

灰色马尔柯夫模型在陕西省苹果产量 预测中的应用

苏哲斌

(西安文理学院 数学系,陕西 西安 710065)

【摘 要】 【目的】将灰色系统理论与马尔柯夫链相结合,建立灰色马尔柯夫预测模型,并对陕西省苹果产量进行预测,为指导苹果产业发展提供科学依据。【方法】首先建立陕西省 1998~2007 年苹果产量的 GM(1,1)预测模型,确定出苹果产量的发展趋势和预测值,在此基础上应用马尔柯夫链理论对预测结果进行修正。【结果】2008 年陕西苹果产量预测值为 769.71 万 t,实例计算表明,灰色马尔柯夫模型更适用于随机波动性较大的数列预测问题。【结论】将灰色 GM(1,1)模型和马尔柯夫链相结合用于苹果产量预测是可行的。

【关键词】 苹果产量;GM(1,1)模型;灰色马尔柯夫模型

【中图分类号】 S661.1

【文献标识码】 A

【文章编号】 1671-9387(2009)07-0135-05

Application of grey Markov model on the prediction of apple yield in Shaanxi Province

SU Zhe-bin

(Department of Mathematics, Xi'an University of Arts and Science, Xi'an, Shaanxi 710065, China)

Abstract: 【Objective】Based on the combination of grey system theory with Markov chain theory, the grey Markov model is established to forecast the apple yield of Shaanxi Province. 【Method】Firstly a dynamic GM (1,1) model of fit is built according to the apple yield of Shaanxi Province from 1998 to 2007 to obtain the evolution trend and predominant value of apple yield. On this basis, the prediction result is modified by use of Markov chain theory. 【Result】The example computed indicates that the apple yield of Shaanxi Province would achieve 7 697.1 thousands ton in 2008 and the grey Markov model is the most suitable for prediction with big fluctuation. 【Conclusion】The combined use of grey GM (1,1) model and Markov chain method in the prediction of apple yield is feasible.

Key words: apple yield; GM (1,1) model; grey Markov model

陕西省是全国苹果最佳优生区和最大的苹果生产基地,2008 年苹果种植面积超过 53.3 万 hm^2 ,产量约占全国总产量的 1/3,占世界总产量的 13%,苹果产业已成为陕西农业经济的重要支柱之一,也是农民收入的重要来源^[1]。影响苹果产量的因素包括树种、树形、灌溉条件、施肥模式、管理措施、生长期气候条件等,其产量形成是一个十分复杂的生物学

和生态学过程,是许多因素综合作用的结果^[2-3]。因而可以把苹果产量的形成过程看作是既含有已知信息,又含有未知信息的灰色动态系统,从苹果产量时间序列本身挖掘有用的信息,建立系统发展变化的 GM(1,1)动态预测模型^[4-5]。由于苹果产量的时间数据序列常常呈现趋势性和较大的波动性,传统的 GM(1,1)灰色预测模型对随机波动性较大的数据

* [收稿日期] 2009-01-16

[基金项目] 国家自然科学基金项目(60603098);西安文理学院科研专项基金项目(KYC200722)

[作者简介] 苏哲斌(1972—),男,陕西蒲城人,硕士,主要从事灰色预测及决策理论与方法研究。E-mail: suzhebin@yahoo.com.cn

序列拟合效果较差,预测精度较低,而马尔柯夫链适合于随机波动性较大的预测问题,因此这 2 种方法的优点可以互补,前者用来揭示预测数据序列的发展变化总趋势,后者用来确定状态的转移规律,把两者有机结合起来,形成一个灰色马尔柯夫预测模型^[6-7],可大大提高随机波动较大数据序列的预测精度。目前,灰色马尔柯夫预测模型已成功应用于社会经济系统等诸多领域^[8-13],而对其在苹果产量预测中的应用研究还未见报道。本研究建立了灰色马尔柯夫预测模型,并用该模型对陕西省苹果产量进行预测,以期为指导陕西省苹果产业发展提供科学的理论依据。

