

doi: 10.3969/j.issn.1006-1576.2010.01.013

基于 SAS 的火控精度时间序列分析

周小程¹, 陈健¹, 马向玲², 范洪达²

(1. 海军航空工程学院 研究生 5 队, 山东 烟台 264001; 2. 海军航空工程学院 兵器科学与技术系, 山东 烟台 264001)

摘要: 为分析火控精度时间序列, 采用 SAS 语言模块, 分别从时间序列数据的预处理、时间序列建模、时间序列的预报方面, 讨论了时间序列分析问题, 并结合火控精度实验中的数据实例探讨了 SAS 的优越性和实用性。结果表明, 利用 SAS 分析火控系统模型, 可大大提高效率, 能实时解决工程中发生的一些问题。

关键词: SAS; 火控精度; 时间序列

中图分类号: TP274 **文献标识码:** A

Analysis of Time Series of Fire Control Precision Based on SAS

ZHOU Xiao-cheng¹, CHEN Jian¹, MA Xiang-ling², FAN Hong-da²

(1. No. 5 Brigade of Postgraduate, Naval Aeronautical & Astronautical University, Yantai 264001, China;

2. Dept. of Ordnance Science & Technology, Naval Aeronautical & Astronautical University, Yantai 264001, China)

Abstract: Adopt SAS to analyze the fire control precision time series. Discuss the time series analysis from pretreatment of time series data, time series modeling, and time series forecasting. Use the data example in fire control precision examines to discuss the superiority and practicality of SAS. The result shows that the method raises the analysis efficiency greatly and it is an effective quantitative analysis method to solve the engineering problems.

Keywords: SAS; Fire control precision; Time series

0 引言

时间序列是一种现代数据分析与处理技术, 是系统辨识与系统分析的重要方法之一^[1]。在对多维时间序列的分析中, 通常用到 C 语言、Fortran 语言和 Matlab 语言^[2]。而 SAS 语言在国际上被誉为标准通用软件, 可用来分析数据编写报告。SAS 语言采用积木式模块结构, 拥有目前最强大的多元统计分析程序集, 库函数极其丰富, 对多维时间序列的分析更加方便灵活。SAS 语言模块作为目前功能最强的多元统计分析模块, 在国际上深受好评^[3]。故对基于 SAS 的时间序列分析问题进行研究。

1 时间序列的分析

1.1 时间序列数据的预处理

时间序列的预处理是时间序列建模的基础, 包括数据的采集, 野点的剔除、平稳性的检验和趋势项的提取。火控系统精度时间序列如图 1。

1.1.1 剔除野点

野点又叫异常点, 是时间序列集合中严重偏离大部分数据所呈现趋势的小部分数据点, 其成因包括: 操作和记录过失、采样环境的变化、一些随机变量的结果^[4]。野点的产生会带来虚假信息, 所以

不能忽略。图 2 为剔除异常值后的时序图。

在理论上, 假定正常的序列值 $\{X_t\}$ 是平滑的, 而异常点是突变的:

$$\bar{X}_t - kS_t < X_{t+1} < \bar{X}_t + kS_t \quad (1)$$

$$\text{其中, } \bar{X}_t = \frac{1}{t} \sum_{j=1}^t X_j, S_t^2 = \overline{X_t^2} - \bar{X}_t^2, \overline{X_t^2} = \frac{1}{t} \sum_{j=1}^t X_j^2,$$

$k \in [3, 9]$, 通常取 $k = 6$ 。

因为这种方法是建立在数理统计的基础上的, 为了满足其理论要求, 时间序列的长度不能太少。

若 X_{t+1} 是个异常点, 可以用 \hat{X}_t 来修正 X_{t+1} , 即:

$$\hat{X}_t = 2X_t - X_{t-1} \quad (2)$$

1.1.2 检验平稳性

主要采用图检验法和自相关函数检验法检验平稳性。

图检验法是根据时间序列图的变化趋势来判断是否平稳, 平稳的时间序列均值函数为常数, 斜方差函数只与时间间隔有关, 图形应围绕某常数上下波动。非平稳的时间序列的变化趋势应该有显著的上升和下降趋势。

