降低；相应地，均衡度降低，总体 Kappa 指数升高。当分 3 类时均衡度最高（300 m 分辨率的均衡度为 0.661 9），其次是分为 5 类时；当分为 2 类时均衡度最低（300 m 分辨率均衡度为 0.067 5）。以 300 m 分辨率为例，分 5 类与分 3 类时，均衡度相差 1.66%；分 7 类与分 3 类时，均衡度相差 37.88%；分 2 类与分 3 类时，均衡度相差 89.80%. 与此相对应，当分为 3 类或 5 类时，总体 Kappa 指数最低；分为 2 类时，总体 Kappa 指数最高。当分类数为 2 类时，均衡度极小，总体精度和总体 Kappa 指数极高。这时将所有用地类型分为建设用地和非建设用地，相对总数量而言，建设用地数量极少，均衡度也极小，表现出的分类效果极好。

总体上看，随着分辨率的降低，均衡度呈上升的趋势（但这种趋势规律并不针对所有分类情况，如

地类分类为7类、3类时并不满足此规律）。

分类数量与均衡度之间并非简单的线性关系，均衡度高低取决于各地类面积间的差异。当分类数为2类时，均衡度极小；当分类数为3类时，均衡度最高；当分类数为5类和7类时，均衡度介于上述两者之间。

4　结论

1）研究以长江三峡库区为例，构建了CART-CA用地变化动态模型。该模型由数据处理模块、综合概率计算模块和动态迭代模块构成。该模型构建的基本思想是：在综合概率计算的基础上，就某栅格而言，动态迭代模块采用概率最大优先转变为某地类原则进行用地转换，直到某地类数量得到满足。模拟结果与基期年用地类型比较，总体Kappa指数达到80.14%；同等情况下，DYNA-CLUE和Logistic-CA模型的总体Kappa指数分别为79.03%和78.26%. 研究表明该模型的模拟精度较高，可用于多用途土地利用格局变化的动态模拟。

2）采用特征重要性分析方法对驱动因子及用地变化结果进行分析。结果显示：研究区域10个驱动因子中高程和坡度两个因素对用地类型变化的影响最大，占所有因子影响力的25.47%和15.14%；其次是距其他河流和距乡镇中心距离，分别占到10.10%和9.55%. 在驱动因子重要性测度的基础上，可用驱动因子逐步删减法进行关键因子筛选以达到既降维、又减少精度损失的效果。在定量获取驱动因子影响力后，用同类变量加权降维的方法与采用特尔菲法降低维度相比，方法更可靠，这也是CART-CA这类用地变化动态模型的建模优势所在。

3）分别采用不同模型和不同分辨率数据对研究区域用地动态变化进行模拟，发现无论何种模型，其用地动态变化模型的尺度效应均明显。就某一区域而言，不是分辨率越低，模拟精度越高，而是存在一最佳分辨率使得模拟精度最高；远离这一分辨率，模拟精度会相应降低；而且这一最佳分辨率不会因为模型的变换而变化，但分类不均衡会引起最佳模拟尺度的变化。该规律的发现可为用地变化模型的基础数据处理、模型的构建提供指导。

4）土地利用结构信息熵方法被引入测度分类的不均衡性，发现分类均衡性对用地变化模型的模拟精度有重要影响。分类不均衡引起模拟精度的变化比因尺度变化引起的变化要大。总体上看，分辨率、不均衡性和模拟精度呈现如下关系：随着分辨率的降低，均衡度总体呈上升趋势；随着地类均衡度的升高，模拟精度降低。

[参考文献] (References)

[1] SANTÉ I, GARCÍA A M, MIRANDA D, et al. Cellular automata models for the simulation of real-world urban processes: a review and analysis[J]. Landscape and Urban Planning, 2010, 96(2): 108-122.

[2] 刘明皓，安广文，李超. 基于动态邻域思想的ACO-CA城市动态模拟——以重庆市沙坪坝区为例[J]. 地理与地理信息科学，2016，32（3）：74-80.
LIU M H, AN G W, LI C. Urban land use simulation based on ACO-CA model embedded dynamic neighborhood theory: a case study of Shapingba district in Chongqing[J]. Geography and Geo-information Science, 2016, 32(3): 74-80. (in Chinese)

[3] 曹敏，史照良. 基于遗传神经网络获取元胞自动机的转换规则[J]. 测绘通报，2010，24（3）：24-27.
CAO M, SHI Z L. Transition rule for GANN-CA[J]. Bulletin of Surveying and Mapping, 2010, 24(3): 24-27. (in Chinese)

[4] 柯新利，边馥苓. 基于C5.0决策树算法的元胞自动机土地利用变化模拟模型[J]. 长江流域资源与环境，2010，19（4）：403-408.

