摘 要

J. H. Wilkinson[?]建立了非奇异矩阵的逆是矩阵元素的连续函数的理论。G. W. Stewart[?]推出了矩阵的广义逆的连续性。为了得到Drazin逆的连续性，本文先给出了$M-$矩阵、$H-$矩阵类的逆的连续性。Campbell和Meyer[?] 也给出了Drazin 逆的连续性性质，但没有给出明显的边界。

Drazin逆对扰动是很不稳定的。然而，在某种特定的扰动条件下，矩阵$(A+E)^D$与A^D的接近程度能够得到量化且也能得到明显的相对误差边界。基于Drazin逆的不同形式，很多科学家和学者从事这一方面的研究。U. G. Rothblum 给出的Drazin逆的以下的表达式：

$$A^D=(A-H)^{-1}(I-H)=(I-H)(A-H)^{-1}$$

其中$H=I-AA^D=I-A^DA$.基于这个表达式,我们在本文中也给出了$\|(A+E)^D-A^D\|_2/\|A^D\|$和$\|(A+E)^\sharp-A^D\|_2/\|A^D\|_2$的范数估计，并与前人的成果进行了比较。

关键词：$M-$矩阵，$H-$矩阵，Drazin逆，Pseudo-Drazin逆，条件数

ABSTRACT

The theory that the inverse of a nonsingular matrix is continuous function of the elements of the matrix was established by J. H. Wilkinson[?]. The continuity of the generalized inverse A^{+} of a matrix A was introduced by G. W. Stewart[?]. In this paper, at first, the continuity of the special matrices inverse, such that $M-$matrices and $H-$matrices, respectively, are provided. Campbell and Meyer[?] also established the continuity properties of Drazin inverse, but the explicit bound was not given.

The Drazin inverse is unstable with respect to perturbation. However, under some specific perturbation , the closeness of the matrices $(A+E)^{D}$ and A^{D} can be proved and the explicit bound the relation error can also be obtained. Based on the different representations of Drazin inverse, many scientists and scholars have worked it research. U. G. Rothblum gave the following representation of Drazin inverse:

$$A^{D} = (A-H)^{-1}(I-H) = (I-H)(A-H)^{-1}$$

where $H = I - AA^{D} = I - A^{D}A$. Based on the representation, we also obtain the norm estimate of $\|(A+E)^{D} - A^{D}\|_2/\|A^{D}\|$ and $\|(A+E)^{\sharp} - A^{D}\|_2/\|A^{D}\|_2$ and compare with the precedent results.

Keywords: $M-$matrices, $H-$matrices, Drazin inverse, Pseudo-Drazin inverse, Condition number

目 录

第一章 绪论

1.1 研究工作的背景与意义

本论文依托于国家“973”项目“卫星故障演变规律与故障反演技术”。本项目基于两颗国防卫星，从2011年至2013年近两年半的数十种遥测信号，在时域、频域、时频联合域及其他变换域，对卫星在轨异变状态进行研究。项目主要研究内容包括，对卫星遥测信号异变状态的特征提取，分析不同特征参数的演变趋势，建立特征参数模型，最后对卫星状态异变进行短期预测。本项目旨在，通过对卫星遥测信号的分析，能够掌握卫星的健康状况，即卫星的异变状态，从而为卫星管理人员提供正确的决策参考。

1.1.1 研究工作的背景

本论文的研究工作，是在国内卫星系统不断扩大，对卫星运行的稳定性和卫星寿命水平的要求不断提高的同时提出的。研究工作不仅仅要求对卫星故障进行诊断和预测，同时对卫星的异变状态的检测和预测也提出了要求。

卫星“状态异变”的含义较“故障信号”更为广泛。在卫星系统中，故障信号指的是卫星的机械装置或者电路系统已经不能完成预设的功能时产生的信号，即已经出现了功能性缺失[?]；而状态异变包括故障信号，它还包括，在发生功能性缺失的前后，设备产生的不同于正常状态下的信号。在此要说明的是，异变信号的出现可能伴随着故障的生成，也有可能不会引起故障，即设备出现扰动，但是并未造成功能性缺失。所以，相较于故障信号，状态异变更具有隐蔽性，随机性。这样对信号特征的提取，建模，乃至预测的难度都大大提高。在本论文及项目中，研究对象状态异变信号使用的都是实际的卫星遥测信号，由于遥测信号的接收、干扰等因素，信号的野值较多，质量较差。并且卫星遥测信号，大多具有非平稳的特性。这些因素对状态异变信号的分析和预测工作提出了更大的挑战。

1.1.2 研究工作的意义

首先，卫星造价昂贵，工艺复杂，制造周期长，且具有唯一性，一旦卫星因异变而损坏废弃，造成的损失将非常严重。所以研究卫星状态的异变并预测，对卫星行业具有极其重要的意义。

对于投资巨大和庞大复杂的卫星系统，可靠性和安全性是迫切需要的。否则，一次故障可能导致灾难性的的后果和巨大的经济损失。如1986年1月美国“挑战

者”号航天飞机失事导致7名宇航员全部遇难；1999年4月9日到5月4日的不到一个月的时间里，美国大力神4B、雅典娜2、德尔它2运载火箭等发射连续失败，损失达十亿美元；1999年11月日本H2运载火箭发射失败，损失2.99亿美元；2000年 11月，印度INSA-2B卫星丧失对地指向功能，造成重大损失；2003年2月1日美国哥伦比亚号航天飞机在返回途中解体，机组人员全部遇难。可见，对卫星系统的故障乃至状态异变进行诊断和预测，对航天事业的健康发展有着重要意义。

其次，目前我国的卫星在数量、型号、功能等各方面不断增加，加上国外势力对我国卫星的干扰等其他因素，传统的异变分析及预测方法将不再完全适用于目前的卫星状况。所以研究更可靠的预测算法必然是目前的趋势。

随着在轨卫星数量、设计寿命和型号种类的不断增加，维持其安全稳定运行变得越来越重要，在轨管理的难度也逐年加大。在轨卫星长期运行在空间环境中，受到多种不确定性因素的作用，其性能与功能可能会出现变化，反映在遥测参数上也会有些变化，如果在轨卫星发生异常，相应的遥测参数的变化趋势也会发生改变。因此，分析在轨卫星的遥测数据变化规律，选择相适应的数据预测方法，对遥测数据进行预测，并在此基础上实现预警，可以在早期及时发现遥测数据的异常变化，有效避免可能发生的重大故障，降低卫星在轨运行的风险，同时为异常的处理赢得宝贵时间，这对于提高卫星在轨运行的安全性和可靠性具有重要的意义[?]。

最后，当前卫星异变状态分析和预测工作通常是一项耗时、重复和劳动强度很大的工作[?]。所以研究自动化的卫星状态异变的预测算法将为卫星管理工作带来极大的便利。

在大部分情况下，卫星管理人员采用手工检查卫星遥测数据的方式来确定卫星是处于健康状态还是异变状态，以及卫星是否存在危险的趋势，以预测卫星是否在不久的将来处于故障状态。应用的具体方法或者是一些统计学估计方法，或者是通过比较卫星遥测参数的实测值与期望值之间的差异方法。这些方法存在严重的缺陷，主要是过分依赖人工经验以及难以实现故障诊断与预测的自动推理[?]。当卫星状态异变出现了新的模式，传统方法的预测结果就会出现较大偏差。研究自动化的卫星状态异变的预测方法，既为卫星管理人员提供了便利，更重要的是，在预测精度上有显著提升，减小预测的虚警和漏警。

1.2 国内外研究历史与现状

1.3 本文的主要贡献与创新

1.4 本文的结构安排

1.5 本章小结

第二章 基于趋势预测算法的理论基础

本论文对卫星在轨异变状态的预测算法基于两种思路进行研究，第一种是基于趋势的预测算法，第二种是基于相似度的预测算法。在本章中，将详细描述基于趋势的预测算法的理论研究，而这两种算法的理论研究和验证试验的余下内容，会在之后的章节分别中逐一介绍。

在此，首先对卫星在轨异变数据进行简要的分析介绍。

2.1 卫星在轨异变数据分析

卫星结构复杂，信号种类繁多，同时收到空间环境中多种因素的干扰，卫星在轨异变数据是一个复杂的动态信号系统。因此先要对卫星信号的特征了解把握。下面就从真实的卫星信号入手，对在轨异变数据的特征进行分析。

2.1.1 低信噪比

卫星在轨异变数据的波动频率和幅度都在不断变化中，即使不存在国外势力对我国卫星恶意干扰的情况，也会因为各种因素的影响呈现小幅的高频的波动，一般将这种高频小幅波动视为噪声。

由于这些干扰因素众多，其中包括空间环境中的各种粒子对卫星工作器件的影响，还有遥测数据在发送、传播、接收过程中的干扰。这些因素对最后遥测信号的噪声影响较大，造成信噪比较低。

但是在对时间序列进行预测时，要用数学方法将噪声尽可能去掉。而由于异变信号时间序列的频率和幅度在不断变化中，只有选用自适应特点的方法对噪声进行处理才能有效。

2.1.2 非平稳性

一个平稳序列的数字特征一般表现为：均值、方差等不随时间的变化而变化，时间序列在各个时间点上的随机性服从一定的概率分布，在图形上往往表现为围绕其均值不断波动。而非平稳时间序列则表现出在不同的时间段具有不同的均值。检验序列平稳性的标准方法是单位根检验，主要有Augmented Dickey-Fuller Test（ADF）， Dickey-Fuller Test with GLS（DFGLS），Philips-Perron，KPSS，ERS及NP 等六种检验方法。本论文采用ADF单位根检验方法进行检验。

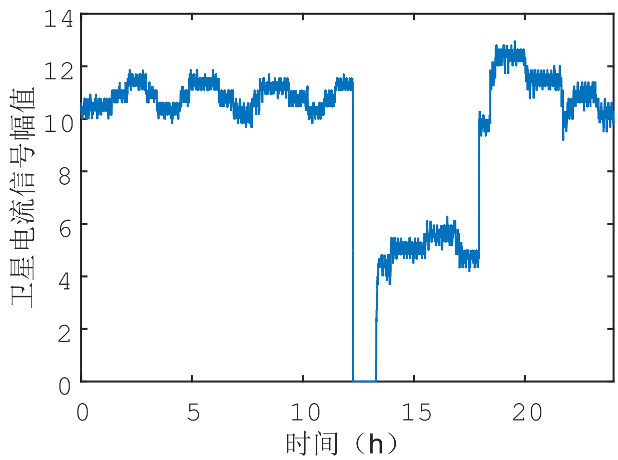


图 2-1 卫星在2011年3月份某一天内的IN4信号时域图

下面，本论文以某卫星在2011年3月份某一天内的负分流调节器分流电流（IN4 ）遥测信号为样例数据，其时域图如图(2-1)所示。

首先，从图(2-1)的时域数据中看来，改时间序列具有非平稳性的特点。

接下来，对该时间序列进行ADF检验，出结果见表2-1。

表 2-1 样本序列的ADF检验值

	T统计量值	P值
ADF	-2.087269	0.7163
1%显著性水平检验临界值	-3.430904	
5%显著性水平检验临界值	-2.861669	
10%显著性水平检验临界值	-2.566880	

从表2-1中可见，不能拒绝存在一个单位根的原假设，ADF统计值大于1%level 、 5%level和10%level的临界值，序列是非平稳的。

从整体的数据看来，卫星在轨异变数据普片具有非平稳性。

2.2 小波分析理论

在1984年，Morlet和Grossman提出小波（wavelet）一词，他们用的是法语词ondelette ，意思就是“小波”。后来在英语里，“onde”被改为“wave”而成了wavelet [?]。小波分析算法作为一种数学理论和方法在科学界引起越来越多的重视，并被运用到各个技术领域中。

由于小波具有可以自由“变焦”的自适应窗口和紧支撑性，特别适合处理卫星信号这种细节分量多而杂的信号。因此，本论文中采用小波算法对卫星在轨异变信号进行处理分析。

2.2.1 小波的定义

小波分析，即小波变换，是从傅里叶变换发展而来，其中还经历了短时傅里叶变换，最后建立小波理论。

Fourier在1807年提出，任何函数都可以用一组正余弦函数的和来表示，这就是傅里叶分析。傅里叶变换是傅里叶分析的抽象，它是将信号分解成不同频率的数学方法。它可以将时域中某一信号变换至频域中，予以定量认识和分析，他能描述信号的整体频谱特征。从实用的角度出发，当人们考虑傅里叶分析的时候，通常指的是（积分）傅里叶变换和傅里叶级数。

时间信号函数$f(t) \in L^1(R)$（其中t表示时间域自变量），满足$\int_{-\infty}^{+\infty} |f(t)|\,\mathrm{d}t < \infty$时，其连续傅里叶变换$F(\omega)$（其中$\Omega$表示时间域自变量），定义为：

$$F(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega t} f(t)\mathrm{d}t \tag{2-1}$$

$F(\omega)$的傅里叶逆变换定义为：

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-it\omega} F(\omega)\mathrm{d}\omega \tag{2-2}$$

为了计算傅里叶变换，需要数值积分，即取$f(t)$在R上的离散点上的值来计算这个积分。下面给出离散傅里叶变换（DFT）的定义。

给定的离散时间序列f_0,f_1,…,f_{N-1}，设该序列绝对可和，即满足$\sum_{n=0}^{N-1} |f_n| < \infty$时，序列$\{f_n\}$的离散傅里叶变换定义为：

$$X(k) = F(f_n) = \sum_{n=0}^{N-1} f_n e^{-i\frac{2\pi k}{N} n} \qquad (k = 0,1,\ldots,N-1) \tag{2-3}$$

