Operations Research and Fuzziology 运筹与模糊学, 2021, 11(4), 407-417
Published Online November 2021 in Hans. http://www.hanspub.org/journal/orf
https://doi.org/10.12677/orf.2021.114045

Hans汉斯

基于灰色马尔科夫模型贵州省水果产量预测方法

吴秋蓉

贵州大学数学与统计学院，贵阳 贵州

收稿日期：2021年9月28日；录用日期：2021年10月28日；发布日期：2021年11月4日

摘　要

水果在人们的膳食中占据了很重要的地位,与人们的身体健康有着千丝万缕的关系，它的营养价值十分丰富，随着国内发展越来越好，家庭对水果的需求也在逐渐增加。贵州省水果产量会受到气候、政策、种植方式、储存方式等等方面的因素的影响，使得产量会具有极强的波动性和随机性。在本文中对比了灰色系统模型和灰色马尔科夫模型对产量预测的结果，发现灰色马尔科夫模型对水果产量预测会更加精准。本文将选取贵州省2004年到2019年的水果产量数据(数据来源：贵州省2020年统计年鉴)进行分析研究。

关键词

水果产量，灰色模型，聚类，马尔科夫模型

Prediction Method of Fruit Yield in Guizhou Province Based on Grey Markov Model

Qiurong Wu

School of Mathematics and Statistics, Guizhou University, Guizhou Guiyang

Received: Sep. 28th, 2021; accepted: Oct. 28th, 2021; published: Nov. 4th, 2021

Abstract

Fruits occupy a very important position in people's diet and are inextricably related to people's health. It is very rich in nutritional value, and with the development getting better and better in China, the demand for fruits by families is gradually increasing. Fruit production in Guizhou province can be affected by factors such as climate, policies, planting methods, storage methods, etc., making the output extremely volatile and random. In this paper, we compare the results of the

文章引用: 吴秋蓉. 基于灰色马尔科夫模型贵州省水果产量预测方法[J]. 运筹与模糊学, 2021, 11(4): 407-417.
DOI: 10.12677/orf.2021.114045

grey system model and grey Markov model for yield prediction, and find that the grey Markov model is more accurate for fruit yield prediction. In this paper, fruit yield data which is from 2004 to 2019 in Guizhou province (data source: Statistical Yearbook of Guizhou Province in 2020) will be selected for analysis and research.

Keywords

Fruit Yield, Gray Model, Clustering, Markov Model

1. 引言

中国的水果产量占据了亚洲水果产量的 16%，由 2020 年统计年鉴可知，全国水果产量 27,400.8 万吨占全国主要农产品产量的 15%，可知水果行业在中国农业经济里占据了重要的地位；根据数据显示，中国水果产量持续增长，这也从侧面体现了中国人民在不断追求美好生活。

通过统计年鉴可知，在贵州果园种植面积从 2015 年的 307.65 (千公顷)增加到 2019 年的 684.5 (千公顷)，在五年时间内水果种植面积翻了 1.2 倍；贵州省水果种植面积在近五年内提升速度特别快，从 2015 年时在全国排十八上升到现在排全国第七；水果产量上，2019 年比 2015 年翻了 1.03 倍；从水果种植面积和产量迅速增长可知，人们对水果的需求日益增长。但是水果产业又是一个靠天吃饭的产业，需要科学地预测水果产量，再科学地规划和调整水果生产种植方式。运用科学的方法预测水果产量，可以使果农们提前做好准备，认清楚市场形式，起到有效的结构性调整作用。

国内有不少的学者对我国或者是各个省份的水果产量进行研究，季洪霄、许峰利用灰色系统模型对中国的主要生产的几种水果进行预测，表明灰色预测的误差较小[1]；姚飞、王波等利用灰色马尔科夫模型预测了中国未来的水果产量，指出灰色马尔科夫模型比单纯的灰色模型的精准度要高[2]；马创、袁野等利用灰色马尔科夫模型对中国的粮食产量进行预测的准确度比传统的灰色模型以及马尔科夫模型的预测精度要高得多[3]；邱颖利用了无偏的灰色预测模型以及马尔科夫理论对预测值进行修正的方法，对陕西省苹果年产量进行预测[4]；在以上学者的研究中都指出了灰色系统与马尔科夫理论相结合的模型预测效果较为显著。