1 灰色马尔柯夫预测模型

1.1 GM(1,1)预测模型的建立^[4]

灰色系统 GM(1,1) 预测模型的基本思路是把一个随时间变化的数据序列通过累加,生成新的数据列,根据灰微分方程的白化微分方程的解,还原后即得灰色 GM(1,1)预测模型。建立模型的基本过程如下。

(1)处理数据:弱化原始数列的随机性,强化规律性。设所要预测的某项指标的原始数据序列为 $X^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$,对原始数据作一次累加生成处理,得到一个新的序列 $X^{(1)}=(x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n))$,其中 $x^{(1)}(k)=\sum_{i=1}^k x^{(0)}(i),k=1,2,\cdots,n$ 。

(2)由生成序列 $X^{(1)}$ 建立 GM(1,1)模型白化形式的微分方程。

$$\frac{dx^{(1)}(t)}{dt}+ax^{(1)}(t)=b。$$

该白化形式的微分方程的离散解为:

$$\hat{x}^{(1)}(k+1)=(x^{(0)}(1)-\frac{b}{a})e^{-ak}+\frac{b}{a},k=0,1,\cdots,n。$$

其中 a,b 为辨识参数,可以通过最小二乘法拟合得到,

$$\hat{a},\hat{b}=\begin{bmatrix} a \\ b \end{bmatrix}=(B^TB)^{-1}B^TY,$$

$$B=\begin{bmatrix} -0.5(x^{(1)}(1)+x^{(1)}(2)) & 1 \\ -0.5(x^{(1)}(2)+x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -0.5(x^{(1)}(n-1)+x^{(1)}(n)) & 1 \end{bmatrix},$$

$$Y=\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}。$$

(3)对 $\hat{x}^{(1)}(k+1)$ 进行累减还原得

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)=$$

$$(1-e^a)(x^{(0)}(1)-\frac{b}{a})e^{-ak},k=1,2,\cdots。$$

当 $k\leq n-1$ 时 $\hat{x}^{(0)}(k+1)$ 为原始数据的拟合值,当 $k\geq n$ 时 $\hat{x}^{(0)}(k+1)$ 为原始序列的预测值。

令 $y(k)=\hat{x}^{(0)}(k+1)$,则曲线 $y(k)$ 反映了原始数据序列总的变化趋势。

1.2 模型状态的划分

为了进行马尔柯夫预测,必须将序列划分成若干状态^[5]。一般是以预测曲线 $y(k)$ 为基准,将坐标平面划分出若干个条形区域,每一个条形区域构成一个状态。区域的个数按照预测的具体情况而定,状态划分时应结合数据序列的特点进行,一般而言,历史数据较多时,状态宜多一些,反之则应少一些。本研究对实际值与 GM(1,1)拟合值的相对误差加以分析,以相对误差所处的不同上下阈值作为状态划分的标准,其中相对误差

$$\delta=(x^{(0)}(k+1)-\hat{x}^{(0)}(k+1))/\hat{x}^{(0)}(k+1),k=1,2,\cdots,n-1。$$

1.3 状态转移概率的计算^[5]

将原始数据划分为 s 种状态,其状态集合为 $E=\{E_1,E_2,\cdots,E_s\}$,系统的 k 步状态转移概率计算公式为:

$$p_{ij}(k)=\frac{m_{ij}(k)}{M_i},i,j=1,2,\cdots,s。$$

式中: $m_{ij}(k)$ 为由状态 E_i 经过 k 步转移到状态 E_j 的原始数据样本数, M_i 为处于状态 E_i 的原始数据样本数。由于数据序列最后的状态转向不确定,故在构建 k 步状态转移概率矩阵过程中,计算 M_i 时要去掉数据序列中最末的 k 个数据,则 k 步状态转移概率矩阵为

$$p(k)=\begin{bmatrix} p_{11}(k) & p_{12}(k) & \cdots & p_{1s}(k) \\ p_{21}(k) & p_{22}(k) & \cdots & p_{2s}(k) \\ \vdots & \vdots & \vdots & \vdots \\ p_{s1}(k) & p_{s2}(k) & \cdots & p_{ss}(k) \end{bmatrix}。$$