KE X L, BIAN F L. Land-ues change CA model based on C5.0 decision tree[J]. Resources and Environment in the Yangtze Basin, 2010, 19(4): 403-408. (in Chinese)

[5] 陈凯，刘凯，柳林，等. 基于随机森林的元胞自动机城市扩展模拟——以佛山市为例[J]. 地理科学进展，2015，34（8）：937-946.
CHEN K, LIU K, LIU L, et al. Urban expansion simulation by random-forest-based cellular automata: a case study of Foshan city[J]. Progress in Geography, 2015, 34(8): 937-946. (in Chinese)

[6] ABDELLAH I, AUGUSTE G, IOULIA T, et al. Scale effect challenges in urban hydrology highlighted with a distributed hydrological model[J]. Hydrology and Earth System Sciences, 2018, 22(1): 331-350.

[7] LUO X B, PENG Y D. Scale effects of the relationships between urban heat islands and impact factors based on a geographically-weighted regression model[J]. Remote Sensing, 2016, 8(9): 760.

[8] HUANG Q P, HUANG J J, YANG X N, et al. Evaluating the scale effect of soil erosion using landscape pattern metrics and information entropy: a case study in the Danjiangkou reservoir area, China[J]. Sustainability, 2017, 9(7): 1243.

[9] 陈睿山，蔡运龙. 土地变化科学中的尺度问题与解决途径[J]. 地理研究，2010，29（7）：1244-1256.
CHEN R S, CAI Y L. Progress in the study of scale issues in land change science[J]. Geographical Research, 2010, 29(7): 1244-1256. (in Chinese)

[10] 徐军亮，章异平. 土地利用中尺度效应研究进展[J]. 安徽农业科学，2009，37（14）：6776-6777 .
XU J L, ZHANG Y P. Progress of scale effects in land use[J]. Journal of Anhui Agricultural Sciences, 2009, 37(14): 6776-6777. (in Chinese)

[11] 郝仕龙，李壁成. 土地利用的尺度和尺度转换[J]. 中国土地科学，2004，18（5）：32-36.
HAO S L, LI B C. Scale and scaling in land-use analysis[J]. China Land Science, 2004, 18(5): 32-36. (in Chinese)

[12] 吴桂平. 不同尺度转换方式对土地利用格局模拟的影响效应研究[J]. 水土保持研究，2010，17（5）：75-79.
WU G P. Exploring effect of different scaling methods on simulation of land use patterns[J]. Research of Soil and Water Conservation, 2010,17(5): 75-79. (in Chinese)

[13] 胡云峰，徐芝英，刘越，等. 地理空间数据的尺度转换[J]. 地球科学进展，2013，28（3）：297-304.
HU Y F, XU Z Y, LIU Y, et al. A review of the scaling issues of geospatial data[J]. Advances in Earth Science, 2013, 28(3): 297-304. (in Chinese)

[14] 徐芝英，胡云峰，刘越，等. 空间尺度转换数据精度评价的准则和方法[J]. 地理科学进展，2012，31（12）：1574-1582.
XU Z Y, HU Y F, LIU Y, et al. A review on the accuracy analysis of spatial scaling data[J]. Progress in Geography, 2012, 31(12): 1574-1582. (in Chinese)

[15] ROBINSON W S. Ecological correlations and the behavior of individuals[J]. American Sociological Review, 1950, 15(3): 351-357.

[16] MA L, LI M C, THOMAS B, et al. Object-based change detection in urban areas: the effects of segmentation strategy, scale, and feature space on unsupervised methods[J]. Remote Sensing, 2016, 8(9): 761.

