

doi: 10.3969/j.issn.1007-855x.2009.04.020

嵌入知识的 GM(1, 1)模型 及其在粮食产量预测中的应用

张 慧¹, 谢长伟², 刘 斌¹

(1. 河南农业大学 信息与管理科学学院, 河南 郑州 450002; 2. 河南农业大学 经管学院, 河南 郑州 450002)

摘要: 利用小样本、贫信息进行数据预测, 传统的数学建模方法并不适合, 灰色系统理论在解决该类问题有着明显的优势. 论文在分析河南粮食产量数据的基础上, 应用嵌入知识的 GM(1, 1)模型, 建立了河南粮食产量的预测模型, 对未来 5 年的河南粮食产量进行了模拟和预测, 进而提出了提高粮食产量的对策建议.

关键词: GM(1, 1); 粮食产量; 预测; 缓冲算子

中图分类号: U412.366; O159 **文献标识码:** A **文章编号:** 1007-855X(2009)04-0092-04

Knowledge - Embedded GM (1, 1) Model and Its Application to Forecasting of Grain Output

ZHANG - hui¹, XIE Chang - wei², LIU Bin¹

(1. College of Information and Management Science, Henan Agricultural University, Zhengzhou 450002, China;

2. College of Economics and Management, Henan Agricultural University, Zhengzhou 450002, China)

Abstract: With its small and inadequate sample, the traditional mathematical modelling is not quite appropriate, while the grey system theory has obvious advantages in dealing with inadequate and uncertain information. Based on the analysis of grain output of Henan province, knowledge - embedded GM (1, 1) model is adopted to establish a forecast model of Henan grain output. The simulation and forecast of grain output of Henan province in the next five years are finally carried out, with some countermeasures to improve grain output provided.

Key words: GM (1, 1); grain output; forecast; buffer operator

0 引言

GM(1, 1)模型是整个灰色理论体系的基础, 如果序列为非负准光滑序列, 则的 1-AGO 序列具有准指数规律^[1]. 这种结论表明对于具有指数变化特征的系统, 灰色预测模型可以取得良好的预测效果. 现实当中的经济系统、生态系统、农业系统等均可以看作广义的能量系统, 而能量的积累和释放一般具有指数规律, 因此以 GM(1, 1)模型为基础的灰色预测体系具有广泛的使用范围.

粮食生产的发展过程符合灰色预测的模式. 粮食产量随着时间的发展总体呈现的是上升的趋势, 为已知的信息, 但是每年上升多少, 出现什么样的波动却是未知的信息. 所以选择灰色预测方法对粮食产量的未来发展趋势做出判断具有可行性. 通过建立的 GM(1, 1)模型对数据进行预测, 如果建立的模型不合格或者精度不理想, 则可以对模型进行修正, 提高模型的有效性^[2]. 论文采用了两种 GM(1, 1)模型, 传统的 GM(1, 1)模型、嵌入知识的新陈代谢 GM(1, 1)模型, 对河南省全年的粮食产量进行了预测对比, 并给出了

收稿日期: 2009 - 05 - 21. 基金项目: 国家“十一五”科技支撑计划: 现代村镇服务业集成示范课题 (编号: 2006BAJ07B09 - 06).

第一作者简介: 张慧 (1975 -), 女, 硕士, 讲师. 主要研究方向: 灰色系统理论、人工智能. E-mail: zhhnau@163.com

未来 5 年的预测值. 并根据预测结果,提出了相关政策建议.

1 GM (1, 1)模型

设原始序列为: $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$

$X^{(0)}$ 的 1 - AGO 序列为:

$$X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$$

其中, $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$,

$Z^{(1)}$ 为 $X^{(1)}$ 的紧邻均值生成序列:

$$Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n));$$

其中, $z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k - 1) \quad k = 2, 3, \dots, n;$

对 GM (1, 1)模型, $x^{(0)}(k) + az^{(1)}(k) = b$, 称为 $\frac{dx^{(1)}}{dt} + ax^{(1)} = b$ 为 GM (1, 1)模型的白化方程.

