

河南省棉花产量的灰色-马尔柯夫预测模型

陈怀亮

(河南省气象科学研究所, 郑州 450003)

提 要

灰色-马尔柯夫预测模型对波动较大序列比较适合, 根据该组合模型, 用 GM(1, 1) 灰色模型预测河南省棉花单产, 再用马尔柯夫模型转移概率预测系统未来发展方向和修正预报产量。经拟合和预报, 效果较好。

关键词: 灰色-马尔柯夫预测模型 棉花 产量

引 言

棉花的生长发育受多个因子影响, 有些因子作用明显、机理清楚; 有些因子属于统计结果, 其作用大小、内在机制等均模糊不清; 还有一些我们根本无法知道的因子。因此, 棉花产量的形成是一个典型的灰色系统问题, 可以用灰色模型 GM(1, 1) 来模拟^[1]。由于 GM(1, 1) 是单调模型, 难以反映棉花产量这样波动较大的序列, 而马尔柯夫过程是一个随机变化的动态过程, 它是以时间序列内部概率分布结构为出发点来预测未来系统的发展方向的, 因而适合于随机波动较大的预测问题^[2]。本文将灰色预测与马尔柯夫预测相结合, 以后者修正前者, 提高了预报准确率。

1 预测模型

1.1 建立 GM(1, 1) 模型

设一时间序列为 $Y_{(t)}^{(0)} (t=1, 2, \dots, n)$, 其相应的 GM(1, 1) 模型时间响应函数为:

$$\hat{Y}_{(t+1)}^{(1)} = (Y_{(1)}^{(0)} - \frac{u}{a})e^{-at} + \frac{u}{a} \quad (1)$$

$$\hat{Y}_{(t+1)}^{(0)} = \hat{Y}_{(t+1)}^{(1)} - \hat{Y}_{(t)}^{(0)} \quad (2)$$

其中, a, u 为模型辨识参数, 可由最小二乘法求得; $\hat{Y}_{(t+1)}^{(0)}$ 为预测值。

1.2 状态划分

对于一个有 m 种状态的马尔柯夫非平稳随机序列的函数 $\hat{Y}_{(t)}^{(0)}$, 其任何一种状态可

表示为:

$$E_i (\otimes_{1i}, \otimes_{2i}) \quad (i=1, 2, \dots, m)$$

$$\otimes_{1i} = \hat{Y}_{(t)}^{(0)} + A_i \quad (3)$$

$$\otimes_{2i} = \hat{Y}_{(t)}^{(0)} + B_i \quad (4)$$

式(3)、(4)中 A_i, B_i 为常数, 但在不同的状态区这两个常数又不相同。 A_i, B_i 可根据序列的具体特点, 且拟合与实况大致相符来确定。

由于灰元 $Y_{(t)}^{(0)}$ 是一个时间函数, 因而灰元 $\otimes_{1i}, \otimes_{2i}$ 也随时间而变。

1.3 状态转移矩阵的构成与考察

系统由状态 E_i 一步转移到状态 E_j 的概率称为转移概率, 记为:

$$p_{ij}^{(1)} = p(E_j | E_i) = M_{ij} / M_i$$

$$(i, j = 1, 2, \dots, m) \quad (5)$$

M_{ij} 为由状态 E_i 一步转移到 E_j 的样本数, M_i 为处于 E_i 状态的样本总数。由各状态的各种转移概率为元素组成的矩阵称为状态转移概率矩阵, 记为:

$$P^{(1)} = (p_{ij}^{(1)}) = \begin{pmatrix} p_{11}^{(1)} & p_{12}^{(1)} & \dots & p_{1m}^{(1)} \\ p_{21}^{(1)} & p_{22}^{(1)} & \dots & p_{2m}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1}^{(1)} & p_{m2}^{(1)} & \dots & p_{mm}^{(1)} \end{pmatrix} \quad (6)$$

系统各种状态转移的统计规律在状态转移概率矩阵 $P^{(1)}$ 中得到了体现。当系统处于 R 状态时, 考察 $P^{(1)}$, 若第 R 行中有 $p_{Rj}^{(1)} = p_{Rj}^{(1)}$, 则系统下一步将转移为 E_j 状态, 这就是

