

# 基于预测模型的中国原木进口数据预测研究

王 斌<sup>1</sup>,杨 抒<sup>2</sup>,王 业<sup>1</sup>,李梦洁<sup>2</sup>,蒙美丽<sup>2</sup>

(1. 新疆农业大学 现代教育技术中心,新疆 乌鲁木齐 830052;2. 新疆农业大学计算机与信息工程学院,新疆 乌鲁木齐 830052)

**摘要:**为了准确地预测未来原木进口量,利用已知中国进口原木数据,使用灰色模型 GM(1,1)模型以及最小二乘法多项式拟合的经验公式对 1993-2013 年的国内原木进口数据进行预测,通过预测值与实际值的对比,分析两组预测模型的准确性。结果表明:基于 GM(1,1)模型的预测结果,在 RMSE、MAE、MAPE 三个指标上均优于基于最小二乘法函数拟合的预测方法。利用 GM(1,1)模型预测 2014 和 2015 年国内原木进口量分别是 4 170.58 万和 4 358.99 万 m<sup>3</sup>。

**关键词:**原木;进口数据;灰色模型 GM(1,1);最小二乘法;预测模型

**中图分类号:**TP311 **文献标识码:**A **文章编号:**1002-2767(2015)02-0127-05 DOI:10.11942/j.issn1002-2767.2015.02.0127

国家原木进口总量是衡量一个国家经济贸易的重要指标。准确地预测未来原木进口量,一方面能帮助相关部门准确把握未来几年原木需求趋势,制定合理政策法规;另一方面,能帮助企业分析未来市场,确保企业的最大收益<sup>[1-3]</sup>。因此,对相关数据的准确预测,对科学地管理我国木业对外贸易活动,保证我国林业资源能得到充分有效的利用等方面都具有非常重要的意义<sup>[3-6]</sup>。

常用的预测模型有最小二乘法拟合、线性自回归模型、差分自回归移动平均模型,稀疏周期自回归模型,霍尔特-温斯特指数平滑模型等<sup>[7]</sup>。国外方面,Askari<sup>[8]</sup>、Hossain<sup>[9]</sup>及 Chi-Jie<sup>[10]</sup>等使用灰色模型、差分自回归移动平均模型预测自己研究领域未来数据趋势。国内方面,王大鹏<sup>[11]</sup>、周晓明<sup>[12]</sup>和周浩<sup>[13]</sup>等使用最小二乘法函数拟合和改进的灰色模型对相关领域数据进行了预测。

本研究使用灰色预测模型和基于最小二乘法的曲线拟合方法预测国内原木未来几年进口量的趋势,使用历史数据分析出现偏差的原因,从而确定合适的预测模型。我国 1993-2013 年的原木进口实际数据的收集是通过文献学习以及网站信息收集等方法获得的。在构建模型的过程中,使用 Matlab7 编程实现。

## 1 数据与方法

### 1.1 研究数据的来源

为检验 GM(1,1)预测方法和最小二乘法预测模型的准确性,使用到的原数据来自中国 1993-2013 年原木进口的实际数据<sup>[14]</sup>。本研究中各年度原木进口数据总量见表 1。

### 1.2 基于 GM(1,1)的预测方法

1982 年,我国学者邓聚龙教授创立了灰色系统理论<sup>[15]</sup>。灰色预测模型是在假设原数据趋势假设不变的情况下对小样本、贫信息及不确定性情况下进行预测的方法<sup>[16]</sup>。灰色模型的定义及具体的设计方法和过程<sup>[17]</sup>;

步骤 1:获得各个时刻原始数据序列

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)) n \geq 1, x^{(0)}(i) \geq 0 \tag{1}$$