定理 1: 设 $X^{(0)}$ 为非负序列, $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)); X^{(1)}$ 为 $X^{(0)}$ 的 1 - AGO 序列, $Z^{(1)}$ 为 $X^{(1)}$ 的紧邻均值生成序列, 若 $\hat{a} = [a, b]^T$ 为参数, 且

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

则 GM (1, 1)模型 $x^{(0)}(k) + az^{(1)}(k) = b$ 的最小二乘估计参数列满足

$$\hat{a} = (B^T B)^{-1} B^T Y$$

定理 2: 设 B, Y, \hat{a} 如定理 1 所述, $\hat{a} = [a, b]^T = (B^T B)^{-1} B^T Y$, 则

白化方程 $\frac{dx^{(1)}}{dt} + ax^{(1)} = b$ 的时间响应函数为:

$$x^{(1)}(t) = (x^{(1)}(1) - \frac{b}{a})e^{-at} + \frac{b}{a}$$

GM (1, 1)模型, $x^{(0)}(k) + az^{(1)}(k) = b$ 的时间响应序列为

$$\hat{x}^{(1)}(k + 1) = (x^{(0)}(1) - \frac{b}{a})e^{-at} + \frac{b}{a}; \quad k = 1, 2, \dots, n$$

还原值

$$\hat{x}^{(0)}(k + 1) = a^{(1)} \hat{x}^{(1)}(k + 1) = \hat{x}^{(1)}(k + 1) - \hat{x}^{(1)}(k) \quad k = 1, 2, \dots, n$$

在应用传统 GM (1, 1)模型进行预测前,需要对原始序列进行准光滑检验,对一次累加生成序列进行准指数规律检验. 原始序列的拟合曲线的指数特征愈明显,采用传统 GM (1, 1)模型预测的效果愈好. 对于传统 GM (1, 1)模型,可以通过对残差考察来判断模型的精度,其中平均相对误差和模拟误差要求越小越好,关联度要求越大越好,均方差比值要求越小越好,小概率误差越大越好.

2 嵌入知识的 GM (1, 1)模型

河南省 1998 年至 2006 年的产量数据如表 1 所示,选取 2002 ~ 2006 年的数据作为初始序列直接应用 GM (1, 1)模型

表 1 河南省粮食产量数据
Tab 1 Henan grain output data (单位:万 t)

年份	1998	1999	2000	2001	2002	2003	2004	2005	2006
产量	4 009. 61	4 253. 25	4 101. 50	4 119. 88	4 209. 98	3 569. 47	4 260. 12	4 582. 28	5 054. 99

有预测模型:

$$\begin{cases} \hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{b}{a})e^{-at} + \frac{b}{a} = 32\,255.75e^{0.1081k} - 28\,045.8 \\ \hat{x}^{(0)}(k+1) = a^{(1)}\hat{x}^{(1)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \end{cases}$$

$x^{(1)}$ 的模拟值:

$$\begin{aligned} x^{(1)} &= (\hat{x}^{(1)}(1), \hat{x}^{(1)}(2), \hat{x}^{(1)}(3), \hat{x}^{(1)}(4), \hat{x}^{(1)}(5)) \\ &= (4\,209.98, 7\,892.412, 11\,995.32, 16\,566.71, 21\,660.09) \end{aligned}$$

还原求出的 $x^{(0)}$ 模拟值,由 $\hat{x}^{(0)}(k+1) = a^{(1)}\hat{x}^{(1)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k)$

$$\begin{aligned} \text{得 } x^{(0)} &= (\hat{x}^{(1)}(1), \hat{x}^{(1)}(2), \hat{x}^{(1)}(3), \hat{x}^{(1)}(4), \hat{x}^{(1)}(5)) \\ &= (4\,209.98, 3\,682.432, 4\,102.908, 4\,571.396, 5\,093.377) \end{aligned}$$

