丁月蓉 郑大伟 编著

天文测量数据的处理方法

南京大学出版社

天文测量数据的处理方法

丁月蓉　郑大伟 编著

南 京 大 学 出 版 社

1990·南京

内 容 简 介

本书共分8章，第1～2章介绍数据处理的最小二乘法及从测量数据中消除干扰的平滑、滤波方法；第3～5章介绍几种统计分析方法，包括回归分析、相关分析和时间序列分析；第6～8章详细给出了各种谱分析、谱估计的概念和方法；最后在附录中给出了15个常用数据处理方法的ＦＯＲＴＲＡＮ子程序及调用程序。

本书介绍的数据处理方法不仅适用于天文观测数据，也适用于地球物理、气象、测绘、地质等学科测量数据的处理．可供这些专业的大学生和科技工作者使用。

天文测量数据的处理方法
丁月蓉　郑大伟 编著
*
南京大学出版社出版
（南京大学校内）
江苏新华印刷厂印刷
江苏省新华书店发行　　各地新华店书经售
*
开本787×1092　1/32　印张14.5　字数326千
1990年4月第1版　1990年4月第1次印刷
印数1—1000
ISBN 7-305-00392-1/P·25　　定价：6.00元

前　言

测量数据的处理历来就是天文学的重要内容之一。随着科学技术的不断发展，特别是电子计算机的广泛应用，数据处理方法无论是在理论方面，或是应用方面，都得到了愈来愈多的重视和发展。一些新的数据处理方法也已在天文学的各个分支中广泛地应用。南京大学天文系于1980年起为本科生开设了“数据处理”课程。

本书是在1982年编写的“数据处理方法”教材的基础上，结合作者多年来的教学实践和科学研究工作，对原教材进行了修改、补充，增加了应用的实际例子，提供了所有方法的FORTRAN子程序及调用程序。

全书共分四部分。第一部分(第一，第二章)介绍如何处理具有误差的测量数据的最小二乘法及如何从测量数据中消除干扰、提取有用信息的若干种新的平滑、滤波方法；第二部分(第三～第五章)介绍几种统计分析方法，包括回归分析，相关分析以及时间序列分析；第三部分(第六～第八章)详细地给出了谱分析的基本概念和近代各种谱估计的方法；第四部分(附录一)给出了常用数据处理方法的 FORTRAN 子程序及调用实例。所介绍的这些数据处理方法带有普适性，它们也适用于地球物理、气象、地质、测绘等学科的测量数据的处理。

本书的第一章、第二章第三节，第三章至第八章由丁月蓉编写；第二章，第五章第六节及附录一由郑大伟编写，郑大伟还提供了各章应用实例。毛昌鑑同志提供了第二章的部

分初稿。

承蒙张承志老师和肖耐园老师分别审阅了本书的第一章、第三章和第二章、第四章、第五章至第八章。魏双林为本书绘制了全部插图，作者在此一并表示感谢。

由于作者水平有限，书中缺点错误在所难免，恳请读者批评指正。

作 者

1988年3月

目　　录

第一章　误差概论和最小二乘法……………………………1
第一节　误差的定义与分类……………………………2
一、绝对误差和相对误差……………………………2
二、系统误差、随机误差和过失误差……………………………3
第二节　测量精度……………………………5
一、精度标准……………………………6
二、误差传递公式……………………………8
三、等精度测量和非等精度测量……………………………10
第三节　直接测量量的最或然值及其精度……………………………13
一、最小二乘准则……………………………13
二、等精度测量列的最或然值及精度……………………………13
三、非等精度测量列的最或然值及精度……………………………15
第四节　间接测量量的最或然值及其精度……………………………17
一、误差方程……………………………17
二、正态方程……………………………19
三、最或然值的标准偏差……………………………22
第五节　正态方程的解法……………………………23
一、最或然值的解……………………………24
二、最或然值的权……………………………27
第六节　矩阵最小二乘法……………………………29