序列$\{X(k)\}$的离散傅里叶逆变换定义为：

$$f_n = \frac{1}{N}\sum_{k=0}^{N-1} X(k) e^{i\frac{2\pi n}{N}} \qquad (n = 0,1,\ldots,N-1) \tag{2-4}$$

虽然傅里叶变换能够将信号的时域特征和频域特征联系起来，并分别观察，但却不能把两者有机的结合起来。这是因为傅里叶变换是一种全局的变换，从式(2-1)和式(2-3)可以看出，它是整个时间域内的积分，没有局部分析信号的功能。也就是说，对于傅里叶谱中的某一频率，不知道这个频率是在什么时候产生的。这样在信号分析中面临一对矛盾：时域和频域的局部化矛盾。

在实际的信号处理过程中，尤其是对非平稳信号的处理中，信号在任一时刻附近的频域特征都很重要。为了克服傅里叶变换在时域上无分辨率能力的缺点，Dennis Gabor引入了短时傅里叶变换（STFT），也称窗口傅里叶变换，其实质是在傅里叶变换中加一个时间窗，以给出信号的谱的时变信息。

短时傅里叶变换的基本思想是把信号划分成许多小的时间间隔，用傅里叶变换分析每一个时间间隔，以便确定该时间间隔存在的频率。其表达式为：

$$S(\omega,\tau)=\int_R f(t)g^*(t-\tau)e^{-i\omega t}\mathrm{d}t \tag{2-5}$$

其中“$*$”表示复共扼，$g(t)$是有紧支集的函数，$f(t)$是进入分析的信号。$e^{-i\omega t}$起频限作用，$g(t)$起时限作用，合在一起起时频分析作用。随着时间τ的变化，所确定的“时间窗”在t轴上移动，使$f(t)$“逐渐”进行分析。因此，$g(t)$被称为窗口函数，$S(\omega,\tau)$大致反映了$f(t)$在时刻τ、频率ω的“信号成分”的相对含量。

由此可见，短时傅里叶变换在一定程度上克服了标准傅里叶变换不具有时域局部分析能力的缺陷，但它也存在着自身不可克服的缺陷，即当窗函数$g(t)$确定后，时间宽度是一个定值，矩形窗口的形状就确定了，τ，ω只能改变窗口在相平面上的位置，而不能改变窗口的形状。可以说短时傅里叶变换实质上是具有单一分辨率的信号分析方法，若要改变分辨率，则必须重新选择窗函数$g(t)$。因此，对于非平稳信号，短时傅里叶变换同样不适合。

所以对于在现实情况中普片存在的非平稳信号，信号分析方法希望拥有一个灵活可变的时间窗。即对于由多种频率分量组成的非平稳信号，当信号尖锐变化时，需要有一个短的时间窗为其提供更多的频率信息当信号变化平缓时，需要一个长的时间窗用于描述信号的整体行为。这就导致了小波分析的出现。

现给出小波变换的定义，对于函数$f\in l^2(R)$的小波变换为：

$$w_f(a,b)=\langle f,\Psi_{a,b}\rangle=|a|^{-\frac{1}{2}}\int_R \bar{\Psi}(\frac{t-b}{a})f(t)\mathrm{d}t \tag{2-6}$$

其中，$a,b\subset R,a>0$，式中a为伸缩（尺度）因子，b为平移（位移）因子，$\langle\ldots\rangle$为内积，$\bar{\Psi}$是Ψ的共轭。

$$\Psi_{a,b}(t)=|a|^{-\frac{1}{2}}\Psi(\frac{t-b}{a}) \tag{2-7}$$

此函数称为小波函数族，式中a为伸缩因子，b为平移因子。

2.2.2 连续小波变换

先给出平方可积的函数空间$L^2(R)$的定义，它是要求在整个实数轴R上的满足$\int_{-\infty}^{+\infty}|f(t)|^2\,\mathrm{d}t<\infty$的可测函数$f(t)$的全体集合，并带有相应的函数运算和内积。工程上常常说成是能量有限的全体信号的集合。

小波（基本小波或母小波）就是函数空间$L^2(R)$中满足下述条件的一个函数或者信号$\Psi(t)$：

$$C_\Psi=\int_{R^*}\frac{\left|\hat{\Psi}(\omega)\right|^2}{|\omega|}\mathrm{d}\omega<\infty \tag{2-8}$$

其中，$R^*=R-0$表示非零实数全体。$\Psi(t)$也称为小波函数，或称母小波。式(2-8)称为容许性条件。

对于任意的实数对(a,b)，其中，参数a必须为非零参数，连续小波的函数定义式如下：

$$\Psi_{a,b}(t)=|a|^{-\frac{1}{2}}\Psi(\frac{t-b}{a})\qquad(a,b\in R;a\neq 0) \tag{2-9}$$

它是由小波母函数$\Psi(t)$经伸缩和平移后生成的依赖于参数(a,b) 的连续小波函数序列，简称为小波。其中：

（1）式(2-9)中a为伸缩因子，b为平移因子。如果小波母函数$\Psi(t)$的傅里叶变换$\hat{\Psi}(\omega)$在原点$\omega=0$是连续的，那么，式(2-8)说明$\hat{\Psi}(0)=0$。这说明，函数$\Psi(t)$有“波动”的特点，另外，小波函数$\Psi(t)$只在原点的附近它的波动才会明显偏离水平轴，在远离原点的地方函数值将迅速“衰减”为零，整个波动趋于平静。这是称函数$\Psi(t)$为“小波”函数的基本原因。

（2）对于任意的参数对(a,b)，显然$\int_R\Psi_{a,b}(t)\mathrm{d}t=0$。但是，这里$\Psi_{a,b}(t)$却是在的附近存在$t=b$明显的波动，而且，有明显波动的范围的大小完全依赖于参a数的变化。当$a=1$时，这个范围和原来的小波函数$\Psi(t)$的范围是一致；的当$a>1$时，这个范围比原来的小波函数班的范围要大一些，小波的波形变矮变胖，而且，当a变得越来越大时，小波的波形变得越来越胖、越来越矮，整个函数的形状表现出来的变化越来越缓慢；当$0<a<1$时，$\Psi_{a,b}(t)$在$t=b$的附近存在明显波动的范围比原来的小波母函数的$\Psi(t)$要小，小波的波形变得尖锐而消瘦。

（3）设初始伸缩系数为a_0，当$a>a_0$时，式(2-9)中$\Psi_{a,b}(t)$定域增大，窗口变大；当$a<a_0$时，则窗口变小，窗的高度增加，窗口面积大小不变（b一定时）。以上体现了小波的“变焦”能力。

为使信号分解及重构的实现在数值上是稳定的，要求小波基函数$\Psi(t)$应该满足下面的约束条件：

（1）$\int_{-\infty}^{+\infty}|\Psi(t)|\mathrm{d}t<\infty$；

（2）$\hat{\Psi}(\omega)$在原点必须等于0，即$\hat{\Psi}(0)=\int_{-\infty}^{+\infty}\Psi(t)\mathrm{d}t=0$；

（3）$A\leqslant\sum_{-\infty}^{+\infty}\left|\hat{\Psi}(2^{-j}\omega)\right|^2\leqslant B$，$(0<A\leqslant B<\infty)$。

对于任意的函数或者信号$f(t) \in L^2(R)$，用小波函数集$\{\Psi_{a,b}(t)\}$进行分解运算，其连续小波变换为：

$$W_f(a,b) = \langle f, \Psi_{a,b} \rangle = |a|^2 \int_R f(t)\bar{\Psi}(\frac{t-b}{a})\mathrm{d}t \tag{2-10}$$

其中，$\langle \ldots \rangle$为内积，$\bar{\Psi}$是Ψ的共轭，W表示小波变换。因为小波母函数$\Psi(t)$只有在原点的附近才会有明显偏离水平轴的波动，在远离原点的地方函数值将迅速衰减为零，整个波动趋于平静，所以，对于任意的参数对(a,b)，小波函数$\Psi_{a,b}(t)$在$t=b$的附近存在明显的波动，远离$t=b$的地方将迅速地衰减到0，因而，从形式上可以看出，式(2-10)的数值$\Psi_f(a,b)$表明的本质上是原来的函数或者信号$f(t)$在$t=b$点附近按$\Psi_{a,b}(t)$进行加权的平均，体现的是以$\Psi_{a,b}(t)$为标准快慢的$f(t)$的变化情况，这样，参数b表示分析的时间中心或时间点，而参数a体现的是以$t=b$为中心的附近范围的大小，所以，一般称参数a为尺度参数，而参数b为时间中心参数。

小波变换的逆变换重构公式为：

$$f(t) = \frac{1}{C_\Psi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_f(a,b)\Psi_{a,b}(t)\frac{\mathrm{d}a\mathrm{d}b}{a^2} \tag{2-11}$$

2.2.3 离散小波变换

工程应用中所采集的信号（如位移量、压力、电压电荷量），都是有限长的离散信号。在实际信号处理中，出于数值计算的可行性和理论分析的简便性考虑，对连续小波$\Psi_{a,b}(t)$和连续小波变换$W_f(a,b)$，加以离散化处理。这一离散化都是针对连续的尺度参数a和连续的平移参数b的，而不是针对时间变量t的。

把连续小波变换式(2-9)中尺度参数a和平移参数b的离散化分别取作$a=a_0^j$，$b=ka_0^jb_0$，$j \in Z$，扩展步长$a_0 \neq 1$是固定值，为方便起见，取定$a_0>1$与$b_0>0$。可得到相应的离散小波集：

$$\Psi_{j,k}(t) = a_0^{-j/2}\Psi(a_0^{-j}t - kb_0) \qquad (j,k \in Z) \tag{2-12}$$

离散化小波变换系数则可以表示为：

$$c_{j,k} = \int_{-\infty}^{+\infty} f(t)\Psi_{j,k}^*(t)\mathrm{d}t \tag{2-13}$$

离散小波变换重构公式为：

$$f(t) = C\int_{j=-\infty}^{+\infty}\int_{k=-\infty}^{+\infty} c_{j,k}\Psi_{j,k}(t) + R \tag{2-14}$$

其中C为一个与信号无关的常数，R为误差项。从上式可以看出，要使$f(t)$能够达到不失真重构，这要求我们选取适当的a_0、b_0及$\Psi(t)$，使$\Psi_{j,k}(t)$构成表示信号空间上的完备集，从而实现小波集$\{\Psi_{j,k}(t)\}$的精确线性表示，即使$R=0$。

基于上述小波变换的变形分析应用特征如下：

（1）小波变换是一种多分辨率分析，在变换中取较小的a_0^j，则时间的分辨率较高，这时可用于分析频率较高的建筑物的动态变形（如高层建筑的在风力作用下的摆），可以揭示建筑物动态变形特征；

（2）当取较大的a_0^j时，则频率分辨率较高，这时宜于分析频率较低的常见的建筑物变形，对分析建筑物规律并进行预报具有较好的稳健性。

设小波为$\Psi(t)$，如果函数族

$$\{\Psi_{j,k}(t) = 2^{-j/2}\Psi(2^{-j}t - k); j,k \in Z\}$$

构成空间$L^2(R)$的标准正交基，即满足下述条件的基：

$$\langle \Psi_{j,k}, \Psi_{l,n} \rangle = \int_R \Psi_{j,k}(t)\bar{\Psi}_{l,n}(t)\mathrm{d}t = \delta(j-l)\delta(k-n) \tag{2-15}$$

则称$\Psi(t)$是正交小波，$\delta(m)$其中符号的定义是：

$$\delta(m) = \begin{cases} 1, & m = 0 \\ 0, & m \neq 0 \end{cases} \tag{2-16}$$

对任何函数或信号$f(t)$，有如下的小波级数展开：

$$f(t) = \sum_{j=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} A_{j,k}\Psi_{j,k}(t) \tag{2-17}$$

其中的系数A_j, k由公式$A_j, k = \langle f, \Psi_j, k \rangle = \int_R f(t)\bar{\Psi}_{j,k}(t)\mathrm{d}t$给出，称为小波系数。容易看出，小波系数$A_{j,k}$正好是信号$f(t)$的连续小波变换$W_f(a,b)$在尺度系数$a$的二进离散点$a_j = 2^j$和时间中心参数$b$的二进整倍数的离散点$b_k = 2^j k$二所构成的点$(2^j, 2^j k)$上的取值，因此，小波系数$A_{j,k}$实际上是信号$f(t)$的离散小波变换。也就是说，在对小波添加一定的限制之一下，连续小波变换和离散小波变换在形式上简单明了地统一起来了，而且，连续小波变换和离散小波变换都适合空间$L^2(R)$ 上的全体信号。

2.3 自回归滑动平均（ARMA）预测模型

ARMA模型是由美国统计学家 George E. P. Box 和英国统计学家 Gwilym Jenkins 在二十世纪七十年代提出的时序分析模型[?]，即自回归滑动平均模型（Auto Regerssive Moving Average Model）。

用ARMA模型模拟随机性时序在实际应用中有许多方便之处。例如它是线性的，只要给出少量参数后，模型就完全确定，便于处理；另外，ARMA模型具有

有理谱密度，它能逼近任何连续谱密度；最后，ARMA模型便于分析数据的结构和内在性质，也便于在线性最小方差意义下进行最佳预报和控制。

2.3.1 ARMA模型的定义

ARMA模型是AR模型和MA模型的组合。

AR模型的预测方式是通过历史的观测值和现在的干扰值的线性组合对数据的未来走势进行预测，自回归模型的数学公式为：

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t \tag{2-18}$$

其中，p为自回归模型的阶数，$\phi_i, (i = 1, 2, \ldots, p)$为模型的待定系数，$\varepsilon_t$为误差，$x_t$为平稳时间序列。