通过对公式(1)灰色序列数据的一次累加生成序列如公式 2 所示,

$$x^{(1)} = \sum_{i=1}^k x^{(0)}(i) \quad k = 1, 2, \dots, n \tag{2}$$

公式(2)称为  $x^{(0)}$  的一次正向累加生成序列,记为 1-AGO。

$$x^{(0)}(k) + ax^{(1)}(k) = u \tag{3}$$

公式(3)是为灰色模型的原始形式,其中  $a$  为发展系数, $u$  为灰色作用量。

步骤 2:均值的生成是用相邻数据的平均值而构造生成的一个新数据。其中  $z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n))$ , 其中  $z^{(1)}(k) = \frac{1}{2}(x^{(1)}(k) + x^{(1)}(k-1)), k = 2, 3, \dots, n$ , 则称

收稿日期:2014-11-08  
基金项目:新疆农业大学前期资助项目(XJAU201116);新疆维吾尔自治区自然科学基金资助项目(2014211B023)  
第一作者简介:王斌(1974-),男,陕西省西安市人,硕士,讲师,从事计算机应用研究。E-mail: wangbin@xjau.edu.cn。  
通讯作者:杨抒(1979-),男,甘肃省酒泉市人,博士,讲师,从事林业信息化研究。E-mail: ys@xjau.edu.cn。

$$x^{(0)}(k) + az^{(1)}(k) = u \tag{4}$$

表 1 1993-2013 各年度中国原木进口实际数据

Table 1 Actual data of log import in China during 1993-2013

年度 Year	原木进口总量/万 m³ Imports of logs	年度 Year	原木进口总量/万 m³ Imports of logs	年度 Year	原木进口总量/万 m³ Imports of logs
1993	345.90	2000	1361.17	2007	3713.26
1994	333.50	2001	1686.38	2008	2956.96
1995	258.30	2002	2433.30	2009	2805.93
1996	318.60	2003	2545.55	2010	3434.75
1997	446.20	2004	2630.85	2011	4232.58
1998	482.30	2005	2936.80	2012	3790.13
1999	1013.57	2006	3215.29	2013	4515.90

数据来源于中国海关统计年鉴<sup>[14]</sup>。  
The data comes from China Customs Statistics Yearbook<sup>[14]</sup>.

公式(4)是灰色模型的基本形式。

步骤 3: 设  $\dot{a}=[a,u]^T$  为参数列, 并且设

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{bmatrix} \tag{5}, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

满足  $\dot{a}=(B^TB)^{-1}B^TY$ , 计算  $a,u$ 。

步骤 4: 根据上式灰色模型  $x^{(0)}(k)+az^{(1)}(k)=u$  的时间响应序列为:

$$\dot{x}^{(0)}(k+1) = (x^{(1)}(1) - \frac{u}{a})e^{-ak} + \frac{u}{a}; k = 0, 1, \dots, n-1 \tag{6}$$

步骤 5: 根据公式(6), 累减求出还原值:

$$\dot{x}^{(0)}(k+1) = \dot{x}^{(0)}(k+1) - \dot{x}^{(1)}(k); k = 1, 2, \dots, n-1 \tag{7}$$

通过这步操作所得到的数据序列即为预测结果序列。

步骤 6: 计算残差:

设残差序列为:  $Q^{(0)}=(q^{(0)}(1), q^{(0)}(2), \dots, q^{(0)}(k), \dots)$  计算残差如下:

$$q^{(0)}(k) = x^{(0)}(k) - \dot{x}^{(0)}(k+1) \tag{8}$$

步骤 7: 拟合误差的计算:

$$N^{(0)}(k) = \frac{-q^{(0)}(k)}{x^{(0)}(k)} \tag{9}$$

1.3 基于最小二乘法函数拟合的预测方法

最小二乘法通过拟合基函数的曲线使得与原数据序列的每个点的平方误差和最小, 也称为曲线拟合法<sup>[18]</sup>。本研究使用多项式进行线性曲线

拟合。最小二乘法曲线拟合的基本原理如下。  
 $l_1, l_2, \dots, l_n$  的最佳值是:

$$L = \frac{1}{n} \sum l \tag{10}$$

残差应该满足平方和  $\sum v_i^2 = \min$ , 对于任何  $L'$ , 都有

$$\begin{aligned} l_i - L' &= l_i - L + (L - L') \\ &= v_i + (L - L') \\ \sum (l_i - L')^2 &= \sum v_i^2 + 2(L - L') \sum v_i + \\ n(L - L')^2 &= \sum v_i^2 + n(L - L')^2 \geq \sum v_i^2 \end{aligned} \tag{11}$$