由于水果生产是具有明显的灰度特征的，因此，虽然可以利用灰色系统模型对其进行产量的预测，但是灰色马尔科夫预测模型预测结果更加精准。本文将选用 2004~2019 年的贵州省水果产量统计数据，对其进行灰色马尔科夫模型预测分析。

2. 模型介绍

2.1. 灰色 GM(1,1)模型

灰色预测模型就是一种灰色系统对某一数列进行预测的方法模型[5]。

灰色预测模型步骤为：设有一原始时间数据数列如下

$$X^{(0)}(k)=\left\{x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\right\},k=1,2,\cdots,n \tag{1}$$

然后对原始数列进行一次累加，生成新的数据序列：

$$X^{(1)}(k)=\left\{x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n)\right\},k=1,2,\cdots,n \tag{2}$$

其中 $X^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i),k=1,2,\cdots,n$。

由累加算子性质可知，新生成的数据序列是近似服从于指数增加的规律，紧邻均值生成的序列如下：

$$Z^{(1)}=\left\{z^{(1)}(2),z^{(1)}(3),\cdots,z^{(1)}(n)\right\} \tag{3}$$

其中 $Z^{(1)}(k)=\frac{1}{2}\left[x^{(1)}(k)+x^{(1)}(k+1)\right],k=1,2,\cdots,n-1$。

由此可以建立灰色微分方程：

$$\frac{\mathrm{d}x^{(1)}(t)}{\mathrm{d}t}+ax^{(1)}(t)=b \tag{4}$$

其中，a 是发展关系数，b 是灰色作用量；其白化方程也就是影子方程为：

$$x^{(0)}(k)+az^{(1)}(k)=b,k=2,3,\cdots,n \tag{5}$$

采用最小二乘法的求解：

$$\hat{A}=\left(\hat{a},\hat{b}\right)^{\mathrm{T}}=\left(B^{\mathrm{T}}B\right)^{\mathrm{T}}B^{\mathrm{T}}Y$$

其中，$B=\begin{Bmatrix}-Z^{(1)}(1) & 1\\ -Z^{(1)}(2) & 1\\ \vdots & 1\\ -Z^{(1)}(n-1) & 1\end{Bmatrix}$，$Y=\begin{Bmatrix}x^{(0)}(2)\\ x^{(0)}(3)\\ \vdots\\ x^{(0)}(n)\end{Bmatrix}$。

将求出的 a、b 代入(4)中就得到灰色 GM(1,1)模型的时间响应函数模型如下：

$$\hat{X}^{(0)}(k)=\hat{X}^{(1)}(k)-\hat{X}^{(1)}(k-1)=\left(\hat{X}^{(0)}-\frac{\hat{a}}{\hat{b}}\right)*\mathrm{e}^{-ab}+\frac{\hat{a}}{\hat{b}}$$

然后原始数据列的预测值需经过一次累减还原得：

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

其中 $\hat{X}^{(0)}(k)=\left\{\hat{x}^{(0)}(1),\hat{x}^{(0)}(2),\cdots,\hat{x}^{(0)}(n)\right\}$。

2.2. 灰色马尔科夫模型

灰色马尔科夫模型简单来说是分成了两个部分；首先利用灰色 GM(1,1)模型对所需要预测的数列的粗略地判断该数列的发展趋势，再利用马尔科夫理论对灰色模型的预测值进行修正，使得预测的精准度提高。

至于利用马尔科夫理论对灰色模型的预测值修正，首先是计算出相对误差，利用相对误差进行状态划分，通过状态转移概率矩阵对灰色预测值进行修正，达到提高准确度的目的。

第一步计算相对误差，相对误差 $\varepsilon(t)$ 是绝对预测误差 $\Delta(t)$ 与实际值 $X(t)$ 的比，表示如下：

$$\varepsilon(t)=\frac{\Delta(t)}{X(t)} \tag{6}$$

相对误差$\varepsilon(t)$是一非平稳的随机序列；然后对$\varepsilon(t)$进行划分状态区间，划分状态区间就是将相对误差$\varepsilon(t)$划分为n个状态区间，状态区间可表示如下：

$$E_i=[e_{1i},e_{2i}],i=1,2,\cdots,n$$

其中，e_{1i},e_{2i}，分别为状态E_i的上限和下限，则相对误差的状态集合为$E=(E_1,E_2,\cdots,E_n)$。

马尔科夫预测模型的一个很关键步骤就是对相对误差的划分，状态划分后计算状态转移概率，状态转移概率的意思是现阶段是一种状态，下一个阶段转移到另外一个状态的概率。任一状态E_i经过m步转移到状态E_j的转移概率，表示如下：