MA模型的预测方式是通过历史的干扰值和现在的干扰值的线性组合对数据的未来走势进行预测。滑动平均模型的数学公式为：

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{2-19}$$

其中，q为自回归模型的阶数，$\theta_i, (i = 1, 2, \ldots, p)$为模型的待定系数，$\varepsilon_t$为误差，$x_t$为一个平稳时间序列。

ARMA(p, q)的定义式为：

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{2-20}$$

在该模型中引入延迟算子，ARMA(p, q)模型简记为：

$$\Phi(B) x_t = \Theta(B) \varepsilon_t \tag{2-21}$$

其中，$\Phi(B) = 1 - \phi_1 B^1 - \cdots - \phi_p B^p$为$p$阶自回归系数多项式；$\Theta(B) = 1 - \theta_1 B^1 - \cdots - \theta_q B^q$为$q$阶移动平均系数多项式。

当p 和q分别等于零时，ARMA模型分别退化成AR模型和MA模型。

2.3.2 ARMA序列的相关函数

在实际中，我们经常研究一对变量之间的夫系。在收集的数据处理中，用相关概念来描述两个变量之间的相依关系。它有两种形式，一种是线性相关，指两个变量之间可用线性方程来描述；另一种是非线性相关，指两变量之间需用非线性方程来描述。不过，这些都是与时间无关的随机变量之间的关联性。在时间序列分析中，需要了解与时间有关的信息在不同时刻的取值有无内在的关联性，这就需要引入相关函数的概念。

2.3.2.1 自相关函数

在t，$t+a$时刻的样本记录$X(t)$，$X(t+a)$的自相关性可以通过在观察时间T上对这两个值的乘积作平均得到。T趋于无穷时，平均乘积的极限将接近一个正确的自相关函数。用公式表示，就是：$R_x(a)=\lim\limits_{T\to\infty}\int_0^T x(t)x(t+a)\mathrm{d}t$。零均值化后的样本自协方差函数和样本自相关函数可作为自协方差函数和自相关函数的估计。样本自协方差函数：

$$\hat{r_k}=\frac{1}{N}\sum_{t=1}^{n-k}x_t x_{t+k} \qquad k=0,1,2,\ldots,N-1 \tag{2-22}$$

样本自相关函数：

$$\hat{p_k}=\frac{\hat{r_k}}{\hat{r_0}} \qquad kk=0,1,2,\ldots,N-1 \tag{2-23}$$

（1）对于AR过程，用x_{t-k}乘以式(2-18)的两端，并取数学期望，可得差分方程：

$$r_k=\phi_1 r_{k-1}+\phi_2 r_{k-2}+\cdots+\phi_p r_{k-p} \tag{2-24}$$

当$k>0$时，x_{t-k} 和 a_t不相关，所以$E[x_{t-k}a_t]=0$。可得：

$$p_k=\phi_1 p_{k-1}+\phi_2 p_{k-2}+\cdots+\phi_p p_{k-p} \tag{2-25}$$

当$k=0$时，因为$E[x_t a_t]=E[{a_t}^2]={\sigma_a}^2$。可得：

$$p_k=C_1{w_1}^{-k}+C_2{w_2}^{-k}+\cdots+C_p{w_p}^{-k},k>-p \tag{2-26}$$

其中，$w_1,w_2,\ldots,w_p$是$\Psi(w)=0$的根，系数$C_1,C_2,\ldots,C_p$由P个初始值来定。

一般来说，平稳时间序列的自相关函数是由指数衰减和衰减正弦波组成。即自相关函数的尾部不可能在延迟某步之后等于零，而只是按负指数律衰减。因此，AR过程的自相关函数序列具有拖尾的特性。

（2）对于MA过程，其自相关函数为：

$$p_k=\begin{cases}\frac{-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q}{1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2} & k=1,2,\ldots,q\\ 0 & k>q\end{cases} \tag{2-27}$$

可见，MA序列的自相关函数，从$k>q$以后全部为零。我们称这一性质为“截尾”，反之，可以证明，若一平稳时间序列的自相关函数为截尾，那么它必是MA(q)序列。

2.3.2.2 偏相关函数

MA(q)过程的自协方差或自相关函数在q步滞后就截止为零，截尾性是MA序列特有的标志，据此可以推测出MA(q)模型的阶数。而AR(p)过程的自相关函数没有截止点，拖尾性是AR和ARMA序列的共有特性。为了构造出某种函数使得AR(p) 过程在滞后P步之后能够截止，引入偏相关函数的概念。

设$\{x_t\}$为平稳、零均值序列，考虑$x_{t-1}, x_{t-2}, \ldots, x_{t-k}$对$x_t$的线性最小方差估计，即选择系数$\Phi_{k1}, \Phi_{k2}, \ldots, \Phi_{kk}$，使得

$$\delta = E[x_t - \sum_{j=1}^{k} \phi k_j x_t - j]^2 = \gamma_0 - 2\sum_{i,j=1}^{k} \phi_{kj}\phi_{ki}\gamma_{j-i}$$

达到极小。为达此目的,对δ的右端分别求偏导数$\frac{\partial\delta}{\partial\Phi_{kj}}, j = 1,2,\ldots,k$，令其为零，便得到$\Phi_{kj}$应该满足的线性方程组：

$$\begin{pmatrix} 1 & p_1 & \cdots & p_{k-1} \\ p_1 & 1 & \cdots & p_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k-1} & p_{k-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix} \tag{2-28}$$

对$k = 1,2,\ldots$，依次利用Cramer法则，可得：

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & p_1 & p_2 & \cdots & p_{k-2} & p_1 \\ p_1 & 1 & p_1 & \cdots & p_{k-3} & p_2 \\ p_2 & p_1 & 1 & \cdots & p_{k-4} & p_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{k-1} & p_{k-2} & p_{k-3} & \cdots & p_1 & p_k \end{vmatrix}}{\begin{vmatrix} 1 & p_1 & p_2 & \cdots & p_{k-1} \\ p_1 & 1 & p_1 & \cdots & p_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{k-1} & p_{k-2} & p_{k-3} & \cdots & 1 \end{vmatrix}} \tag{2-29}$$

可以看出，当$k > p$时，AR(p)过程的行列式ϕ_{kk}将变为零，因为其中的任一行可以写成其它行的线性组合。我们称$\phi_{kk}, (k = 1,2,\ldots)$为$x_t$的偏相关函数。显然，偏相关函数$\phi_{kk}$是利用自相关函数$p_k$构造出来的。对于$p$阶自回归过程AR($p$)，出现了偏相关函数序列的截尾特性：$\phi_{kk} \neq 0, k \leqslant p$；$\phi_{kk} = 0, k > p$。

这就是说AR(p)模型的偏相关函数是p步截尾的，并且也只有AR过程的偏相关函数具有截尾的特性。

故此，我们可以依据样本的自相关、偏相关函数的截尾性和拖尾性，辨别模型类别，识别模型的阶数。

2.3.3 ARMA模型的参数估计

模型参数估计是基于ARMA 模型的参数识别一个关键，它既是识别的主要目标，又是进行模型检验的基础。选择适当的时间序列模型建模，通过考察样本自相关系数与偏自相关系数的性质进行初步识别：

- 若平稳时间序列的偏相关函数是截尾的，而自相关函数是拖尾的，则可断定此序列适合AR模型。
- 若平稳时间序列的偏相关函数是拖尾的，而自相关函数是截尾的，则可断定此序列适合MA模型。
- 若平稳时间序列的偏相关函数和自相关函数均是拖尾的，则 ARMA 模型适合此序列。

根据上述原则，总结的识别方法可见表2-2。

表 2-2 ARMA模型识别方法

自相关系数	偏相关系数	模型识别
拖尾	p阶截尾	AR(p)模型
q阶截尾	拖尾	MA(q)模型
拖尾	拖尾	ARMA(p,q)模型

按照上述方法选取的模型和参数，往往并不能确认是否完全符合线性关系。所以一般也难以仅仅按照机理方法建立模型。因而必须要利用一种可行的估计模型的方法，用有限数量的参数最逼真的描述某一时间序列。故通过上述方法选取的参数，对序列$\{x_t\}$由低阶到高阶逐一拟合模型ARMA(p,q)，并经有关统计量的检验选优。

有关统计量方法主要有FPE 准则（最终预测误差准则）方法、AIC 准则方法、F 检验临界值定阶准则方法等。但是，由于实际问题多种多样、随机过程千变万化，从而使统计分析计算存在误差。在本论文中，使用的是AIC 准则方法。下面就这一方法准则介绍。

AIC（Akaike’ s Information Criterion）准则方法是 Akaike 教授把极大似然函数用于时序模型假设检验而推导出相应的 AIC 准则：

$$\mathrm{AIC} = -2\ln(\text{模型的极大似然度}) + 2(\text{模型的独立参数个数})$$

对于ARMA(p,q)模型而言，其极大似然度和残差方差的对数有关，因此，其 AIC 函数为：

$$\mathrm{AIC} = N\ln(\hat{\sigma_\varepsilon}^2) + (p+q+l) \tag{2-30}$$

在不同模型参数做比较时，希望AIC指标越小越好。

2.4 支撑向量机回归（SVR）预测模型

支持向量机（SVM）是一种基于统计学习理论的新兴的数据挖掘和机器学习方法，由Cortes Corinna和Vapnik Vladimir等发明，在1995年引入机器学习领域后[?]，得到了广泛关注和深入发展，现在已成为数据挖掘领域的标准工具。SVM是建立在统计学习理论的VC 维理论和结构风险最小原理基础上的，它通过一个非线性映射ϕ将数据x_i映射到一个高维特征空间，并在空间进行线性回归，从而将低维特征空间的非线性回归问题转化为高维特征空间线性回归问题解决。

2.4.1 SVR的定义

支持向量回归（SVR）是支持向量机用于回归中的情况。

先考虑线性情况，设给定训练样本集：

$$S = \{(x_1,y_1),(x_2,y_2),\ldots,(x_l,y_l)\} \subset R^n \times R$$

样本集S是ε线性近似的，如果存在一个超平面$f(x) = \langle w,x\rangle + b$，其中$w \in R^n, b \in R$，下面的式子成立[?]：

$$|y_i - f(x_i)| \leqslant \varepsilon, i = 1,2,\ldots,l \tag{2-31}$$

图2-2显示了一个典型的线性近似。

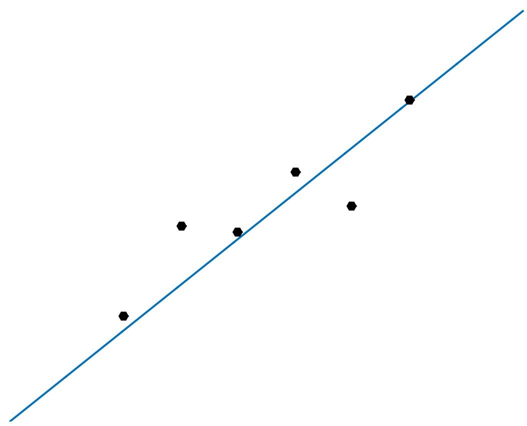


图 2-2 线性近似

引入d_i表示点$(x_i,y_i)\in S$到超平面$f(x)$的距离：

$$d_i=\frac{|\langle w,x_i\rangle+b-y_i|}{\sqrt{1+\|w\|^2}} \tag{2-32}$$

因为S集是ε线性近似的，所以有

$$|\langle w,x_i\rangle+b-y_i|\leqslant\varepsilon;i=1,2,\ldots,l$$

可得：

$$\frac{|\langle w,x_i\rangle+b-y_i|}{\sqrt{1+\|w\|^2}}\leqslant\frac{\varepsilon}{\sqrt{1+\|w\|^2}},1=1,2,\ldots,l$$

于是有：

$$\frac{d_i\leqslant\varepsilon}{\sqrt{1+\|w\|^2},i=1,2,\ldots,l} \tag{2-33}$$

式2-33表明，$\frac{\varepsilon}{\sqrt{1+\|w\|^2}}$是$S$中的点到超平面的距离的上界。

然而ε线性近似集S的最优近似超平面是通过最大化S中的点到超平面距离的上界而得到的超平面[?]。

图2-3 显示了最优近似超平面。 由这个定义能够得出最优近似超平面是通过

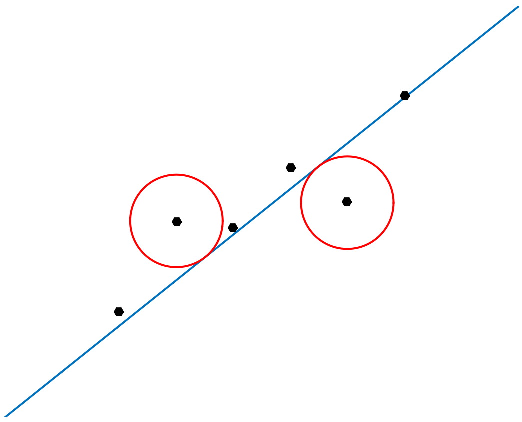


图 2-3 最优近似超平面

最大化 $\frac{\varepsilon}{\sqrt{1+\|w\|^2}}$ （即最小化$\sqrt{1+\|w\|^2}$）得到的。因此，只要最小化$\|w\|^2$就可以得到最优近似超平面。于是线性回归问题就化为求下面的优化问题：

$$\min\frac{1}{2}\|w\|^2 \tag{2-34}$$

约束为

$$|\langle w,x_i\rangle+b-y_i|\leqslant\varepsilon;i=1,2,\ldots,l \tag{2-35}$$

2.4.2 SVR的优化

考虑到可能存在一定误差,因此引入两个松弛变量：

$$\xi_i, \xi_i^* \geqslant 0, i = 1, 2, \ldots, l$$

损失函数采用是ε不敏感函数，它的定义为：

$$|\xi_\varepsilon| = \begin{cases} 0 & \text{如果}\, |\xi| \leqslant \varepsilon \\ |\xi| - \varepsilon & \text{其他} \end{cases}$$

