

江苏省耕地面积变化模型的比较分析

曹蕾 (淮阴师范学院地理系, 江苏淮安 223001)

摘要 采用多元线性回归和灰色动态模型两种建模方法, 分别对江苏省耕地面积短期变化进行分析, 并比较了不同模型分析的效果。结果表明, 两者都可用于预测, 多元线性回归模型拟合精度明显好于灰色动态模型, 但多元线性回归模型是经过修正的, 且对数据要求较高, 而灰色动态模型对数据没有特殊要求。

关键词 耕地变化; 模型; 比较; 江苏省

中图分类号 F301.21 文献标识码 A 文章编号 0517-6611(2007)27-08635-02

Comparative Analysis of the Models of Cultivated Land Change in Jiangsu

CAO Lei (Dept. of Physics, Huaiyin Normal College, Huai'an, Jiangsu 223001)

Abstract The scale and quantity analysis of cultivated land is the basis in the research on the change rule of land and in the protection of land. In the article with two models of Grey Systematic model and Multiple Linear Regression the change of cultivated land in short period in Jiangsu was analyzed. And based on the history statistical data, the forecast ability of two models was compared. The result showed that the two models are all predictive. However, Multiple Linear Regression model is the superior in precision to Grey systematic model, but the former was re-mediated, with the more critical over the data, whereas the later had not any particular requirement on the data.

Key words Cultivated land change; Model; Comparison; Jiangsu Province

江苏省是典型的人多地少省份, 耕地资源非常紧张。2006 年全省人均耕地面积不足 0.067 hm^2 , 仅为全国平均水平的 70%。作为经济强省, 江苏面临耕地保护和经济发展的双重压力, 对耕地保护工作十分重视。江苏省将“十一五”期末耕地保有量 471.670 万 hm^2 作为约束性指标纳入了“十一五”规划纲要, 还明确要求将地方耕地保护工作列为各市市长工作业绩考核的重要内容。而开展耕地变化研究, 定量分析耕地变化规律, 是科学有效地实施耕地保护工作的前提。笔者针对近 10 年来江苏耕地面积短期变化数据的特点, 采用不同建模方法进行定量分析, 通过比较模型特点和分析结果确定更适合的研究路线。

1 研究方法和数据说明

1.1 研究方法 在对耕地面积变化进行数理研究中, 常采用多元线性回归方法^[1-2]。该方法是通过研究耕地面积变化与各类影响要素间的变动关系, 建立经验性的多元线性回归方程。同时针对短期变化统计数据少、影响因素复杂的特点, 选取了灰色动态模型方法。在模型建立分析的过程中, 笔者尝试对模型进行修正, 以期获得更好的分析效果。

1.2 数据说明 模型中所使用的相关社会经济数据均取自《江苏统计年鉴》、《江苏农业年鉴》。由于《江苏统计年鉴》中年末实有耕地面积的统计数据从 1996 年起与农业普查数据接轨, 与之前年份的统计数据变化规律明显不符, 不利于进行连续时间序列的分析和解释。为了研究的科学性, 笔者只选取 1996~2005 年的相关统计数据, 进行短期变化的建模。江苏省 1996~2005 年末实有耕地面积分别为: 506.170 万、505.567 万、503.654 万、502.422 万、500.839 万、497.412 万、490.502 万、485.834 万、497.519 万、478.037 万 hm^2 。

2 多元线性回归模型分析

2.1 耕地面积变化驱动因子选择(图1) 经分析发现, 耕地数量的变化与特定的社会经济发展阶段和区域发展定位相联系, 既有经济发展和人口增长的驱动、相关政策背景变