1.4 预测值的计算

系统各种状态转移的统计规律在状态转移概率矩阵 $p(k)$ 中得到了反映。通过考察状态转移概率矩阵 $p(k)$,可预测系统未来的发展变化。预测时需要先列出预测表,表的编制方法是^[11]:选取距离预测时刻最近的 j 个时刻,按照距离预测时刻由近到

远的次序,转移步数分别定为 $1,2,\cdots,j$,在转移步数所对应的转移矩阵中,取起始状态对应的行向量,从而组成新的状态转移概率矩阵,对新的状态转移概率矩阵列向量求和,其和最大的列向量的状态即为系统预测时刻最有可能所处的状态。

确定了系统未来的状态 E_i ,也就确定了预测值的变动区间 $[E_{1i},E_{2i}]$,则可采用下式计算最有可能的预测值。

$$\hat{y}(k)=\hat{x}^{(0)}(k+1)(1+\bar{\delta}\%)。$$

式中: $\bar{\delta}=\frac{1}{2}(\delta_1+\delta_2)$, $\delta_1、\delta_2$ 分别为某一特定状态的下限和上限。

2 陕西省苹果产量的灰色马尔柯夫预测

2.1 GM(1,1)预测模型的建立

以陕西省 1998~2007 年苹果产量^[14]作为原始数据序列 $X^{(0)}=(347.35,399.27,388.57,391.27,392.16,461.79,555.21,560.12,649.98,701.60)$ (单位:万 t),则一次累加生成序列 $X^{(1)}=(347.35,746.62,1\ 135.19,1\ 526.46,1\ 918.62,2\ 380.41,$

$2\ 935.62,3\ 495.74,4\ 145.72,4\ 847.32)$,求得辨识参数为 $a=-0.087\ 1,b=297.871\ 3$,进而得到灰色微分方程的离散解为:

$$\begin{aligned}\hat{x}^{(1)}(k+1)&=3\ 767.227\ 2e^{0.087\ 1k}-3\ 419.877\ 2,\\k&=0,1,2,\cdots,n。$$

累减还原后得到原始数据的拟合值为

$$\begin{aligned}\hat{x}^{(0)}(1)&=x^{(0)}(1),\hat{x}^{(0)}(k+1)=314.241\ 6e^{0.087\ 1k},\\k&=1,2,\cdots,n。$$

则预测曲线 $y(k)=\hat{x}^{(0)}(k+1)=314.241\ 6e^{0.087\ 1k}$,
 $k=1,2,\cdots。$

根据 GM(1,1)预测模型,所得 1998~2007 年的数据序列拟合值和相对误差结果见表 1,同时可以得到 2008 年陕西省苹果产量的预测值为 750.94 万 t。

2.2 状态的划分

根据马尔柯夫链的分析方法,以 GM(1,1)模型预测结果的相对误差为对象,选取 $-15\%, -5\%, 5\%, 15\%$ 为相对误差阈值,可以将 1998~2007 年陕西苹果产量划分为以下状态:状态 1 为 $(-15\%, -5\%]$,状态 2 为 $(-5\%, 0]$,状态 3 为 $(0, 5\%]$,状态 4 为 $(5\%, 15\%)$,原始序列状态划分结果见表 1。

表 1 1998~2007 年陕西省苹果产量(万 t)及 GM(1,1)预测模型拟合结果
Table 1 Apple yield (ten thousand ton) and fitted result of Shaanxi 1998—2007