此时的优化方程为：

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{2-36}$$

约束为：

$$\begin{cases} f(x_i) - y_i \leqslant \xi_i^* + \varepsilon & i = 1, 2, \ldots, l \\ y_i - f(x_i) \leqslant \xi_i + \varepsilon & i = 1, 2, \ldots, l \\ \xi_i, \xi_i^* \geqslant 0 & i = 1, 2, \ldots, l \end{cases} \tag{2-37}$$

优化式(2-36)中第一项使函数更为平坦，从而提高泛化能力，第二项则为减小误差，常数C对两者做出折中。ε为一正常数，$f(x_i)$与y_i的差别小于ε时不计入误差，大于ε时误差计为$|f(x_i) - y_i| - \varepsilon$。

引入拉格朗日函数：

$$\begin{aligned} L(w, b, \xi, \alpha, \alpha^*, \gamma) = \ & \tfrac{1}{2}\|w\|^2 + C\textstyle\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ & - \textstyle\sum_{i=1}^{l}\alpha_i[\xi_i + \varepsilon - y_i + \langle w, x_i\rangle + b] \\ & - \textstyle\sum_{i=1}^{l}\alpha_i^*[\xi_i + \varepsilon + y_i - \langle w, x_i\rangle - b] \\ & - \textstyle\sum_{i=1}^{l}\gamma_i(\xi_i + \xi_i^*) \end{aligned}$$

其中$\alpha, \alpha^*, \gamma_i > 0, i = 1, 2, \ldots, l$

函数L的极值应满足条件：

$$\frac{\partial}{\partial w}L = 0, \frac{\partial}{\partial b}L = 0, \frac{\partial}{\partial \xi_i^{(*)}}L = 0$$

于是得到下面的式子：

$$\begin{cases} w = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i \\ \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ C - \alpha_i - \gamma_i = 0 & i = 1, 2, \ldots, l \\ C - \alpha_i^* - \gamma_i^* = 0 & i = 1, 2, \ldots, l \end{cases} \tag{2-38}$$

将式(2-38)代入式(2-31)中，得到优化问题的对偶形式为：

$$\max -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\langle x_i,x_j\rangle+\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)y_i-\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)\varepsilon \tag{2-39}$$

约束为：

$$\begin{cases} sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0 \\ 0\leqslant \alpha_i,\alpha_i^*\leqslant C \qquad\qquad i=1,2,\ldots,l \end{cases} \tag{2-40}$$

对于非线性回归，同分类情况一样，首先使用一个非线性映射ϕ把数据映射到一个高维特征空间，然后在高维特征空间进行线性回归。由于在上面的优化过程中只考虑到高维特征空间中的内积运算，因此用核函数$K(x,y)$代替$\langle\phi(x),\phi(y)\rangle$就可以实现非线性回归。于是，非线性回归的优化方程为最大化下面的函数：

$$\begin{aligned} W(\alpha_i,\alpha_i^*)= & -\tfrac{1}{2}\textstyle\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(x_i,x_j) \\ & +\textstyle\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)y_i-\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)\varepsilon \end{aligned} \tag{2-41}$$

约束条件为式(2-40)。解出α的值后，可得$f(x_0)$的表达式为：

$$f(x)=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)K(x_i,x)+b \tag{2-42}$$

通常情况下，大部分α_i 和 α_i^*的值将为零，不为零的α_i或α_i^* 所对应的样本被称为支持向量。

根据KKT条件，在鞍点有：

$$\begin{cases} \alpha_i[\xi_i+\varepsilon-y_i+f(x_i)]=0 & i=1,2,\ldots,l \\ \alpha_i^*[\xi_i+\varepsilon+y_i-f(x_i)]=0 & i=1,2,\ldots,l \\ (C-\alpha_i)\xi_i=0 & i=1,2,\ldots,l \\ (C-\alpha_i^*)\xi_i^*=0 & i=1,2,\ldots,l \end{cases}$$

于是可得b的计算式如下：

$$b=\begin{cases} y_j-\varepsilon-\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)K(x_j,x_i) & \text{当}\alpha_j\in(0,C) \\ y_j+\varepsilon-\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)K(x_j,x_i) & \text{当}\alpha_j^*\in(0,C) \end{cases} \tag{2-43}$$

用任意一个支持向量就可以计算出b的值，也可以采用取平均值的方法。

2.4.3 SVR参数的选取

在的实现过程中，参数的选取直接关系最终回归结果的准确率。目前关于SVR参数的选取问题，一般采用的方法有经验选择法、实验比较法、梯度下降法、交叉验证法、贝叶斯算法，分布估算法等等。在综合考虑了运算速度和可信度等因素之后，本论文决定采用交叉验证法来进行参数的选择。

交叉验证思想（Cross validation，简称CV），是用来验证分类器性能的一种统计分析方法，其基本思想是把在某种意义下将原始数据进行分组，一部分作为训练集，另一部分作为验证集，先用训练集对分类器进行训练，再用验证集来测试训练得到的模型，以得到的回归准确率作为评价分类器的性能指标。

2.5 本章小结

本章主要对基于趋势的预测算法的理论进行描述。

首先，对卫星在轨异变信号的特征进行分析，并根据这些特征，给出基于小波分析、ARMA 预测模型、SVR 预测模型的算法融合的预测方法。

接下来，逐一描述了小波分析、ARMA 预测模型、SVR 预测模型的基本算法理论，并给出仿真实验需要用到的参数选择方法等必要理论叙述。

第三章 基于相似度预测算法的理论基础

如同第二章所说的，本论文对卫星在轨异变状态的预测算法基于两种思路进行研究，第一种是基于趋势的预测算法，第二种是基于相似度的预测算法。在本章中，将详细描述基于相似度的预测算法的理论研究。

3.1 相似度预测基础理论

3.2 分形算法理论

3.2.1 分形算法概述

分形最初概念起始于科学家Mandelbrot于1967年在《Science》上发表的一篇文章，提出了一个著名问题："英国的海岸线有多长？"对于任何非内陆国家都有海岸线，其长度都与一个具体的数值相对应，但这个数值是如何得到的？一般情况下，采用折线近似方法得到，选择单位长度ε对其进行度量，共需要$N(\varepsilon)$个单位长度，则海岸线长度为$L(\varepsilon)=N(\varepsilon)\cdot\varepsilon$，这样通过计算可以得到其长度，但是实际计算得到结果却不够准确。因为海岸线自身自然结构的原因，包括很多与非平滑曲线类似的海湾，当选用尺度ε较大时，对于很多细节无法准确测量，使得测量得到的结果不存在较大误差，不够准确；若选用相对较小的测量尺度，也不能得到准确结果。随着测量尺度的减小，会发现实际上需要测量的成分会逐步增多，导致测量最终结果趋于无穷。

在实验中，Mandelbrot发现，$N(\varepsilon)\cdot\varepsilon^D=8\mathrm{p}$，$D$不随单位长度发生变化，$D$一般也是非整数，$D$就被定义为其维数。

$$N(\varepsilon)\cdot\varepsilon^D=L_0$$

化简之后可得：

$$L(\varepsilon)=N(\varepsilon)\cdot\varepsilon=L_0\cdot\varepsilon^{1-D}$$

两边同时取对数可得：

$$\ln L(\varepsilon)=(1-D)\ln\varepsilon+\ln L_0$$

可以利用测得的实际长度对数为纵坐标，选用尺度对数为横坐标，绘出直线图，通过计算斜率的方式可以求出维数D。这是对于D的最原始的求法。例如对于单位长度的线段，选取尺寸长度为ε对其进行测量，当$\varepsilon=1$时，测量次数$N=1$；

当$\varepsilon = 0.5$时，测量次数$N = 2$；随着测量尺度ε的不断减小，N会越来越大，但是$N(\varepsilon) \cdot \varepsilon^D$始终为常数，不发生变化。

分形理论被用于异常状态诊断和预测领域主要是基于相同状态下信号的自相似性理论基础上，即同一工作状态下的信号彼此之间空间填充能力类似，分形特性非常接近。自然界中物体的自相似性不是绝对意义上的相等，而是统计意义上的相等，即允许在一定范围内存在适度波动，这种特性使得分形在诸多方面得到广泛应用，例如医院对人体血液内各项指标进行判定，通常采取适量血液对人体整体情况进行估量；在天文学中基本粒子和诸多星系之间的运动也存在着自相似性。

分形维数的自相似性只存在特定的区域，即信号自身存在一定的波动性使得维数在一个合理的区间内波动，使得相同状态下计算得到的分形维数之间彼此非常接近，这是基于分形实现状态检测的重要理论基础。

分形维数反映了信号的非线性特性和空间填充能力，在传统观念中，一般物体的维数以自然数形式存在，但是在实际情况中大部分物体都是以非整数的形式存在。

3.2.2 典型的分型维数

单重分形主要依靠计算单重分形维数来实现，在描述信号分形特性方面，常使用以下几种维数：容积维数（盒维数）、信息维数、关联维数、自相似维数，在故障诊断方面，通常使用盒维数与关联维数来实现。

3.2.2.1 盒维数

在讨论盒维数之前，不妨先考虑这样一个问题：如何计算一个面积为S的平面区域 Z 的维数。假设采用边长为ε的正方形覆盖这个平面区域 Z，所需的正方形数量为 $N(\varepsilon)$，则可以得到：

$$N(\varepsilon) = \frac{S}{\varepsilon^2}$$

两边取对数，得：

$$\ln N(\varepsilon) = \ln S + 2\ln(\frac{1}{\varepsilon})$$

由此可以得到 Z的维数为：

$$D(Z) = \lim_{\varepsilon \to 0} \frac{\ln N(\varepsilon)}{\ln(1 varepsilon)} = 2 \tag{3-1}$$

前文提到过，对于维数的测量存在这样一个公式： $N(\varepsilon)\cdot\varepsilon^{D}=8\text{p}$。也可以得到一个类似的等式：

$$N(\varepsilon)\cdot\varepsilon^{2}=S \tag{3-2}$$

其中，S为二维区域 Z的面积，是一个常数。

基于这种用集的覆盖来定义维数的思想，数学家 Kolmogorov 于 1958 年和1959 年提出了盒维数的概念，又称之为 Kolmogorov 容量维。盒维数是以尺度为ε 的超立方体去覆盖集合 X的方式来定义的。

设 X是R^{n}的非空有界子集，如果可以采用 $N(\varepsilon)$个边长为ε的超立方体将其覆盖，则有：

$$D_{B}(X)=\lim_{\varepsilon\to 0}\frac{\ln N(\varepsilon)}{\ln(1varepsilon)} \tag{3-3}$$

上式即为分形盒维数的定义表达式。

在式(3-3)中，采用的 ε 覆盖是一个一般的集类，在应用于具体问题时，可以根据需要选择不同的集类作为 ε 覆盖，这在定义上是等价的。在实际应用中，$N(\varepsilon)$可以取下面的几种形式：

（1）覆盖X的半径为ε的最少闭球数；

（2）覆盖 X的边长为 ε的最少立方体数；

（3）与 X相交的ε坐标网立方体个数；

（4）覆盖 X的直径最大为ε 的集的最少个数；

（5）球心在 X上，半径为ε 的互不相交的球的最多个数。

在实际的分形盒维数计算中，广泛地采用（2）和（3）中所定义的 $N(\varepsilon)$，这两种形式可以使分形盒维数的计算更加有效、简便。

3.2.2.2 信息维数

根据盒维数的定义，在计算集合 X的盒维数时，用边长为 ε 的超立方体进行覆盖，在这些超立方体中，有的是空的，有的不是空的。如果将超立方体编号，设集合 X落在第 i个超立方体中的几率为$P_{i}(\varepsilon)$，则用边长为 ε 的超立方体进行测量，所得到的信息量用香侬（Shannon）公式表示为：

$$I(\varepsilon)=-\sum_{i}^{N(\varepsilon)}P_{i}(\varepsilon)\ln P_{i}(\varepsilon) \tag{3-4}$$

如果用$I(\varepsilon)$来代替盒维数定义中的$\ln N(\varepsilon)$，则可以得到信息维数：

$$D_{I}=\lim_{\varepsilon\to 0}\frac{I(\varepsilon)}{\ln(\frac{1}{\varepsilon})} \tag{3-5}$$

即为：

$$D_I = \lim_{\varepsilon\to 0} \frac{\sum_i^{N(\varepsilon)} P_i(\varepsilon)\ln P_i(\varepsilon)}{\ln\varepsilon} \tag{3-6}$$

在特定的情况下，如果集合X落在每个超立方体中的概率都相等，则有$P_i(\varepsilon) = \frac{1}{N(\varepsilon)}$，那么可得到：

$$\begin{aligned} I(\varepsilon) &= -\textstyle\sum_i^{N(\varepsilon)} P_i(\varepsilon)\ln P_i(\varepsilon) \\ &= -\textstyle\sum_i^{N(\varepsilon)} \frac{1}{N(\varepsilon)} \ln\frac{1}{N(\varepsilon)} \\ &= -N(\varepsilon)\cdot\frac{1}{N(\varepsilon)}\cdot[-\ln N(\varepsilon)] \\ &= \ln N(\varepsilon) \end{aligned} \tag{3-7}$$