最小二乘法拟合实现的预测使用 Matlab 绘制数据 x、y 的曲线关系图, 观察与此图相近的基函数曲线图, 然后得出最佳的经验公式, 具体步骤为:

步骤 1: 输入  $Yx_i, y_iY$  的数据并绘制数据图。

其中, 因变量  $x$  为年度,  $y$  为实际的原木进口数据。

步骤 2: 观察绘制好图形以选择适合的基函数。

本研究采用观察法将绘制的图形与典型的数学基函数图进行比较, 将近似的基函数公式作为  $Yx_i, y_iY$  公式类型。本研究在函数曲线拟合的过程中发现, 该图形与型  $y=ax^b$  相似, 所以选择  $y=ax^b$  为经验公式。由最小二乘法得

$$\begin{aligned} Q(a,b) &= \sum (y - ax^b)^2 = \min \\ a &= \frac{\sum y_i x_i^b}{\sum x_i^{2b}} \end{aligned} \tag{12}$$

步骤 3:通过多次多项式拟合找到最合适的多项式参数

对于多项式,可用函数 polyfit 来做多项式函数的拟合之后计算系数值 a、b、c、d。经过比较发现,当算出 a 的特定值后,不断改变 b 的值,计算 b 的最小值,尽可能找到能使得  $f(x)$  的值最小的 b 的取值。以此类推,计算出成为最佳的拟合曲线和其系数值。根据上述描述,执行多项式拟合之后的系数为  $a = -0.758\ 5$ 、 $b = 23.611$ 、 $c = 29.17$ 、 $d = 48.324$ ,因此可得到多项式公式:

$$f(x) = 10.758\ 5x^3 + 23.611x^2 + 29.17x + 48.324$$

步骤 4:根据拟合的曲线方程可与得到数据的未来发展的曲线,从而估算未知数据。

1.4 预测结果的评价标准

为了计算预测的准确性。预测结果用 3 个不同的统计方法验证,分别是均方根误差(RMSE),绝对平均误差(MAE)和绝对百分比误差(MAPE)。

$$\begin{aligned} RMSE &= \sqrt{\sum_{i=1}^n (P_i - A_i)^2 / n} \\ MAE &= \sum_{i=1}^n |P_i - A_i| / n \\ M. APE &= \sum_{i=1}^n |(P_i - A_i) / A_i| / n \times 100 \end{aligned} \tag{13}$$

式中: P(Prediction)和 A(Actual values)分别代表预测值和实际值; n 代表数据的数量。

在 Lewis 定义了相关准确性的衡量标准<sup>[19]</sup>(见表 2)。

表 2 预测模型预测的 MAPE 值的分类

MAPE/%	预测能力 Predictive ability
>50	弱且不精确的预测
20~50	合理的预测
10~20	好的预测
<10	高精度预测

2 结果与分析

2.1 GM(1,1)模型预测结果分析

以 GM(1,1)估测结果见图 1。模型为基础,估测出 2013 年的原木进口量为 4 358.89 万 m<sup>3</sup> (95%置信区间),与实际值相差 157.01 万 m<sup>3</sup>,有一定的误差,然而从图 1 可以看出,近 20 年,我国原木进口量逐年增长。从预测结果中也可以看出

结果是持续上升的趋势。对于 GM(1,1)预测模型的应用需要对原始数据的变化趋势有所了解,另外需要大致掌握未知数据对预测结果的影响,在此基础上进行预测,对预测结果的分析后就可得出有用的决策信息。