$$p_{ij}(m)=\frac{M_{ij}(m)}{M_i}$$

其中，$M_{ij}(m)$为样本状态E_i经过m步转移到状态E_j的次数，M_i是状态E_i在样本中出现的次数。则可计算出m步状态转移概率矩阵可表示为：

$$P(m)=\begin{vmatrix}p_{11}(m) & p_{12}(m) & \cdots & p_{1n}(m)\\ p_{21}(m) & p_{22}(m) & \cdots & p_{2n}(m)\\ \vdots & \vdots & & \vdots\\ p_{n1}(m) & p_{n2}(m) & \cdots & p_{nn}(m)\end{vmatrix}$$

而且在状态转移矩阵中需知各行元素之和等于1；在一步状态转移矩阵中每一行的最大元素不止一个时，则需要继续计算二步状态转移矩阵甚至是计算到第n步状态转移概率矩阵，这样做只是为了使状态转移矩阵的每一行的元素的最大值只有一个，计算k步状态概率转移矩阵的计算方式如下：

$$P(k)=[P(1)]^k$$

其中，$P(1)$为一步状态转移矩阵。

最后一步就是对灰色模型的预测值进行修正，由于预测值的修正与下一个转移状态是相关的，则当所修正的值转移到下一个状态时E_j，则灰色预测值的修正公式就表示如下：

$$\hat{Y}(k)=\frac{\hat{X}(k)}{1\pm 0.5|e_{1j}+e_{2j}|}(1)$$

其中e_{1j},e_{2j}，分别为状态E_j的上限和下限，当预测值比实际值高时则取“+”号进行修正，预测值比实际值低时取“−”号进行修正。

3. 实证分析

本文将基于2004~2019年的贵州省水果产量数据(数据来源：2020年贵州省年鉴)进行实证分析，数据如表1所示：

Table 1. Guizhou 2004~2019 fruit yield data
表1. 贵州2004~2019年水果产量数据

年份	产量/万吨	年份	产量/万吨
2004	87.17	2012	139.89
2005	95.96	2013	159.23

Continued

2006	109.19	2014	187.34
2007	111.19	2015	216.89
2008	110.95	2016	235.84
2009	114.75	2017	280.14
2010	117.21	2018	369.01
2011	121.81	2019	441.98

3.1. 灰色模型建立

利用上述数据以及模型进行演算，得到灰色模型 GM(1,1)的估计参数发展系数−a 以及灰色做用量 b 分别为：$-a=0.1312277, b=42.345$，因此灰色模型的预测公式为：

$$\hat{x}^{(0)}(k+1)=42.345\mathrm{e}^{0.13128k}, k=1,2,\cdots,n \tag{8}$$

得到灰色模型预测实际情况如表 2 所示：

Table 2. Grey model predicted values
表 2. 灰色模型预测值

年份	实际产量/万吨	预测产量/万吨	绝对预测误差/万吨	相对误差	平均相对误差%	后验差比值	精度
2004	87.17	87.17	0				
2005	95.96	57.47266	38.48734	0.401076907			
2006	109.19	65.5319	43.6581	0.399836066			
2007	111.19	74.72127	36.47277078	0.328010121			
2008	110.95	85.19923	25.74781336	0.232073002			
2009	114.75	97.1465	17.6049534	0.15341813			
2010	117.21	110.7691	6.439710705	0.054942207			
2011	121.81	126.3019	4.492837812	0.036884266	13.67%	0.2350814	86.33%
2012	139.89	144.0129	4.123566628	0.029477348			
2013	159.23	164.2075	4.976852794	0.031255621			
2014	187.34	187.2339	0.106265469	0.000567233			
2015	216.89	213.4892	3.395944436	0.015657801			
2016	235.84	243.4262	7.584955899	0.032161278			
2017	280.14	277.5612	2.575322915	0.009193099			
2018	369.01	316.4829	52.528127	0.142348394			
2019	441.98	360.8625	81.121757	0.183539924			

经过计算，方差比 $C = 0.23 < 0.35$，小残差概率 $P = 1$，关联度为 $r = 0.705 > 0.6$，可以看到预测效果不错。但精度为 86.33%，有空间可以提高。拟合效果图如下：