误差检验如表 2 所示,且平均相对误差为: $= \frac{1}{4} \sum_{k=2}^5 \epsilon_k = 1.96\%$

表 2 误差检验表

Tab 2 Error test table

序号	实际数据 $x^{(0)}(k)$	模拟数据 $\hat{x}^{(0)}(k)$	相对误差 $\epsilon_k = \frac{ x^{(0)}(k) - \hat{x}^{(0)}(k) }{x^{(0)}(k)}$
2	3 569.47	3 682.43	3.16%
3	4 260.12	4 102.91	3.69%
4	4 582.28	4 571.40	0.23%
5	5 054.99	5 093.38	0.76%

由原始数据一阶累加生成的数据光滑性较好,而且准指数规律也很明显,对原始数据拟合的平均相对误差为 1.96%,很符合数据预测的原则. 如果选择这个模型进行预测,得到 2007~2011 年的年平均增长率为 4.9%. 但是仔细分析原始数据发现,从 2002~2006 年,河南的粮食产量有一个负增长点:2003 年的增长率为 -15.2%,对比 1998~

2001 年间的数(4 009.61, 4 253.25, 4 101.5, 4 119.88),2003 年的粮食产量是一个突变的超低产量,由于 2003 年产量的突变,使得 2004 年的增长率达到了 19.3%,而 2005 年和 2006 年的增长率分别为 7.56% 和 10.32%. 2003 年产量的突变是由于我国遭遇了“非典”,并且 2003 年秋河南省部分地区遭受了严重的自然灾害,使得秋季种植业生产受损过重,所以 2003 年的粮食产量数据可以看作是一个噪声值. 因此用前期数据的预测结果进行代替,以剔除粮食产量的突变性,提高系统的“鲁棒性”^[3,4].

同时,由于 2004 年以后河南省全面贯彻落实中央关于农业和农村工作的一系列方针政策,积极推进了农村的各项改革,全省农村经济呈现多年少有的好形势. 尤其是“1 减 3 补”政策的实施,极大的调动了农民的种粮积极性,农民不断加大生产投入,强化田间管理,使得粮食产量不断创历史新高^[5]. 对比这两个阶段的增长率,1998~2002 年产量平均增长率为 1.28%,2004~2006 年产量平均增长率为 12.72%,两个平均增长率相差达到 10 倍,系统波动性较大. 为了较真实的反映实际情况,剔除政策性因素,论文对原始数据进行一阶平滑,以期弱化系统的波动性.

引入一阶弱化算子 D ,则原始序列的一阶缓冲算子为:

$$X^{(0)}D = (x^{(0)}(1)d, x^{(0)}(2)d, x^{(0)}(3)d, x^{(0)}(4)d, x^{(0)}(5)d,)$$

$$\text{其中 } x^{(0)}(k)d = \frac{1}{5-k+1} (x^{(0)}(k) + x^{(0)}(k+1) + \dots + x^{(0)}(5)), k = 1, 2, \dots, 5$$

类似的可建立二阶弱化缓冲算子,三阶弱化缓冲算子等. 论文依据对生产情况的定性分析选择了一阶弱化算子.

$$\begin{aligned} \text{设 } X^{(0)} &= (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), x^{(0)}(4), x^{(0)}(5)) \\ &= (4\,472.491, 4\,538.118, 4\,632.333, 4\,818.5, 5\,054.99) \end{aligned}$$

对 $X^{(0)}$ 作 1-AGO,得:

$$X^{(1)} = (x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), x^{(1)}(4), x^{(1)}(5)) = (4\,472.491, 9\,010.609, 13\,642.94, 18\,461.44,$$

23 516. 44)

对 $X^{(0)}$ 和 $X^{(1)}$ 作准光滑性检验和准指数规律检验如表 3 所示.

