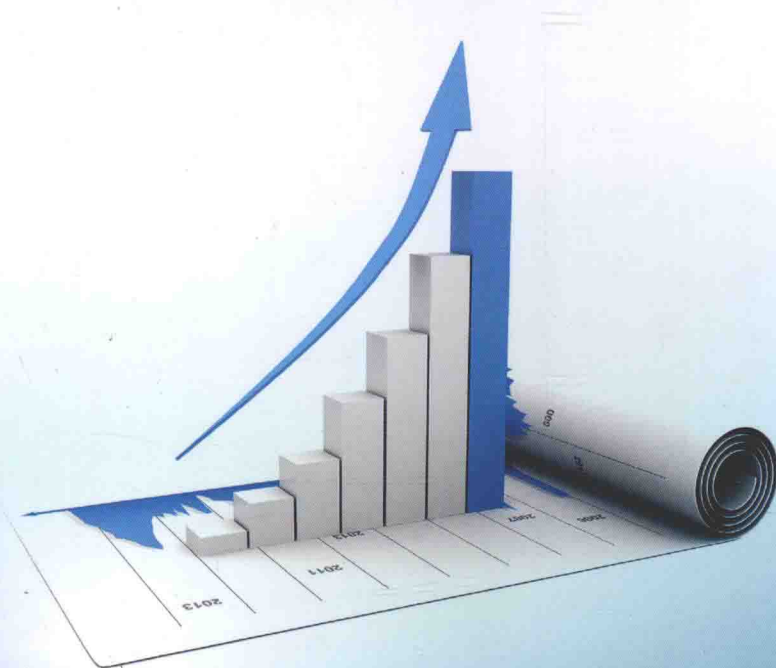


应用型本科院校系列规划教材

数据处理 及模型化方法

SHUJU CHULI JI MOXINGHUA FANGFA

付小宁 主编



西北农林科技大学出版社

应用型本科院校系列规划教材

数据处理及模型化方法

主 编 付小宁

副主编 程为彬 李智奇



西北农林科技大学出版社

图书在版编目(CIP)数据

数据处理及模型化方法/付小宁主编. —杨凌:西北农林科技大学出版社,2016.6

ISBN 978 - 7 - 5683 - 0118 - 3

I. ①数… II. ①付… III. 测量—数据处理 IV. ①P207

中国版本图书馆 CIP 数据核字(2016)第 128337 号

数据处理及模型化方法

付小宁 主编

出版发行 西北农林科技大学出版社

地 址 陕西杨凌杨武路 3 号 邮 编:712100

电 话 总编室:029 - 87093105 发行部:87093302

电子邮箱 press0809@163.com

印 刷 陕西天地印刷有限公司

版 次 2016 年 6 月第 1 版

印 次 2016 年 6 月第 1 次

开 本 787 mm × 1092 mm 1/16

印 张 17.75

字 数 380 千字

ISBN 978 - 7 - 5683 - 0118 - 3

定价: 38.00 元

本书如有印装质量问题,请与本社联系

测量数据处理是对测量所获得的一系列数据进行深入的分析,以确认各变量之间相互制约、相互联系的依存关系,往往还需要用数学建模的方法,推导出各变量之间未知的函数形式。只有经过科学的处理,才能去粗取精、去伪存真,获得反映被测对象的物理状态和物质属性的有用信息,这才是测量数据处理的最终目的。

本书介绍了测量数据处理及其模型方法。首先介绍了测量数据建模的数学基础。接着,论述了数据集映射与测量建模,包括经验模型的建立;重点讲述了数据关系的回归分析。然后,介绍了系统的时间序列模型、软测量建模。最后,给出了模型评价准则和方法。

本书由 8 个章节构成——

第一章 绪论。强调了测量模型的客观性和重要性。

第二章 测量系统的误差及抑制。测量误差是测量结果是否精确的评判依据,涉及测量系统的多个环节。

第三章 测量数据的表达与数据映射,测量过程与数据集映射、正交变换、数据的频域变换、小波变换域、Radon 变换、无模式映射。这些映射对测量数据的提纯及测量结果的分析是必要的。

