

## 目录

1 ca 模块介绍.....	2
2 使用步骤 .....	2
2.1 主配置界面参数设置 .....	2
2.2 土地利用类型和栅格数值对应关系设置.....	5
3.3 转换控制矩阵设置 .....	8
2.4 输出界面 .....	8
2.5 kappa 系数计算界面.....	10
3.案例 .....	12
3.1 示例数据介绍 .....	12
3.2 验证结果 .....	13
3.2.1 随机森林 ca.....	14
3.2.2 神经网络 ca.....	15
4.2.3 决策树 ca.....	16
3.2.4 logistic ca.....	17

# 1 ca 模块介绍

GeoSOS 的元胞自动机(ca)土地利用变化模拟模块，是在 GeoSOS 系统上实现了常用的 ca 算法，kappa 精度计算等功能。能够满足用户一般的 ca 使用需求。

ca 模块分为三个子模块，c++ ca 模块，c# ca 模块，算法模块，其中最重要的是 c# ca 模块。

- 1) c# ca 模块，细分为多类土地利用模拟和二类土地利用模拟模块。
- 2) 算法模块，提供了 kappa 系数计算和 logistic 概率图层生成模块。
- 3) c++ ca 模块是为了提高运行效率的实验模块，它和 c#模块相比，缺少了模拟结果动态展示的功能。

表 1 c# ca 模块

大类	小类
多类土地利用变化模拟	随机森林 ca
	神经网络 ca
二类土地利用变化模拟	Logistic ca
	决策树 ca
	多准则 ca（ahp 层次分析法）

## 2 使用步骤

本部分以随机森林 ca 的使用介绍具体的使用步骤。

### 2.1 主配置界面参数设置

点击“RandomForest”按钮，将弹出一个配置界面。

The image shows a software window titled "RandomForestSetUpForm". It is divided into three main sections:

- 起始和终止年份数据 (Start and End Year Data):** Contains two text input fields labeled "起始年份影像" (Start Year Image) and "终止年份影像" (End Year Image), each with a "打开" (Open) button to its right.
- 驱动数据 (Driving Data):** Contains a list box on the right and two buttons on the left: "添加" (Add) and "移除" (Remove).
- 参数 (Parameters):** Contains several input fields:
  - 采样数目 (Sampling Number): 10000
  - 模拟次数 (Simulation Times): 10
  - 邻域大小 (Neighborhood Size): 5
  - 随机因子 (Random Factor): 2
  - 树数目 (Number of Trees): 100
  - 样本使用率(0-1) (Sample Usage Rate): 0.66
  - 是否计算变量重要性 (Whether to calculate variable importance): ☐
  - 城市发展概率修正 (Urban Development Probability Correction): 0
  - 目标城市栅格数目 (Target Urban Grid Count): 2

At the bottom, there are two buttons labeled "设置" (Settings) under the headings "土地利用类型对应关系" (Land Use Type Correspondence) and "转换控制矩阵" (Conversion Control Matrix), and a large "确定" (Confirm) button at the bottom right.

图 2-1 ca 设置界面 （初始）

配置界面分为四个小部分。分别为，土地利用数据设置，驱动数据设置，参数设置，元数据设置。

在土地利用数据设置区，用户点击文本框右侧的按钮，在文件系统中选择两期土地利用数据，或者直接在文本框中输入数据文件的路径。

在驱动数据设置区，用户通过点击左侧的“添加”按钮，在文件系统中选择驱动数据文件。如果用户错误添加了数据文件，也可以先在右边的列表框中选中要删除的项目，然后点击“移除”按钮移除。在参数设置区，用户需要根据 ca 的类型，研究区域的特点，凭借其它的一些先验知识输入参数。

表 2-1 参数说明（这里说明了所有参数，并非每个模型都需要使用全部参数）

参数名称	说明
采样数目	模型的训练样本数目，采样使用分层等比采样法
模拟次数	元胞自动机模拟的控制参数，控制模拟的最大迭代次数
邻域大小	元胞自动机的邻域半径
随机因子	随机因子决定要在模型中加入多大的随机性
目标城市栅格数目	元胞自动机模拟的控制参数，当城市栅格的数目超过设定值，模型运行终止
树数目	随机森林模型使用的 CART 决策树数目
样本使用率	随机森林训练每一棵决策树时随机采样样本的比例
是否计算变量重要性	如果勾选，模型将使用随机森林的 OOB 数据来计算变量的重要性。这一项将重复训练和变量数目等同的随机森林模型，使模型的训练时间变长
城市发展概率修正	允许用户根据其它知识输入对城市发展概率的修正，它将在模型的归一化概率上，对城市发展概率进行修改，并且归一化其它概率
训练次数	神经网络的训练次数，本模型使用的是 3 层的 BP 神经网络，使用梯度下降的方法来优化权重，训练次数代表优化的次数，一般训练到一定次数后，正确率将稳定，用户可以通过设定一个较大的值来实验训练次数
转换阈值	控制土地利用类型的变化阈值，较低的转换阈值会使模型快速稳定并结束运行

The screenshot shows a software window titled "RandomForestSetUpForm" with several sections for configuring a random forest model:

- 起始和终止年份数据 (Start and End Year Data):**
  - 起始年份影像 (Start Year Image): C:\tasks\geosos\GEOSOS201705, with an "打开" (Open) button.
  - 终止年份影像 (End Year Image): C:\tasks\geosos\GEOSOS201705, with an "打开" (Open) button.
- 驱动数据 (Driver Data):**
  - Buttons: "添加" (Add) and "移除" (Remove).
  - Text area containing a list of file paths:
 

```
C:\tasks\geosos\GEOSOS20170505\ca_data\city
C:\tasks\geosos\GEOSOS20170505\ca_data\high
C:\tasks\geosos\GEOSOS20170505\ca_data\rail
C:\tasks\geosos\GEOSOS20170505\ca_data\road
C:\tasks\geosos\GEOSOS20170505\ca_data\town
```
- 参数 (Parameters):**
  - 采样数目 (Sampling Number): 10000
  - 样本使用率 (0-1) (Sample Usage Rate): 0.66
  - 模拟次数 (Simulation Times): 300
  - 是否计算变量重要性 (Calculate Variable Importance): ☒
  - 邻域大小 (Neighborhood Size): 5
  - 城市发展概率修正 (Urban Development Probability Adjustment): 0
  - 随机因子 (Random Factor): 2
  - 目标城市栅格数目 (Target City Grid Count): 100000
  - 树数目 (Number of Trees): 100
- 底部控制:**
  - 土地利用类型对应关系 (Land Use Type Correspondence): 设置 (Settings)
  - 转换控制矩阵 (Conversion Control Matrix): 设置 (Settings)
  - 确定 (Confirm)

图 2-2 ca 设置界面(结果)

