

# 卷烟烟气焦油量的 ARMA 预测模型研究

陆鸣<sup>1</sup>, 李东亮<sup>2</sup>, 许自成<sup>2</sup>, 刘秀彩<sup>1</sup>, 许寒春<sup>1</sup>

(1. 福建中烟技术中心, 福建厦门 361022; 2. 河南农业大学国家烟草栽培生理生化研究基地, 河南郑州 450002)

**摘要** [目的]为计算机辅助配方研究提供理论依据。[方法]以同一品牌不同批次卷烟的烟气焦油量为研究对象, 采用时间序列分析法, 建立了卷烟烟气焦油量的预测模型, 并进行了模型预测验证。[结果]ARMA(2,2)模型的 AIC、LF 和 FPE 值在各模型中均最小, 所以选择 ARMA(2,2)为卷烟烟气焦油量的预测模型, 即:  $(1-1.622q^{-1}+0.844q^{-2})y(t)=(1-1.836q^{-1}+1.02q^{-2})e(t)$ 。根据对模型残差序列进行的白噪声检验判定, 建立的 ARMA(2,2)模型是显著有效模型, 模型预测验证表明模型预测精度达 99.51%, 平均相对误差为 0.49%, 属于一级(优等)模型。时间序列一般用于短期预测, 不能用于长期预测。[结论]该研究建立的卷烟烟气焦油量的 ARMA(2,2)预测模型的预测精度高、误差小, 可以用于卷烟烟气焦油量的短期预测。

**关键词** 卷烟; 烟气焦油量; ARMA 模型; 时间序列分析

中图分类号 S572 文献标识码 A 文章编号 0517-6611(2008)14-05696-03

## ARMA Forecast Model of Tar of Cigarette Smoke

LU Ming et al (Technology Center of China Tobacco Fujian Industrial Corporation, Xiamen, Fujian 361022)

**Abstract** [Objective] The study aimed to provide a base for computer-aided formula design. [Method] With the tar of cigarettes from different batches of cigarettes with the same trademark as tested materials and by using time series sequence method, the forecast model of tar of cigarette smoke was built and validated. [Result] Among forecast models, the values of AIC, LF and FPE were all least in ARMA(2,2) model, namely  $(1-1.622q^{-1}+0.844q^{-2})y(t)=(1-1.836q^{-1}+1.02q^{-2})e(t)$ , so it was selected as forecast model for tar of cigarette smoke. The established ARMA(2,2) model was significantly effective model according to autocorrelation check of the residuals on the model residual sequence. The forecast precision of ARMA(2,2) model reached 99.51% and the average relative error was 0.49%, so the ARMA(2,2) model belonged to the first grade model (excellent model). However, time series model was usually used for short-term forecast but not for long-term forecast. [Conclusion] The forecast model of tar of cigarettes smoke built in this study had a better precision, lower error and could be used for short-term forecast of the tar of cigarettes smoke.

**Key words** Cigarettes; Tar of cigarette; ARMA model; Time series analysis

卷烟烟气焦油量的预测模型是计算机辅助配方的一项基础研究工作, 有助于卷烟配方的维护和焦油量的控制。已有研究<sup>[1-3]</sup>利用误差反传前向网络(BP神经网络)、多元回归等统计方法在烟丝中总糖、总氮、氯和烟气焦油量之间或烟叶焦油量和烟气焦油量之间建立了因果预测模型, 但由于影响卷烟焦油量的因素众多<sup>[4]</sup>, 如何选择进入预测模型的自变量成为烟气焦油模型构建成败的关键。在缺乏被研究对象的必要相关资料而无法运用因果预测时, 时间序列是用于预测的一个有效工具<sup>[5]</sup>。利用时间序列不需要知道预测变量的因果关系, 在系统的动态性较强, 关于影响预测变量的决定性因素信息很少, 且有足够多的数据量可以用来构成一个合理长度的时间序列情况下, 运用时间序列模型往往可达到很好的效果<sup>[6-9]</sup>。时间序列通过变量的自身历史观测值来反映有关因素对预测目标的影响和作用, 不受模型变量相互独立的假设条件约束, 所构成的模型可以消除一般回归预测方法中由于自变量选择、多重共线性等造成的困难<sup>[10-12]</sup>。为此, 笔者选取同一品牌不同批次卷烟的烟气焦油量作为研究对象, 采用时间序列分析法, 分析和构建了卷烟烟气焦油量的预测模型, 旨在为计算机辅助配方研究提供理论依据。