[17] 陶舒，周旭，程滔，等. 地理国情普查数据的空间尺度转换研究[J]. 测绘科学，2015，40（11）：89-95.
TAO S, ZHOU X, CHENG T, et al. Spatial scaling transformation of data from national geographical census[J]. Science of Surveying and Mapping, 2015, 40(11): 89-95. (in Chinese)

[18] 邓祥征，战金艳. 中国北方农牧交错带土地利用变化驱动力的尺度效应分析[J]. 地理与地理信息科学，2004，20（3）：64-68.
DENG X Z, ZHAN J Y. Scale-effect analysis of LUCC driving forces in the farming-pasturing interlocked area in northern China[J]. Geography and Geo-information Science, 2004, 20(3): 64-68. (in Chinese)

[19] 邱炳文. 福建省龙海市土地利用空间分布影响因子的尺度效应分析[J]. 自然资源学报，2007，22（1）：70-78.
QIU B W. Scale effect analysis of driving forces of land-use patterns in Longhai city of Fujian province[J]. Journal of Natural Resources, 2007, 22(1): 70-78. (in Chinese)

[20] 刘明皓，田降生，尚尉. 基于 DBN 与 GeoCA 相结合的城市动态模型构建与多方案对比研究——以成都市为例[J]. 西南大学学报（自然科学版），2018，40（1）：140-148.
LIU M H, TIAN J S, SHANG W. The construction of an urban dynamic model based on DBN in combination with GeoCA and a multi-program comparison research: a case research of Chengdu[J]. Journal of Southwest University (Natural Science Edition), 2018, 40(1): 140-148. (in Chinese)

[21] 黎夏，刘小平，李少英. 智能式 GIS 与空间优化[M]. 北京：科学出版社，2010.
LI X, LIU X P, LI S Y. Intelligent GIS and space optimization[M]. Beijing: Science Press, 2010. (in Chinese)

[22] 张大川，刘小平，姚尧，等. 基于随机森林 CA 的东莞市多类土地利用变化模拟[J]. 地理与地理信息科学，2016，32（5）：29-36.
ZHANG D C, LIU X P, YAO Y, et al. Spatiotemporal change of multiple land use types in Dongguan by using random forest based on cellular automat[J]. Geography and Geo-information Science, 2016, 32(5): 29-36. (in Chinese)

[23] 邓华，邵景安，王金亮，等. 多因素耦合下三峡库区土地利用未来情景模拟[J]. 地理学报，2016，71（11）：1979-1997.
DENG H, SHAO J A, WANG J L, et al. Land use driving forces and its future scenario simulation in the Three Gorges Reservoir area using CLUE-S model[J]. Acta Geographica Sinica, 2016, 71(11): 1979-1997. (in Chinese)

[24] 翟俨伟. 三峡库区生态环境面临的主要问题及治理对策[J]. 焦作大学学报，2012，26（1）：90-93.
ZHAI Y W. Main problems and countermeasures of ecological environment in Three Gorges Reservoir area[J]. Journal of Jiaozuo University, 2012, 26(1): 90-93. (in Chinese)

[25] 刘明皓，陶媛，夏保宝，等. 邻域因子对城市土地开发强度模拟效果的影响分析——基于 BP 人工神经网络模拟的结果对比[J]. 西南师范大学学报（自然科学版），2014，39（2）：39-46.
LIU M H, TAO Y, XIA B B, et al. On neighborhood factor impact upon simulation effect of urban land development intensity: comparative analysis of simulation results based on BP artificational neural network method[J]. Journal of Southwest China Normal University (Natural Science Edition), 2014, 39(2): 39-46. (in Chinese)

[26] 刘明皓，李东鸿. ANN-CA 模型改进及城市用地动态模拟[J]. 重庆邮电大学学报（自然科学版），2014，26（5）：720-726.
LIU M H, LI D H. Simulation of urban land use change based on the improved ANN-CA[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2014, 26(5): 720-726. (in Chinese)

[27] 郭荣中，杨敏华. 基于信息熵的长株潭区域土地利用结构分析[J]. 农业现代化研究，2013，34（1）：73-76.
GUO R Z, YANG M H. Studies on land use structures of city group based on information entropy[J]. Research of Agricultural Modernization, 2013, 34(1): 73-76. (in Chinese)

[28] 苏金树，张博锋，徐昕. 基于机器学习的文本分类技术研究进展[J]. 软件学报，2006，17（9）：1849-1859.
SU J S, ZHANG B F, XU X. Advances in machine learning based text categorization[J]. Journal of Software, 2006, 17(9): 1849-1859. (in Chinese)

[29] 谢汀，伍文，高雪松，等. 基于信息熵与偏移-份额模型的建设用地结构变化分析——以成都市为例[J]. 资源科学，2014，36（4）：722-730.
XIE T, WU W, GAO X S, et al. Changes in construction land structure for Chengdu based on an information entropy model and shift-share model[J]. Resources Science, 2014, 36(4): 722-730. (in Chinese)

（责任编辑：刘楠）