## 2.2 土地利用类型和栅格数值对应关系设置

在元数据设置去，用户需要设置“土地利用类型和栅格数值的对应表”告知系统和栅格值对应的土地利用类型。点击按钮后，将弹出新的设置界面：

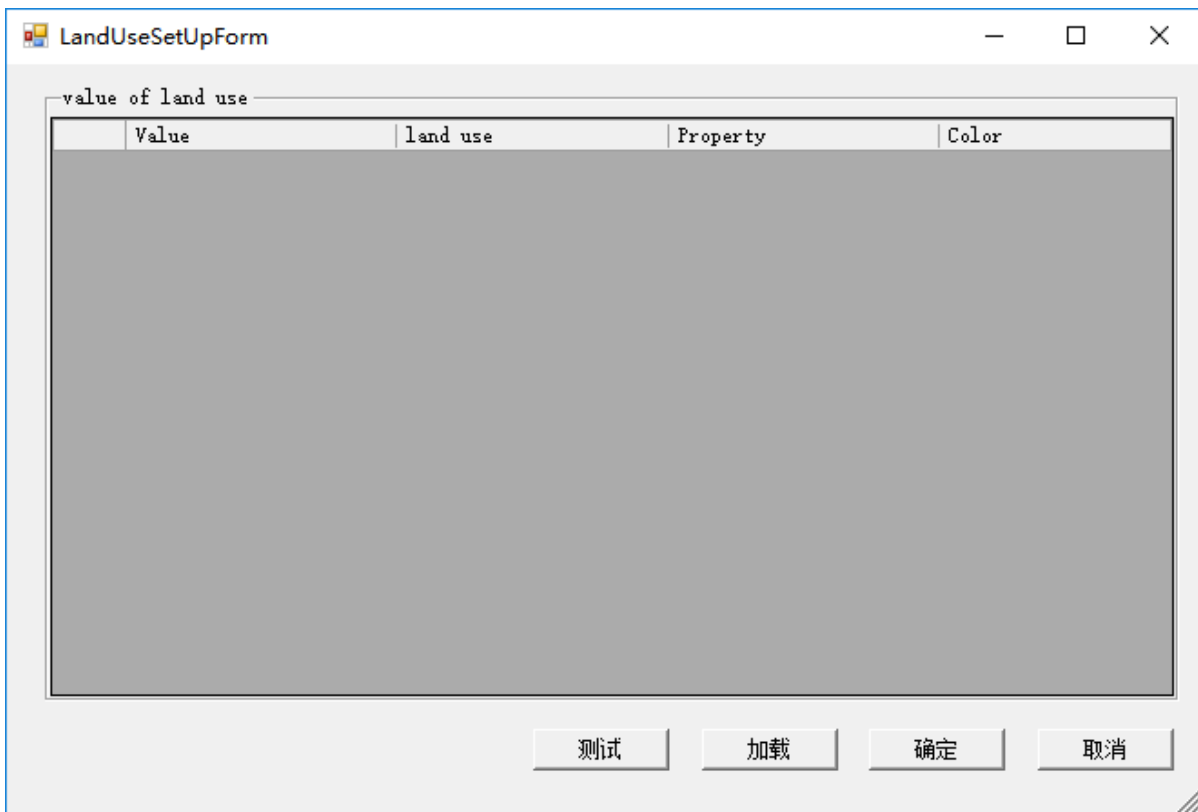


图 2-3 土地利用类型设置界面（初始）

在这个界面，首先点击加载按钮，在文件系统中选择文件加载土地利用数据。

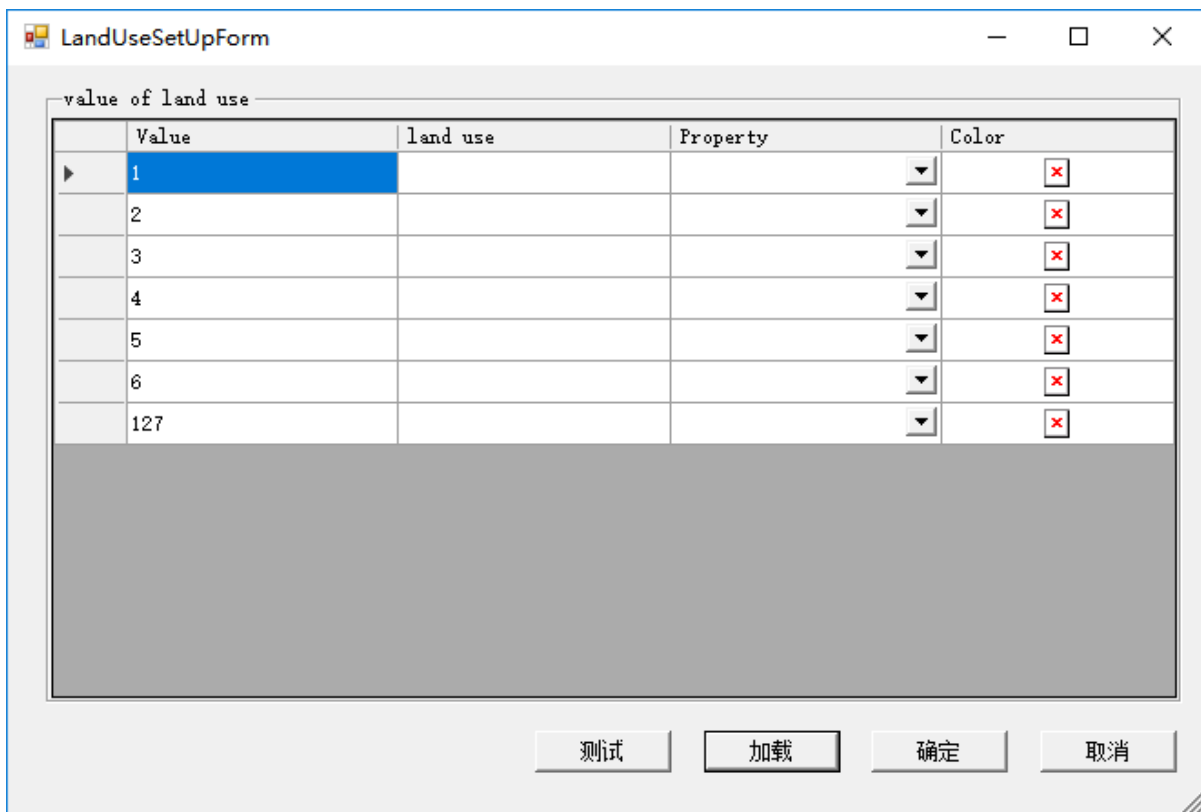


图 2-4 土地利用类型设置界面（数据加载后）

这个界面枚举了输入的土地利用数据中的栅格数值，用户需要在“land use”列，输入土地利用类型，在“Property”列中选择土地利用的性质，在“Color”列中选择土地类型的显示颜色。

