

基于Matlab的时间序列分析和动态数据建模

董言治¹, 刘松涛¹, 尉志华², 沈同圣¹, 周晓东¹

(1. 烟台海军航空工程学院自动控制系, 烟台 264001; 2. 大连深蓝泵业有限公司, 大连 116031)

摘 要: 介绍了Matlab语言的特点, 结合火控精度测定数据实例着重讨论了其在时间分析和动态数据建模方面的优越性以及应用。在实际工作中使用Matlab可以大大提高试验的效率。

关键词: Matlab; 时间序列; 动态数据建模

Analyses and Modeling of Time Series Based on Matlab

DONG Yanzhi¹, LIU Songtao¹, WEI Zhiping², SHEN Tongsheng¹, ZHOU Xiaodong¹

(1. Autocontrol Department, Yantai Naval Aeronautical Engineering Institute, Yantai 264001;

2. Dalian Deepblue Pump CO. Ltd., Dalian 116031)

【Abstract】 This paper introduces characteristic of Matlab language, discusses its superiority when it is applied to analyses and model of time series by example of fire control system accuracy measure data emphatically. The use of Matlab in real work can greatly raise test's efficiency.

【Key words】 Matlab; Time series; Dynamic data modeling

时间序列是指有序的随机数据, 实际上是离散的随机过程, 有时又称为动态数据^[1]。时间序列的滤波、平滑、去卷、预报和控制的基础和前提是建模。建模是时间序列分析中的重要分支。在建模中, 往往要进行大量的数学计算。目前流行用Fortran、C语言等编制计算程序, 既需要对有关算法有深刻的了解, 还需要熟练地掌握所用语言的语法及编程技巧。对多数科学工作者而言, 同时具备这两方面才能有一定困难。Matlab语言被称为是一种“演草纸式的科学计算语言”^[2], 它的强大计算和模拟功能使得许多应用领域的各种计算、演算、模拟等工作变得相当简单, 是一个实时进行建模和仿真的有力工具。本文着重讨论了基于Matlab的时间序列分析和动态数据建模问题, 并结合火控精度测定数据实例讨论了Matlab在这方面的优越性和实际应用。

1 Matlab语言的特点

被称为第四代计算机语言的Matlab, 利用其丰富的函数资源, 使编程人员从烦琐的程序代码中解放出来。Matlab用更直观的、符合人们思维习惯的代码, 代替了Fortran和C语言的冗长代码, 给用户带来最简洁的程序开发环境。Matlab语言简洁紧凑, 使用方便灵活, 库函数极其丰富, 程序书写形式自由, 利用其丰富的库函数避开繁杂的子程序编程任务, 压缩了一切不必要的编程工作。由于库函数是由本领域的专家编写的, 用户不必担心函数的可靠性^[3,4]。可以说, 用Matlab进行科学开发是站在专家的肩膀上。

2 时间序列分析和动态数据建模

2.1 采样数据的检验和预处理^[5]

在实际问题中, 合理地选择采样间隔得到大量的量测数据之后, 还要进行初步的整理和必要的检验, 以期去粗取精, 为进一步深入分析提供较好的依据和参考。这些工作都属于预处理。包括对量测数据进行均值、方差和概率直方图分析、数据的正态性检验、数据的独立性检验、平稳趋势的检验、剔除野点和提取趋势项。

2.1.1 野点剔除

文中采用的野点剔除方法是Tukey提出的, 其基本思想

是产生一个曲线的平滑估计, 然后把它从数据中减掉, 这样识别野点就容易得多了。该方案利用“中位数”是均值的鲁棒(robust)估计这个事实。其步骤为:

(1)从 $x(i)$ 构造一个新序列 $x_1(i)$, 方法是取 $x(1), \dots, x(5)$ 的中位数作为 $x_1(3)$, 然后舍去 $x(1)$ 加入 $x(6)$ 取中位数得 $x_1(4)$, 依此类推, 直到加入最后一个数据。换言之, 总是在相邻5个数据中择取中位数。显然, $x_1(i)$ 的项数比真 $x(i)$ 项数少4项。

(2)用类似的方法在 $x_1(i)$ 的相邻的3个数据中择取中位数构成序列 $x_2(i)$ 。

(3)最后是由序列 $x_2(i)$ 按如下方式构成序列 $x_3(i)$:

$x_3(i) = x_2(i-1)/4 + x_2(i)/2 + x_2(i+1)/4$ (这是一海宁平滑滤波器)。

分析序列 $x(i)-x_3(i)$, 看是否有 $|x(i)-x_3(i)| > k(\text{预定值})$, 如果有, 则用一内插值代替 $x(i)$ 。剔除野值后的仿真图像见图1。

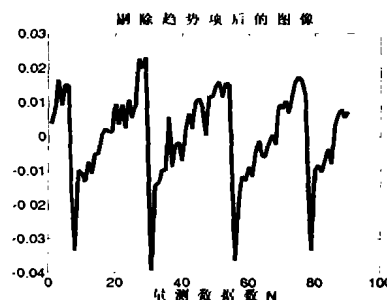


图1 趋势项剔除后图

2.1.2 量测数据的均值、方差和概率直方图分析

量测数据的均值、方差和概率直方图分析可以直接调用Matlab函数: mean.m、var.m 和histfit.m, 过程非常简洁有效, 仿真结果见图2。

2.1.3 提取趋势项

在许多实际问题中, 由量测直接得到的随机序列大多数

作者简介: 董言治 (1972-), 男, 博士生, 研究方向为红外成像、智能制导、图像处理等; 刘松涛, 博士; 尉志华, 本科生; 沈同圣, 博士、副教授; 周晓东, 教授