第四章 测量数据建模的数学基础,介绍了待估函数的多项式表示、样条函数表示,概述了获得经验公式的途径。

第五章 强调数据关系的回归分析,以实例说明拟合具有正确性,阐述拟合的要求与实现,拟合的注意事项,最后讲述有助于提升拟合效果的数学方法。

第六章 介绍系统的时间序列模型方法:包括时间序列概述和分类,正态性检验、独立性检验、平稳性检验, $AR(p)$ 、 $MA(q)$ 、 $ARMA(p,q)$ 模型的参数估计。

第七章 软测量建模。讲述了软测量技术基本原理、技术分类和一些典型的软测量建模方法。涉及基于工艺机理分析、回归分析、状态估计、人工神经网络、回归支持向量机、模糊数学和其他方法的软测量建模。

第八章 模型评价。在论述模型评价准则的基础上,简述了评价方法,并给出了几种重

要的技术方法的案例。

作者撰写此书的目的是通过基本数学理论的介绍和实际应用的案例教学,使得学员能够较好地掌握相关方法和能力。本书内容脱胎于作者的教学讲稿,融入了本领域的研究成果,取材比较广泛。

本书由付小宁、程为彬、李智奇编写,其中付小宁编写了第1、2、5、8章,程为彬编写了第3、4章,李智奇编写了第6、7章,全书由付小宁统稿。研究生单兰鑫、雷新忠、赵蓓也参加了本书4个章节的编写,书中部分素材出自小木虫学术论坛 <http://emuch.net/bbs/> 的具体问题讨论,在此一并表示感谢。

由于时间和水平有限,书中必定存在疏漏和不足,恳请读者不吝指正,以便今后逐步改进完善。联系邮箱:xning_fu@163.com。

编者

2016年3月



第1章 绪论

1.1 测量的概念	1
1.2 测量是人类探知自然界的必要手段	1
1.2.1 测量与测量技术	2
1.2.2 测量技术的应用	2
1.2.3 测量与科学技术相伴相生	3
1.2.4 测量过程与方法	3
1.2.5 测量技术新发展的影响	4
1.3 测量模型是客观的存在	5
1.4 测量模型处于不断完善的过程	7
1.5 许多发现和发明始于仪器和测量模型的突破	9
1.5.1 卡文迪什扭称	9
1.5.2 质谱仪	10
1.5.3 电子显微术	11
1.5.4 CT 技术	11
1.5.5 DNA 双螺旋结构	11

第2章 测量系统的误差及抑制

2.1 测量系统的概念	13
2.2 测量误差	15
2.2.1 系统误差	15
2.2.2 偶然误差(随机误差)	16
2.2.3 绝对误差	17

2.2.4	相对误差	17
2.3	计算误差	17
2.3.1	从一个实例说起	17
2.3.2	算法稳定性很重要	18
2.4	模型误差与最佳测量方案	19
2.4.1	误差合成公式	19
2.4.2	误差间相关性的讨论	20
2.4.3	最佳测量方案	21
2.5	误差的抑制	24
2.5.1	基于测量方法的误差抑制	24
2.5.2	基于数据处理的误差抑制	25

第3章 基于数据集的参数表达及数据集映射

3.1	测量数据集与数据集变换	30
3.1.1	测量数据集的概念	30
3.1.2	数据集的分类	31
3.1.3	测量过程与数据集映射	32
3.1.4	几种正交变换	37
3.2	数据的频域变换	48
3.2.1	傅里叶变换家族	48
3.2.2	希尔伯特变换	55
3.3	小波变换域	58
3.3.1	小波变换的基本理论	58
3.3.2	离散小波变换/反变换	60
3.3.3	小波包与小波变换软件	62
3.4	Radon 变换	64
3.4.1	Radon 变换定义	64
3.4.2	Radon 反变换	65
3.5	无模式映射举例	67
3.6	本章小结	71