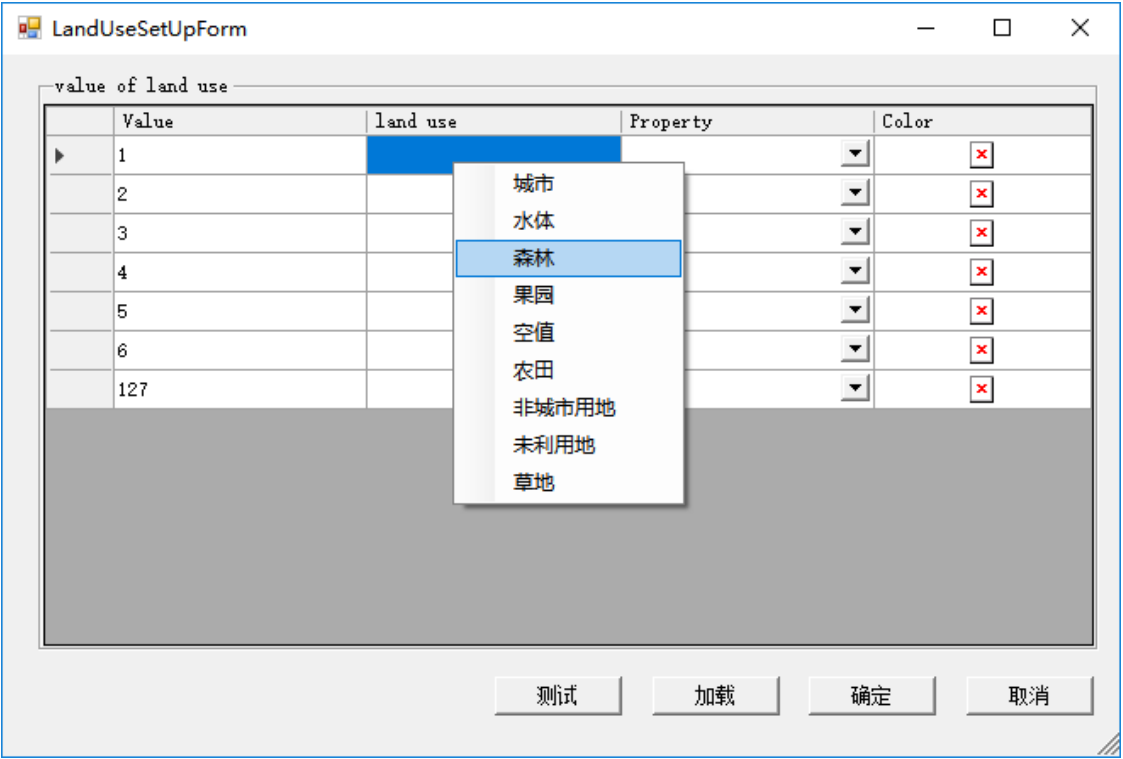


图 2-5 土地利用类型设置界面（操作）

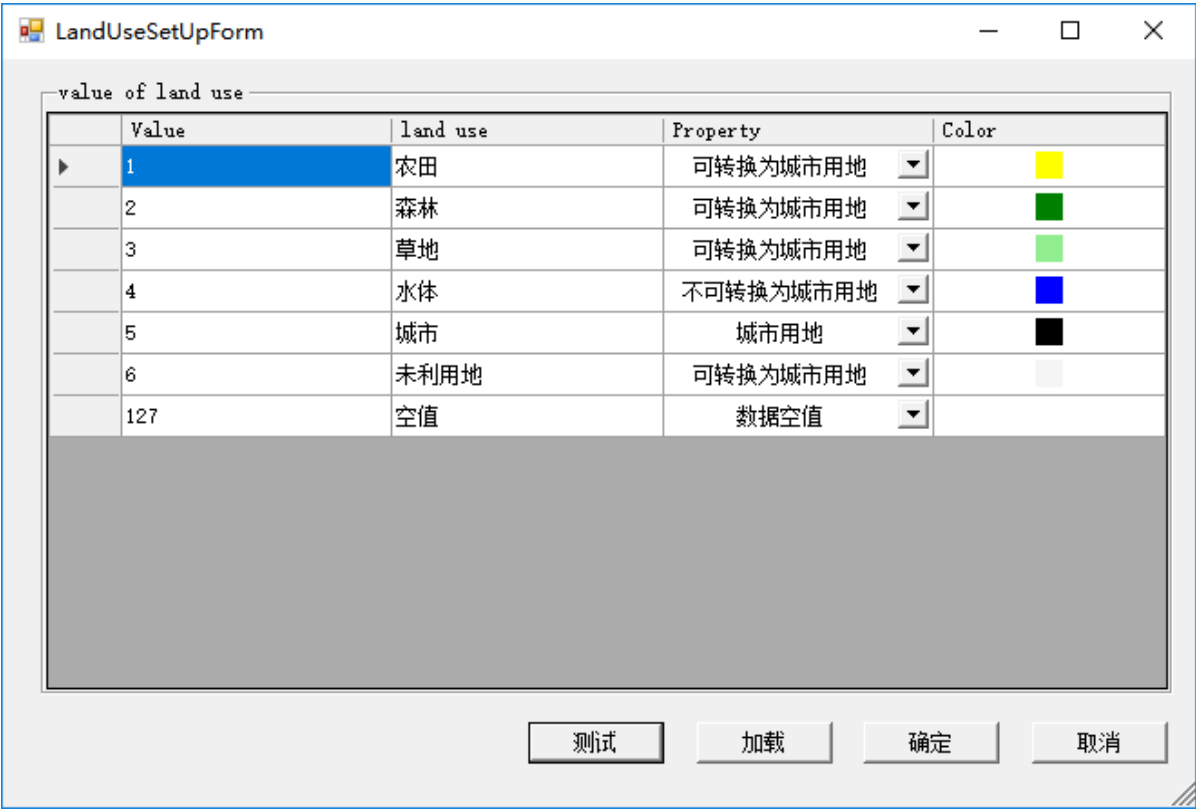


图 2-6 土地利用类型设置界面（结果）

用户完成相应的设置后，点击“确定”按钮，完成设置。

3.3 转换控制矩阵设置

完成上述设置后，点击“转换矩阵设置”按钮将弹出“转换控制矩阵设置界面”，这个界面，让用户根据先验知识，设置各种土地利用类型之间是否可以互相转换，‘1’代表允许转换，‘0’代表不允许转换。

TransformControlForm

	农田	森林	草地	水体	城市	未利用地
▶ 农田	1	1	1	1	1	1
森林	1	1	1	1	1	1
草地	1	1	1	1	1	1
水体	1	1	1	1	1	1
城市	1	1	1	1	1	1
未利用地	1	1	1	1	1	1
*						