收稿日期: 2002-06-20

并不平稳,而是呈现出明显的趋势性或季节性。这时需要用更一般的模型来描述,即取 $X_t=U_t+Y_t$,其中 U_t 表示 X_t 中随时间变化的均值,它往往可以用多项式、指数函数、正弦函数等描述,而 Y_t 表示零均值平稳过程,可以用ARMA模型拟合。做了这样的分解之后,可以用处理平稳过程的方法来分析 Y_t 了。Matlab提供了提取趋势项函数detrend.m和trend.m。使用它们可以方便地提取趋势项。如果不仅要分析 Y_t ,而且也需要知道 U_t 的具体形式,那么就可以采用Matlab提供的多项式函数polyfit.m去拟合量测数据 U_t 。然后对残差进行分析,该残差序列可认为是平稳过程,通过对平稳序列的分析处理,可以得出 Y_t 的估计值,于是就可以对 X_t 进行估计了。仿真的残差图像见图3。

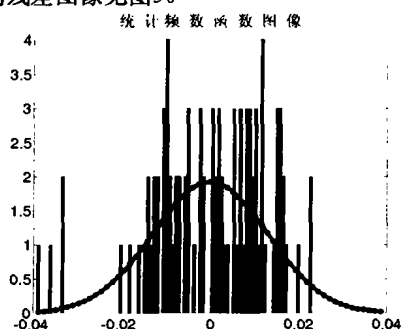


图2 统计频谱函数图

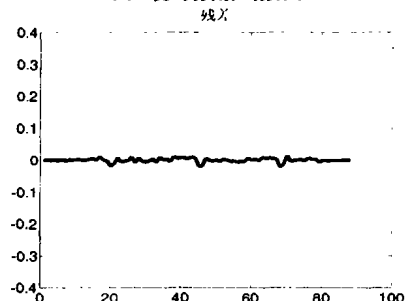


图3 残差图像

2.2 时间序列时间域模型的参数估计和模型确定以及实用性的检验

文中采用AR模型辨识^[5],其原理是由低阶开始对系统建模然后逐渐增加模型的阶数,并用F检验进行自动筛选,同时确定系统的模型参数。辨识步骤如下:

(1)利用递推最小二乘法估计参数和模型阶数

其思想是:

对于AR(P)模型 $x(t)=a_1*x(t-1)+a_p*x(t-p)+e(t)$

可得递推最小二乘法公式:

$$\hat{\theta}(t+1) = \hat{\theta}(t) + K(t+1)[x(t+1) - \varphi'(t+1)\hat{\theta}(t)]$$

(2)利用F检验进行筛选,确定模型的阶数

在F检验中, H_0 : AR(p-1)模型是合适的(即 $a_p=0$);

H_1 : AR(p-1)模型是不合适的(即 $a_p \neq 0$);

统计量 $F=(A_1-A_0)/(N-p)/A_0$,服从分布 $F(1, N-p)$,其中 A_0 是AR(p)的残差平方和, A_1 是AR(p-1)的残差平方和, N 是量测数据数目。给定显著性水平 α ,查F分布表可得临界值 F_α 。若 $F > F_\alpha$,则拒绝 H_0 ,即AR(p-1)模型是不合适的,否则接受 H_0 。前者称为F检验显著,后者为不显著。文章中给出了Matlab的实现程序及仿真图像(图4)。

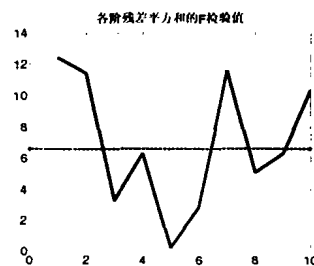


图4 F检验图

2.3 ARMA模型的特性和时间序列的预报

从系统分析与数据处理来讲,主要有4个特性:(1)Green函数(单位脉冲响应函数);(2)频率特性;(3)自协方差函数;(4)自谱函数。在Matlab中没有相应的Green函数,需要自行构建。然后再用构建的函数结合语言自身的函数进行其余3个特性的分析和时间序列的预报。运用Matlab程序仿真的特性图像见图5~图8。

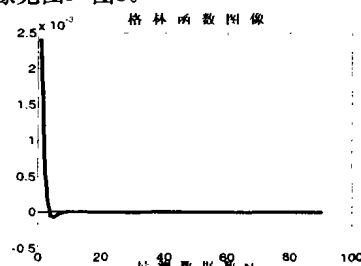


图5 格林函数图像

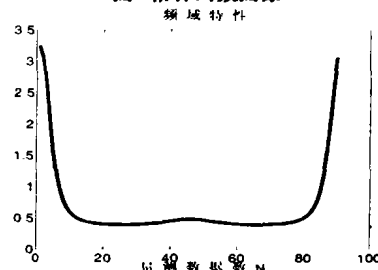


图6 频域特性

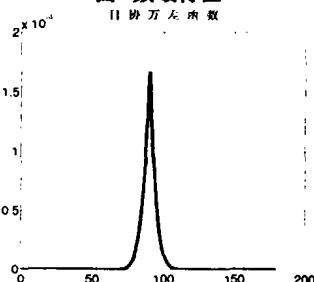


图7 自协方差函数

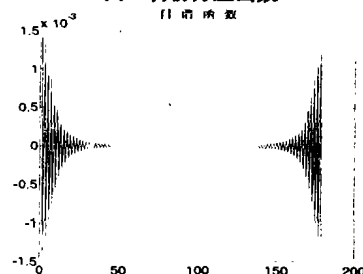


图8 自谱函数