第二章　平滑与滤波……………………………33
第一节　曲线拟合……………………………33
一、多项式拟合……………………………34
二、周期函数拟合……………………………35

第二节　滑动平均……………………………………………37
第三节　数字滤波……………………………………………39
一、数字滤波的一般原理……………………………………39
二、预测误差滤波……………………………………………41
第四节　高斯权函数平滑法…………………………………45
一、卷积平滑…………………………………………………45
二、高斯权函数平滑法………………………………………46
第五节　Vondrak平滑法……………………………………48
一、Vondrak平滑法的基本原理……………………………49
二、Vondrak平滑法计算公式推导…………………………50
三、方程组的解算方法………………………………………53
四、Vondrak平滑法原理的改进……………………………57
第六节　Vondrak平滑法和高斯权函数平滑法的频率滤波器……………………………………………59
第七节　平滑的两端效应……………………………………63
第八节　选取平滑因子的几种方法…………………………65
一、测量误差法………………………………………………65
二、平滑误差法………………………………………………66
三、频率响应法………………………………………………66
四、交叉证认法………………………………………………66
第九节　平滑曲线的插值方法………………………………70
第三章　回归分析……………………………………………75
第一节　引言…………………………………………………75
第二节　一元线性回归………………………………………76
一、数学模型及其参数估计…………………………………77
二、回归方程的显著性检验…………………………………79
三、回归方程的误差和因变量的预测………………………86
四、曲线回归分析……………………………………………89

第三节 多元线性回归……93
一、多元线性回归方程的求解……94
二、多元回归的显著性检验……98
第四节 逐步回归分析……106
一、最优回归方程的选择……106
二、线性方程组的求解求逆紧凑变换……108
三、逐步回归的计算步骤……114
四、关于逐步回归算法的几点说明……119
五、逐步回归的应用……120
第五节 逐步回归分析实例……122
第四章 相关分析……133
第一节 相关系数、协方差函数……133
第二节 协方差函数的性质和应用……137
一、自协方差函数的性质和应用……137
二、互协方差函数的性质和应用……141
第三节 协方差函数的计算……143
一、自协方差函数的计算……143
二、互协方差函数的计算……151
第五章 时间序列分析……154
第一节 时间序列……155
第二节 ARMA模型……156
一、ARMA模型的定义……156
二、ARMA 模型的传递形式和逆转形式……159
三、ARMA(p,q)序列的自相关函数和偏相关函数……163
第三节 ARMA模型的建立……170
一、模型参数的估计……171
二、模型的识别与定阶……186

第四节　时间序列的预报 ……………………………………192
一、平稳线性最小方差预报 ……………………………………192
二、各类模型的预报公式 ……………………………………196
第五节　ARIMA 模型的季节性回归模型 ……………207
一、ARIMA模型 ……………………………………208
二、季节性模型 ……………………………………211
第六节　非线性时间序列的门限自回归模型 ………213
一、引言 ……………………………………213
二、门限自回归模型的定义 ……………………………………215
三、选取门限模型参数的AIC准则 ……………………………216
四、建模过程 ……………………………………217
五、时间序列的非线性检验 ……………………………………221
六、实例 ……………………………………233
第七节　时间序列分析的应用实例 ……………………226
第六章　谱分析基础及快速傅里叶变换 ………………231
第一节　傅里叶级数和离散频谱 ……………………231
第二节　傅里叶变换与连续频谱 ……………………234
一、傅里叶变换的基本公式 ……………………………………234
二、傅里叶变换的性质 ……………………………………237
三、傅里叶变换的实例 ……………………………………240
第三节　δ函数 ……………………………………241
第四节　卷积 ……………………………………247
第五节　离散化公式 ……………………………………252
一、有限离散傅里叶变换 ……………………………………252
二、离散卷积 ……………………………………256
第六节　快速傅里叶变换 ……………………………………261
一、FFT的基本原理 ……………………………………262
二、实序列的FFT ……………………………………268
三、运用FFT进行频谱分析 ……………………………………272