确定

图 2-7 转换矩阵设置界面

**注意：**由于这个矩阵的设置使用了土地利用信息，所以需要先完成前面的“土地利用类型和栅格数值的对应关系”设置。

2.4 输出界面

点击确定以后即开始模拟，弹出模拟界面。



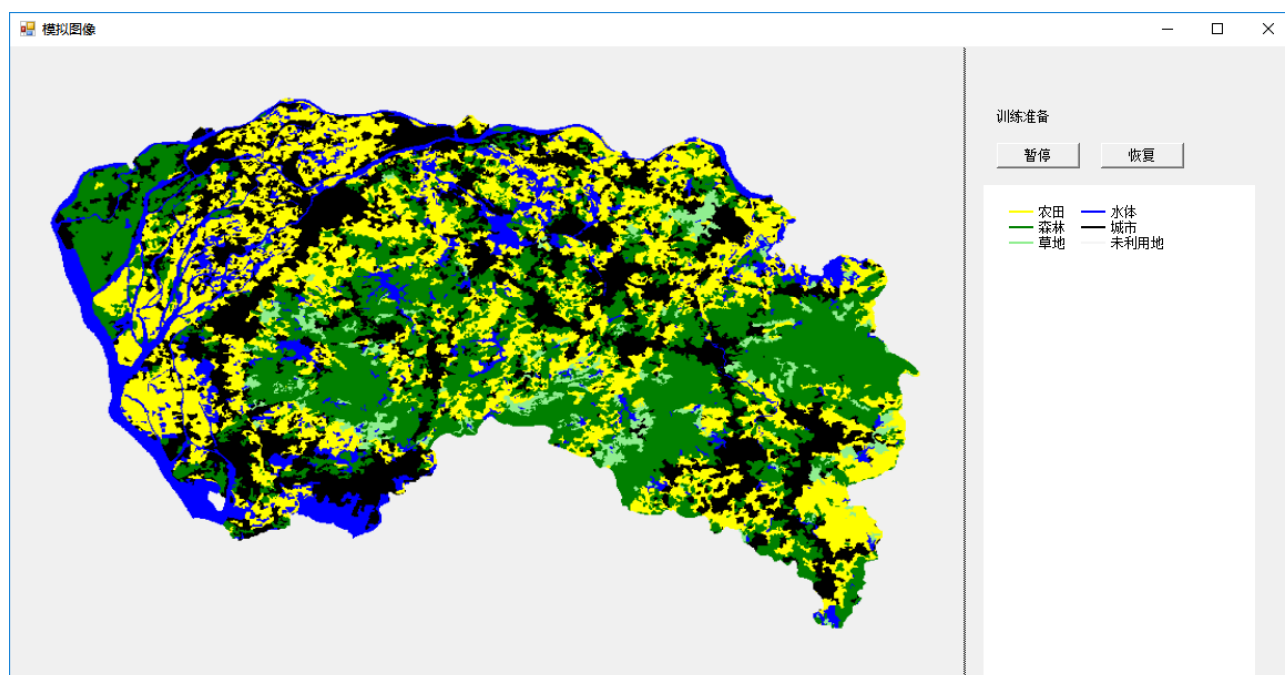


图 2-8 模拟监控界面（训练中）

在模拟界面上显示“训练准备”，代表正在进行模型训练，用户可以在控制台界面观察训练的进度。

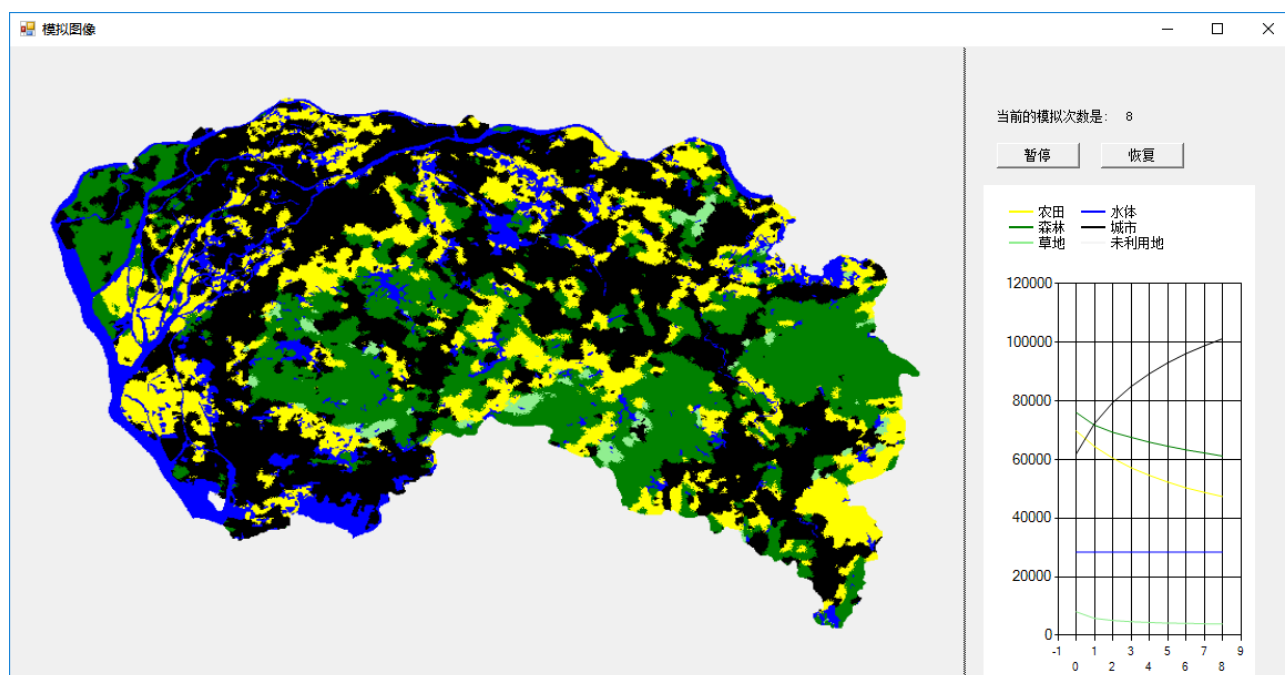


图 2-9 模拟监控界面(运行中)

开始模拟后，界面上将显示当前模拟的次数，左边的栅格图像将随着模拟进行变化，右边的图表也将随着模拟进行变化。用户可以观察模拟的状况。

```
起始的城市数目:61920
终止的城市数目:107554
-----
平均误差5.30106666666734%
变量1C:\tasks\geosos\GEOSOS20170505\ca_data\city_dist.tif的重要性分数oobrmerror是0.0207422815986332
变量2C:\tasks\geosos\GEOSOS20170505\ca_data\highway_dist.tif的重要性分数oobrmerror是0.00801648776344738
变量3C:\tasks\geosos\GEOSOS20170505\ca_data\railway_dist.tif的重要性分数oobrmerror是0.009727589969378
变量4C:\tasks\geosos\GEOSOS20170505\ca_data\road_dist.tif的重要性分数oobrmerror是0.00292972321128515
变量5C:\tasks\geosos\GEOSOS20170505\ca_data\town_dist.tif的重要性分数oobrmerror是0.00502157640159834
---当前城市数目: 72354
转化了:25693
---当前城市数目: 79560
转化了:16791
---当前城市数目: 85004
转化了:12560
---当前城市数目: 89424
转化了:10309
---当前城市数目: 93132
转化了:8862
---当前城市数目: 96292
转化了:7945
---当前城市数目: 98990
转化了:7215
---当前城市数目: 101428
转化了:6689
模拟结束
kappa值 : 0.73238667841194
```