## 1 材料与方法

### 1.1 材料

1.1.1 供试样品。选取厦门卷烟厂某品牌(品牌 A)的 15 个批次卷烟产品, 每批样品各 5 条。分别编号为 200601、200602、200603、200604、200605、200606、200607、200608、200609、

200610、200611、200612、200701、200702 和 200703。其中编号为 200601~200612 的卷烟样品用作模型构建, 200701~200703 的卷烟样品用作模型预测验证。

1.1.2 仪器。SM400 型 20 通道吸烟机(英国费尔创纳公司); HP6890 气相色谱仪(美国惠普公司); GFL-3017 振荡器(德国); AT261 电子天平(瑞士梅特勒); IM1000 型恒温恒湿箱(澳大利亚格雷升公司); 恒温恒湿机(美国利伯特公司)。

### 1.2 方法

1.2.1 数据预处理。按照文献[13-17]规定的方法分别对 15 个批次卷烟的烟气焦油量进行检测, 全部数据的处理采用 MATLAB 统计方法<sup>[18-23]</sup>进行。首先对数据进行预处理, 即对它的平稳性和纯随机性进行检验, 以得到零均值平稳时间序列, 从而满足时序建模的前提要求<sup>[5]</sup>。作卷烟批次 200601~200612 的 12 批卷烟焦油量的时序图, 然后对数据进行零均值光滑化处理, 并作数据的自相关图。

1.2.2 模型的结构判定、参数估计。作卷烟批次 200601~200612 的 12 批卷烟零均值光滑化后焦油量的偏自相关图, 先依据自相关图和偏自相关图的分布特征确定模型的类别, 再探索模型的合适阶次<sup>[20-23]</sup>。

1.2.3 模型检验。在确定序列模型后, 还需对模型残差序列进行白噪声检验。对模型的残差序列作自相关图。

1.2.4 模型预测。应用建立的 ARMA(2,2)模型对 200701~200703 的 3 批次卷烟烟气焦油量进行预测并与实际测定值进行比较。

## 2 结果与分析

2.1 原始数据检测结果 检测结果见表 1。

2.2 数据预处理结果 平稳序列的时序图应该显示出该序列始终在一个常数值附近随机波动, 而且波动的范围有界的

作者简介 陆鸣(1972-), 男, 上海人, 工程师, 从事烟草化学、卷烟工艺研究。\*通讯作者, 工程师, E-mail: xuhc@xmjyc.com。

收稿日期 2008-03-10

表 1 不同批次卷烟烟气焦油量检测结果  
Table 1 Detection results of smoke tar amount in cigarettes with different batches

品牌 A Brand	平均重量 Average weight//g	平均吸阻 Average draw resistance//kpa	湿焦 Wet coke//mg/支	水分 Moisture//mg/支	烟碱 Nicotine//mg/支	口数 Number//口/支	焦油 Tar//mg/支
200601	0.91	1.140	17.95	2.21	1.27	7.7	14.5
200602	0.92	1.100	17.86	1.81	1.31	8.2	14.7
200603	0.92	1.190	16.40	1.64	1.40	7.8	13.4
200604	0.94	1.220	16.49	1.53	1.43	8.0	13.5
200605	0.93	1.170	15.69	1.54	1.41	7.7	12.7
200606	0.92	1.180	16.49	1.51	1.44	7.7	13.5
200607	0.92	1.190	15.94	1.51	1.29	7.8	13.1
200608	0.93	1.210	15.04	1.26	1.32	8.0	12.5
200609	0.91	1.170	18.49	2.41	1.46	8.0	14.6
200610	0.93	1.190	17.21	2.30	1.47	7.9	13.4
200611	0.91	1.170	17.32	2.20	1.37	8.0	13.8
200612	0.92	1.210	17.51	1.83	1.40	7.6	14.3
200701	0.91	1.040	18.35	1.50	1.50	7.2	15.4
200702	0.91	1.000	17.66	1.55	1.35	7.3	14.8
200703	0.91	1.090	17.50	1.63	1.33	7.2	14.5

特点。如果显示出有明显的趋势性或周期性,则其通常不是平稳序列<sup>[6]</sup>。图 1 表明,该观察值序列基本围绕在 13.7 mg/支附近随机波动,没有明显的趋势性和周期性,基本可以视为平稳序列。为了稳妥起见,还需要利用自相关图进一步辅助识别。图 2 表明,该序列的自相关系数很快衰减为零,而且都落在 2 倍标准差范围内,所以,结合时序图和自相关图的分析结果,可以判定该序列为平稳序列<sup>[6]</sup>。应用自相关图对序列纯随机性进行检验,自相关图显示,该序列为非纯随机序列<sup>[6]</sup>,蕴含着值得提取的相关信息,可以对该序列进行进一步的分析研究。

