

基于改进的 MGM 模型的我国耕地面积预测研究

门可佩, 朱璋亭 (南京信息工程大学, 江苏南京 210044)

摘要 根据《中国农业发展报告》和2007年《国土资源公报》关于耕地面积的最新统计数据,通过建立改进的MGM模型,对我国耕地面积进行了实证分析和预测研究。结果表明,到2008年我国耕地面积将减至1.2137亿 hm^2 ,按该趋势发展,到2012年我国耕地面积1.2亿 hm^2 存量底线将有失守的危险。指出应切实加大耕地保护力度。

关键词 耕地;灰色建模;改进的MGM模型IMGGM;预测

中图分类号 F201 文献标识码 A 文章编号 0517-6611(2009)05-01867-02

Study on the Prediction of Cultivated Land Area in China Based on the Improved MGM Model

MEN Ke-pi et al (Nanjing University of Information Science & Technology, Nanjing, Jiangsu 210044)

Abstract According to the latest data in China Agricultural Development Report and Bulletin of China Land and Resources in 2007, the empirical analysis and prediction study were made on the cultivated land area in China through establishing the improved MGM model. The results showed that the cultivated land area in China in 2008 would reduce to 121.370 million hm^2 . According to the developmental trend, the cultivated land area in 2012 would be less than 120 million hm^2 . It was pointed out that the cultivated land protection should be strengthened.

Key words Cultivated land; Grey modeling; Improved MGM model (IMGGM); Prediction

保护耕地直接关系到国家粮食安全,关系到国家长治久安。我国人多地少,人均资源相对贫乏,人地矛盾十分突出。自20世纪80年代以来,我国耕地面积逐年减少。据2007年《国土资源公报》^[1]显示,2007年全国耕地面积为1.2176亿 hm^2 ,净减少土地4.067万 hm^2 ,减幅同比下降0.22个百分点。截至2007年底,我国人均耕地仅0.092 hm^2 ,不到世界平均水平的40%,优质耕地仅占全部耕地的1/3。“十分珍惜和合理利用土地,切实保护耕地”是必须长期坚持的一项基本国策。党中央、国务院历来十分重视耕地保护工作,先后制定了一系列重大方针政策,一再强调加强土地管理,切实保护耕地。今后5~10年是中国全面建设小康社会、加快推进社会主义现代化建设第三步发展战略的重要时期,切实强化耕地保护对促进国民经济持续稳定健康发展具有重大意义。耕地总量的预测是土地利用规划的必要前提。灰色预测是灰色系统理论的基本内容之一,通过建立GM(1,1)动态模型探索事物未来发展变化的规律。它把难以描述的综合量作为灰色量来处理,通过对数据的生成和开发弱化随机因素的干扰,提取有用信息,实现对系统运行规律的正确认识和有效控制^[2]。灰色模型所需数据量较少,建模灵活方便,预测精度较高,其实用性强,从而在社会、经济、科技、农业、生态等各领域得到广泛应用。因此,笔者根据2001~2007年我国耕地指标最新修正数据,运用改进的MGM模型(IMGGM)对未来几年耕地面积进行分析和预测,旨在为我国社会经济可持续发展和粮食安全提供重要依据。

1 改进的 MGM 模型的建立

耕地面积的减少是人与自然众多复杂因素相互影响制约、协调发展的共同结果,适宜采用灰色模型去研究和挖掘系统的内在规律。为了提高GM(1,1)模型的拟合和预测精度,很多科技工作者通过努力,提出了许多可行的方法,如对背景值的优化^[3-5]。但是该法在实际操作中存在2个问题:

一是多次迭代会出现较大的累积误差;二是多次迭代得出的发展系数序列,目前还没有理论能保证其一定收敛。另外, $x^{(1)}(1)$ 与 $x^{(0)}(1)$ 在表达式无法相等时也是灰色模型产生误差的一个原因。笔者参照文献[6]的方法,先将待定系数响应式还原成预测式后,再利用初始条件 $x^{(0)}(n) = x^{(0)}(n)$ 确定待定系数的新方法。以下称为改进的MGM模型。