马氏预测的基础。若 $P^{(1)}$ 中某行有两个或两个以上相等或相近的 $p_{ij}^{(1)}$, 系统未来状态的转向难以确定, 这时需进行二步或二步以上转移概率计算。高阶转移概率矩阵 $P^{(t+1)}$ 计算公式为:

$$p_{ij}^{(t+1)} = \sum_{R=1}^m p_{iR}^{(t)} \cdot p_{Rj} = \sum_{R=1}^m p_{iR} \cdot p_{Rj}^{(t)} \quad (7)$$

$$P^{(t+1)} = P^{(t)} \cdot P^{(1)} = P^{(1)} \cdot P^{(t)} \quad (8)$$

当用 $P^{(1)}$ 难以确定系统未来状态的转向时, 可用高阶状态转移概率矩阵来确定。同时, 通过考察高阶状态转移概率矩阵, 可作出状态数步后转移方向预测。

1.4 预测值计算

确定了系统的未来转向后, 也就确定了灰元 \otimes_{1t} 、 \otimes_{2t} , 由于 \otimes_{1t} 、 \otimes_{2t} 分别为某状态区的下限和上限, 则确定的预测值可取其平均值, 故:

$$\hat{y}_{(t)} = \frac{1}{2} (\otimes_{1t} + \otimes_{2t}) = \hat{Y}_{(t)}^{(0)} + \frac{1}{2} (A_i + B_i) \quad (9)$$

2 河南省棉花单产预测

2.1 河南省棉花单产 GM(1,1) 预测模型

河南省棉花单产 $\hat{Y}_{(t)}^{(0)}$ ($t=1, 2, \dots, 42$; 单位: kg/亩) 的 GM(1,1) 模型为:

$$\hat{Y}_{(t+1)}^{(1)} = 257.3514e^{0.04035871t} - 247.3514 \quad (10)$$

则据式(2)可求出各年预测值 $\hat{Y}_{(t+1)}^{(0)}$ 。

2.2 状态划分

根据河南省棉花单产实际情况, 参照产量预报业务的要求, 将河南省棉花产量序列划为如下 5 种状态 E_i ($i=1, 2, 3, 4, 5$):

$$E_1(\text{歉}): \otimes_{11} = \hat{Y}_{(t)}^{(0)} - 13.0 \quad (11)$$

$$\otimes_{21} = \hat{Y}_{(t)}^{(0)} - 7.0 \quad (12)$$

$$E_2(\text{偏歉}): \otimes_{12} = \hat{Y}_{(t)}^{(0)} - 7.0 \quad (13)$$

$$\otimes_{22} = \hat{Y}_{(t)}^{(0)} - 2.0 \quad (14)$$

$$E_3(\text{平}): \otimes_{13} = \hat{Y}_{(t)}^{(0)} - 2.0 \quad (15)$$

$$\otimes_{23} = \hat{Y}_{(t)}^{(0)} + 2.0 \quad (16)$$

$$E_4(\text{偏丰}): \otimes_{14} = \hat{Y}_{(t)}^{(0)} + 2.0 \quad (17)$$

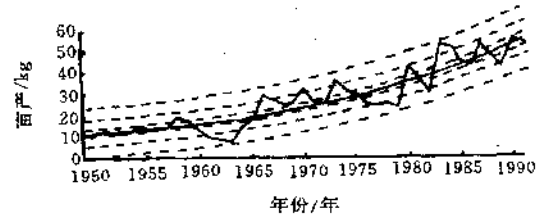
$$\otimes_{24} = \hat{Y}_{(t)}^{(0)} + 7.0 \quad (18)$$

$$E_5(\text{丰}): \otimes_{15} = \hat{Y}_{(t)}^{(0)} + 7.0 \quad (19)$$

$$\otimes_{25} = \hat{Y}_{(t)}^{(0)} + 13.0 \quad (20)$$

式中, $\hat{Y}_{(t)}^{(0)}$ 为 t 时刻按式(2)求出的棉花单产预测值。各产量状态区的常数是参照河南省棉花产量的丰、平、歉年的具体情况确定的, 而且依上式拟合, 拟合出的产量年型与实况大体相符。

上述状态划分形成 5 个围绕 $\hat{Y}_{(t)}^{(0)}$ 平滑单调曲线的条形区域, 每个区域代表一种状态 (附图)。实产 $Y_{(t)}^{(0)}$ 落在哪个区域, 该年产量就划为哪种状态。据此状态, 再结合状态转移概率矩阵, 就可以预测出下步的产量状态。



附图 河南省棉花单产状态划分
光滑曲线, $\hat{Y}_{(t)}^{(0)}$ 实折线, $Y_{(t)}^{(0)}$

2.3 状态转移概率矩阵构成与考察

按式(5)~(8), 计算各状态间的 1、2、3 步状态转移概率矩阵分别为:

$$P^{(1)} = \begin{bmatrix} 0.50 & 0.13 & 0.13 & 0.00 & 0.25 \\ 0.40 & 0.20 & 0.00 & 0.20 & 0.20 \\ 0.13 & 0.13 & 0.56 & 0.06 & 0.06 \\ 0.00 & 0.14 & 0.29 & 0.43 & 0.14 \\ 0.00 & 0.00 & 0.50 & 0.33 & 0.17 \end{bmatrix}$$