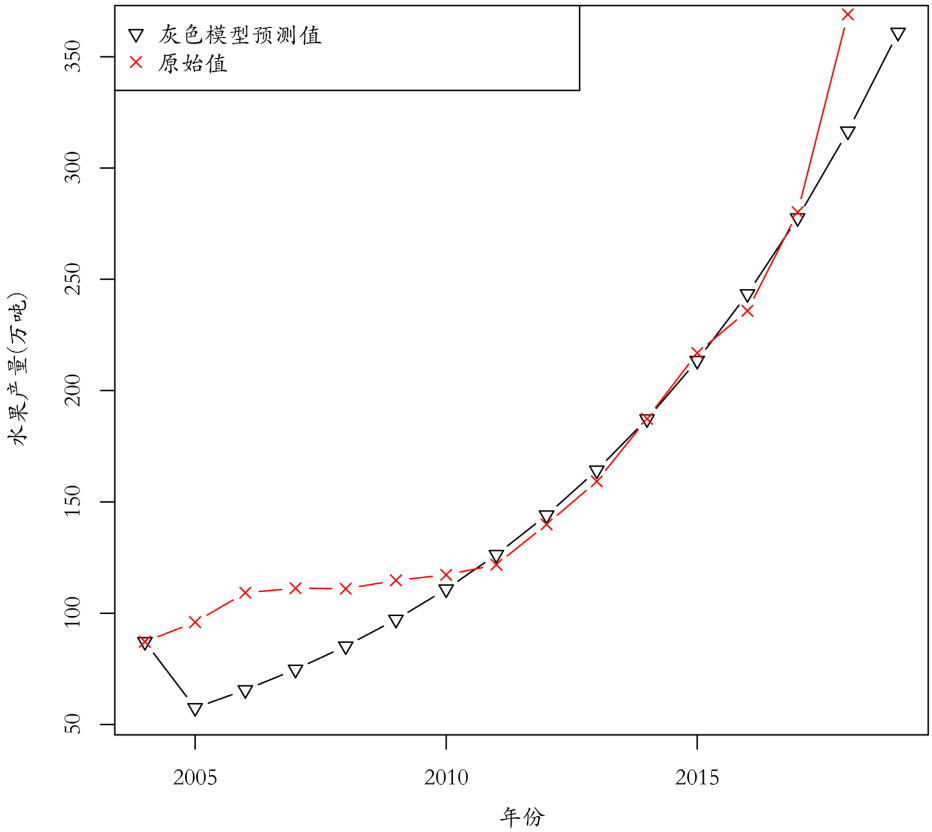

Figure 1. The fitting diagram of predicted value (▽) and actual value (×) of grey model
图 1. 灰色模型预测值(▽)与实际值(×)拟合图

从上图 1 中可观察到在前期和后期的预测中，灰色模型预测值比实际值要偏低，中间的拟合效果较好。

3.2. 灰色马尔科夫模型建立

状态空间的划分是很重要的一步，常见的划分方法有样本空间等距法，聚类分析法、均值法以及常数划分法，由于状态划分的不同会使得状态转移矩阵会存在不同，状态划分不同会使得预测结果出现差异，状态划分不准确也会使得预测结果不准确。

通过 GM(1,1)模型得到的灰色预测值，然后得到的相对残差，对相对残差进行排序。得到序列 $\varepsilon(t)$ = (0, 0.0005664747, 0.00920529, 0.01567989, 0.02947272, 0.031259770, 0.03216678, 0.03687669, 0.05495186, 0.142346, 0.1534074, 0.1835321, 0.2320934, 0.3279857, 0.3998361, 0.4010769) (后面都将保留四位小数)，建立相对误差范围，若是将 $\varepsilon(t)$ 是等距的区间，对于本文的数据会使得状态分布不均匀，本文利用了聚类的方法，对相对误差进行聚类如图 2 所示：

在图示中“single”、“complete”、“median”、“average”分别表示最短距离法、最长距离法、中间距离法、类平均法这四类聚类的计算方法，在这四张图下的聚类，都有显示，状态区间分为三组比较合适；因此，本文将状态区间划分如表 3 所示：