设原始非负时间序列为 $X^{(0)} = [x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)]$,再对时间序列 $X^{(0)}$ 进行一阶累加生成,得生成序列 $X^{(1)}$,即 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) (k=1, 2, \dots, n)$

1.1 发展系数 a 初始值的确定 令初始值 $a_0 = \ln\{\frac{1}{n-1} \sum_{k=2}^n \frac{x^{(0)}(k-1)}{x^{(0)}(k)}\}$, $\theta = \frac{1}{a_0} - \frac{1}{e^{a_0} - 1}$, 序列 $Z^{(1)} = [z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)]$, 其中 $z^{(1)}(k) = \theta x^{(1)}(k-1) + (1-\theta)x^{(1)}(k)$, 建立优化的灰色GM(1,1)模型:

$$x^{(0)}(k) + a_1[\theta x^{(1)}(k-1) + (1-\theta)x^{(1)}(k)] = b_1$$

其中 a_1, b_1 为待定系数,分别称为发展系数和灰色作用量, a_1 的有效区间是 $(-2, 2)$,应用最小二乘法可经下式求得:

$$a = (a_1, b_1)^T = (B^T B)^{-1} \cdot B^T \cdot Y$$

$$B = \begin{bmatrix} -[\theta x^{(1)}(1) + (1-\theta)x^{(1)}(2)] & 1 \\ -[\theta x^{(1)}(2) + (1-\theta)x^{(1)}(3)] & 1 \\ \dots & \dots \\ -[\theta x^{(1)}(n-1) + (1-\theta)x^{(1)}(n)] & 1 \end{bmatrix}$$

$$Y = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)]^T$$

其中 $B =$

$$\dots \dots$$

$$-[\theta x^{(1)}(n-1) + (1-\theta)x^{(1)}(n)] \quad 1$$

$$Y = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)]^T$$

若响应模式精度不够理想,可以多次迭代求 a_2, a_3, \dots ,直至理想为止。一般情况无需进行迭代。

1.2 新的预测表达式 由原模型求出的时间响应序列及还原值:

$$x^{(1)}(k) = c \cdot e^{-a_1 \cdot (k-1)} + \frac{b_1}{a_1}$$

$$x^{(0)}(k) = c \cdot (1 - e^{-a_1}) \cdot e^{-a_1 \cdot (k-1)}$$

其中 c 是待定的,可通过取任意的 $x^{(0)}(k) = x^{(0)}(k)$ 代

基金项目 国家统计局2008年全国统计科研计划重点项目(2008LZ022)研究成果之一;南京信息工程大学科研基金资助项目。

作者简介 门可佩(1949-),男,江苏南京人,教授,硕士生导师,从事信息预测理论与应用统计分析研究。

收稿日期 2008-12-03

入来确定c,得到相应的预测表达式。为了避免累积误差,同时又不影响预测效果,根据灰色系统中新信息优先原理,令 $x^{(0)}(n) = x^{(0)}(n)$,即可求得c,由此得到白化响应式:

$$x^{(0)}(k) = x^{(0)}(n) e^{-a_1 \cdot (k-1)}$$

这种新组合预测公式对于标准指数序列而言,没有模型误差,只有计算误差。笔者采用 Matlab 进行编程计算,对我国耕地总量进行预测研究。

2 我国耕地面积预测分析

2.1 2007 年我国耕地面积检验性预测 为筛选合适的模型,根据近年我国耕地的实际,经综合分析选取6 维短序列,即以2001~2006 年的耕地面积数据,建立我国耕地预测的普通GM(1,1)模型、GM(1,1)优化模型(即MGM模型)、初始点固定的离散灰模型(SDGM模型)、终点固定的离散灰模型(EDGM模型)以及改进的MGM模型(IMGM),对2007 年我国耕地面积进行检验性预测,计算得各种灰色模型预测值见表1。比较检验结果可知,IMGM模型比其他模型具有更高的预测精度。

表1 各类模型关于2007 年我国耕地面积的检验性预测比较结果

Table 1 The test prediction comparison of cultivated land area in China in 2007 by using different models