图 2-10 控制台界面

在模拟进行的同时，控制台上也将打印相应的信息。平均误差是随机森林的分类误差率。如果用户勾选了计算变量重要性，将显示每个变量的重要性分数，用户可以据此对每个变量对土地利用类型的重要性做出判断，数值代表 OOB 数据的 RMSE 的减少值，数值越大说明变量越重要。当前城市数目代表当前模拟周期的城市栅格数目。转换数目，代表当前模拟周期土地利用类型发生了变化了的栅格数目。

模拟结束后，数据将自动加载到系统中，用户可以使用系统的其它空间分析功能进行进一步研究。同时数据也会保存到输出目录，数据包含有地理变化和空间参照信息，可以直接在其它通用的地理信息系统中使用。

## 2.5 kappa 系数计算界面

使用系统提供的 kappa 系数计算模块来计算精度。点击菜单上的 kappa 按钮弹出设置界面。



图 2-11 kappa 系数计算界面（初始）

这个界面的设置分成两部分，分别是输入设置模块，和元数据设置模块。和元胞自动机的设置界面相同。

用户选择对照的验证图像和模型输出的模拟图像，然后设置土地利用类型和栅格数值对应表之后，点击生成即可生成 kappa 系数，总体进度，FoM 和混淆矩阵。

KappaSetUpForm

数据

真实影像

C:\tasks\geosos\GEOSOS201705

打开

模拟影像

C:\tasks\geosos\GEOSOS201705

打开

土地利用类型对应关系

设置

生成

kappa: 0.72392912904566

正确率: 0.727531618378144

FoM: 0.608527681028632

(注: 行代表真实情况, 列表模拟情况)

	农田	森林	草地	水体	城市	未利用地
农田	25525	4439	226	265	14155	0
森林	3815	44423	986	184	11269	0
草地	586	2559	2283	34	857	0
水体	560	286	12	23982	246	150
城市	12905	8487	649	3921	81592	0
未利用地	0	0	0	0	0	3
*						

图 2-12 kappa 系数计算界面（结果）

### 3.案例

#### 3.1 示例数据介绍

表 3-1 示例数据清单

类型	文件名	数据说明	用途
土地利用数据	dg2000.tif	2000 年东莞市 100 米分辨率土地利用数据	初始年份土地利用数据，模型输入
	dg2005.tif	2005 年东莞市 100 米分辨率土地利用数据	在神经网络 ca 中它是网络训练数据，在所有 ca 中用于 kappa 系数计算
土地利用变化驱动数据	city_dist.tif	到市中心距离	计算适应性概率
	town_dist.tif	到城镇中心距离	
	railway_dist.tif	到铁路距离	
	road_dist.tif	到主干道距离	
	highway_dist.tif	到高速公路距离	

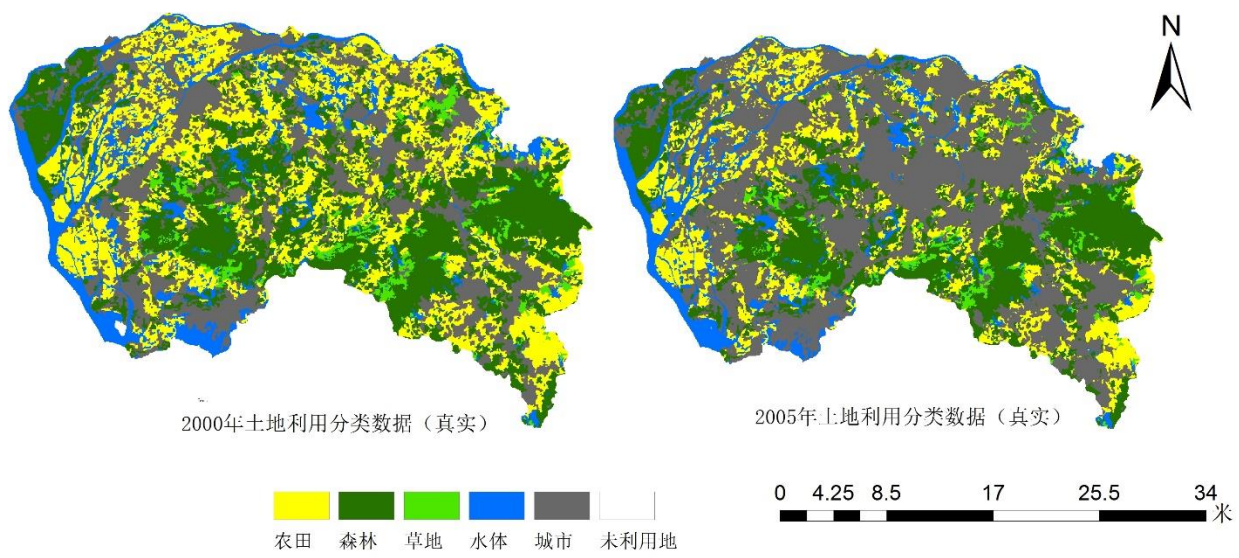


图 3-1 2000 年和 2005 年真实土地利用分类数据

3.2 验证结果

3.2.1 随机森林 ca

参数设置:

表 3-2 随机森林 ca 参数设置

参数名称	参数值
采样数目	10000
模拟次数	100
邻域大小	5
随机因子	2
目标城市栅格数目	110000
树数目	100
样本使用率	0.66
城市发展概率修正	0