第七章　谱分析的数字化问题 ……………………275
第一节　引言 ……………………275
第二节　频谱混叠效应 ……………………276
一、采样定理 ……………………276
二、频谱混叠效应 ……………………278
第三节　截断效应 ……………………280
第四节　平滑窗 ……………………286
一、平滑窗的作用 ……………………286
二、几种常用的平滑窗 ……………………287
三、窗函数的选择 ……………………298
第八章　随机信号的谱估计 ……………………303
第一节　功率谱密度函数 ……………………304
一、几个定义 ……………………304
二、几个例子 ……………………308
第二节　功率谱估计法 ……………………301
第三节　相关功率谱估计 ……………………311
一、自功率谱的相关变换法 ……………………312
二、互相关功率谱估计 ……………………318
第四节　周期图估计 ……………………319
一、周期图估计式 ……………………319
二、平均周期图 ……………………322
三、平滑周期图 ……………………323
第五节　最大熵谱估计 ……………………327
一、最大熵谱密度估计式 ……………………328
二、最大熵谱估计的伯格算法 ……………………334
三、最大熵谱估计的马波算法 ……………………345
四、最大熵谱估计阶数的选取 ……………………352
第六节　ARMA模型信号谱估计 ……………………354

第七节　功率谱估计分辨率的检验 ……………………357

附录一　测量数据处理中的常用程序 ……………………364
一、产生满足正态分布的随机数 ……………………365
二、产生满足均匀分布的随机数 ……………………368
三、Householder变换的最小二乘法 ……………………369
四、谐波分析计算 ……………………375
五、线性方程组的全主元高斯消去法 ……………………380
六、逐步回归分析 ……………………383
七、三次样条插值 ……………………392
八、高斯权函数平滑法 ……………………396
九、Vondrak 平滑法 ……………………400
十、高阶Vondrak 滤波器 ……………………406
十一、周期图估计 ……………………412
十二、FFT算法(矩阵法) ……………………417
十三、时间序列AR模型的参数估计 ……………………424
十四、Burg算法的AR谱估计 ……………………428
十五、Marple算法的AR谱估计 ……………………433

附录二
表 1　F分布临界值表 ……………………446
表 2　相关系数表 ……………………452
表 3　正态分布函数表 ……………………454

第一章 误差概论和最小二乘法

天文学的很多理论是以天文观测为基础的，地球自转理论、行星运动理论、人造卫星运动理论等等都离不开天文观测．通过对某一天文量的直接或间接观测，获得大量数据，我们的目的就是要从这大量的测量数据中求出被测量值的可靠结果．

任何一种观测或测量都不可避免地具有误差．因此，当人们在使用由测量得到的结果时，必须要提出这样一个问题：这个结果是否可靠？或说这个结果精确到什么程度？这个精度是否满足要求？如果这个结果精度是高的，能满足人们提出的要求，就可以大胆放心地去使用；如果这个结果不能满足所提出的精度要求，则不能使用，就需要设法找出或消除影响结果精确度的各种误差的原因，加以改正或重新测量，直到满足要求为止．

最小二乘法是用来处理具有误差的测量数据的一种极有效的方法，也是最早用于天文观测资料处理的一种数学工具．早在1794年，高斯(Gauss)为了利用小行星坐标的多次观测准确的推算小行星的轨道，第一次应用了最小二乘法．1805年勒让德(Legendre)应用测量平差方法确定了彗星的轨道和地球子午线弧长．1809年高斯又推证了误差的概率定律，从而使最小二乘法高度完善化，成为数据处理中应用最广的一个分支．随着统计学和矩阵理论的发展以及电子计算机的广泛应用，也使最小二乘法进入近代数据处理方法的行列．