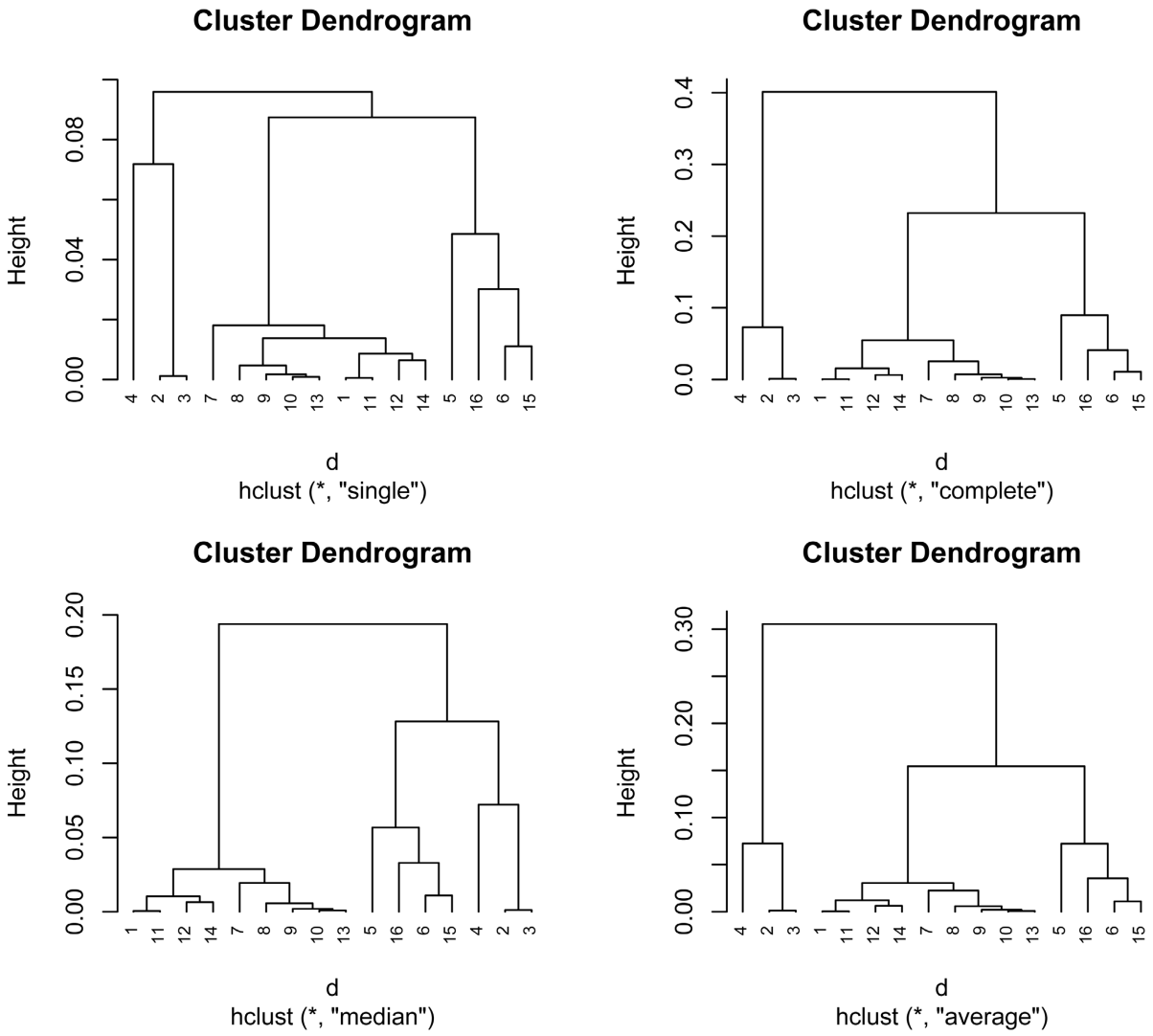

Figure 2. Clustering demonstration chart
图 2. 聚类演示图

Table 3. Markov Model division
表 3. 马尔科夫模型划分

状态划分	E_1	E_2	E_3
相对误差范围	[0.3280, 0.4011]	[0.1423, 0.2321]	[0.0006, 0.0549]

因此原始数据的状态具体情况如表 4 所示：

Table 4. Raw data specific status display
表 4. 原始数据具体状态展示

灰色模型		
年份	实际值与预测值比较	状态
2004		

Continued

2005	低估	1
2006	低估	1
2007	低估	1
2008	低估	2
2009	低估	2
2010	低估	3
2011	高估	3
2012	高估	3
2013	高估	3
2014	低估	3
2015	低估	3
2016	高估	3
2017	低估	3
2018	低估	2
2019	低估	2