使用缓冲算子后的数据原始序列的拟合曲线的指数特征更明显,故可对 $X^{(1)}$ 建立 GM (1, 1)模型.

$$\begin{cases} \hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{b}{a})e^{-at} + \frac{b}{a} = 120\,461.02e^{0.0367k} - 115\,988.53 \\ \hat{x}^{(0)}(k+1) = a^{(1)}\hat{x}^{(1)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \end{cases}$$

得 $X^{(0)} = (\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \hat{x}^{(0)}(3), \hat{x}^{(0)}(4), \hat{x}^{(0)}(5)) = (4\,472.49, 4\,501.6, 4\,669.9, 4\,844.4, 5\,025.4)$

检验误差如表 4 所示,且平均相对误差: $= \frac{1}{4} \sum_{k=2}^5 k = 0.93\%$

为了使预测结果能够更加准确的反映生产产量的动态变化,采用灰色系统理论的新陈代谢模型以改进现有的预测结果.两种预测模型的结果见表 5 所示.可以发现,嵌入知识的新陈代谢 GM (1, 1)在模拟中有较佳的效果,且河南省粮食的平均增长率符合定性分析结论.

3 结论与政策建议

通过各项检验,发现应用了缓冲算子的新陈代谢模型有较高的精度.用其预测表明:未来 5 年,河南省的粮食产量增长速度大致在 2.17%, 和实际情形比较符合.但是粮食产量的增长受诸如耕地面积、天气、自然灾害、政策等多种因素的制约.因此,在土地资源逐年减少,农民的积极性已经被调动的情况下,如何继续保证粮食的高产,是我们共同关注的问题.

当前和今后一个时期河南省应充分认识并防范粮食产量剧烈波动的出现,做好防御和平抑粮食产量超常规波动的各项工作,尤其要在以下几方面予以加强:

1)继续稳定、完善和加强粮食生产资金、技术和服务等方面的支持,强化政策支持能力.现阶段“2 减免、3 补贴”等农业支持政策对提高粮食综合生产能力、缓解粮食产量滑坡起到了积极作用.因此,国家农业政策的制定与实施要继续坚持“多予、少取、放活”的方针,做到支农政策的持续、稳定和战略性,确保给予农民的实惠不减少,支持农业的力度不断加大.

表 3 准光滑性检验和准指数规律检验对比表

Tab 3 Comparison table of quasi-smooth and quasi-exponential law tests

	准光滑性			准指数规律		
	(3)	(4)	(5)	⁽¹⁾ (3)	⁽¹⁾ (4)	⁽¹⁾ (5)
原始数据	0.54	0.38	0.3	1.54	1.38	1.3
使用缓冲算子后的数据	0.52	0.35	0.27	1.52	1.35	1.27

表 4 使用缓冲算子后预测模型的误差检验表

Tab 4 Error table of prediction model test after the use of a buffer operator

序号	实际数据 $x^{(0)}(k)$	模拟数据 $\hat{x}^{(0)}(k)$	相对误差 $k = \frac{ x^{(0)}(k) - \hat{x}^{(0)}(k) }{x^{(0)}(k)}$
2	4 583.118	4 501.677	1.78%
3	4 632.333	4 669.906	0.81%
4	4 818.5	4 844.422	0.54%
5	5 054.99	5 025.461	0.58%

表 5 2 种预测模型的预测结果对比

Tab 5 Results comparison of two prediction forecast models

年份	嵌入知识的新陈代谢 GM (1, 1)模型	
	GM (1, 1)模型	嵌入知识的新陈代谢 GM (1, 1)模型
2007	5 306.2	5 213.2
2008	5 565.1	5 400.8
2009	5 836.6	5 527.8
2010	6 121.4	5 612.8
2011	6 420.0	5 679
平均相对误差	4.9%	2.3%
平均增长率	4.88%	2.17%

(下转第 100 页)

采用加权平均的反模糊化策略:

$$u(k) = \frac{\sum_n \mu(n) \times U(n)}{\sum_n \mu(n)} \quad (10)$$

可得到溢流粒度所对应的给矿量的设定值和排矿水量.