第一节　误差的定义与分类

误差不仅存在于测量值中，计算时采用近似的理论模型，计算公式中一些理论常数的不准确等也会在计算结果中产生误差。而我们下面介绍的误差是仅对观测误差而言的。

某个被测量值的误差定义为该量的观测值（或测量值）与其真值之差。真值是被测量本身所具有的真实大小，它是一个理想的概念。一般说来，真值是未知的，通常用约定值来代替。某一系统的天文常数也可以看作为相应量值的真值。测量值与真值（约定值）之差常被称为偏差。

从不同的角度出发，误差有各种不同的分类方法。按误差的表达形式可分为绝对误差和相对误差；按误差的性质及产生原因可分为系统误差、随机误差和过失误差。

一、绝对误差和相对误差

设x_i为某个被测量的测量值，a为其真值，则**绝对误差**$\varDelta_i$由下式定义：

$$\varDelta_i = x_i - a, \qquad i = 1, 2, \cdots, N \tag{1.1}$$

从绝对误差的定义式（1.1）不难看出，绝对误差和被测量具有相同的量纲。因此，若说星位误差为$0''.1$，测时的记录误差为$0^s.001$都是指的绝对误差。

在有的情况下用绝对误差来表示测量的精度是不恰当的。譬如，目前卫星激光测距的准确度（测量值与被测量真值之间的偏离程度）已达30cm级，卫星的距离一般为10^3km量级，但如果我们测定的是地球和恒星之间的距离（1或几个

天文单位），则如得到绝对误差为30cm 的测量值，那结果是准确得令人难以置信.这两种情况绝对误差都相同，但由于被测量本身大小相差很大，用绝对误差并不能描述测量值的精度，为此引入了相对误差的概念。

被测量的绝对误差Δ与其真值a之比定义为这个量的**相对误差**，并用下式表示。

$$\gamma=\frac{\Delta}{a} \tag{1.2}$$

对上面的例子，它们的相对误差分别为1×10^{-4}和1×10^{-14}.显然，在这里应用相对误差来表示测距精度要比应用绝对误差更明了。

二、系统误差、随机误差和过失误差

由观测的环境因素差异、仪器性能、不同的观测者等因素造成的按某种规律变化的误差称为**系统误差**。系统误差通常可以用某种函数关系式来近似地描述它的变化过程，必须注意仅靠多次观测是不能消除或减弱系统误差的影响的.因此，若存在某项系统误差而不消除，这是很不利的。所以人们总是首先设法估计出系统误差的值，在观测结果中加以改正。

由某些难以控制的随机因素造成的，绝对值和符号的变化时大时小、时正时负，以不可预测的方式变化的误差称为**随机误差**.当我们对被测量在相同条件下进行多次重复测量，并已消除了系统误差的影响后，测量值会出现一些无规则的变化，这就是由随机误差造成的。

随机误差产生的原因很多，观测时环境因素的微小变化是产生随机误差的一个重要原因.虽然就其个体而言，随机误差没有规律、不可预料，但就其总体而言，随着测量次数的增

加，它又服从某种统计规律。下面我们从概率论的角度出发讨论随机误差的统计规律。

古典误差理论认为，随机误差服从正态分布，这个结论从统计的角度来说也是正确的，因此我们可以用正态分布密度曲线来表征随机误差。随机误差的概率密度曲线（见图1-1）可表为

$$p(\varepsilon)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{\varepsilon^2}{2\sigma^2}\right] \tag{1.3}$$