第4章 测量数据建模的数学基础

4.1	引言	75
-----	----------	----

4.2	多项式表示待估函数	76
4.2.1	均匀逼近和平方逼近	76
4.2.2	最佳一致逼近	77
4.2.3	最小偏差逼近代数多项式的逼近阶	80
4.2.4	待估函数多项式表示的基底	81
4.2.5	最佳平方逼近	86
4.3	样条函数表示待估函数	87
4.3.1	多项式样条函数的定义	87
4.3.2	三次样条函数	88
4.3.3	B 样条函数	90
4.4	用常微分方程的通解表示待估函数	93
4.4.1	一元常微分方程	94
4.4.2	欧拉(Euler)法与后退欧拉法	94
4.4.3	龙格-库塔方法	97
4.4.4	二阶微分方程	101
4.5	经验公式	106
4.5.1	费米问题	106
4.5.2	如何建立经验公式?	108
4.5.3	典型的经验公式介绍	109
4.5.4	有偏估计与无偏估计	121
4.5.5	待估参数的选择	121

第5章 数据关系的回归表示

5.1	拟合具有正确性	124
5.2	拟合的要求与实现	125
5.2.1	已知模型的拟合	125
5.2.2	已知模型的校验	126
5.2.3	自己总结规律的拟合	131
5.3	拟合的注意事项	133
5.3.1	数据点的充分性	133
5.3.2	数据噪声或测量精度	133
5.3.3	模型(公式)的正确性或者适用性	138
5.3.4	算法是否先进或适用	140

5.3.5	参数初值是否合理	140
5.3.6	迭代计算的终止判断指标高低	143
5.4	几种增强回归效果的方法	144
5.4.1	逐步线性回归	144
5.4.2	主成分分析	144
5.4.3	偏最小二乘	147
5.4.4	几种方法的对比	147
5.5	本章小结	151

第6章 系统的时间序列模型

6.1	时间序列简介	154
6.1.1	时间序列概述	154
6.1.2	时间序列的定义及分类	154
6.1.3	时间序列模型的应用	155
6.2	时序观测数据的检验	155
6.2.1	正态性检验	155
6.2.2	独立性检验	156
6.2.3	平稳性检验	156
6.3	系统模型的时间序列表示	158
6.3.1	零极点匹配法	158
6.3.2	双线性变换法	159
6.4	平稳时间序列模型的参数估计	159
6.4.1	AR(p)模型的参数估计	160
6.4.2	MA(q)模型的参数估计	168
6.4.3	ARMA(p, q)模型的参数估计	177
6.4.4	基于某型车载 GPS 所得北向定位数据的 ARMA 模型的参数估计	181
6.5	平稳时间序列建模	182
6.5.1	模型的选择	182
6.5.2	模型的检验	183
6.5.3	两个例子	184
6.6	非平稳时间序列	189
6.6.1	时间序列的非平稳性	189
6.6.2	ARMA 模型	190

6.6.3	PAR 模型	191
6.6.4	PAR 模型拟合及参数估计	193
6.6.5	非平稳时间序列的建模与预测的一个例子	195
6.7	小结	197

第7章 软测量建模

7.1	软测量的概念	200
7.1.1	概述	200
7.1.3	软测量技术基本原理	201
7.1.4	软测量技术分类	201
7.2	基于工艺机理分析的软测量建模	202
7.2.1	工艺机理分析	202
7.2.2	电厂汽包水位软测量模型的建立	202
7.2.3	汽包水位软测量系统的实现	204
7.3	基于回归分析的软测量建模	207
7.3.1	多元线性回归 (MLR, Multivariate Linear Regression)	207
7.3.2	多元逐步回归法 (MSR)	208
7.3.3	主元分析与主元回归 (PCA、PCR)	209
7.3.4	基于最小二乘法的糖液过饱和度软测量	211
7.4	基于状态估计的软测量建模	212
7.4.1	状态估计	212
7.4.2	基于状态估计的软测量的基本原理	213
7.4.3	基于状态估计的软测量在管道煤气控制系统中的应用	214
7.5	基于人工神经网络的软测量建模	215
7.5.1	人工神经网络概念	215
7.5.2	神经元特征函数	216
7.5.3	人工神经网络模型和分类	217
7.5.4	典型神经网络模型	218
7.5.5	基于神经网络的软测量建模的应用	223
7.6	基于回归支持向量机的方法	225
7.6.1	引言	225
7.6.2	支持向量机回归	226
7.6.3	最小二乘支持向量机	228