自相关函数检验法的依据是: 平稳时间序列的自相关函数要么是截尾的, 要么是拖尾的, 如果一

收稿日期: 2009-07-22; 修回日期: 2009-10-29

作者简介: 周小程 (1984-), 男, 辽宁人, 海军航空工程学院在读硕士, 从事动态数据建模研究。

个时间序列零均值化以后的自相关函数出现了缓慢衰减或周期性的衰减情况，则说明序列可能存在某种趋势和周期性。

根据图 1 可见：火控系统精度时间序列存在某

种趋势项，显然是非平稳时间序列。为进一步确定其平稳性，编写命令做出该序列的自相关图如图 3，图 3 显示了自相关函数出现缓慢的衰减，表明该火控精度时间序列存在某种趋势。

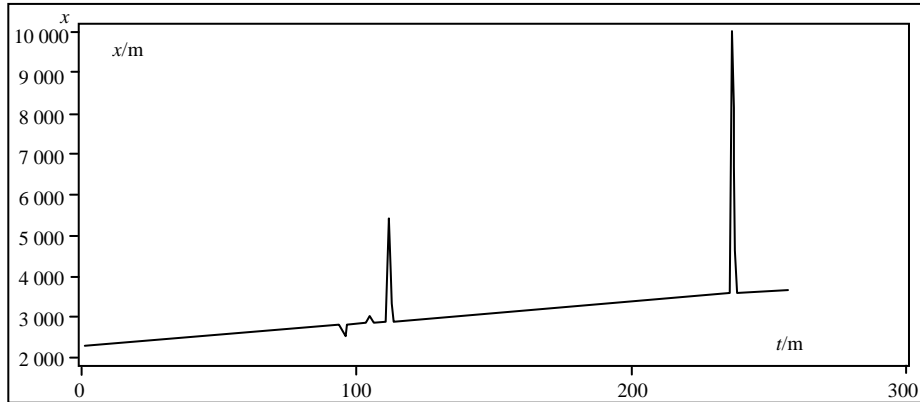


图 1 火控系统精度时间序列图

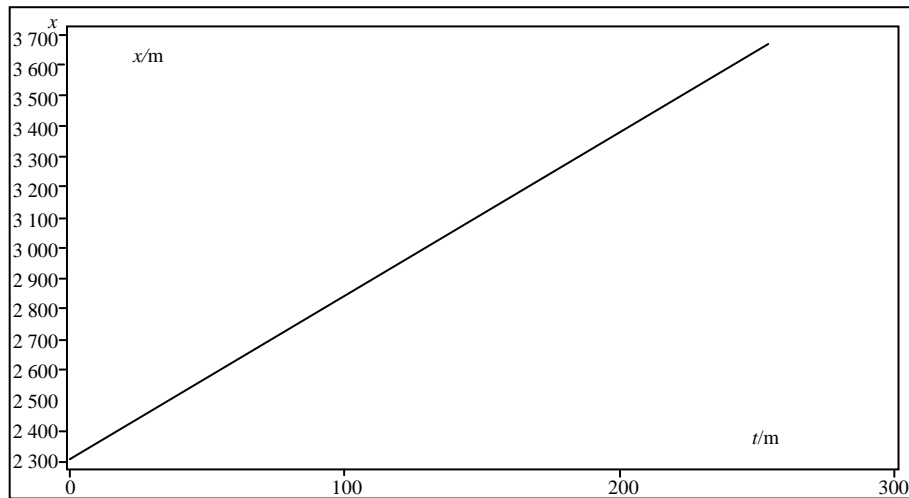


图 2 剔除异常值后的时序图

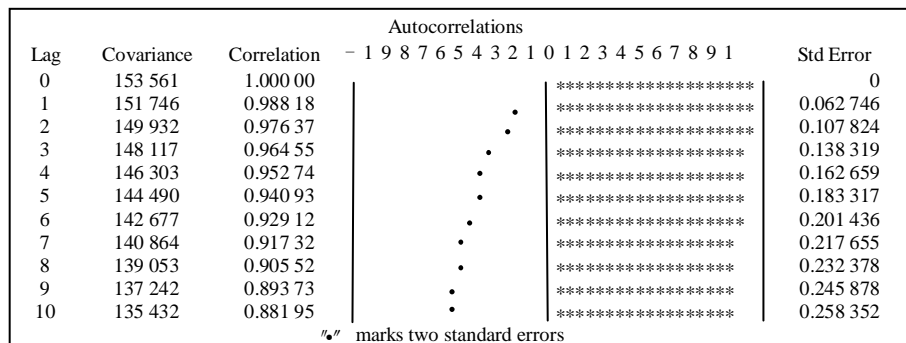


图 3 火控精度的自相关图

1.1.3 提取趋势项

在许多直接测量的时间序列模型中，大多数并不平稳，而是呈现出一定的趋势性和季节性，确定性的因素导致的非平稳，通常显示出非常明显的规律性，这种规律性比较容易提取，根据这种性质可以将非平稳时间序列分解为两部分，一部分是由多

项式决定的确定性的趋势成分，另一部分是平稳的零均值误差成分。简单记为：

$$X_t = U_t + \varepsilon_t \tag{3}$$

其中， ε_t 为零均值白噪声序列，这种方法就称为确定性分析方法。