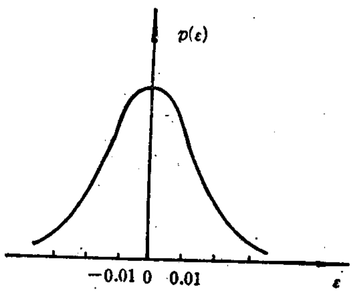

图 1-1 随机误差的理论概率密度曲线

式中σ^2为随机误差的方差。当随机误差的样本足够大时，从它的概率密度曲线不难得到随机误差有下列统计特征：

（1）绝对值相等、符号相反的正负误差数近于相等。因此，随机误差的算术平均值随着观测次数的增加愈来愈小，以零为极限。

（2）误差的概率与误差的大小有关，绝对值小的误差出现的机会比绝对值大的误差出现的机会多，绝对值很大的误差出现的概率极小。

根据随机误差的这些特征，当不存在系统误差的影响时，多次测量结果的算术平均值将更接近于真值。

前面我们分别介绍了系统误差和随机误差的定义、特点及产生的原因，但实际上系统误差与随机误差之间并没有明显的界限。有时，我们把一些具有复杂规律、暂时还掌握不到的系统误差都当作随机误差处理，而随着人们对误差及其规律的认识和加深，就有可能把这些以往认识不到因而归之于

随机误差的这类误差确认为系统误差.反之,在一个较短的时期内可能呈现出某种规律,故而归为系统误差,但经过一段较长时期的测量,发现这种变化规律破坏了,而呈现出随机性.这就是说,随着时间的推移,两种不同性质的误差是可能互相转化的。这就是系统误差与随机误差之间的辩证关系.

过失误差是指测量结果与事实明显不符的一种误差。如观测时对错星、记错时间或观测过程中望远镜螺旋松动等过失原因造成的结果异常,这些都是由观测人员工作粗枝大叶引起的。这种误差一般比较容易发现,而且只要观测人员态度认真,操作细心,是完全可以避免的.

第二节 测 量 精 度

数据处理工作中一个很重要的问题是评定一列测量值的精度.测量精度是指测量结果的可靠程度。一般说来,在消除系统误差和过失误差之后,测量的精度由随机误差的大小来衡量。因此,一列测量值精度高低必须从全列测量值的误差来衡量,而不能只根据个别值的误差来判断。例如,天文观测中,测量者的熟练程度、整个观测期间仪器及外界环境的稳定性都会造成测量结果的精度差异。所以我们不能只从个别测量值的好坏来决定精度高低.

另外,测量的目的是要从一列测量值中确定(直接地或间接地)被测量的真值,但由于测量手段和测量次数的限制,真值是测不到的,只能得到真值的一个近似值或估计值。通常,我们把最接近于被测量的真值的值称为它们的**最或然值**.因此,数据处理的又一个很重要的问题是给出被测量的最或然值及其精度。最或然值的精度是衡量测量结果的精度和处

理方法的有效性的综合指标。

在这一节里，我们首先给出衡量一列测量值精度高低的精度标准，然后介绍表征几种不同的测量序列的精度公式，最或然值的精度估计将在下面一节介绍。

一、精度标准

标准偏差（又称**均方差**）是用来衡量一列测量值精度高低的一个较好的指标，有时也把它叫作**中误差**。

设$\{x_i\}$，$i=1\sim N$为被测量的一组测量值，a为被测量的真值，且x_i中只包含随机误差，则$\varDelta_i=x_i-a$称为x_i的真误差，我们定义真误差的平方的算术平均值的平方根为这列测量值的标准偏差，并用σ表示：

$$\sigma=\sqrt{\frac{\sum\varDelta_i^2}{N}}=\sqrt{\frac{\sum(x_i-a)^2}{N}} \quad (1.4)$$ ①

这里定义的标准偏差和统计学中从方差的正平方根定义的标准差是一致的，因为从概率论的角度来说，x_i的真值可用其数学期望表示。

下面我们来说明如何从标准偏差σ的大小来衡量一列测量值的质量。

由正态分布的性质可知，测量值x_i出现在$(a-\sigma,a+\sigma)$区间上的概率，或说误差$\varDelta_i$出现在$(-\sigma,\sigma)$范围内的概率为

$$P\{(a-\sigma<x_i<a+\sigma)\}=68.3\% \quad (1.5a)$$

或

$$P\{(-\sigma<\varDelta_i<\sigma)\}=68.3\% \quad (1.5b)$$

---

① 当不发生误解时，求和$\sum\limits_{i=1}^{N}$也表达成$\sum\limits_{i}$或$\sum$，以下类同。