混淆矩阵:

表 3-3 随机森林 ca 混淆矩阵

年份	土地利用类型	模拟						
		农田	森林	草地	水体	城市	未利用地	
2010	实际	农田	24430	4409	236	265	15270	0
		森林	3957	44051	744	184	11741	0
		草地	573	2589	2285	34	838	0
		水体	540	267	14	23982	317	116
		城市	12194	8046	627	3921	82766	0
		未利用地	0	0	0	0	0	3
	kappa	0.72						
	整体精度	0.73						
	FoM	0.61						

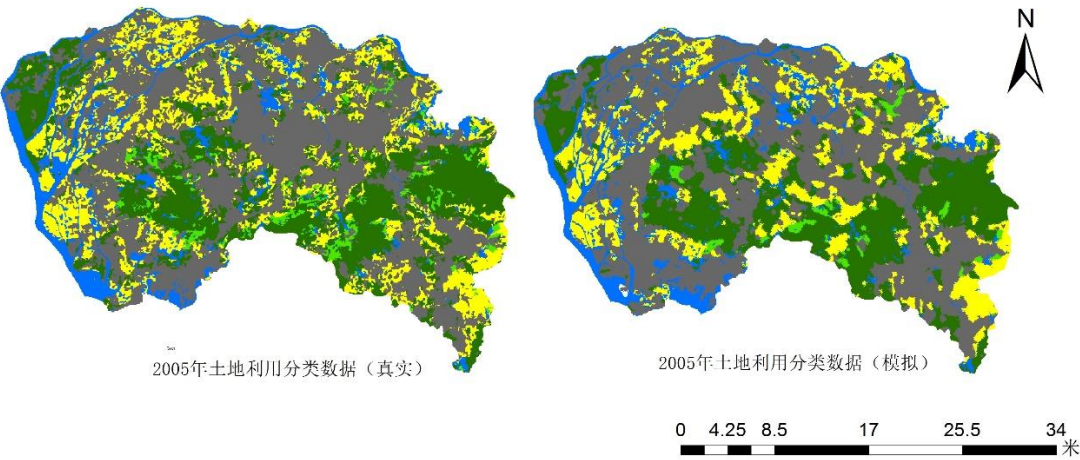


图 3-2 随机森林 ca 模拟结果与实际数据对比



3.2.2 神经网络 ca

参数设置:

表 3-3 神经网络 ca 参数设置

参数名字	参数值
采样数目	10000
模拟次数	500
邻域大小	5
随机因子	1.1
目标城市栅格数目	107554
转换阈值	0.75

混淆矩阵:

表 3-4 神经网络 ca 混淆矩阵

年份	土地利用类型	模拟						
		农田	森林	草地	水体	城市	未利用地	
2010	实际	农田	37115	224	32	267	6972	0
		森林	376	42105	353	191	17652	0
		草地	35	111	5013	313	847	0
		水体	591	201	27	23994	236	187
		城市	16371	4606	712	3967	81898	0
		未利用地	0	0	0	0	0	3
		kappa			0.79			
		整体精度			0.79			
		FoM			0.60			

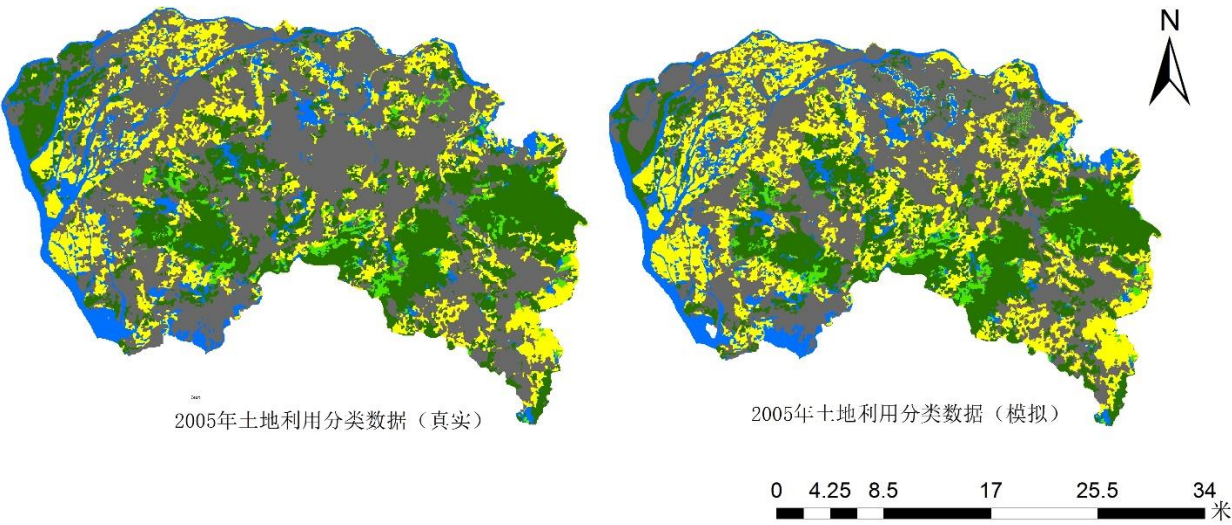


图 3-3 神经网络 ca 模拟结果与实际数据对比

4.2.3 决策树 ca

参数设置:

表 3-5 决策树 ca 参数设置

参数名称	参数值
采样率	0.01
模拟次数	100
邻域大小	5
目标城市栅格数目	107554

混淆矩阵:

表 3-6 决策树 ca 混淆矩阵

年份	土地利用类型	模拟						
		农田	森林	草地	水体	城市	未利用地	
2005	实际	农田	30011	266	36	265	14032	0
		森林	386	48529	296	184	11282	0
		草地	35	127	5296	34	827	0
		水体	592	267	38	23982	196	161
		城市	12639	8223	877	3921	81894	0
		未利用地	0	0	0	0	0	3
	kappa		0.77					
整体精度		0.78						
FoM		0.61						

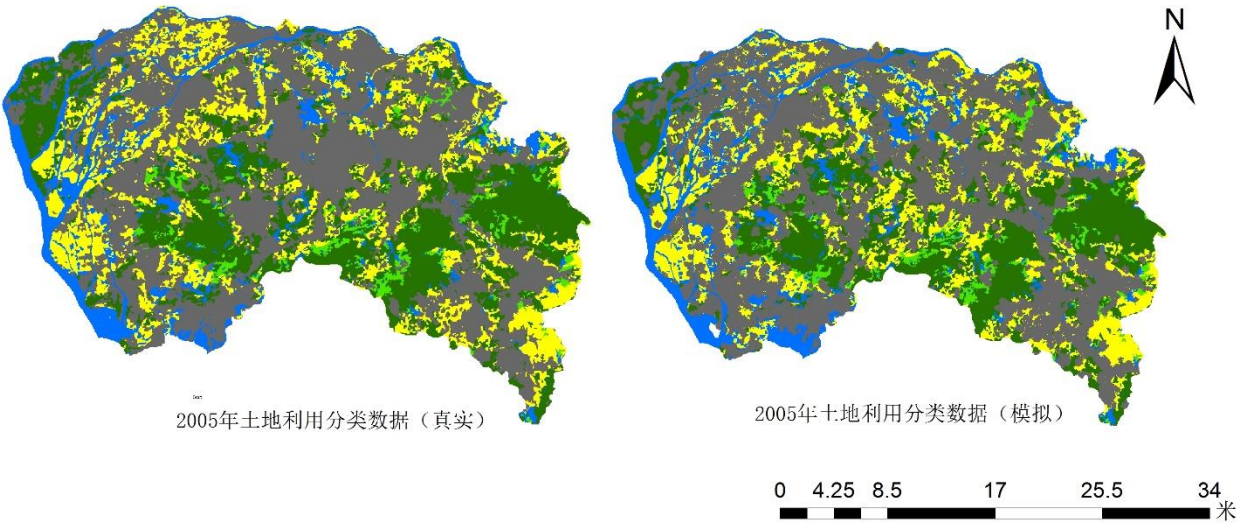


图 3-4 决策树 ca 模拟结果与实际数据对比



3.2.4 logistic ca

参数设置:

表 3-7 logistic ca 参数设置

参数名称	参数值
采样数目	10000
模拟次数	500
邻域大小	3
目标城市栅格数目	107554
转换阈值	0.7
随机因子	1.1

混淆矩阵:

表 3-8 logistic ca 混淆矩阵

年份	土地利用类型	模拟						
		农田	森林	草地	水体	城市	未利用地	
2005	实际	农田	30728	279	37	265	13301	0
		森林	440	50850	296	184	8907	0
		草地	34	133	5723	34	395	0
		水体	602	268	38	23982	159	187
		城市	12867	9065	998	3921	80703	0
		未利用地	0	0	0	0	0	3
	kappa	0.78						
整体精度	0.79							
FoM	0.62							

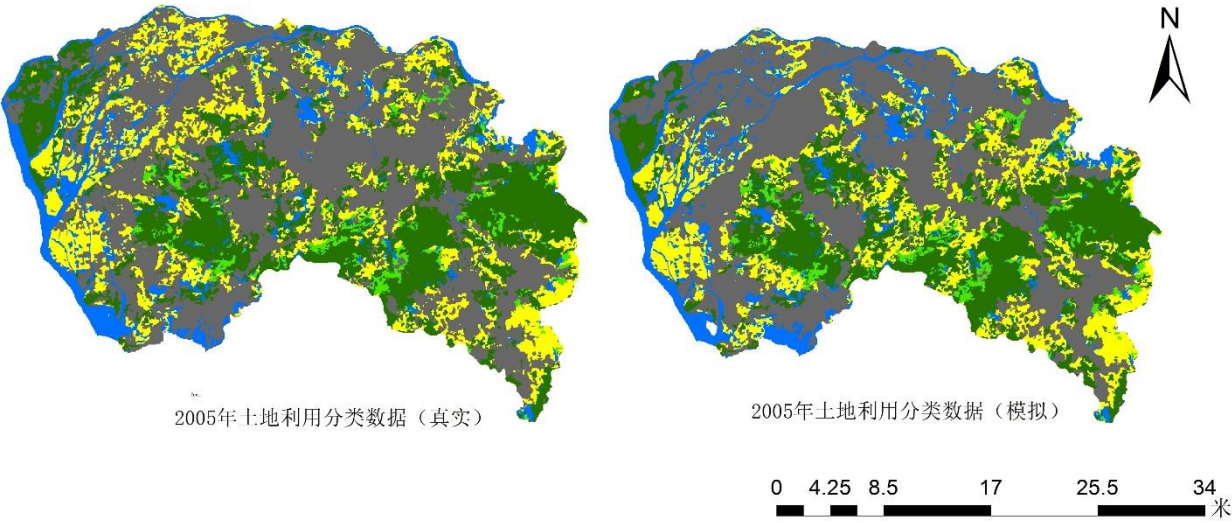


图 3-5 logistic ca 模拟结果与实际数据对比