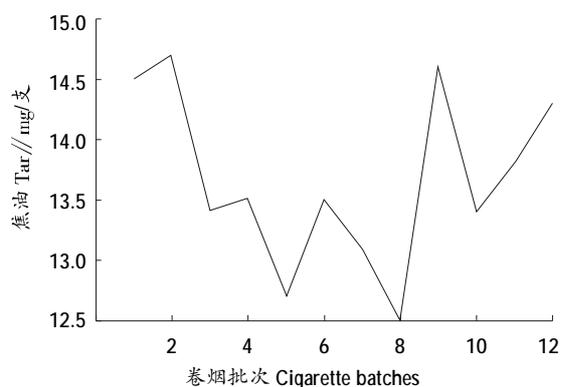


图 1 卷烟烟气焦油量的时序关系

Fig. 1 Sequence diagram of smoke tar amount in cigarettes

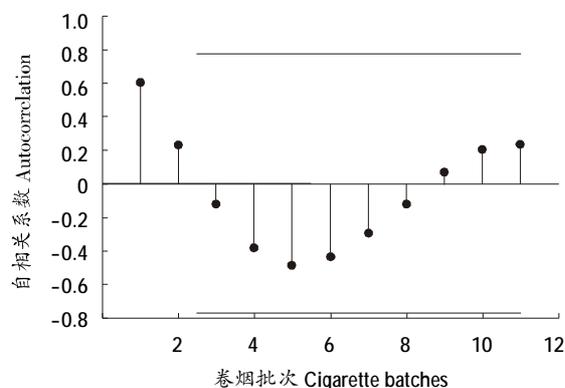


图 2 平滑零均值化后卷烟烟气焦油量的自相关关系

Fig. 2 Autocorrelation of smoke tar amount in cigarettes after smoothing zero equalization

2.3 模型结构判定和参数估计结果 平稳时间序列模型有自回归移动平均(Auto Regression Moving Average, ARMA)、

自回归(Auto Regression, AR)、移动平均(Moving Average, MA)模型,ARMA 模型的自相关系数和偏自相关系数都呈现拖尾性<sup>[6]</sup>。图 2-3 表明,平滑零均值化后卷烟烟气焦油量的自相关、偏自相关系数都拖尾,所以选择 ARMA 模型为该序列的合适模型。ARMA 模型结构一般为  $A(q)y(t)=C(q)e(t)$ , 其中,  $A(q)=1+a_1q^{-1}+a_2q^{-2}+\dots+a_nq^{-n}$ ,  $C(q)=1+c_1q^{-1}+c_2q^{-2}+\dots+c_mq^{-m}$ ,  $q$  为延迟算子。选择 ARMA(1,1)、ARMA(1,2)、ARMA(2,1)、ARMA(2,2)、ARMA(2,3)、ARMA(3,2) 进行模型阶次的选择和判定。

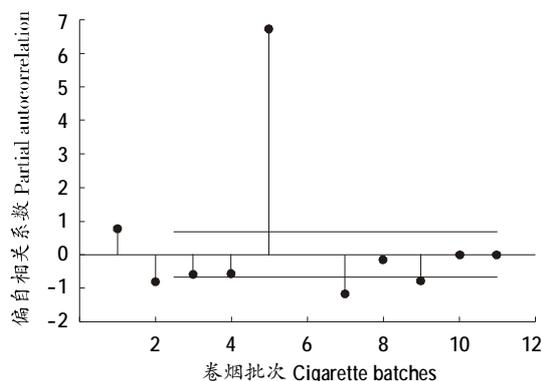


图 3 平滑零均值化后卷烟烟气焦油量的偏自相关关系

Fig. 3 Partial autocorrelation of smoke tar amount in cigarettes after smoothing zero equalization

一个拟合模型的好坏可以从用来衡量拟合程度的使然函数值和模型中未知参数的个数 2 方面去考察。一个好的拟合模型应该是拟合精度和未知参数个数的综合最优配置,采用最小信息量准则(An Information Criterion, AIC)、最终预测误差(Final Prediction Error, FPE)和函数损失(Loss Function, LF)对模型的适用性进行综合判定和检验<sup>[5-6]</sup>。  $AIC=-2\ln(\text{模型的极大似然函数值})+2(\text{模型中未知参数个数})$ , FPE, LF, 3 个评价指标在各自的准则函数取得最小值时的模型为适用模型。表 2 表明, ARMA(2,2)模型的 AIC、LF 和 FPE 在各模型中为最小,所以选择 ARMA(2,2)为该序列的预测模型,即卷烟烟气焦油量的预测模型为  $(1-1.622q^{-1}+0.844q^{-2})y(t)=(1-1.836q^{-1}+1.02q^{-2})e(t)$ 。