此时，信息维数与盒维数相等。

3.2.2.3 关联维数

关联维数是通过点对点的关联性表示集合中所有点的相关程度。设集合S表示点数为N的时域序列$\{x_1, x_2, \ldots, x_n\}$，对任意常数$r(r>0)$，若任意两点距离小于$r$的点对则是有关联点对。关联点对占有的比例称为相关函数，定义如下：

$$C(r) = \frac{1}{N^2}\sum_{i\neq j}^{N} H(\mu) \tag{3-8}$$

其中，H是单位阶梯函数（Heaviside function）：

$$H(\mu) = \begin{cases} 1 & \mu \geqslant 0 \\ 0 & \mu < 0 \end{cases}$$

$$u = r - |x_i - x_j|$$

当r较大时，μ始终大于0，则相关函数为1；当r较小时，μ始终小于0，相关函数为0；这两种极端情况不能客观反映系统的真实性质。所以对r的取值范围有一定的限制，当r满足一定条件时，有如下关系：

$$\lim_{r\to 0} C(r) = r^{D_c} \tag{3-9}$$

变换后可以得到：

$$\lim_{r\to 0} \frac{\ln(C(r))}{\ln r} \tag{3-10}$$

其中D_c即为关联维数。

在针对单重分形的异常状态检测的实际应用中，盒维数与关联维数应用最为广泛。其中关联维数对信号的不均匀性反映敏感，可以有效区分不同的工作状态；

但是关联维数计算相对复杂，计算时通过相空间重构方法实现，参数变量多。因此分形盒维数常常对于快速的诊断故障，关联维数用于敏感测量。

3.2.3 分型维数计算方法

3.2.3.1 盒维数的计算方法

用于描述信号的分形特征使用的主要特征参数是分形维数，对于计算机而言，直接通过盒维数的定义来对其进行计算是非常不方便的，网格维数与盒维数等效，其计算方法适合计算机处理。

对实际采集的信号序列$x_1, x_2, \ldots, x_n$，采样频率f_s ,采样间隔 $T_s = \frac{1}{f_s}$，该信号的网格维数为：

$$d = -\frac{\ln(\sum |x_{i+1} - x_i|)/\Delta t}{\ln \Delta t} \tag{3-11}$$

网格维数从盒维数的测度上反映了信号的分形特性，由于不同状态下的信号具有不同的分形特性，因此网格维数可以用于特征量，从而对不同状态进行判定。

3.2.3.2 关联维数的计算方法

关联维数在状态空间中是重要的几何参数，是对于描述该空间所需变量个数的描述。分形维数对于事物分形特征可以进行定量描述，非常重要，这是分形能够广泛应用于诸多领域的重要理论基础。

设$\{X_k | k = 1, 2, \ldots, n\}$是采样得到的信号序列，对信号序列进行相空间重构，其结果记为$X_k(m, \tau) = (x_n, x_{n+\tau}, \ldots, x_{n+(-1)\tau})$，$m$是重构相空间的维数，$\tau$是时间延迟。定义重构相空间的关联维数

$$D = \lim_{r \to 0} \frac{d \ln C_r}{d \ln r} \tag{3-12}$$

因此对于关联维数的计算步骤是：

（1）对时间序列进行规格化处理，依据相关原则确定重构相空间嵌入维数m，构建m维空间的相应矢量。

（2）求得两个矢量间的距离 $r_{ij} = |y_i - y_j|$。

（3）对任意$r \in (0, 1]$，通过比较所有点对的距离值和r值，计算距离小于r的点对所占比例：

$$C_r = \frac{1}{N(N-1)} \sum_{i=1}^{N-m+1} \sum_{j=1}^{N-m+1} H(r - |X_i - X - j|) \tag{3-13}$$

合理调整r的取值，使 $C_r \propto r^{D_k}$

（4）通过关联维数计算公式

$$D_h = \lim_{r\to 0} \frac{d\ln C_r}{d\ln r}$$

就可以求得对应信号序列的关联维数。

关联维数对信号不均匀性反映敏感，更能反映信号的动态结构，对于信号的分形特征分析一般也通过计算关联维数来实现。

3.2.4 分型维数应用于异变数据

3.3 模糊聚类算法理论

模糊聚类方法是基于 Zadeh 教授在1965年提出的模糊理论与聚类分析相结合的产物[?]。模糊聚类方法能够对类与类之间有交叉的数据样本集进行有效的聚类，所得的聚类结果明显优于硬聚类方法。由于模糊聚类建立了数据样本对于类别的不确定性的描述，表达了样本类属的模糊性，因此能够更客观地反应现实世界，成为聚类分析研究的热点和主流。

3.3.1 模糊聚类的分析步骤

模糊聚类分析步骤可以概括为：数据标准化，建立模糊相似矩阵，聚类。

3.3.1.1 数据标准化

在实际课题中，不同的数据可能有不同的量纲。为了使不同量纲的数据也能进行比较，需要对数据进行适当变换。根据模糊矩阵的要求将数据压缩到区间$[0,1]$。设论域$U=[u_1,u_2,\ldots,u_n]$为被分类的对象，每个元素又由m个数据表示，对第i个元素有

$$u_i = [x_{i1}, x_{i2}, \ldots, x_{im}] \qquad (i=1,2,\ldots,m)$$

这时的原始数据矩阵为：

$$\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}$$

其中：x_{nm}表示第n个分类对象的第m个指标的原始数据。

（1）标准差变换

$$x'_{ik} = \frac{x_{ik} - \bar{x_k}}{s_k} \qquad (i = 1,2,\ldots,n; k = 1,2,\ldots,m) \tag{3-14}$$

其中，

$$\bar{u_k} = \frac{1}{n}\sum_{i=1}^{m} x_{ik}, s_k = \sqrt{\frac{1}{n}\sum_{i=1}^{m} x_{ik} - \bar{x_k}^2}$$

经过变换后，每个变量的均值为0，标准差为1，并可以消除量纲的影响。

（2）极差变换

$$x''_{ik} = \frac{x'_{ik} - \min\limits_{1\leqslant i\leqslant n} x'_{ik}}{\max\limits_{1\leqslant i\leqslant n} x'_{ik} - \min\limits_{1\leqslant i\leqslant n} x'_{ik}} \tag{3-15}$$

经过极差变换后有$0 \leqslant x''_{ik} \leqslant 1$，并且消除了量纲的影响。

（3）对数变换

$$x'_{ik} = \lg x_{ik} \qquad (i = 1,2,\ldots,n; k = 1,2,\ldots,m) \tag{3-16}$$

取对数以缩小变量间的数量级。

3.3.1.2 建立模糊相似矩阵

建立模糊相似矩阵又称为标定，即标出衡量被分类对象间相似程度的统计量$r_ij (i,j = 1,2,\ldots,n)$。

设论域$U = u_1, u_2, \ldots, u_n$，其中每一个元素为一个样本，建立$U$上的相似关系$R$，$R$表示相似矩阵$r_ij = R(u_i, u_j)$。每个样本为$m$维向量，$u_i = x_i1, x_i2, \ldots, x_im$。

计算r_ij可以用以下方法：

（1）相似系数法

（a） 数量积法

$$r_{ij} = \begin{cases} 1 & i = j \\ \frac{1}{M}\sum_{k=1}^{m} x_{ik}x_{jk} & i \neq j \end{cases} \tag{3-17}$$

其中，

$$M = \max_{i\neq j}\left(\sum_{k=1}^{m} x_{ik}x_{jk}\right)$$

显然$|r_{ij}| \in [0,1]$，如果r_{ij}中出现负数，需要再进行变换：

$$r'_{ij} = \frac{r_{ij}+1}{2}$$

（b） 夹角余弦法

$$r_{ij}=\frac{\left|\sum_{k=1}^{m}x_{ik}x_{jk}\right|}{\sqrt{\sum_{k=1}^{m}x_{ik}^{2}}\sqrt{\sum_{k=1}^{m}x_{jk}^{2}}} \tag{3-18}$$

（c） 相关系数法

$$r_{ij}=\frac{\sum_{k=1}^{m}\left|x_{ik}-\bar{x_i}\right|\left|x_{jk}-\bar{x_j}\right|}{\sqrt{\sum_{k=1}^{m}(x_{ik}-\bar{x_i})^2}\sqrt{\sum_{k=1}^{m}(x_{jk}-\bar{x_j})^2}} \tag{3-19}$$

其中，$\bar{x_i}=\frac{1}{m}\sum_{k=1}^{m}x_{ik}$，$\bar{x_j}=\frac{1}{m}\sum_{k=1}^{m}x_{jk}$。

（d） 指数相似系数法

$$r_{ij}=\frac{1}{m}\sum_{k=1}^{m}\exp(-\frac{3}{4}\cdot\frac{(x_{ik}-x_{jk})^2}{s_k^2}) \tag{3-20}$$

其中，$s_k^2=\frac{1}{n}\sum_{i=1}^{n}(x_{ik}-x_{jk})^2$，$\bar{x_k}=\frac{1}{n}\sum_{k=1}^{n}x_{ik}$。

（e） 最大最小法

$$r_{ij}=\frac{\sum_{k=1}^{m}\min(x_{ik},x_{jk})}{\sum_{k=1}^{m}\max(x_{ik},x_{jk})} \tag{3-21}$$

（f） 算术平均最小法

$$r_{ij}=\frac{\sum_{k=1}^{m}\min(x_{ik},x_{jk})}{\frac{1}{2}\sum_{k=1}^{m}(x_{ik}+x_{jk})} \tag{3-22}$$

（g） 几何平均最小法

$$r_{ij}=\frac{\sum_{k=1}^{m}\min(x_{ik},x_{jk})}{\sum_{k=1}^{m}\sqrt{x_{ik}x_{jk}}} \tag{3-23}$$

（2）距离法

（a） 绝对值倒数法

$$r_{ij}=\begin{cases}1 & i=j\\ \frac{M}{\sum_{k=1}^{m}\left|x_{ik}-x_{jk}\right|} & i\neq j\end{cases} \tag{3-24}$$

其中，M需要适当选取，使$0\leqslant r_{ij}\leqslant 1$。

（b） 绝对值指数法

$$r_{ij}=\exp(-\sum_{k=1}^{m}\left|x_{ik}-x_{jk}\right|) \tag{3-25}$$

（c） 直接距离法

$$r_{ij}=1-d(u_i,u_j) \tag{3-26}$$

其中，c为适当选取系数，使得$0\leqslant r_{ij}\leqslant 1$。$d(u_i,u_j)$为距离，经常使用的距离有以下几种：

●海明距离：

$$d(u_i,u_j)=\sum_{k=1}^{m}\left|x_{ik}-x_{jk}\right| \tag{3-27}$$

●欧式距离：

$$d(u_i,u_j)=\sqrt{\sum_{k=1}^{m}(x_{ik}-x_{jk})^2} \tag{3-28}$$

●切比雪夫距离：

$$d(u_i,u_j)=\max\left|x_{ik}-x_{jk}\right| \qquad (1\leqslant k\leqslant m) \tag{3-29}$$

（3）主观评分法请专家直接对u_i和u_j的相似度评分，也是一种有效方法。

（a） 百分制

采用百分制时，将评出的总分数除以 100，即得闭区间$[0,1]$的一个r_ij。为降低其主观性，可以请多个专家参与评分，再取平均定出r_ij。

（b） 相似度和自信度

假定请N个专家组成专家组，这时有

$$r_{ij}=\frac{\sum_{k=1}^{N}r_{ij}(k)a_{ij}(k)}{\sum_{k=1}^{N}a_{ij}(k)} \tag{3-30}$$

式中，$r_{ij}(k)$为第k个专家所给出u_i和u_j的相似度，$a_{ij}(k)$是专家对自己给出相似度时的自信度。r_{ij}和a_{ij}都是在$[0,1]$区间的数值。

3.3.1.3 基于模糊等价关系的聚类方法

（1）传递闭包法

根据标定所建立的模糊矩阵$\tilde{R}$；一般来说仅具有自反性和对称性，不满足传递性，只是模糊相似矩阵，只有当$\tilde{R}$是模糊等价矩阵时才能聚类，故需要将$\tilde{R}$改造成模糊等价矩阵。

（2）直接聚类

直接聚类是在建立模糊相似矩阵之后，不用传递闭包$t(\tilde{R})$，而直接从模糊相似矩阵进行聚类。步骤如下：

（a） 取$\tilde{R}$中的最大值$\lambda_1=1$，对每个μ_i作相似类$[\mu_i]_{\tilde{R}}$，且

$$[\mu_i]_{\tilde{R}}=\{\mu_i|r_ij=1\}$$

（b） 将满足$r_ij=1$的x_i和x_j放在一类，构成相似类。不同的相似类可能存在公共元素，此时将有公共元素的相似类合并，即可得$\lambda_1=$

1水平上的等价分类；

（c） 取λ_2为次大值，从中$\tilde{R}$找到满足λ_2的等价类，再与（b）的λ_1水平上的等价类进行合并，最终得出对应λ_2的等价分类；

（d） 取λ_3为第三大值，从$\tilde{R}$中找出相似度λ_3的等价类，然后与（c）的λ_2 水平上的等价类合并，得出λ_3水平上的等价分类；

以此类推，直到合并到U成为一类为止。

（3）分类的F检验

从上面方法可知，模糊聚类分析是动态的，对于不同的$\lambda \in [0,1]$，可以获得不同的分类。随着λ的变化而形成的多种分类对全面了解样本情况是有利的。但在实际应用中需要选择阈值λ，从而给出一个较明确的分类。用统计学的F检验方法可以剔除一些不够格的类，使分类变为较为清晰。