模型 Models	实际值 Actual value $\times 10^3 \text{hm}^2$	预测值 Predictive value $\times 10^3 \text{hm}^2$	相对误差 Relative error %	拟合精度 Fitting accuracy %
普通GM(1,1)模型 Common GM(1,1) model	121 759.3	120 268.8	1.227 4	98.772 6
MGM模型 MGM model	121 759.3	120 265.4	1.226 9	98.773 1
EDGM模型 EDGM model	121 759.3	120 262.9	1.229 0	98.771 0
SDGM模型 SDGM model	121 759.3	120 262.9	1.229 0	98.771 0
IMGM模型 IMGM model	121 759.3	120 851.6	0.745 5	99.254 5

2.2 2008~2012 年我国耕地面积的预测 由上述对比分析,笔者选用2002~2007 年我国耕地面积资料^[7-8],建立6 维IMGM预测模型:

$$x^{(1)}(k) = -38\,573\,515.967\,7e^{-0.003\,202\,4(k-1)} + 38\,620\,662.228\,5$$

$$x^{(0)}(k) = 121\,759.300\,0e^{-0.003\,202\,4(k-n)}$$

经检验,模型的平均拟合精度 $\bar{q} = 99.42\%$,模型通过检验,模型精度为一级(优)。据该模型计算得2008~2012 年中国耕地面积预测(模拟)值见表2。

3 结论与讨论

(1) 原始灰预测响应式是误差产生的原因之一,也掩盖了运用原始模型响应式去预测时首项误差其实为假“0”的事实。笔者选用IMGM模型,针对原预测模型响应式存在的问题,改进了利用初始条件决定响应式,并改进了MGM模型中求发展系数a的方法,大大减少了迭代次数。IMGM模型弱化了干扰因素,揭示了系统的运行规律,所建模型更为稳定,因而具有更高的预测精度。实证分析表明,这种新型灰色模

型不受样本量的限制,建模方便,计算量不大,易于调控,对

表2 IMGM模型关于2008~2012 年我国耕地面积的预测

Table 2 The prediction of cultivated land area in China during 2008-2012 by using IMGM model

年份 Year	实际值 Actual value $\times 10^3 \text{hm}^2$	预测(模拟)值 Predictive value $\times 10^3 \text{hm}^2$	相对误差 Relative error %	拟合精度 Fitting accuracy %
2002	125 929.6	123 724.6	1.751 0	98.249 0
2003	123 392.2	123 329.0	0.051 2	99.948 8
2004	122 444.3	122 934.7	0.400 5	99.599 5
2005	122 066.7	122 541.6	0.389 1	99.610 9
2006	121 800.0	122 149.8	0.287 2	99.712 8
2007	121 759.3(终点固定)			
2008		121 370.0		
2009		120 982.0		
2010		120 595.1		
2011		120 209.6		
2012		119 825.2		

注:2007 年数据来自文献[9];其余来自文献[8]。

Note: The data in 2007 are from reference [9] and other data are from reference [8].

短期增长较快的时间序列的预测效果优于一般的灰色模型,是从事社会经济预测研究十分理想的新工具。

(2) 若依照系统目前状态运行,根据笔者的预测结果,2008~2012 年我国耕地面积将分别为:1.213 700 亿、1.209 820 亿、1.205 951 亿、1.202 096 亿、1.198 252 亿 hm^2 。按照我国农业生产发展状况,到2010 年若要实现1996 年我国政府发布的《中国的粮食安全问题》白皮书所确定的粮食95%的自给目标,需要耕地面积1.216 亿 hm^2 ,而由IMGM模型所得的预测值比该目标约少100 万 hm^2 。如按该趋势发展,到2012 年耕地预测值仅有1.198 252 亿 hm^2 ,必须保证的1.2 亿 hm^2 耕地存量底线终将失守。而实际上对于我国这样一个人口占世界总数1/5 的大国,从保障粮食安全和社会稳定的要求来看,1.2 亿 hm^2 耕地绝不是一个宽裕的数字。为此,必须严守1.2 亿 hm^2 耕地。