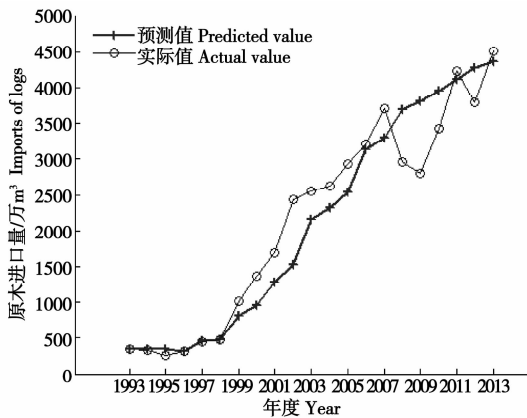


图 1 基于灰度模型的中国原木进口数据预测  
Fig. 1 The prediction of log import data in China using GM(1,1)

2.2 基于最小二乘法函数拟合预测结果分析

本研究在基于最小二乘法函数拟合的预测模型中使用原木进口数据进行了 3、9 和 15 次多项式曲线拟合(见图 2)。

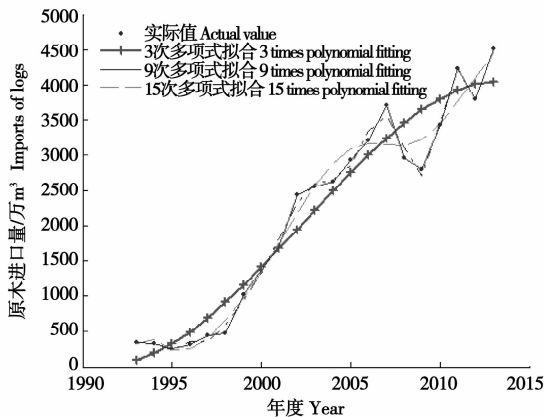


图 2 基于最小二乘法对实际数据 3、9、15 次的拟合  
Fig. 2 The prediction Log import data in China using LSM

从图 2 可以看出,3 次拟合图形更加通用,因此以 3 次多项式拟合的方程进行预测,然而所得的函数图形和实际图形仍然有较大偏差。以相关模型为基础,根据所拟合的方程式估测出 2013 年的原木进口量为 4 848.88 万 m<sup>3</sup> (95%置信区间),与实际值相差 332.98 万 m<sup>3</sup>,误差较大。因为实际数据的分布很难通过拟合曲线函数模型确定。在这种情况下,要分析多个可能的函数模型,

通过和实际值比较再确定模型。本研究得出了二项式拟合外还进行了指数函数的拟合,得到的相关方程无法分析出最佳的取值范围,拟合出的曲线图与原数据的趋势相差很大,因此放弃了以指数函数作为基函数的拟合。

### 2.3 两种预测模型结果比较

本研究对两种预测模型从 RMSE、MAE 和 MAPE3 个统计指标进行了比较,发现 GM(1,1) 在 3 个指标中都优于基于最小二乘法函数拟合的方法(见表 3)。

表 3 预测模型的准确性比较

Table 3 The accuracy of forecasting models

指标 Index	GM (1,1)	最小二乘法函数拟合 The least squares fitting function
RMSE	42.03	71.67
MAE	166.53	237.00
MAPE	9.01	22.72

由于 GM(1,1)模型预测结果准确,本文利用该模型预测 2014-2020 年我国原木进口量分别是 4 170.58 万、4 358.99 万、4 558.12 万、4 670.58 万、4 832.34 万、5 001.76 万和 5 112.53 万 m<sup>3</sup>。

## 3 结论与讨论

### 3.1 两种模型的准确性分析

对于预测的原始数据,为了对比更加明显,本研究使用的原始数据为原木进口数据中 1993-2013 年的实际数据进行最小二乘法的曲线拟合。原始数据也符合最小二乘法曲线拟合的数据要求,对于这种预测模型的原数据可以是多样的数据序列变化趋势,因为原理较为简单,但是实现较为复杂。总的来说,20 年间使用 GM(1,1)预测的结果与实际值中数据相差不大,说明灰色预测模型比最小二乘法函数拟合的预测模型预测精度更高。值得注意的是,对具体的研究对象进行预测时,模型的选择不是一成不变的,需要对相关模型多重检验后再评判其是否能满足要求,当精度满足一定的要求时,才能选择和使用这个模型。