2.4 模型检测结果 一个模型是否显著有效主要看它提取的信息是否充分,一个好的拟合模型应该能够提取观察值序列中几乎所有样本的相关信息,即拟合残差项中将不再蕴含任何相关信息,残差序列应该为白噪声序列,这样的

表 2 各模型参数和 LF、FPE 和 AIC  
Table 2 Model parameters, LF, FPE and AIC

模型 Model	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	c <sub>1</sub>	c <sub>2</sub>	c <sub>3</sub>	LF	FPE	AIC
ARMA(1,1)	-0.678 8	-	-	0.863 8	-	-	0.040 1	0.066 8	-2.717 3
ARMA(1,2)	-1.003 0	-	-	0.413 4	-0.602 4	-	0.059 7	0.145 0	-1.984 8
ARMA(2,1)	-1.657 0	0.899 8	-	-1.010 0	-	-	0.014 8	0.035 9	-3.381 8
ARMA(2,2)	-1.622 0	0.844 0	-	-1.836 0	1.020 0	-	0.009 1	0.027 2	-3.701 6
ARMA(2,3)	-1.566 0	0.783 9	-	-0.725 4	-0.073 7	-0.215 4	0.018 4	0.092 0	-2.662 0
ARMA(3,2)	-0.539 2	-0.750 9	0.795 0	-0.758 0	-0.254 6	-	0.015 3	0.076 7	-2.843 8

模型为显著有效模型<sup>[5-6]</sup>。

图 4 表明,该残差序列的自相关系数一直都较小,始终控制在 2 倍的标准差范围以内,可以认为该残差序列自始至终都在零轴附近随机波动,符合白噪声序列的特征,所以判定建立的 ARMA(2,2)模型为显著有效模型。

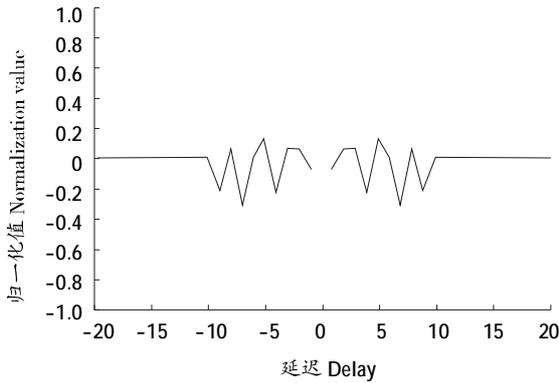


图 4 卷烟烟气焦油量拟合误差的自相关关系  
Fig. 4 Partial autocorrelation of fitting error of smoke tar amount in cigarettes

2.5 模型预测结果 预测结果见图 5 和表 3。表 3 表明,平均相对误差<1%,均方差比值<0.35,小误差概率>0.95,说明建立的卷烟焦油量 ARMA(2,2)预测模型为一级(优等)模型<sup>[24]</sup>,是模型等级中最高预测模型。

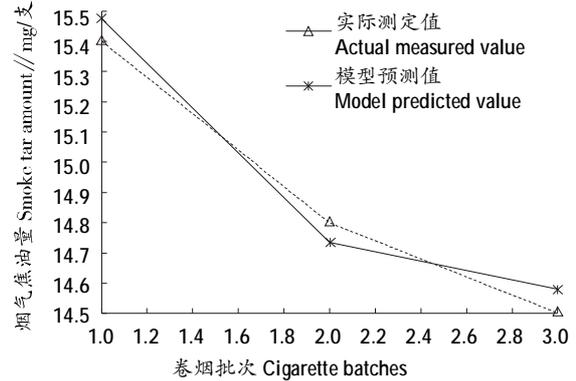


图 5 卷烟烟气焦油量实际测定值与模型预测值的比较  
Fig. 5 Comparison of actual measured value and model predicted value of smoke tar amount in cigarettes

表 3 ARMA(2,2)模型焦油实际测定值与模型预测值的比较结果

Table 3 Comparative results of actual measured value and model predicted value of ARMA(2,2)model tar