状态之间的转移状况如下表 5 所示：

Table 5. Statistics of numbers of state transitions
表 5. 状态转移个数统计

		转移后状态		
		状态 1	状态 2	状态 3
实际状态	状态 1	2	1	0
	状态 2	0	3	1
	状态 3	0	1	2

由上述状态可以得到以下状态转移概率：

$$P(1)=\begin{vmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & \frac{1}{8} & \frac{7}{8} \end{vmatrix}$$

$$P(2)=\left[P(1)\right]^2=\begin{vmatrix} 0.444 & 0.472 & 0.084 \\ 0 & 0.594 & 0.406 \\ 0 & 0.203 & 0.797 \end{vmatrix}$$

$$P(3)=\left[P(1)\right]^3=\begin{vmatrix}0.296 & 0.513 & 0.191\\ 0 & 0.496 & 0.503\\ 0 & 0.252 & 0.748\end{vmatrix}$$

$$P(4)=\left[P(1)\right]^4=\begin{vmatrix}0.198 & 0.507 & 0.295\\ 0 & 0.435 & 0.565\\ 0 & 0.282 & 0.718\end{vmatrix}$$

3.3. 预测值修正

利用公式(7)对灰色模型的预测值进行修正，例如2005年的修正值为：

$$\frac{57.47}{1-0.5\left|0.3280+0.4011\right|}=90.4398$$

其它年份的具体修正情况如下表6所示：

Table 6. Comparison of the two models
表6. 两种模型的对比情况

		灰色模型		灰色马尔科夫模型	
年份	实际产量/万吨	预测产量/万吨	相对误差	预测产量/万吨	相对误差
2004	87.1700	87.1700	0.0000		
2005	95.9600	57.4727	0.4011	90.4440	0.0575
2006	109.1900	65.5319	0.3998	103.1268	0.0555
2007	111.1940	74.7213	0.3280	117.5880	0.0575
2008	110.9470	85.1992	0.2321	104.8219	0.0552
2009	114.7515	97.1465	0.1534	119.5208	0.0416
2010	117.2088	110.7691	0.0549	114.2480	0.0253
2011	121.8091	126.3019	0.0369	122.5697	0.0062
2012	139.8893	144.0129	0.0295	139.7573	0.0009
2013	159.2306	164.2075	0.0313	159.3551	0.0008
2014	187.3402	187.2339	0.0006	193.1142	0.0308
2015	216.8851	213.4892	0.0157	220.1941	0.0153
2016	235.8412	243.4262	0.0322	236.2329	0.0017
2017	280.1365	277.5612	0.0092	286.2784	0.0219
2018	369.0110	316.4829	0.1423	389.3736	0.0552
2019	441.9843	360.8625	0.1835	443.9745	0.0045

从表6中可以看到修正后的数据更接近真实数据。将修正后的数据、实际数据以及灰色模型的预测值做拟合图如图3所示，在图3中可以直观的看到修正后的预测值更加贴合实际数据。