化和调整的影响, 又有环境变迁和自然灾害等因素的影响, 多种因素综合作用形成了一个复杂系统^[3]。

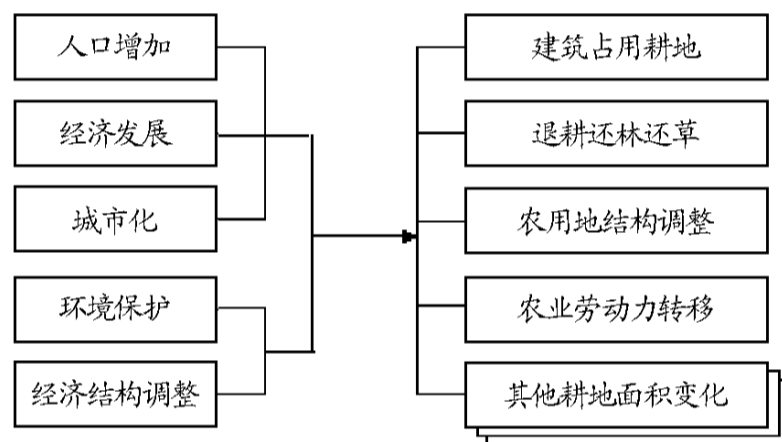


图1 江苏耕地面积变化的驱动因子

江苏省自然条件优越, 自然灾害、环境变迁等原因造成的耕地面积变化所占比重较小, 社会经济因素起到决定性作用。由于耕地面积变化与社会经济发展之间的关系非常复杂, 选择代表江苏经济发展状况的指标时, 应尽可能全面。但选择变量太多会造成分析问题难度加大, 同时由于很多变量之间可能存在多重共线性, 会给回归系数带来不合理的解释。依据定性分析和参照专家意见, 初步选定总人口、GDP、全社会固定资产投资、第三产业产值、城市化水平(以第二、第三产业从业人员比重为测度)、粮食单产水平、人均居住面积 7 个因子作为自变量。

2.2 多元线性回归模型分析 数学模型为:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m \quad (1)$$

式中: $b_0, b_1 x_1, b_2 x_2, \dots, b_m x_m$ 为回归系数。当回归模型引入变量过多时, 将导致回归系数难以解释, 方程的稳定性差、精度低。为了提高模型可靠性, 使用逐步回归法选取配合较好和方差贡献大的自变量组建回归方程。经过逐步回归分析, 有人口(x_1)、全社会固定资产投资(x_2)、第三产业产值(x_3)、城市化水平(x_4)、粮食单产水平(x_5) 5 个因子进入了方程。得到的耕地面积变化(y) 与驱动因子(x_i) 的多元线性回归模型为:

$$y = 6391.7688 + 0.0293x_1 + 0.0212x_2 + 5.4924x_3 - 31.7028x_4 - 69.6357x_5$$

作者简介 曹蕾(1977-), 男, 江苏徐州人, 硕士, 讲师, 从事地理信息科学和数学建模在地理学中的应用等方面的教学和研究工作。

收稿日期 2007-05-14

相关系数 R 为 0.998, 判定系数 R^2 为 0.997, 调整的判定系数 R^2 为 0.994, 估计值的标准误差为 20.675。由以上各值可以看出, 因变量和自变量存在较强的线性关系, 进入方程的 5 个因子可以解释 99.4% 的因变量。

回归方差分析结果表明, 由于显著性概率小于 5%, 认为回归系数不为 0, 即回归方程是有意义的。方程的拟合效果非常好, 可以利用该方程进行相关分析。在模型的多重共线性检验中, 发现各自变量的方差膨胀系数 (VIF) 都大于 10, 说明自变量之间存在较强的多重共线性。

为了消除方程的“病态性”, 笔者尝试使用有偏回归分析中的岭回归方法加以修正。其中, 分别取 $B = 0.001, \dots, 1$ 代入运算, 当 $k = 0.005$ 时, b 值开始稳定, 得到新回归方程:

$$y = 6\,746.361\,8 - 0.089\,x_1 + 0.008\,7x_2 + 6.153\,9x_3 - 23.393\,4x_4 - 0.606\,3x_5$$

($F = 148.516\,3, P = 0.000\,1$)

3 灰色动态模型分析

3.1 灰色系统和灰色动态模型^[5] 耕地面积动态变化本身没有客观物理原型, 作用机理不明确, 各类驱动因子难以辨别, 很多影响因素无法量化, 是一个典型的灰色系统。灰色系统分析方法对样本数据及其分布没有特别要求和限制, 适用范围广。因此, 针对该研究样本数据较少、影响因素多和驱动机制不明确的特点, 选用灰色动态 $GM(n, h)$ 建模与线性回归模型相比较。

灰色动态 $GM(n, h)$ 模型大体可归为 $GM(1, 1)$ 、 $GM(2, 1)$ 和 $GM(1, h)$ 三类。其中 $GM(1, 1)$ 、 $GM(2, 1)$ 主要用于预测, 而 $GM(1, h)$ 主要用于动态分析。

3.2 灰色数列 $GM(1, h)$ 建模分析 $GM(1, h)$ 模型是多元一阶线性动态模型, 其微分方程为:

$$\frac{dx_1^{(1)}}{dt} + ax_1^{(1)} = b_1x_2^{(1)} + b_2x_3^{(1)} + \dots + b_{h-1}x_h^{(1)} \quad (2)$$