通常，利用差分和平滑的方法减肥平稳时间序

列化成平稳时间序列, 再利用该序列的自相观和偏相关函数的性质建立 $ARMA(p, q)$ 模型。通过对原始

序列进行 10 步和 1 次差分, 得到自相关函数和偏相关函数, 如图 4。

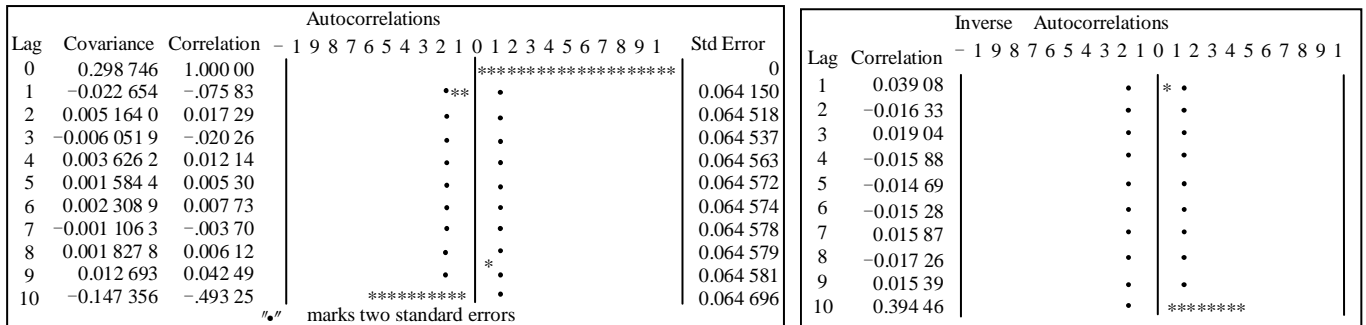


图 4 差分后序列的自相关函数和偏相关函数

1.2 时间序列建模

在对数据进行预处理后, 对时间序列进行建模, 其中包括模型的定阶, 参数的估计, 参数显著性的检验和模型的适应性的检验。

1.2.1 模型的定阶

在对原始时间序列进行模型拟合时, 需要确定一个准则函数来判定拟合对原始数据的接近程度, 建模时按准则函数的取值来确定模型的优劣, 决定模型的取舍, 使准则函数达到极小的就是最佳模型, 即模型的定阶。采用 BIC 准则对模型进行定阶^[5]。

BIC 准则定义如下:

$$BIC(n) = \ln \hat{\sigma}_\varepsilon^2(n) + \frac{n}{N} \ln N \quad (4)$$

其中, n 为参数个数。若某一阶数 n'_0 满足:

$$BIC(n'_0) = \min_{1 \leq n \leq M(N)} BIC(n) \quad (5)$$

其中, $M(N)$ 等于 $\lceil \sqrt{N} \rceil$ 或 $\lceil \frac{N}{10} \rceil$, 则 n'_0 为最佳阶数。编写 SAS 程序得到结果, 如图 5。