(3) 目前,我国有限的耕地资源还在快速地减少,这其中自然有自然灾害的原因,更多的是人为因素。自2003 年来,以“开发区热”、“房地产热”和部分行业产能过剩突出为典型的新一轮经济过热和失衡现象,造成了耕地的大量流失;而耕地控制不严,又使得经济过热、盲目投资等占用了大量土地。在国土严重流失的现实面前,很多人不但没有吸取教训,还利用《全国土地利用总体规划纲要》修编的机会,从规划上变相地扩张投资用地。虽然2007 年我国耕地面积减少速度趋缓,但其中一个主要原因是生态退耕和灾毁耕地的大幅减少及农业结构调整净减耕地大幅下降。考虑到这些因素,我国耕地保护形势依然严峻。

(4) 2007 年全国土地日的宣传主题是:“节约集约用地,坚守耕地红线”^[7]。节约和集约用地是既保护耕地、控制土地供应量,又保障经济发展合理用地需求的唯一出路。为此,必须充分发挥土地利用规划、计划的调控作用;按照依法依规、有保有压的原则,严把土地供应闸门;大力推进节约集约用地,按照“管住总量、严控增量、盘活存量、节约集约”的

2.2 系统聚类分析结果 按“1.4.6.2”方法, 经过 SPSS10.0 统计软件的分析, 得聚类分析结果(图3), 分类结果见表3。

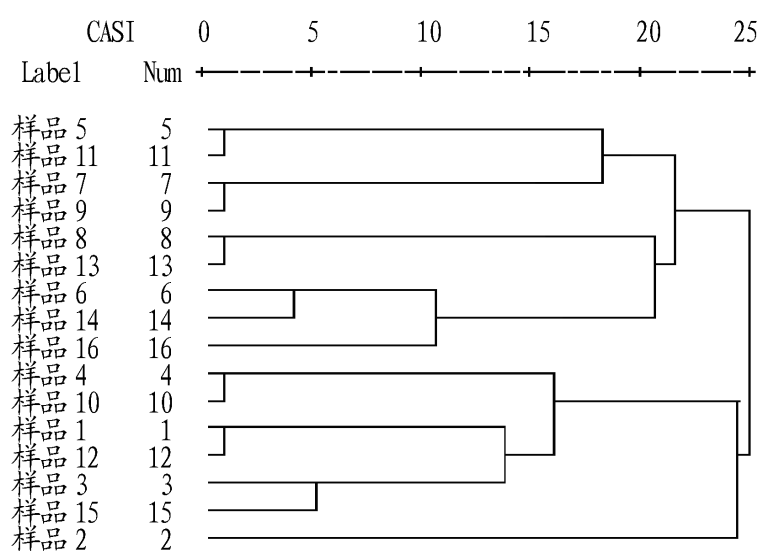


图3 聚类分析

Fig.3 The clustering analysis

表3 聚类分析结果

Table 3 The clustering analysis results

类别 Types	样品号 No. of samples
	5, 11, 7, 9
	8, 13
	6, 14, 16
	4, 10, 1, 12, 3, 15
	2

由图3、表3可知, 在标尺20处, 16个样品分为5类。首先, 蓼属植物区别很大。珠芽蓼的8[#]、9[#]峰峰面积明显大于其他4种植物相应峰的峰面积, 而14[#]峰的峰面积又小于其他4种植物相应峰的峰面积。蒺藜的1[#]、2[#]峰峰面积明显大于其他4种植物相应峰的峰面积。其次, 6、14号西伯利亚蓼样品距离较近, 而16号西伯利亚蓼样品与它们距离较远, 比较3个样品的气相色谱图, 发现6、14号西伯利亚蓼样品的12[#]峰峰面积明显大于16号西伯利亚蓼样品12[#]峰峰面积, 而15[#]峰则恰恰相反。这可能是影响他们差异的主要原因。银川的蒺藜与其他采集地的蒺藜化学成分与含量相比差别

(上接第1868页)

原则, 充分利用闲置土地, 提高土地利用效率, 保障必要的建设用地; 充分调动农民群众保护耕地的积极性; 严格落实耕地保护目标责任制。地方各级政府和国土资源部门, 要充分认识节约集约用地的重要性和紧迫性, 全面贯彻落实党的十七届三中全会精神, 认真实施《全国土地利用总体规划纲要(2006~2020年)》, 将约束性的耕地保护指标量化到地方政府目标责任, 切实转变用地观念, 转变经济发展方式, 调整优化经济结构, 科学规划用地, 着力内涵挖潜, 实现经济社会的可持续发展。