### 3.2 出现偏差的原因分析

利用 1993-2013 年的原木进口数据进行预测的过程中,从图 1 和图 2 可以看出使用 GM(1,1) 和基于最小二乘法函数拟合的方法进行预测的结果与实际值都有偏差,但总体相近。从 1993-2013 年,这 20 年中由于国家经济持续发展,国内经济建设需要大量木材,同时国家也出台了相关

政策鼓励向外国进口原木,因此原木进口数量在 20 年中比率持续增加。然而在 2008 年由于国际金融危机的影响,原木进口总量下降,在 2008 年后原木进口总量又开始回升,使用 GM(1,1)的预测结果基本相同;2011 年开始,由于国内经济增长放缓以及原木出口国对原木出口限制等影响下,中国原木进口数据下滑较快,通过两种预测模型得到的预测结果比实际数据大。

本研究通过两种预测模型对中国原木进口量进行了预测。通过对两种模型预测结果精度比较发现,使用 GM(1,1)预测国内原木进口量是有效的,因此该模型更适合原木进口总量的预测。利用 GM(1,1)模型进行预测的数据来看,中国在未来几年中原木进口量增长将趋于放缓。

### 参考文献:

- [1] Kayacan E, Ulutas B, Kaynak O. Grey system theory-based models in time series prediction[J]. Expert Systems with Applications, 2012, 37(2): 1784-1789.
- [2] Zhang Feifei, Wu Jie, Lv Zhihui. Survey on prediction models in cloud resource management schemes[J]. Computer Engineering and Design, 2013, 34(9): 421-424.
- [3] 郑悦, 刘通, 史常亮. 我国林业经济增长影响因素的灰色关联度分析[J]. 中国林业经济, 2011(5): 11-13.
- [4] 杨抒. 基于 WEB 的林产品信息资源整合方法研究[D]. 北京: 北京林业大学, 2011.
- [5] 程念, 刘俊昌. 中国主要林产品进口来源变动分析[J]. 林业经济, 2006(3): 44-52.
- [6] 陈晓倩, 宿海颖, 王光祈. 中国主要林产品海外市场贸易政策变动分析[J]. 世界林业研究, 2013, 26(5): 88-92.
- [7] 王忠桃. 灰色预测模型相关技术研究[D]. 西安: 西安交通大学, 2005.
- [8] Mehdi Askari, Abdolvahhab Fetanat. Long-term Load Forecasting in Power System: Grey System Prediction-based Models [J]. Journal of Applied Sciences, 2011, 11: 3034-3038.
- [9] Altaf Hossain, Mohammed Nasser. Comparison of the finite mixture of ARMA-GARCH, back propagation neural networks and support-vector machines in forecasting financial returns[J]. Journal of Applied Statistics, 2011, 38(3): 533-551.
- [10] Lu Chi-Jie, Wu Jui-Yu, Chiu Chih-Chou, et al. Predicting stock index using an integrated model of NLICA, SVR and PSO[J]. International Symposium on Neural Networks, 2011, 9: 228-237.
- [11] 王大鹏. 灰色预测模型及中长期电力负荷预测应用研究[D]. 武汉: 华中科技大学, 2013.
- [12] 周晓明, 罗文柯, 李润球. 改进 GM(1,1)预测模型对我国煤炭消费需求的预测分析[J]. 矿业工程研究, 2010, 25(2): 65-68.
- [13] 周浩, 郭淑珍, 唐道洲, 等. 灰色模型对细菌性痢疾发病率预测[J]. 中国初级卫生保健, 2010, 24(10): 59.

[14] 中华人民共和国海关总署. 中国海关统计年鉴[R]. 北京: 中国海关杂志社, 2013.