Minimum Information Criterion						
Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	11.851 02	11.189 08	11.182 17	11.162 83	11.124 02	11.061 93
AR 1	3.353 461	-1.091 35	-1.420 74	-1.537 92	-1.588 53	-1.612 39
AR 2	-1.116 92	-1.683 04	-1.660 97	-1.638 9	-1.616 83	-1.612 26
AR 3	-1.435 75	-1.660 96	-1.669 66	-1.647 58	-1.625 51	-1.625 36
AR 4	-1.547 67	-1.638 94	-1.647 58	-1.645 47	-1.623 39	-1.623 38
AR 5	-1.595 19	-1.616 94	-1.625 49	-1.623 39	-1.618	-1.618

Error series model: AR(5)
Minimum Table Value: BIC(2,1) = -1.683 04

图 5 火控精度的 BIC 值

1.2.2 估计参数

利用最小二乘估计^[6], 其基本原理是使残差平方和达到最小的那组参数值。对于 $ARMA(p, q)$ 模型, 记 $\beta = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)^T$, 则使目标函数:

$$J = \min_{t=\max\{p,q\}}^N \sum \varepsilon_t^2 = \sum_{t=\max\{p,q\}}^N (X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q})^2 \quad (6)$$

达到最小的 $\hat{\beta} = (\hat{\varphi}_1, \dots, \hat{\varphi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)^T$ 为 $\beta = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)^T$ 的最小二乘估计。这种方法充分应用了每个观测值所提供的信息, 因而其估计精度高, 通过 SAS 程序得到参数估计结果如图 6。

Conditional Least Squares Estimation						
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag	
MU	2 557.3	16.322 55	156.67	<.000 1	0	
MA1, 1	0.014 17	3.261 41	0.00	0.996 5	1	
AR1, 1	1.035 45	3.259 58	0.32	0.751 0	1	
AR1, 2	-0.035 45	3.260 10	-0.01	0.991 3	2	

图 6 模型的参数估计及检验

1.2.3 检验参数显著性

所谓参数显著性的检验是要检验模型中每个未知参数是否显著为非零, 检验的目的就是为了使模型最精简。经检验后, 将参数不显著的变量从模型中删除, 再重新建模, 得到最简约而又精确的模型, 对于上述最小二乘估计后得到的参数, 令 $S(\hat{\beta})$ 为 $\hat{\beta}$ 的标准差, 当样本容量充分大时, 可使用统计量:

$$t = \frac{\hat{\beta}}{S(\hat{\beta})} \sim t(N - k) \quad (7)$$

其中, k 为模型的参数个数, 对假设

$$H_0: \beta = 0 \longleftrightarrow H_1: \beta \neq 0, \quad (8)$$

进行检验, 当统计量 t 的 p 值大于 0.05, 接受 H_0 , 否则就拒绝 H_0 。

1.2.4 模型的适应性检验

模型的适应性是指一个完全能反映系统动态特性的 $ARMA(p, q)$, 其 ε_t 是独立的。适应性的检验实

质上就是 ϵ_t 的独立性检验, 检验方法如下:

原假设: 残差序列为白噪声序列, 即:

$$H_0: \epsilon_1 = \epsilon_2 = \dots = \epsilon_m = 0, \forall m \geq 1 \quad (9)$$

备择假设: 残差序列为非白噪声序列

$$H_1: \text{至少存在某个 } \epsilon_k \neq 0, \forall m \geq 1, k \leq m \quad (10)$$

检验统计量为:

$$N(N+2) \sum_{k=1}^m \frac{\hat{\epsilon}_k^2}{N-k} \sim \chi^2(m), \quad \forall m > 0 \quad (11)$$

如果拒绝原假设, 说明残差序列中还残留着相关信息, 拟合模型不显著, 如果没有拒绝原假设, 就说明拟合模型显著有效。通过编写 SAS 程序得到模型的残差检验, 从图 7 中可以看出, 残差模型是白噪声序列, 所以, 上述模型是适应的。