设论域$U = u_1, u_2, \ldots, u_n$是样本数为n的样本空间，而每个样本$_i$有$m$个特征，即$_i = (x_{i1}, x_{i2}, \ldots, x_{im})$，由此得到原始的数据矩阵。

总体样本的中心向量为

$$\bar{u} = (\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k, \ldots, \bar{u}_n) \tag{3-31}$$

其中$\bar{u}_k = \frac{1}{n}\sum_{i=1}^{n} x_{ik}, (k = 1, 2, \ldots, m)$。

设对应于λ值得分类数为r。其样本标记为$u_1^{(j)}, u_2^{(j)}, \ldots, u_n^{(j)}$，第$j$类的样本数为$n_j$。第$j$类的聚类中心为：

$$\bar{u}^{(j)} = (\bar{u}_1^{(j)}, \bar{u}_2^{(j)}, \ldots, \bar{u}_k^{(j)}, \ldots, \bar{u}_n^{(j)}) \tag{3-32}$$

其中，$\bar{u}_k^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ik}^{(j)} \qquad (k = 1, 2, \ldots, m)$作$F$统计量

$$F = \frac{\sum_{j=1}^{r} n_j \left\| \bar{u}^{(j)} - \bar{u} \right\|^2 / (r-1)}{\sum_{j=1}^{r}\sum_{i=1}^{n_j} \left\| u_i^{(j)} - \bar{u}^{(j)} \right\|^2 / (n-r)} \tag{3-33}$$

其中$\left\| \bar{u}^{(j)} - \bar{u} \right\| = \sqrt{\sum_{k=1}^{m} (\bar{u}^{(j)} - \bar{u})^2}$。

F统计量服从自由度$r-1$，$n-r$的F分布。分子表示类与类之间的距离，分母表示类内样本间的距离。F值越大，说明类与类之间的距离越大，表示类与类之间差异越大，分类越明显。

3.3.2 模糊C均值聚类算法

文献中研究最多、实际中应用最广的是基于目标函数的模糊聚类方法，该类方法把聚类问题描述为一个带约束的优化问题，通过求解优化问题的解来确定数

据集的模糊划分和聚类结果。此类算法设计简单易于应用且聚类性能良好，并借助于经典的数学非线性规划理论来求解优化问题，容易编程实现。因此，基于目标函数的模糊聚类成为新的研究主流。

基于目标函数的聚类算法中模糊C均值（Fuzzy C-Means， FCM）算法的理论最为完善、应用最为广泛。FCM 算法首先是由 E.Ruspini 提出来的，后来 J.C.Dunn 与 J.C.Bezdek 将 E.Ruspini 算法从硬聚类算法推广成模糊聚类算法。FCM 算法是基于对目标函数优化的一种逐步迭代的数据聚类方法，每一步的迭代都沿着目标函数减小的方向进行。聚类结果是每一个数据点对聚类中心的隶属程度，该隶属程度用一个数值来表示。为了借助目标函数法求解聚类问题，人们利用均方逼近理论构造了带约束的非线性规划函数，从此类内平方误差和（WGSS）成为聚类目标函数的普遍形式。为极小化该目标函数而采取 Pikard 迭代优化方案就是著名的硬C均值算法和ISODATA（Iterative Self-Organizing Data Analysis Technique A）算法。模糊划分的概念提出后，Dunn 首先把 WGSS 函数扩展到了一类内加权平均误差和函数，后来Bezdek 又引入了一个参数，把如推广到一个目标函数的无限族，并给出了交替优化（AO，Alternative Optimization）算法，即为人们所熟知的 FCM 算法。从此，奠定了 FCM 算法在模糊聚类中的地位。FCM 算法在模式识别、语音识别、图像处理以及空间导航系统等领域得到广泛应用，是模糊聚类的最主要的实用算法之一。

3.3.2.1 FCM算法的基本原理

FCM聚类算法是一种能自动对样本点进行分类的方法，通过优化准则函数来确定每个样本点对类中心的隶属度，从而决定样本所属的类。其中为各样本与其所在类中心的误差平方和。FCM 算法实现原理如下。

设样本集合$X=\{x_1,x_2,\dots,x_i,\dots,x_n\}$中元素$x_i$有$s$个特征，即$x_i=\{x_{i1},x_{i2},\dots,x_{is}\}$，要把$X$分为$c$类，$(2\leqslant c\leqslant n)$。

设有c个聚类中心$V=\{v_1,v_2,\dots,v_c\}$，取d_{ik}为样本，x_k与聚类中心v_i的欧式距离，记作：

$$d_{ik}=\|x_k-v_i\|=\sqrt{\sum_{j=1}^{s}x_{kj}-v_{ij}{}^2} \tag{3-34}$$

聚类准则是使得下列目标函数达到最小值

$$\min J_{FCM}(X,U,V)=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}^{m}(d_{ik})^2 \tag{3-35}$$

其中： u_{ik}表示第k个样本在第i类中的隶属度，且满足

$$\sum_{i=1}^{c} u_{ik} = 1 \tag{3-36}$$

m是模糊因子用来决定聚类结果模糊度的权重指数，$m \in [1,\infty)$，取其经验值范围为$1.5 \leqslant m \leqslant 2.5$。结合应用 Lagrange 乘数法，结合式(3-36)的约束条件对式(3-35)求解，可得模糊划分矩阵U和聚类原型V更新公式如下：

$$u_{ik} = \begin{cases} \frac{\|x_k - v_i\|^{-\frac{2}{m-1}}}{\sum_{j=1}^{c} \|x_k - v_j\|^{-\frac{2}{m-1}}} & \|x_k - v_j\| > 0 (1 \leqslant j \leqslant c) \\ 1 & \|x_k - v_i\| = 0 (1 \leqslant i \leqslant c) \\ 0 & \exists j, j \neq i, \|x_k - v_j\| = 0 \end{cases} \tag{3-37}$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m \cdot x_k}{\sum_{k=1}^{n} u_{ik}^m} \tag{3-38}$$

3.3.2.2 FCM算法的流程及具体步骤

FCM算法流程图如图3-1所示：

FCM算法的具体步骤如下：

（1）初始化：给定聚类类别数$c(2 \leqslant c \leqslant n)$，$n$ 为样本个数；给定模糊权数m（一般取2）；设定迭代停止阈值ε（一般取0.001至 0.01）；设置迭代计数次数l，初始化聚类原型$V^{(l)}(l=0)$；

（2）根据$V^{(l)}$，按照公式(3-37)更新模糊划分矩阵U^{l+1}得：

$$u_{ik}^{(l+1)} = \left(\frac{d_{ik}^{(j)}}{\sum_{j=1}^{c} d_{jk}^{(j)}}\right)^{-\frac{2}{m-1}} \tag{3-39}$$

（3）根据$U^{(l)}$，按照公式(3-38)计算新的聚类中心矩阵V^{l+1}得：

$$v_i^{(l+1)} = \frac{\sum_{k=1}^{n} (u_{ik}^{(j)})^m \cdot x_k}{\sum_{k=1}^{n} (u_{ik}^{(j)})^m} \tag{3-40}$$

（4）判定阈值：根据给定的阈值ε，如果$\left\| V^{(j+1)} - V^{(j)} \right\|$，则停止迭代，否则$l = l + 1$，转到（2）。

3.3.2.3 FCM算法的优点

FCM算法的优点主要表现在以下几个方面：

（1）FCM算法是硬C均值算法的推广，由于硬C均值函数在理论上的研究已经相当完善，这就为FCM 算法的研究提供了条件；

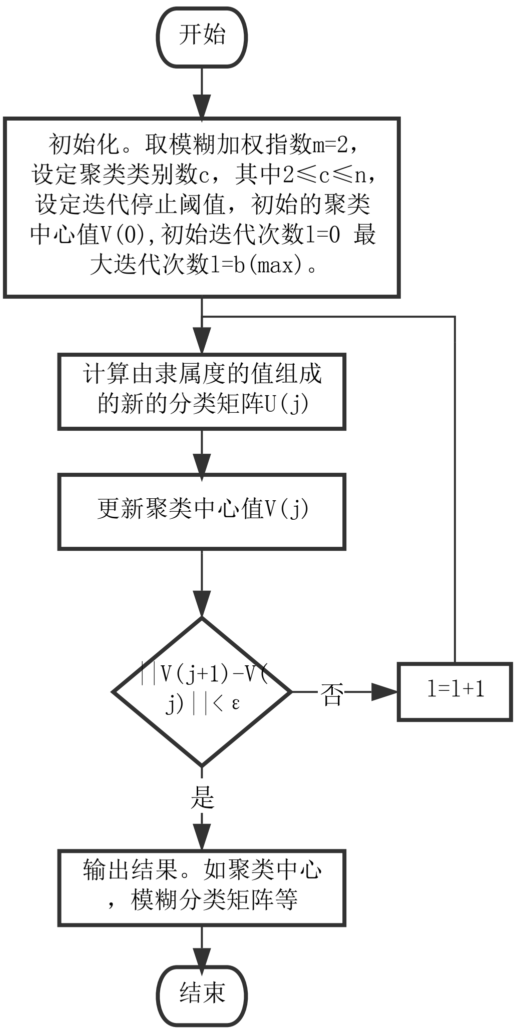


图 3-1 FCM算法流程图

（2）FCM算法结构简单，可转化为目标函数的优化问题，可借助最优化理论进行研究；

（3）FCM算法的目标函数与希尔伯特空间结构有密切的关系，因此具有深厚的数学基础，且复杂度低，应用效果好；

（4）FCM算法研究成熟，算法稍做改动即可适用于某些特定的领域，目前已经在很多领域获得了非常成功的应用。

3.4 本章小结

第四章 基于趋势预测算法的实验仿真

在第二章中讲解了基于趋势预测算法的理论，而在本章中将根据第二章中的理论，进行仿真实验的描述。

4.1 实验方案及预测结果评判指标

4.1.1 实验方案

在此先给出基于趋势预测算法的实验方案图，如图4-1所示。

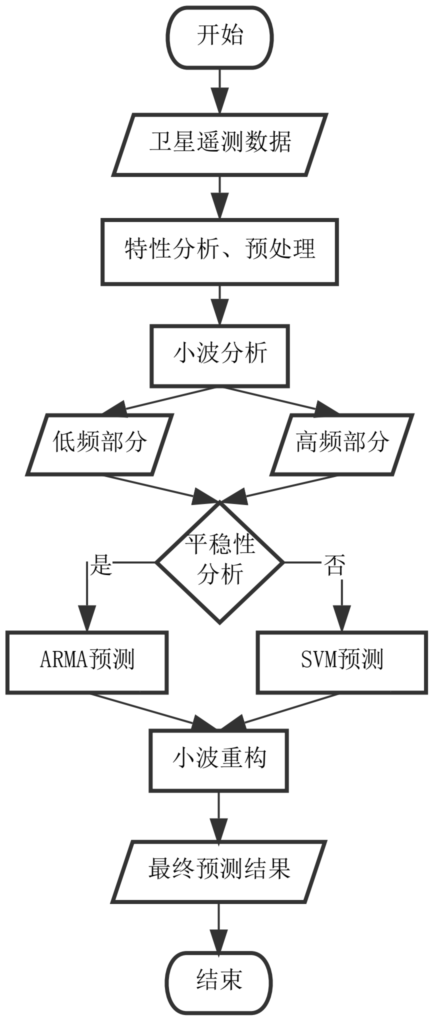


图 4-1 基于趋势预测算法的实验方案

基于趋势预测算法的实验方案，是采取算法融合的方式来对卫星在轨异变数据进行预测，主要用到小波分析算法，ARMA预测模型，SVR预测模型等方法。根据实验方案图，下面将给出具体的算法步骤：

（1）根据实测卫星信号$Y:\{y_1,y_2,...\}$的时频特征，选取一个小波对信号序列进行多尺度分解，对信号的趋势进行分析，确定分解尺度n和最大分解层数J。

（2）信号经小波分解方法处理，得到高频序列$d_{n,i}$和低频序列$c_{J,i}$；

（3）分别对高频序列$d_{n,i}$和低频序列$c_{J,i}$进行数据平稳性检验。一般情况下，高频序列是满足平稳性的，而低频序列则要视情况而定。

（4）如果低频序列$c_{J,i}$和$d_{n,i}$满足平稳性验证，则对其经行ARMA预测，如果不满足，则对其进行SVR预测。

（5）对得到的高频预测值$d'_{n,i+1}$和低频预测值$c'_{J,i+1}$进行小波重构，即进行加权计算，得到最终结果y'_{i+1}：

$$y'_{i+1}=d'_{1,i+1}+d'_{2,i+1}+\cdots+d'_{J,i+1}+c'_{J,i+1}$$

同理，根据t_i时刻之前的数据预测第k步状态值，则有：

$$y'_{i+k}=d'_{1,i+k}+d'_{2,i+k}+\cdots+d'_{J,i+k}+c'_{J,i+k} \tag{4-1}$$

4.1.2 预测结果评判指标

一般时间序列预测结果的评判指标有以下三种方式：

（1）均方根误差：

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2} \tag{4-2}$$

（2）平均绝对百分比误差：

$$\mathrm{MAPE}=\frac{1}{n}\sum_{t=1}^{n}\frac{|y_t-\hat{y}_t|}{|y_t|} \tag{4-3}$$