实际值 Actual value//mg/支	预测值 Predicted value//mg/支	残差 Residual error//mg/支	相对误差 Relative error//%	平均相对误差 Average relative error//%	均方差比值 Average deviation ratio	小误差概率 Small error probability	平均相对精度 Average relative accuracy//%
15.4	15.477 9	-0.077 9	0.505 8	0.49	0.180 7	1.000 0	99.51
14.8	14.733 7	0.066 3	0.448 0	-	-	-	-
14.5	14.576 4	-0.076 4	0.526 9	-	-	-	-

3 结论

(1)建立了卷烟烟气焦油量的 ARMA(2,2)预测模型,模型预测精度达到了 99.51%,平均相对误差为 0.49%,属于一级(优等)模型,预测精度优于文献[1-3]所建立的因果模型。

(2)利用时间序列建立预测模型可以避开模型自变量的选择这一难题,它不受模型变量相互独立的假设条件约束,所构成的模型可以消除一般回归预测方法中由于自变量选择、多重共线性等造成的困难。但时间序列模型也有其局限性,由于随着序列长度的增加,变量之间的自相关性迅速衰减,所以时间序列一般用于短期预测,不能用作长期预测。

参考文献

[1] 杨再波,叶冲,韩伟,等.基于 BP 神经网络在卷烟焦油预测中的应用[J].黔南民族师范学院学报,2006(6):5-8.  
 [2] 张志刚,王二彬,苏东赢.卷烟常规化学成分与焦油的线性回归分析[J].烟草科技,2003(11):32-33.  
 [3] 徐雅静,汪远征,王建民.卷烟焦油含量预测的数学模型[J].郑州轻工业学院学报:自然科学版,2005,20(3):35-38.  
 [4] 彭传新,尤长虹,李兵役,等.影响卷烟焦油量的因素探讨[J].烟草科技,2000(11):5-8.  
 [5] 王燕.应用时间序列分析[M].北京:中国人民大学出版社,2005.  
 [6] GEORGE E P B.GWILYM M J.GREGORY C R.时间序列分析预测与控制[M].顾岚,译.北京:中国统计出版社,1997.  
 [7] 徐国祥.统计预测与决策[M].上海:上海财经大学,2000.  
 [8] 胡永红,吴志峰,李定强,等.基于 ARIMA 模型的区域水生态足迹

时间序列分析[J].生态环境,2006,15(1):94-98.  
 [9] 朱丹.非典型的时间序列预测法的选择与应用[J].广东职业技术学院学报,2000(1):47-52.  
 [10] 杨志钦,顾岚.时间序列分析与动态数据建模[M].北京:北京工业学院出版社,1986.  
 [11] 崔敬巍,谢里阳.自相关过程的统计控制方法研究[J].机械制造,2006,44(5):66-68.  
 [12] JIANG WEI, TSUI K L.A new SPC monitoring method:the ARMA chart[J].Technometrics,2000,42(4):399-410.  
 [13] GB/T 16447-2004/ISO3402:1999,烟草及烟草制品调节和测试的大气环境[S].1999.  
 [14] GB/T 16450-2004,常规分析用吸烟机定义和标准条件[S].2004.  
 [15] GB/T 19069-2004,卷烟用常规分析用吸烟机测定和焦油[S].2004.  
 [16] YC/T 157-2001,卷烟总粒相物中水分的测定气相色谱法[S].2001.  
 [17] YC/T 156-2001,卷烟总粒相物中烟碱的测定气相色谱法[S].2001.  
 [18] 张圣勤.MATLAB7.0 应用教程[M].北京:机械工业出版社,2006.  
 [19] DUANE HANSELMAN, BRUCE LITTLEFIELD.精通 MATLAB7[M].北京:清华大学出版社,2006.  
 [20] 王国锋,王子良,王太勇,等.Matlab 在时间序列分析中的应用[J].应用科技,2003,30(5):36-38.  
 [21] 曹鲜花,吴美平,胡小平.Matlab 在光纤陀螺随机漂移建模中的应用[J].航天控制,2007,25(1):18-20,25.  
 [22] 张善文,雷英杰,冯有前.MATLAB 在时间序列分析中的应用[M].西安:西安电子科技大学出版社,2007.  
 [23] 韩路跃,杜行检.基于 MATLAB 的时间序列建模与预测[J].计算机仿真,2005,22(4):105-107,182.  
 [24] 刘思峰,郭天榜,党耀国,等.灰色系统理论及其应用[M].北京:科学出版社,1999.