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	----- Autocorrelations -----					
6	10.11	3	0.017 6	0.066	0.084	0.084	0.084	0.084	0.083
12	20.58	9	0.014 7	0.083	0.082	0.082	0.081	0.081	0.081
18	30.67	15	0.009 7	0.080	0.080	0.079	0.079	0.079	0.078
24	40.38	21	0.006 7	0.078	0.077	0.077	0.076	0.076	0.076
30	49.73	27	0.004 9	0.075	0.075	0.074	0.074	0.074	0.073
36	58.66	33	0.003 9	0.073	0.072	0.072	0.071	0.071	0.070
42	67.20	39	0.003 3	0.070	0.070	0.069	0.069	0.068	0.068
48	75.40	45	0.003 0	0.068	0.067	0.067	0.066	0.066	0.066

图 7 模型的残差检验

1.3 时间序列的预报

对于 $ARMA(p,q)$ 时间序列的预报问题, 采用条件期望预测, $ARMA(p,q)$ 可用逆转形式表示为:

$$X_t = \sum_{j=1}^{\infty} I_j X_{t-j} + \epsilon_t \quad (12)$$

$$\hat{X}_t(l) = E(X_{t+l} | X_t, X_{t+1}, \dots) = \sum_{j=1}^{\infty} I_j E(X_{t+l-j} | X_t, X_{t-1}, \dots)$$

$$= \sum_{j=1}^{l-1} I_j \hat{X}_t(l-j) + \sum_{j=l}^{\infty} I_j X_{t+l-j} \quad (13)$$

从式 (13) 可见, 预测要用到过去 X_t 的信息。实际上, $ARMA(p,q)$ 的可逆性保证了 I_j 构成了收敛级数, 按预定的精度要求, 可取某 k 值, 当 $j > k$ 时, 令 $I_j = 0$, 即忽略了 X_{t+l} 对 X_{t+l-j} 的依赖性, 得出预测值。编写 SAS 程序得预报区域, 如图 8。实线为预报值, 虚线为置信系数为 95% 的置信上限和下限。

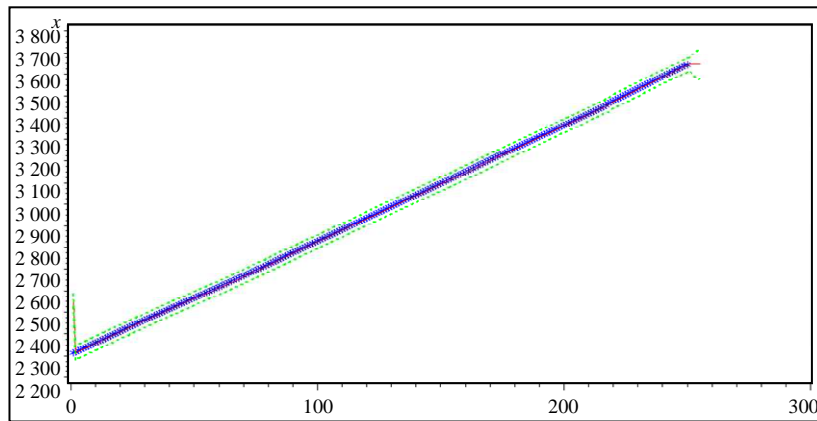


图 8 火控系统精度预报区域

2 结论

试验证明, 该方法可以对火控精度的时间序列进行快速准确的分析, 能实时解决工程中的一些问题。通过 SAS 分析火控系统模型, 大大提高了效率, 为火控精度的定量分析提供了一种有效的方法。

参考文献:

[1] 袁景山, 杨建华. 变形观测数据时间序列建模中几个重要问题的研究[J]. 地矿测绘, 2005, 21(3): 30-31.

[2] 董言治, 等. 基于 Matlab 的时间序列分析和动态数据建模[J]. 计算机工程, 2003, 29(12): 170-173.
 [3] 阮桂海, 等. SAS 统计分析大全[M]. 北京: 清华大学, 2003.
 [4] 胡绍林, 孙国基. 靶场外测数据野值点的统计诊断技术[J]. 宇航学报, 1999, 20(2): 68-69.
 [5] Bruce L. Bowerman, Richard T. O'Connell. Forecasting and time series[M]. United States American, 1993.
 [6] 王振龙. 时间序列分析[M]. 北京: 中国统计出版社, 1999.