（3）平均绝对误差：

$$\mathrm{MAE}=\frac{1}{n}\sum_{t=1}^{n}|y_t-\hat{y}_t| \tag{4-4}$$

4.2 实验流程

4.2.1 实验数据的选取

本实验选取的数据是某卫星在2011年5月初某一天多内，转向轮子合成的Y轴角动量（Hy）的遥测数据。其时域图如图4-2所示。

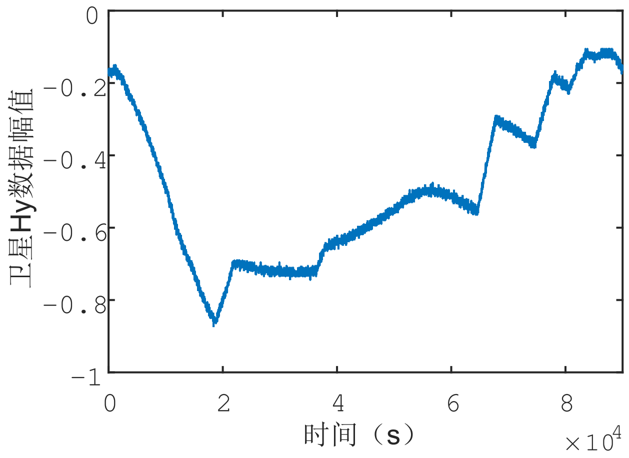


图 4-2 实验数据Hy信号时域图

从图中可以看出，数据变化趋势不固定，并且数据毛刺较多。

本实验的目标是对这段数据进行在轨实时短期预测。

4.2.2 小波分解

接下来对数据进行小波分解，综合多种因素，本实验选取db5小波，阶数选择为7。下图是得到的小波分解结果如图4-3。

从图中可以看出，随着分解层数的提高，小波分解的低频分量越平滑，高频分量展现了一些规律性。

4.2.3 对低频分量的平稳性检验及预测

在本实验中需要对低频分量进行平稳性检验，如果信号满足平稳性，则可以使用ARMA预测算法，进行预测；如果信号不满足平稳性，则需要使用SVR预测算法，进行预测。在此采用Augmented Dickey-Fuller test（ADF）单位根验证方法，利用Eviews软件来实现。

首先给出卫星遥测信号Hy的小波分解低频分量HyA7的时域图如图4-4。 从时域图可以看出，低频分量HyA7是非平稳的。

本实验对序列进行ADF检验。序列不存在明显的趋势，所以选择对常数项，不带趋势的模型进行检验，检验结果如表4-1 所示。 结果表明不能拒绝存在一个

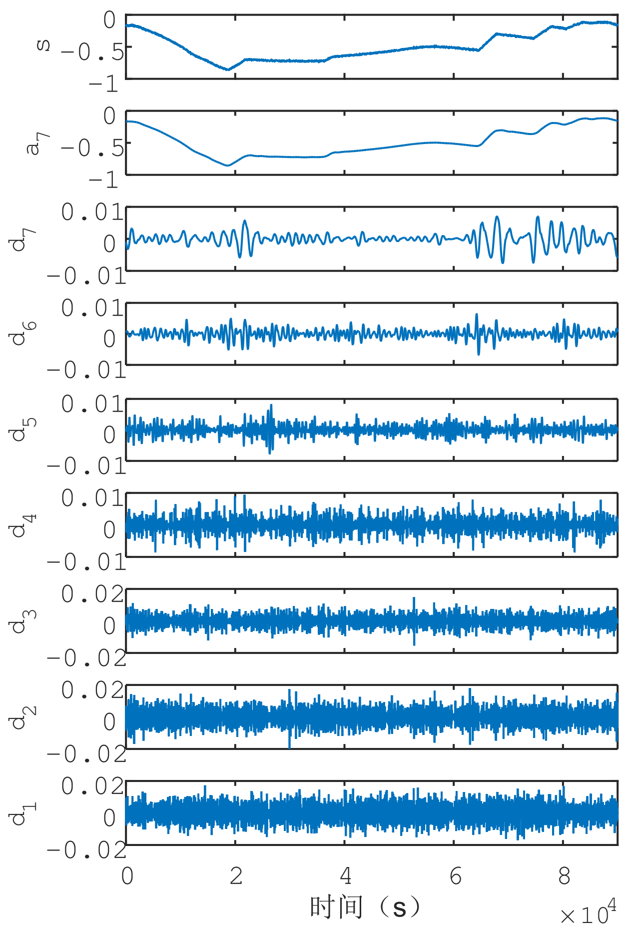


图 4-3 Hy信号的小波分解图

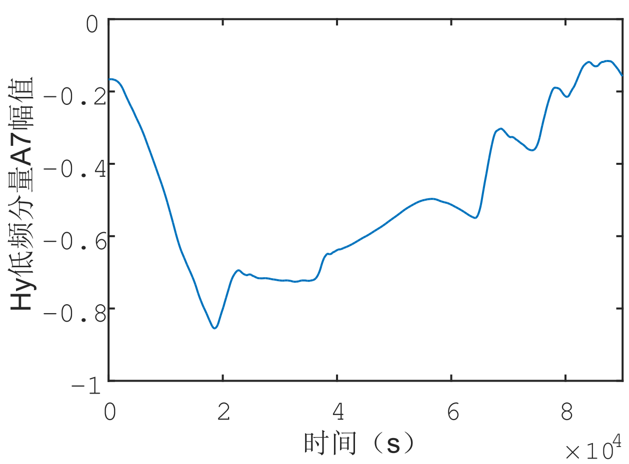


图 4-4 Hy信号小波分解低频分量HyA7时域图

表 4-1 HyA7的ADF检验结果表

	T统计量值	P值
ADF	-1.049255	0.7367
1%显著性水平检验临界值	-3.430904	
5%显著性水平检验临界值	-2.861669	
10%显著性水平检验临界值	-2.566880	

单位根的原假设，ADF统计值大于1%level、5%level和10%level的临界值，序列是非平稳的。因此，本试验将采用非稳系统分析的SVR模型对低频分量进行建模预测。

以下是对低频分量HyA7进行SVR建模预测过程。全数据段为9000个数据点，取前500个数据点对未来的10个数据点进行预测，并且采用的是在轨的动态预测方法。在轨SVR预测结果图如图4-5 所示。

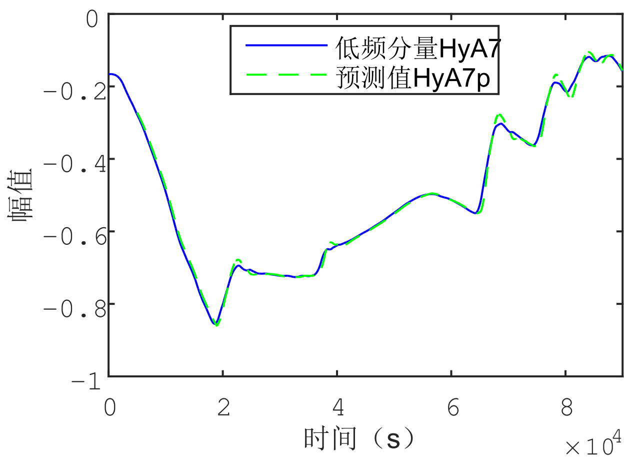


图 4-5 Hy信号小波分解低频分量HyA7时域图

在图中，蓝色的实线是原始的时域数据曲线，绿色的虚线是预测结果。可以从图中直观的看到预测结果比较理想。

4.2.4 对高频分量的平稳性检验及预测

在此，本实验取卫星遥测数据Hy的小波分解后的其中一个高频分量HyD1经行演示。HyD1的时域波形如图4-6所示。

从图中可以看到，信号是满足平稳性的。并且卫星遥测信号的小波分解高频分量一般都具有平稳性的，但是在此，我们同样要使用Eviews进行平稳性检验，并来确定ARMA的模型参数。

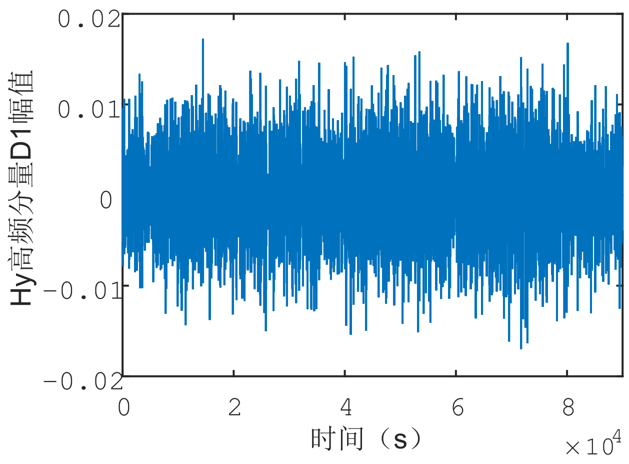


图 4-6 Hy信号小波分解低频分量HyA7时域图

试验中对高频分量HyD1进行ADF检验，检验结果如下表4-2所示，ADF统计值远小于三个比较值，序列是平稳的，可以对其进行ARMA建模预测。

表 4-2 HyA7的ADF检验结果表

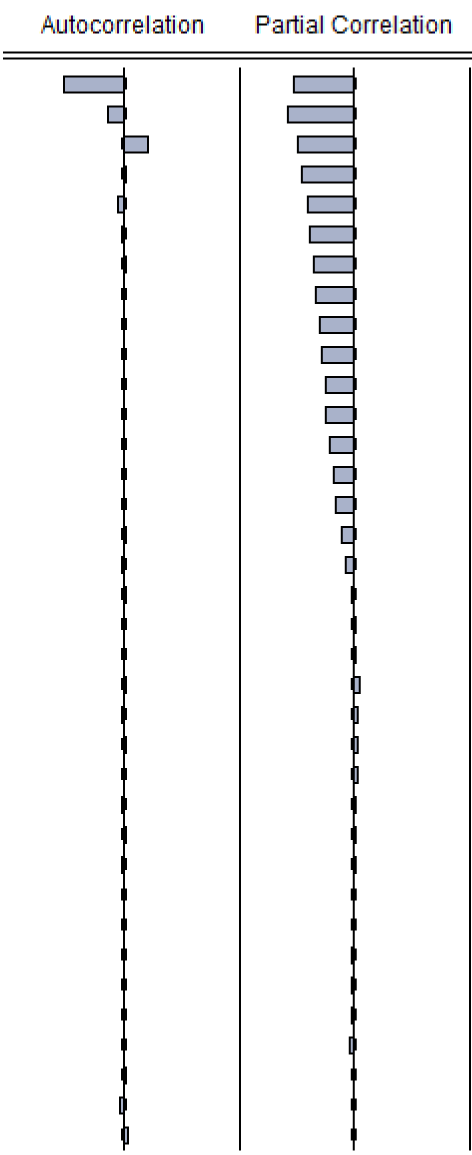
	T统计量值	P值
ADF	-50.83602	0.0001
1%显著性水平检验临界值	-3.430904	
5%显著性水平检验临界值	-2.861669	
10%显著性水平检验临界值	-2.566880	

接下来确定HyD1的ARMA模型参数。先求得HyD1高频分量的相关图，如图4-7所示。

根据序列相关图中的自相关系数和偏相关系数估计模型参数，序列的自相关系数在3阶后截尾，ARMA模型参数p可以考虑选择2，偏自相关系数在5阶后拖尾，ARMA模型参数q可以考虑选择5。

但是在选择ARMA模型参数时，通过自相关系数和偏相关系数估计出的参数只能作为参考，而目前常用的信息准则为AIC、BIC、FPE等。AIC准则是最小信息准则，它的目的是判断预测目标的发展过程与哪一随机过程最为接近。在应用中，设置模型的阶数在一定范围内从低到高，对不同的阶数的模型计算AIC值，AIC值最小对应的阶数就是合适的阶数。

根据自相关系数和偏相关系数估计出的参数作为范围，算出AIC指标如表4-3所示。



	AC	PAC	Q-Stat	Prob
1	-0.537	-0.537	2598.5	0.000
2	-0.142	-0.606	2780.4	0.000
3	0.218	-0.511	3206.7	0.000
4	0.010	-0.474	3207.7	0.000
5	-0.053	-0.424	3233.4	0.000
6	-0.013	-0.398	3234.9	0.000
7	0.021	-0.364	3239.0	0.000
8	-0.001	-0.337	3239.0	0.000
9	-0.005	-0.315	3239.2	0.000
10	0.004	-0.285	3239.3	0.000
11	-0.004	-0.260	3239.5	0.000
12	0.001	-0.246	3239.5	0.000
13	0.006	-0.221	3239.8	0.000
14	-0.007	-0.178	3240.2	0.000
15	-0.006	-0.157	3240.5	0.000
16	0.018	-0.100	3243.6	0.000
17	-0.018	-0.073	3246.5	0.000
18	0.011	-0.020	3247.5	0.000
19	-0.007	0.016	3247.9	0.000
20	0.001	0.022	3247.9	0.000
21	0.012	0.045	3249.1	0.000
22	-0.019	0.040	3252.3	0.000
23	0.010	0.035	3253.2	0.000
24	0.003	0.037	3253.3	0.000
25	-0.010	0.018	3254.3	0.000
26	0.011	0.023	3255.5	0.000
27	-0.010	0.010	3256.4	0.000
28	0.006	-0.000	3256.8	0.000
29	0.001	-0.007	3256.8	0.000
30	-0.005	-0.009	3257.0	0.000
31	0.002	-0.013	3257.1	0.000
32	0.003	-0.018	3257.1	0.000
33	-0.003	-0.038	3257.2	0.000
34	0.008	0.007	3257.8	0.000
35	-0.026	-0.004	3263.9	0.000
36	0.032	0.007	3273.4	0.000