年份 Year	实际值/万 t Actual value	预测值/万 t Predominant value	差额/万 t Deviation	相对误差/% Relative error	状态 State
1998	347.35	347.35	—	—	—
1999	399.27	342.85	56.42	14.13	4
2000	388.57	374.06	14.51	3.74	3
2001	391.27	408.10	−16.83	−4.30	2
2002	392.16	445.25	−53.09	−13.53	1
2003	461.79	485.78	−23.99	−5.19	1
2004	555.21	529.99	25.22	4.54	3
2005	560.12	578.24	−18.12	−3.23	2
2006	649.98	630.87	19.11	2.94	3
2007	701.60	688.29	13.31	1.90	3

2.3 状态转移概率矩阵的建立

根据状态划分结果可知,1998~2007 年陕西省苹果产量处于状态 E_1 的年份个数 $M_1=2$,而由状态 E_1 一步转移到状态 E_1, E_2, E_3, E_4 的个数分别为 $m_{11}=1,m_{12}=0,m_{13}=1,m_{14}=0$,根据状态转移概率的计算公式可以求出状态 E_1 一步转移到状态 E_1, E_2, E_3, E_4 的概率分别为 $p_{11}(1)=1/2, p_{12}(1)=0, p_{13}(1)=1/2, p_{14}(1)=0$,同理可求出其他一步转移概率的值,得到一步转移概率矩阵和多步转移概率矩阵分别为:

$$\begin{aligned}p(1)&= \begin{bmatrix} 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 2/3 & 1/3 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \\p(2)&= \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},\end{aligned}$$

$$p^{(3)} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$
$$p^{(4)} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}。$$

2.4 编制状态预测计算表并计算预测值

利用状态转移概率矩阵,选择离预报时刻(2008 年)最近的 4 个时段,其转移步数分别为 1,2,3 和 4,在转移步数所对应的状态转移概率矩阵中,取起始状态所对应的行向量,得到状态预测结果见表 2。

从表 2 可以看出,在离预报时刻(2008 年)最近

的 4 个初始年份中,由各年份所处初始状态经过不同步数的状态转移后,在 2008 年到达各个状态的累积状态转移概率分别为 0.5,0.67,2.83 和 0,其中到达状态 3 的累积转移概率最大,因而在 2008 年陕西省苹果产量最有可能处于状态 3。因此,根据 GM(1,1)预测模型得到的预测值为 750.94 万 t,利用灰色马尔柯夫预测模型得到 2008 年陕西省苹果产量介于[750.94,788.49](万 t),最有可能的预测值为 $750.94 \times (1 + 0.5(0 + 5\%)) = 769.71$ 万 t,这与陕西省经济作物气象服务台预测 2008 年苹果产量 770 万 t 极为相近。同理可得,2007 年与 2006 年陕西省苹果产量的预测值分别为 705.49 和 646.64 万 t,相对误差分别为-0.55%和 0.51%,预测精度较高。

表 2 状态预测计算表
Table 2 State prediction calculation

初始年份 Initial year	转移步数 Transition times	初始状态 Initial state	状态 State			
			1	2	3	4
2007	1	3	0	0.67	0.33	0
2006	2	3	0.5	0	0.5	0
2005	3	2	0	0	1	0
2004	4	3	0	0	1	0
合计 Total	—	—	0.5	0.67	2.83	0

3 结 论

本研究用 GM(1,1)模型预测曲线来反映系统宏观发展规律,获得苹果产量发展的趋势和预测值,再利用随机过程理论的马尔柯夫链寻求系统的微观波动规律,对 GM(1,1)的预测值进行修正,结果表明 2 种模型的结合可以充分利用历史数据给予的信息,对既具有趋势性又有一定波动性的苹果产量进行模拟和预测。该方法计算简便,精度较高,能够为指导陕西省苹果产业发展提供技术支持。

[参考文献]

[1] 李艳花,薛江伟. 基于 SWOT 分析的陕西苹果国际竞争力提升策略 [J]. 干旱区资源与环境,2008,22(9):30-36.
Li Y H, Xue J W. The countermeasures of raising international competitive power of Shaanxi apple products based on SWOT [J]. Journal of Arid Land Resources and Environment, 2008, 22 (9): 30-36. (in Chinese)