[15] Deng J L. Control problems of grey systems[J]. Systems and Control Letters, 1982, 5: 288-294.

[16] Deng J L. Introduction to grey system theory[J]. The Journal of Grey System, 1989(1): 1-24.

[17] Li G, Wang T. A new model for information fusion based on grey theory[J]. Journal of Information Technology, 2011, 10: 189-194.

[18] 王可, 毛志伋. 基于 Matlab 实现最小二乘数曲线拟合[J]. 北京广播学院学报: 自然科学版, 2012, 12(2): 52-56.

[19] Lewis C D. Industrial and business forecasting methods: a practical guide to exponentials smoothing and curve fitting[J]. Butterworth Scientific London, UK, 1982, 33(2): 85-92.

## Prediction of Log Import Data in China Using GM(1,1) and LSM Models

WANG Bin<sup>1</sup>, YANG Shu<sup>2</sup>, WANG Ye<sup>1</sup>, LI Meng-jie<sup>2</sup>, MENG Mei-li<sup>2</sup>

(1. Modern Educational Technology Center, Xinjiang Agriculture University, Urumqi, Xinjiang 830052; 2. College of Computer and Information Engineering, Xinjiang Agriculture University, Urumqi, Xinjiang 830052)

**Abstract:** Through the existing data, the grey model GM (1,1) prediction model and minimum two-way method of polynomial fitting empirical formulas, domestic timber import data of 1993-2013 years was predicted, through the predicted and actual values were compared, two kinds of forecasting model were analyzed. The results showed that based on the GM (1,1) model prediction in RMSE, MAE and MAPE were better than the prediction method based on least squares fitting. The use of GM(1,1) model to forecast in 2014 and 2015 domestic log import volume was 41.705 8 million and 43.589 9 million cubic meter respectively.

**Keywords:** log import data; gray model; least squares method; forecasting model

(上接第 107 页)

[17] Guoping Z, Motohiro F, Hitoshi S. Genotypic differences in effects of cadmium on growth and nutrient composition in wheat[J]. Journal of Plant Nutrition, 2000, 9: 1337-1350.

[18] Laszlo S. Cadmium accumulation and distribution in sunflower plant[J]. Journal of Plant Nutrition, 1998, 21(2): 341-352.

[19] 邹晓霞, 陈红, 陈磊, 等. 秸秆和猪粪施用对樱桃萝卜的效果比较及对土壤性状的影响[J]. 中国农学通报, 2009, 25(5): 165-172.

[20] 敖俊华, 黄振瑞, 江永, 等. 石灰施用对酸性土壤养分状况和甘蔗生长的影响[J]. 中国农学通报, 2010, 26(15): 266-269.

## Effect of TM on Growth, Quality of Flowering Chinese Cabbage and Physicochemical Properties of Soil Contaminated with Cadmium

PANG Qiang-qiang, CHEN Wang, CHEN Ri-yuan, LIU Hou-cheng, SONG Shi-wei, SU Wei, SUN Guang-wen

(College of Horticulture, South China Agricultural University, Guangzhou, Guangdong 510642)

**Abstract:** In order to improve vegetable quality and control cadmium pollution of vegetables, effect of TM on growth and quality of flowering chinese cabbage and physicochemical properties of soil contaminated with cadmium were studied. The results showed that compared with CK, after application of TM, the fresh weight of flowering chinese cabbage increased by 27.33% dramatically, the total cadmium content of shoot decreased by 9.57%, the content of nitrate decreased significantly by 48.08%, the content of VC increased significantly by 31.92%. Compared with CK, TM treatment could decrease the content of total cadmium by 0.65%, the pH, total nitrogen contents of soil increased significantly, total phosphorus and alkali-hydro nitrogen contents decreased markedly, contents of total potassium, available phosphorus and available potassium had no notably change.

**Keywords:** TM; flowering chinese cabbage; cadmium; growth; quality; soil; physicochemical properties