图 4-7 HyD1相关结果图

表 4-3 ARMA模型参数AIC结果表

p	q	AIC
2	5	-9.572221
2	4	-9.405544
2	3	-9.402324
2	**2**	**-9.672930**
2	**1**	**-9.971523**
1	5	-9.437207
1	4	-9.546885
1	**3**	**-9.626855**
1	2	-9.511113
1	1	-9.299825

根据不同模型参数算出的AIC指标，得知模型ARMA（2,2）、ARMA（2,1）和ARMA（1,3）的AIC指标都比较小。经过进一步筛选和比较AIC指标，选取ARMA（2,1）模型，其估计结果如图4-8所示。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.435015	0.010285	-42.29413	0.0000
AR(2)	-0.533300	0.009318	-57.23381	0.0000
MA(1)	-0.999145	0.016900	-59.12255	0.0000

R-squared	0.791110	Mean dependent var	-4.03E-07
Adjusted R-squared	0.791063	S.D. dependent var	0.004203
S.E. of regression	0.001921	Akaike info criterion	-9.671523
Sum squared resid	0.033197	Schwarz criterion	-9.669154
Log likelihood	43515.18	Hannan-Quinn criter.	-9.670717
Durbin-Watson stat	2.900231		

Inverted AR Roots	-.22+.70i	-.22-.70i
Inverted MA Roots	1.00	

图 4-8 HyD1的ARMA（2,1）结果图

实验结果可以看出，系数高度显著。可以看出ARMA（2,1）模型是较优选择。确定模型后，对HyD1进行ARMA动态预测。预测结果如图4-9所示。

根据以上的方法与流程，对其他高频分量HyD2至HyD7建立预测模型并进行短期预测。

得到的预测模型参数如表4-4所示。

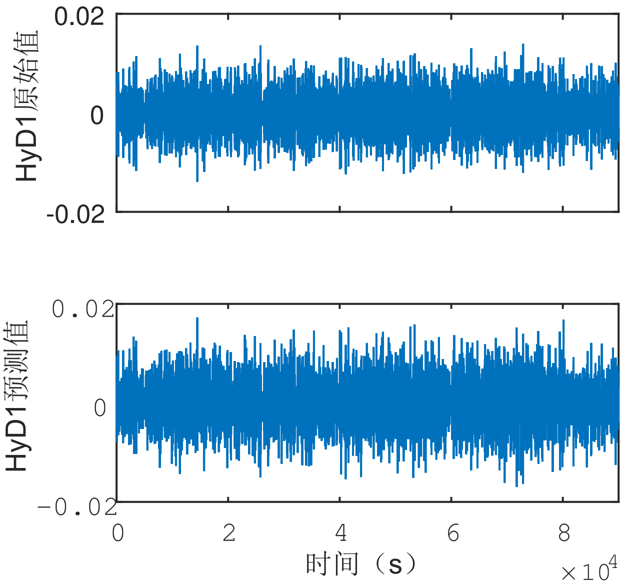


图 4-9 高频分量HyD1及其预测值比较图

表 4-4 高频分量的预测模型参数表

高频分量	预测模型参数
HyD2	ARMA（2,3）
HyD3	ARMA（2,4）
HyD4	ARMA（3,4）
HyD5	ARMA（2,3）
HyD6	ARMA（2,4）
HyD7	ARMA（4,4）

根据以上模型参数建立的ARMA模型对各个高频分量进行预测。得到的预测结果与原始时域值比较图如图4-10所示。

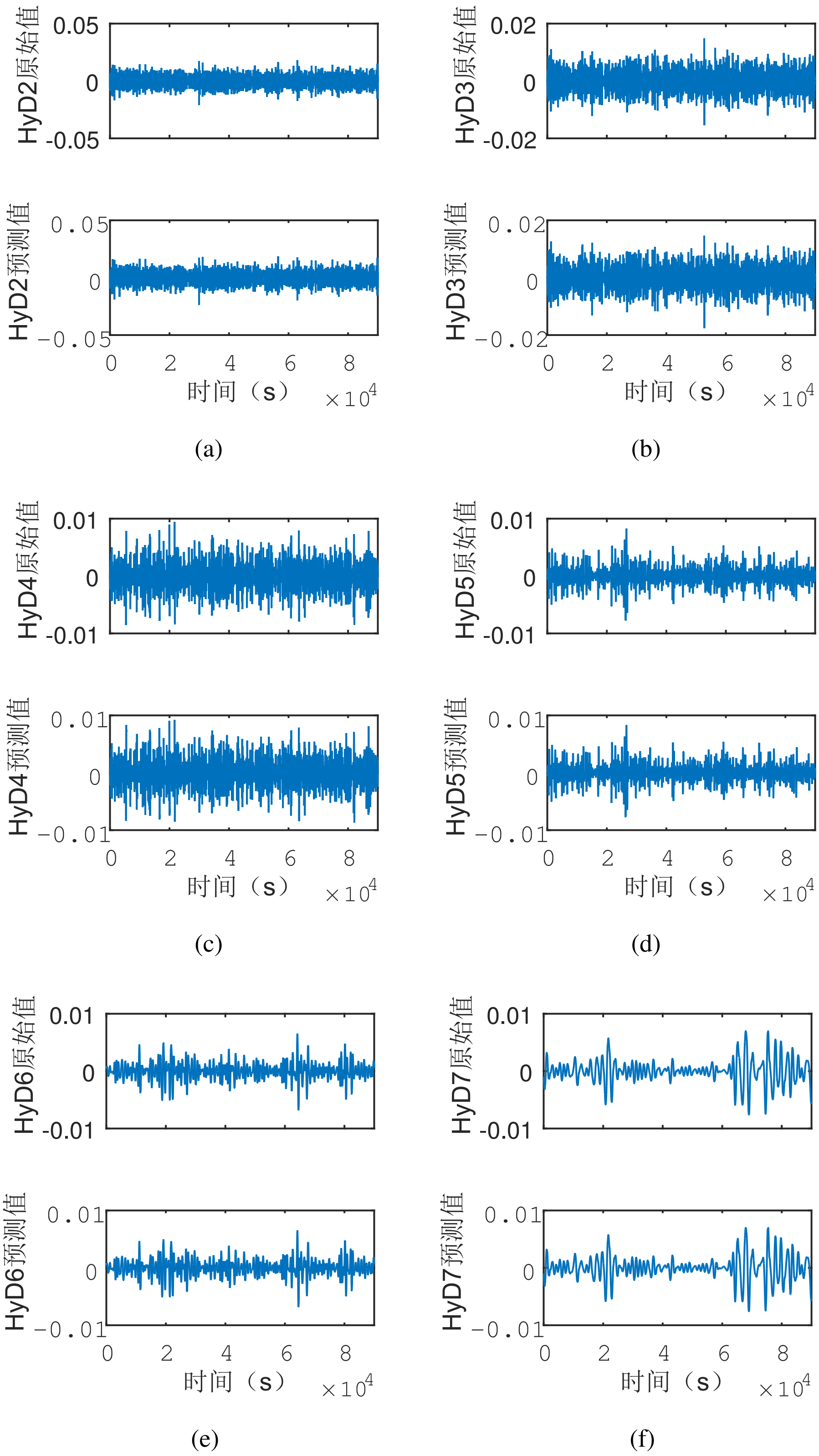


图 4-10 Hy信号高频分量的预测值与原始值比较图

(a)HyD2及其预测值；(b)HyD3及其预测值；(c)HyD4及其预测值；

(d)HyD5及其预测值；(e)HyD6及其预测值；(f)HyD7及其预测值

4.2.5 对预测值进行小波重构

根据式(4-1)对得到的高频预测值$d'_{n,i+1}$和低频预测值$c'_{J,i+1}$进行小波重构，即进行加权计算，得到最终结果如图4-11所示。

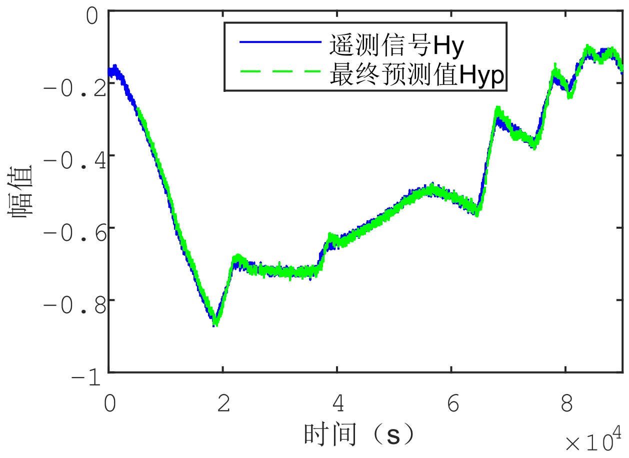


图 4-11 遥测信号Hy及其预测值Hyp

计算得出该预测结果的均方根误差$\mathrm{RMSE}=0.0109$，平均绝对百分比误差：$\mathrm{MAPE}=0.0236$，平均绝对误差$\mathrm{MAE}=0.0079$。

4.3 实验结果比较、分析

下面就一些其他方法和本论文方法结果进行比较。现行比较常用的方法有，将原始时域数据直接用SVR预测方法进行预测，或者将原始数据进行小波去噪之后，使用SVR方法进行预测。用这两者方法得到的预测结果图分别如图4-12和图4-13所示。

比较上述两个方法和本实验方法的预测指标如表4-5所示。

表 4-5 高频分量的预测模型参数表

预测方法	RMSE	MAPE	MAE
SVR预测方法	0.0160	0. 0347	0. 0115
去噪SVR预测方法	0.0132	0.0282	0.0101
本论文方法：基于趋势预测方法	0.0109	0.0236	0.0079

从预测指标看来，本论文的方法在每项指标上都比其他两个预测模型，预测的结果要好不少。

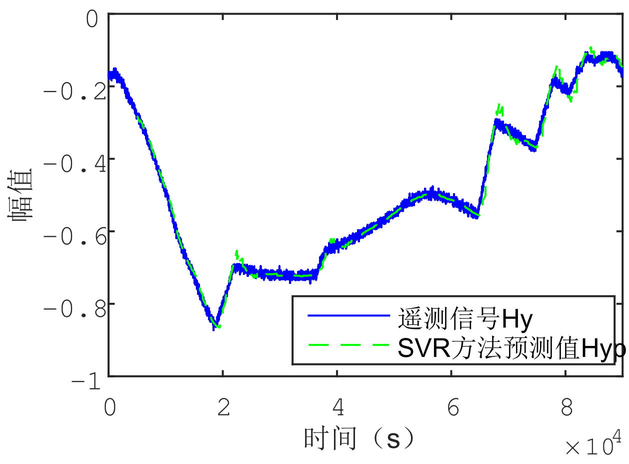


图 4-12 遥测信号Hy及通过SVR方法预测值Hyp

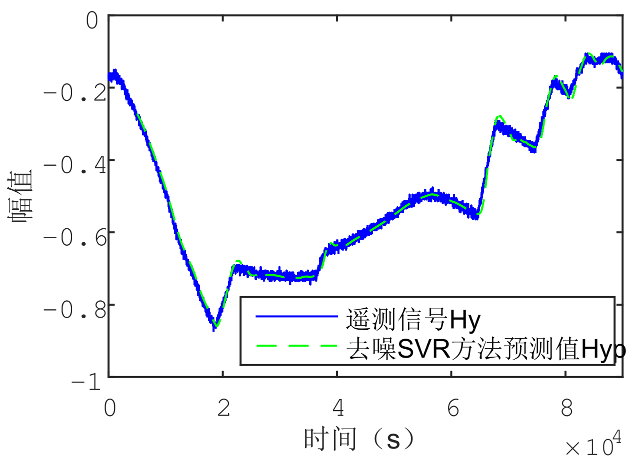


图 4-13 遥测信号Hy及通过去噪SVR方法预测值Hyp

4.4 本章小结

本章主要对基于趋势的预测算法的仿真实验进行描述。

首先，提出了基于趋势的预测算法的实验方案预测结果的评判指标；接下来，逐步讲解了仿真实验的每一步操作，并给出了逐步实验结果；最后给出了本论文算法方案与其他预测方法结果的比较，并进行分析。

通过对实际卫星遥测数据的仿真实验，可以看出基于趋势的预测算法具有优良的现实效果，对理论算法是一次有效的验证。

致 谢

历时将近两个月的时间终于将这篇论文写完，在论文的写作过程中遇到了无数的困难和障碍，都在同学和老师的帮助下度过了。尤其要强烈感谢我的论文指导老师—XX老师，她对我进行了无私的指导和帮助，不厌其烦的帮助进行论文的修改和改进。另外，在校图书馆查找资料的时候，图书馆的老师也给我提供了很多方面的支持与帮助。在此向帮助和指导过我的各位老师表示最中心的感谢！

感谢这篇论文所涉及到的各位学者。本文引用了数位学者的研究文献，如果没有各位学者的研究成果的帮助和启发，我将很难完成本篇论文的写作。

感谢我的同学和朋友，在我写论文的过程中给予我了很多你问素材，还在论文的撰写和排版灯过程中提供热情的帮助。由于我的学术水平有限，所写论文难免有不足之处，恳请各位老师和学友批评和指正！