参考文献

[1] 国土资源部. 2007年国土资源公报 [EB/OL]. <http://www.mnr.gov.cn/>

不大, 这可能是因为它们生长的环境条件相似, 使得其内在主要成分相对稳定。

3 结论与讨论

(1) 试验以柱温70^o, 保留3 min后, 程序升温(3^o/min)至250^o, 运行时间60 min为色谱条件, 发现22 min后气相色谱图几乎成直线, 没有明显的色谱峰, 足见运行时间为22 min的GC分析效果可以满足试验要求; 待测样品色谱图在数据处理时除去了溶剂峰, 因此总峰面积以每个样品4~22 min的所有色谱峰的峰面积计算。

(2) 试验采用了GC-MS法进行色谱分离, 1次进样可得较多的色谱峰(16个样品中共计有近1000个不同保留时间的峰应用到聚类分析中), 分离分析的效率较高, 且其保留时间相对稳定, 得到的气相色谱图具有一定的稳定性与可控性。在此试验的基础上, 可对不同省份蓼属植物进行分类研究, 进一步验证气相色谱-系统聚类分析方法在蓼属植物分类中的应用。

(3) 在对样品进行聚类分析时发现, 用不同的系统聚类方法进行聚类, 所得到的结果不同。看一个聚类方法是否可取, 主要取决于这种方法所得的结果是否正确。该试验没有采用常用的欧式距离或欧式平方距离, 这2种方法的聚类结果都不理想, 而采用夹角余弦, 分类结果较好。

运用指纹图谱与聚类分析相结合的方法能更全面、综合和准确地区分不同种类的蓼, 作为药材质量检测的一种辅助手段, 该法有一定应用价值。

参考文献

- [1] 戚欢阳, 杨永建, 赵汝能. 甘肃省蓼属药用植物[J]. 中药材, 2002, 25(3): 164-166.
- [2] 戚欢阳, 张朝凤, 张勉, 等. 毛脉蓼化学成分及抑菌活性的研究[J]. 中国药学杂志, 2005, 40(11): 819-822.
- [3] 吕晴, 秦军, 陈桐. 气相色谱-质谱法分析荷花穗挥发油化学成分[J]. 理化检测化学分册, 2004, 40(7): 405-407.
- [4] 李康兰, 沈序维, 郑尚珍, 等. 珠芽蓼挥发油化学成分的研究[J]. 西北师范大学学报: 自然科学版, 1999, 35(3): 68-70.
- [5] 杨再, 陈佳铭, 黄晓兰, 等. 天然植物有效成分的提取新技术——超临界流体萃取技术[J]. 饲料博览, 2005(6): 30-33.
- [6] 林枫, 刘利军. 高效液相色谱-系统聚类分析方在麻黄属药材分类中的应用[J]. 化学与生物工程, 2006, 23(2): 60-62.

xwd/tpxw/200804/t20080416_101192.htm

- [2] 刘思峰, 党耀国, 方志耕. 灰色系统理论及其应用[M]. 北京: 3版. 科学出版社, 2004.
- [3] 谢乃明, 刘思峰. 离散GM(1,1)模型与灰色预测模型建模机理[J]. 系统工程理论与实践, 2005, 25(1): 93-99.
- [4] 穆勇, 李放. 灰色GM(1,1)优化模型[J]. 济南大学学报: 自然科学版, 2001(12): 341-343.
- [5] 党耀国, 刘思峰, 刘斌. 以 $x^{(0)}(t)$ 为初始条件的GM模型[J]. 中国管理科学, 2005(2): 132-135.
- [6] 余逗, 魏勇. 发展系数与预测模型初始值确定的新方法[J]. 统计与决策, 2008(2): 42-44.
- [7] 林英. 不可逾越的红线: 坚守18亿亩耕地[N]. 光明日报, 2007-06-25.
- [8] 中华人民共和国农业部. 中国农业发展报告-2007[M]. 北京: 中国农业出版社, 2007.
- [9] 国家统计局. 中华人民共和国2007年国民经济和社会发展统计公报 [EB/OL]. (2008-02-28) <http://www.stats.gov.cn/> 2008.2.28.