$$P^{(2)} = \begin{bmatrix} 0.32 & 0.10 & 0.26 & 0.12 & 0.20 \\ 0.28 & 0.12 & 0.21 & 0.19 & 0.20 \\ 0.18 & 0.12 & 0.38 & 0.11 & 0.11 \\ 0.09 & 0.13 & 0.35 & 0.28 & 0.13 \\ 0.06 & 0.11 & 0.46 & 0.23 & 0.11 \end{bmatrix}$$

$$P^{(3)} = \begin{bmatrix} 0.23 & 0.11 & 0.32 & 0.15 & 0.17 \\ 0.21 & 0.11 & 0.31 & 0.19 & 0.17 \\ 0.19 & 0.11 & 0.32 & 0.13 & 0.13 \\ 0.14 & 0.12 & 0.36 & 0.21 & 0.13 \\ 0.13 & 0.12 & 0.39 & 0.18 & 0.12 \end{bmatrix}$$

根据某年棉花实产状态, 由 $P^{(1)}$ 、 $P^{(2)}$ 、 $P^{(3)}$ 可分别预测未来 1、2、3 年的棉花产量趋势。

2.4 预测值的计算

河南省1990年棉花单产为53kg/亩,划为 E_3 状态,从 $P^{(1)}$ 第3行可以看出,下一步转向 E_3 状态的概率最大, $p_{33}^{(1)}$ 达0.56,因此可以初步判断出1991年也将划为 E_3 状态。根据GM(1,1)模型预测, $\hat{Y}_{(1991)}^{(0)}$ 为55.4kg/亩;再根据式(9),可得1991年全省棉花单产可能处在53.4kg/亩—57.4kg/亩之间,即55.4kg/亩左右。该年实产为53kg/亩,预测准确率(预测准确率=[1-(预测值-实况值)/实况值]×100%)为95.5%。类似可得出1992、1993、1994年的预测值(见附表)。

附表 两种预测方法结果比较表

年份	实产	GM(1,1)模型		灰色-马尔柯夫模型	
		预测值	准确率/%	预测值	准确率/%
1980	43	35.6	82.7	45.6	94.0
1981	37	37.0	100.0	37.0	100.0
1982	29	38.6	67.0	28.6	98.6
1983	53	40.1	75.8	50.2	94.7
1984	50	41.8	83.6	51.8	96.4
1985	45	43.5	96.7	43.5	96.7
1986	43	45.3	94.6	40.8	94.9
1987	53	47.2	89.0	51.7	97.5
1988	46	49.1	93.2	44.6	97.0
1989	42	51.1	78.2	41.1	97.9
1990	55	53.3	96.8	53.3	96.8
1991	53	55.4	95.5	55.4	95.5
1992*	35	57.4	36.0	47.4	64.6
1993	45	56.3	75.4	46.3	97.6
1994	43	56.8	68.8	46.8	91.9

注:1991—1994年为试报和预报;

*:1992年,由于棉铃虫严重发生,致使预测准确率下降

从附表可以看出,灰色-马尔柯夫预测的平均准确率为96.8%,比单纯用GM(1,1)模型预测准确率提高了9.7%。从4年的试报和预报看,前者又比后者的准确率高18.5%。因此可以认为,灰色-马尔柯夫预测模型比较适合随机波动较大的时间序列的预测。

3 讨论

3.1 灰色-马尔柯夫模型兼有GM(1,1)预测模型和马尔柯夫转移概率预测的长处,适合随机波动较大的序列预测。

3.2 灰色-马尔柯夫预测仍属一种统计预测,在实际工作中应注意结合具体的农业气象条件变化的特点,对一些突发性灾害天气或病虫害等因素给予特别重视。

3.3 样本越多,转移概率越能反映真实情况。因此,在可能的情况下,要尽量多选一些样本进行转移概率分析。此外,状态划分的合适与否,也将直接影响到预测的精度。一般,在历史数据较少时,状态区域应分大一些,状态数目少一些;当历史数据较多时,可增加状态数目,状态区域可划分小一些。

参考文献

- 1 陈怀亮,祝新建. 获嘉县小麦产量灰色预测预报模型. 中国农业气象,1991,1:44—46.
- 2 阎文义,刘文雄. 灰色-马尔柯夫预测模型及其应用. 系统科学与综合研究,1992,8(1):64—66.

The Grey-Markov Forecasting Model on Cotton Yield in Henan Province

Chen Huailiang

(Meteorological Institute of Henan, Zhengzhou 450003)

Abstract

The grey-Markov forecasting model is suited to the succession with larger variation. The GM(1,1) grey model has been applied to forecasts of cotton yield in Henan province. The transition possibility of Markov model has been used in predicting the future direction development of the system and amending the forecasted yield. The results are better in practice of the simulation and forecasting.

Key Words: grey-Markov forecasting model cotton yield