[2] 侯丽薇,孙立城,穆维松. 苹果产量的神经网络预测模型 [J]. 中国农业大学学报:社会科学版,2001,42(1):51-53.
Hou L W, Sun L C, Mu W S. Neural network model for forecasting apple yield [J]. Journal of China Agricultural University: Social Sciences Edition, 2001, 42(1): 51-53. (in Chinese)

[3] 姚 聪,王 俊. 中国苹果产量预测模型比较分析 [J]. 果树学

报,2007,24(5):682-684.
Yao C, Wang J. A comparative study on forecast models of apple output in China [J]. Journal of Fruit Science, 2007, 24(5): 682-684. (in Chinese)

[4] 邓聚龙. 灰理论基础 [M]. 武汉:华中科技大学出版社,2002: 221-238.
Deng J L. A textbook of grey system theory [M]. Wuhan: Huazhong University of Science and Technology Press, 2002: 221-238. (in Chinese)

[5] 刘思峰,党耀国,方志耕. 灰色系统理论及其应用 [M]. 3 版. 北京:科学出版社,2004:126-133,206-209.
Liu S F, Dang Y G, Fang Z G. Grey system theory and its application [M]. 3rd ed. Beijing: Science Press, 2004: 126-133, 206-209. (in Chinese)

[6] 李志俊,范斐斐,陈绵云,等. 灰色最小二乘马尔可夫链预测算法 [J]. 武汉理工大学学报:信息与管理工程版,2008,30(1): 59-62.
Li Z J, Fan P P, Chen M Y, et al. Grey-least square method-markov chain algorithm for forecasting [J]. Journal of Wuhan University of Technology: Information & Management Engineering Edition, 2008, 30(1): 59-62. (in Chinese)

[7] 陈有孝,林晓言. 灰色-马尔可夫链改进的预测方法 [J]. 统计与决策,2005(16):36-38.
Chen Y X, Lin X Y. An improved predictive method based on grey-markov chain [J]. Statistics and Decision, 2005(16): 36-38. (in Chinese)

[8] 程 伟,刘国璧. 灰色系统理论在粮食产量预测中的应用 [J].

湖南工程学院学报,2008,18(3):64-67.

Cheng W, Liu G B. Application of gray system to grain yield prediction [J]. Journal of Hunan Institute of Engineering, 2008,18(3):64-67. (in Chinese)

[9] 贺福利,胡 勇,陈 淳. 用灰色马尔柯夫链预测模型对我国粮食产量的预测 [J]. 数学的实践与认识,2003,33(12):45-47. He F L, Hu Y, Chen C. Forecast of grain output in China with grey markov forecasting model [J]. Mathematics in Practice and Theory, 2003,33(12):45-47. (in Chinese)

[10] 王艳玲. 基于灰色马尔可夫链的农民平均收入实证研究 [J]. 河南师范大学学报:自然科学版,2008,36(1):16-18. Wang Y L. Empirical study on per capita annual net income of rural residents based on grey markov model [J]. Journal of Henan Normal University: Natural Science Edition, 2008, 36 (1):16-18. (in Chinese)

[11] 刘 安,赵 妹,张燕平. 基于灰色-马尔可夫模型的粮食产量预测 [J]. 计算机技术与发展,2007,17(6):191-193. Liu A, Zhao S, Zhang Y P. Yield forecast based on grey-markov model [J]. Computer Technology and Development, 2007,17(6):191-193. (in Chinese)

[12] 南都国,吴溪涌. 作物产量灰色马尔柯夫链预测模型 [J]. 中国农业气象,1997,18(1):44-48. Nan D G, Wu X Y. Grey markov forecasting model of crop yield [J]. Chinese Journal of Agrometeorology, 1997,18(1):44-48. (in Chinese)