其基本计算方程是:

$$x_1^{(1)}(t+1) = [x_1^{(0)} - \sum_{i=2}^h \frac{b_{i-1}}{a}(t+1)]e^{-at} + \sum_{i=2}^h \frac{b_{i-1}}{a}x_i^{(1)}(t+1) \quad (3)$$

为实现模型效果对比, 选取人口 (x_2)、全社会固定资产投资 (x_3)、第三产业产值 (x_4)、城市化水平 (x_5)、粮食单产水平 (x_6) 5 个因子与耕地面积数量 (x_1) 一起构造原始数据序列。

经过计算, 得到如下方程:

$$x_1(t+1) = (5\,061.7 - 1.503\,0x_2 - 0.008\,6x_3 + 71.353\,6x_4 + 57.131\,9x_5 + 215.393\,8x_6) e^{-1.8810t} + 1.503\,0x_2 + 0.008\,6x_3 - 71.353\,6x_4 - 57.131\,9x_5 - 215.393\,8x_6$$

通过后验比 c 和小误差概率 P 对模型进行检验, $c = 0.267\,4, P = 1.000\,0$ 。参照表 1 中评价标准, 模型精度等级为一级 (好)。通过模型各因子的传递函数, 可以看出江苏省耕地变化将在 2~3 年内趋于稳定, 并可能出现轻微的波动。

4 江苏耕地面积短期变化模型的选择和比较

为了比较以上模型的拟合能力, 分别计算各模型拟合值

的误差百分比 (PE)、相对平均误差百分比 (MPE) 和绝对平均误差百分比 (MAPE)。

表 1 $GM(1, h)$ 模型精度等级评价

模型精度等级	精度	等级
一级 好	> 0.95	< 0.35
二级 合格	> 0.8	< 0.50
三级 勉强	> 0.7	< 0.65
四级 不合格	0.7	> 0.65

$$PE = \frac{y_i - \hat{y}_i}{y_i} \quad (4)$$

$$MPE = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \hat{y}_i}{y_i} \quad (5)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (6)$$

其中, y_i 为观察值, \hat{y}_i 为预测值。

从表 2 可见, 两个模型的 MAPE 值均小于 10, 模型拟合精度都很好, 可用于分析预测。同时, 从表 2 可以直观地看出, 多元线性回归方程的拟合效果要更好一些。使用同样的自变量, 多元线性回归方程精度要比 $GM(1, h)$ 要好。经过计算, 即使是使用单变量的 $GM(2, 1)$ 模型和 $GM(1, 1)$ 模型对近 10 年江苏耕地面积变化数据进行建模, 拟合精度也不如多元线性回归方程。因此, 在研究近年来江苏省耕地面积短期变化时, 使用多元线性回归方法可以获得更好的效果。需要指出的是, 表 2 中多元线性回归精度数据是由经过修正的线性回归方程产生的, 直接通过逐步回归分析获得的结果并不令人满意。同时, 线性回归对数据的要求比灰色模型方法要高, 需要更大的数据量和线性分布。选择不同影响因子对模型使用效果的影响还需在以后的研究中深入探讨。

表 2 模型精度比较

年份	精度比较		%
	多元线性回归	$GM(1, h)$	
1997	PE	0.114 7	12.269 3
1998		- 0.112 7	- 10.797 5
1999		0.015 1	- 2.497 1
2000		- 0.099 2	- 2.346 3
2001		0.231 5	- 0.142 0
2002		- 0.164 9	0.290 7
2003		0.076 4	0.570 7
2004		- 0.066 6	- 0.931 2
2005		0.028 5	0.728 2
	MPE	- 0.000 1	- 0.317 2
	MAPE	0.093 3	3.397 0

参考文献

- [1] 周丙中, 赵其国. 江苏省耕地变化及其驱动机制的数理探讨[J]. 土壤学报, 2003(5): 665 - 671.
- [2] 汤江龙, 赵小敏. 江苏省经济发展与耕地数量动态变化及对策分析[J]. 江西农业大学学报, 2004(4): 54 - 58.
- [3] 毛良祥, 孙习稳, 杜新波. 当前耕地保护形势及政策优化分析[J]. 资源与产业, 2005, 7(4): 31 - 33.
- [4] 徐建华. 现代地理学中的数学方法[M]. 北京: 高等教育出版社, 2004.
- [5] 邓聚龙. 灰色系统理论教程[M]. 武汉: 华中理工大学出版社, 1990.
- [6] 江苏省统计局. 江苏统计年鉴 2006 [M]. 北京: 中国统计出版社, 2006.