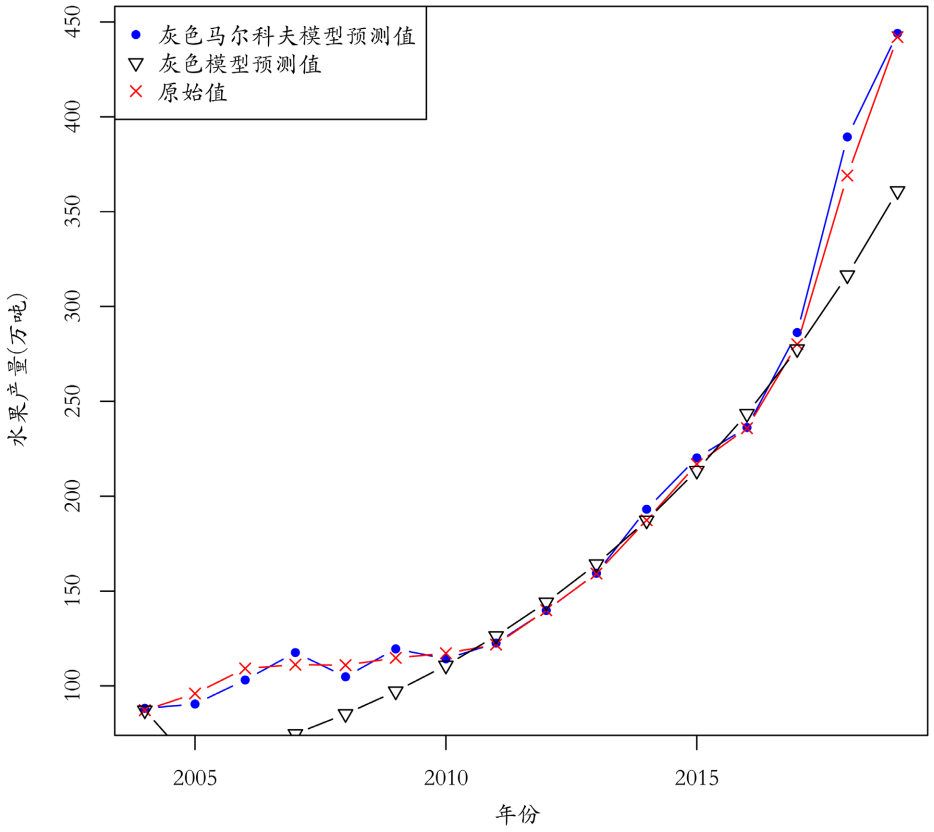

Figure 3. Predicted after Markov model correction (●), actual (×) and grey model predicted (▽) values after Markov model correction
图 3. 马尔科夫模型修正后的预测值(●)、实际值(×)和灰色模型预测值(▽)

4. 数据预测

本文将再对 2020~2023 年的水果产量先利用灰色模型 GM(1,1)进行预测，再利用公式对其进行修正。

对 2020 年的水果产量预测需要考虑 2020 年处于哪一个状态，见表 7，由于 2019 年的处于状态 2，2020 年的水果产量处于状态 1 的概率为 0，处于状态 2 的概率为 3/4，处于状态 3 的概率为 1/4；利用加权平均法取修正预测值：

Table 7. Guizhou Province fruit yield probability forecast table in 2020
表 7. 2020 年贵州省水果产量概率预测表

起始年	所在状态	起始步数	状态		
			1	2	3
2019	2	1	0.00	0.750	0.250
2018	2	2	0.00	0.594	0.406
2017	3	3	0.00	0.252	0.748
2016	3	4	0.00	0.282	0.718

在 GM(1,1)预测 2020 年的水果产量为 411.4653 (万吨)，再通过公式对其修正，得到以下结果：

$$\left(411.4653/\left(1-0.5\times 0.3744\right)\right)\times 0.75+\left(411.4653/\left(1-0.5\times 0.0609\right)\right)\times 0.25=485.7709\left(\text{万吨}\right)$$

因此得到以下表 8：

Table 8. Two models prediction results
表 8. 两种模型预测结果

模型	2020 年产量/万吨	2021 年产量/万吨	2022 年产量/万吨	2023 年产量/万吨
灰色模型预测值	411.4653	469.164	534.9535	609.9687
灰色马尔科夫模型预测值	485.7709	539.3313	578.569	663.3401

5. 结论

研究水果产量波动规律对省内农业安全问题具有一定的实际意义，本文利用 2004~2019 年的贵州省水果产量的数据(贵州省 2020 年统计年鉴)，先通过灰色模型对水果产量进行预测，然后利用马尔科夫模型对预测值修正，在修正中本文没有使用等距的方法划分状态区间，而是利用了聚类的方法对相对误差分成了三类，再继续计算得到最终结果。在本文中可知灰色模型的平均相对误差值为 13.66971%，精度为 86.33%；而灰色马尔科夫模型的平均相对误差值为 2.87%，误差大幅度地减小了，相对精度 97.13%，也比灰色模型预测的精度提高了。说明灰色马尔科夫模型能更好地预测短期水果产量。

参考文献

[1] 季洪霄, 许峰. 基于灰色系统模型的水果产量预测[J]. 科技展望, 2015, 25(21): 235-236.
[2] 姚飞, 王波, 吴天魁. 基于灰色马尔科夫模型的我国水果产量预测分析[J]. 农村经济与科技, 2014, 25(11): 113-115.
[3] 马创, 袁野, 尤海生. 基于灰色——马尔科夫模型的农产品产量预测方法[J]. 计算机科学, 2020, 47(S1): 535-539.
[4] 邱颖. 基于改进的灰色马尔科夫模型的陕西省苹果年产量预测[J]. 西部皮革, 2020, 42(8): 75.
[5] 刘思峰. 灰色系统理论及其应用[M]. 第 5 版. 北京: 科学出版社, 2010: 55-58.