[13] 盖春英,裴玉龙. 公路货运量灰色模型——马尔可夫链预测方法研究 [J]. 中国公路学报,2003,16(3):113-116. Gai C Y, Pei Y L. Study of the GM(1,1)-markov chain model on highway freight forecast [J]. China Journal of Highway and Transport, 2003,16(3):113-116. (in Chinese)

[14] 中华人民共和国国家统计局. 中国统计年鉴 1998~2007 [M]. 北京:中国统计出版社,1998~2007. National Bureau of Statistics of China. China statistic year-book 1998—2007 [M]. Beijing:China Statistics Press,1998—2007. (in Chinese)

(上接第 134 页)

[9] 徐小彪,陈 华,张秋明. 美味猕猴桃基因组 DNA 的高效提取 [J]. 中国农学通报,2004,20(4):41-43. Xu X B, Chen H, Zhang Q M. Efficient methods for genomic DNA extraction from *Actinidia deliciosa* [J]. Chinese Agricultural Science Bulletin, 2004,20(4):41-43. (in Chinese)

[10] 陈昆松,李 方,徐昌杰,等. 改良 CTAB 法用于多年生植物组织基因组 DNA 的大量提取 [J]. 遗传,2004,26(4):529-531. Chen K S, Li F, Xu C J, et al. An efficient macro-method of genomc DNA isolation from *Actinidia chinensis* leaves [J]. Hereditas, 2004,26(4):529-531. (in Chinese)

[11] 卢启泉. 四类平菇种质资源的分子标记分析 [D]. 福州:福建农林大学,2007. Lu Q Q. Application of molecular marker in germplasm resource of four pleurotus species [D]. Fuzhou: Fujian Agriculture and Forestry University, 2007. (in Chinese)

[12] 揭褐成,周青文,陈佩度. 苎 麻栽培品种亲缘关系的 RAPD 分析 [J]. 作物学报,2002,28(2):254-259. Jie Y C, Zhou Q W, Chen P D. Genetic relation analysis of ramie cultivars with RAPD mark [J]. Acta Agron Sin, 2002, 28(2):254-259. (in Chinese)

[13] Cmhurst R N, Lints R, Gardner R C. Who was Hayward's mother [C]. Proceedings of NZ kiwifruit Authority National Research Conferince, NZ Kiwifruit Sped. Publ New Zealand, 1988(2):3-4. Kokudo K, Beppu K, Kmaa I, et al. Phylogenetic classification of introduced and indigenous *Actinidia* in Janan and indentification of interspecific hybrids using RAPDA analysis [C]. Proceedings of the Fifth Intemational Symposium on Ki-wifruit. Wuhan, China, 2002(9):15-20.

[14] 栗 琪,李作洲,黄宏文. 猕猴桃野生居群的 SSR 分析初报 [J]. 武汉植物学研究,2004,22(2):175-188. Li Q, Li Z Z, Huang H W. Preliminary study on SSR analysis in natural populations of *Actinidia* [J]. Journal of Wuhan Botanical Research, 2004,22(2):175-188. (in Chinese)

[15] Cipriani G, Bella R, Testolin R, et al. Screening RAPD primers for molecular taxonomy and cultivar fingerprinting in the genus *Aetinidia* [J]. Euphytica, 1996(2):169-174.

[16] 陈万秋,李思光,罗玉萍. 分子标记技术在猕猴桃属植物中的研究进展 [J]. 江西科学,2001,19(3):162-165. Chen W Q, Li S G, Luo Y P. Application of molecular markers on *Actinidia* [J]. Jiangxi Sci, 2001,19(3):162-165. (in Chinese)

[17] 刘亚令,李作洲,张鹏飞,等. 猕猴桃自然居群 SSR 遗传变异的空间自相关分析 [J]. 生物多样性,2006,14(5):421-434. Liu Y L, Li Z Z, Zhang P F, et al. Spatial genetic structure in natural populations of two closely related *Actinidia* species (*Actinidiaceae*) as revealed by SSR analysis [J]. Biodiversity Science, 2006,14(5):421-434. (in Chinese)