6 结论

文中针对磨矿分级系统是具有特性多变、非线性、大滞后的特点,设计了双输入双输出的二维模糊控制器,采用模糊控制的方法获得了给矿量设定值和排矿水量,从而满足了对溢流浓度和溢流粒度的控制,排除了以往人工观察和经验操作的主观性.同时达到了节能降耗的效果.另外整个控制系统采用了串级控制的方式,提高了系统抗干扰能力,增加了灵敏度.

参考文献:

- [1] 于军琪. 磨矿分级系统溢流浓度的模糊智能控制 [J]. 交通大学学报: 1999, 33(9): 30 - 34
- [2] 李士勇. 模糊控制和智能控制理论与应用 [M]. 哈尔滨: 哈尔滨工业大学出版社, 1990: 78 - 90
- [3] 李启衡. 碎矿与磨矿 [M]. 曹胜利, 译. 北京: 冶金工业出版社, 2002: 68 - 70, 161 - 164
- [4] Craig, I. K, Hulbert, D. G. Optimized multivariable control of an industrial run-of-mine milling circuit [J]. Journal of South African inst. of mine and Metall, 1992, 192(6): 12 - 16
- [5] 叶彦斐. 磨矿过程优化控制及信息集成的应用研究 [D]. 成都: 电子科技大学, 2004: 27 - 28
- [6] 苏霞. 选矿自动化 [M]. 2版北京: 冶金工业出版社, 1995: 87 - 98
- [7] 方仕雄. 磨矿分级给矿模糊控制的设计和应用 [J]. 微计算机信息: 2002, 18(9): 18 - 24.

(上接第 95 页)

2) 依法保护耕地, 稳定粮田面积. 耕地面积及其质量的大幅变动会加重粮食产量的波动程度. 因此要认真贯彻和真正落实有关法律法规, 实行最严格耕地保护制度, 在县级以上的地方政府建立地籍档案, 建设“专家+政府+农户”的耕地质量信息沟通系统和耕地质量监测网络, 以确保基本农田总量不减、用途不变、质量不降.

3) 加快农业产业化进程. 走集约发展道路, 发展规模经济, 提高土地产出率; 延伸农产品链条, 提高附加值; 完善政府支持农业发展的政策体系, 建立结构合理、分工明确、运转高效的农业服务体系.

4) 加快农业现代化进程. 建立采用现代科学技术、运用现代工业装备、推行现代管理理念和方法的农业综合体系. 建立病虫害疫情监控、旱涝监控、农药监控等构建生态农业、高技术农业、信息农业、“精准农业”等新型农业发展模式.

5) 充分挖掘农业生产潜力, 对中低产田进行集中开发. 加快中低产田改造, 实施新一轮“沃土工程”, 建设高标准农田, 确保粮食生产的根基稳定. 事实证明, 推进中低产田的集中连片开发, 既避免了过去农业综合开发资金使用分散、效果不好的弊端, 提高了农业生产的综合效益, 也有利于建立优势和特色农产品基地, 推进农业区域化布局、专业化生产和产业化经营, 促进农业结构调整和增长方式转变, 加快现代农业产业体系建设.

参考文献:

- [1] 刘思峰, 郭天榜, 党耀国. 灰色系统理论及其应用 [M]. 北京: 科学出版社, 1999.
- [2] 刘斌, 刘思峰, 党耀国, 等. 基于灰色系统理论的时序数据挖掘技术 [J]. 中国工程科学, 2003, 3(9): 32 - 35.
- [3] 杨军, 刘斌, 尚曼龙. 中国大豆进口的预测与分析 [J]. 系统工程理论与实践, 2006, 26(6): 141 - 144.
- [4] 刘斌, 刘思峰, 翟振杰, 等. GM(1, 1) 时间响应函数的最优化 [J]. 中国管理科学, 2003, 11(4): 54 - 57.
- [5] 张越杰, 王军. 吉林省粮食产量波动分析 [J]. 农业技术经济, 2007, 3: 75 - 79.