7.6.3	基于最小二乘支持向量机的软测量建模	229
7.7	基于模糊数学的软测量建模	230
7.7.1	模糊数学的基本概念	230
7.7.2	基于模糊数学的软测量	231
7.8	其他方法	235
7.8.1	基于过程层析成像的软测量方法	235
7.8.2	基于模式识别的软测量方法	236
7.8.3	基于相关分析的软测量建模	237
7.8.4	基于现代非线性信息处理技术的软测量建模	237
7.9	本章小结	237

第8章 模型评价

8.1	评价准则	241
8.2	评价方法	242
8.2.1	正确性的评价	242
8.2.2	稳定的评价	246
8.2.3	灵敏度的评价	248
8.2.4	足够的线性度及动态范围的评价	249
8.2.5	数学美感和高计算效率的评价	252
8.2.6	测量系统成本的评估	253
8.2.7	适用性评价	253
8.3	蒙特卡洛仿真	253
8.3.1	定义	253
8.3.2	实现方法	253
8.3.3	应用	254
8.4	正交实验法	257
8.4.1	定义	257
8.4.2	实现方法	258
8.4.3	应用	260
8.5	克拉美-罗下界	264
8.5.1	克拉美-罗下界定义	264
8.5.2	实现方法	265
8.5.3	类似的下界	267

第1章 绪论

测量数据是测量活动的科学记录,它包含了被测系统的信息,并因此成为人类认知事物的主要源头和验证资料。一般情况下,测量数据难免会受到测量方案、测量工具或设施、及人员操作的影响。

1.1 测量的概念

300多年前,伽利略从望远镜中仰望苍穹,宣布了一个科学家的远大抱负:“有必要测量一切可测的,并努力使那些尚不可测的成为可测。”

测量是一种基于人类感知能力的行为,随着人类的认知水平和辅助工具的发展而系统化成一门技术,最终上升为一门科学。测量活动中的辅助工具或设施称为测量仪器(或科学仪器)。现代测量是借助于测量仪器研究事物自然属性及变化规律的一门科学。作为测量者,人始终是测量活动的主体。

著名科学家门捷列夫说:“科学是从测量开始的,没有测量就没有科学,至少是没有精确的科学、真正的科学”。测量科学的先驱开尔文又说,一个事物你如果能够测量它,并且能用数字来表达它,你对它就有了深刻的了解;但如果你不知道如何测量它,且不能用数字表达它,那么你的知识可能就是贫瘠的、是不令人满意的。测量是知识的起点,也是引领人类进入科学殿堂的开端。

1.2 测量是人类探知自然界的必要手段

一切科学都建立在精确的数据上,自然科学是如此,人文科学也是如此。而精确数据的获得依靠的就是测量。正如著名科学家钱学森先生曾指出的,“信息技术包括测量技术、计算机技术和通信技术。测量技术是关键和基础。”测量水平的高低直接反映一个国家科学水平的高低。

1.2.1 测量与测量技术

测量是人类认识事物本质的不可缺少的过程,是人类对事物获得定量概念及其内在规

律的过程。

测量技术可分为传感器技术、测量方法和仪器仪表技术等三部分。其中,测量方法包括测量模型、数据处理,是测量系统的理论支撑;仪器仪表技术是测量系统的硬件支撑。科学研究和工程技术中所要测量的参数大多为非电量,这些参数的物理特性或化学特性千差万别,正是因为传感器技术的发展,拓宽了电子测量技术与系统的内涵及应用。例如,放射性现象是人类肉眼看不见的,要靠特定的传感器(传感技术)来证明它的实际存在。

测量数据是测量系统的输出,见图 1.1。

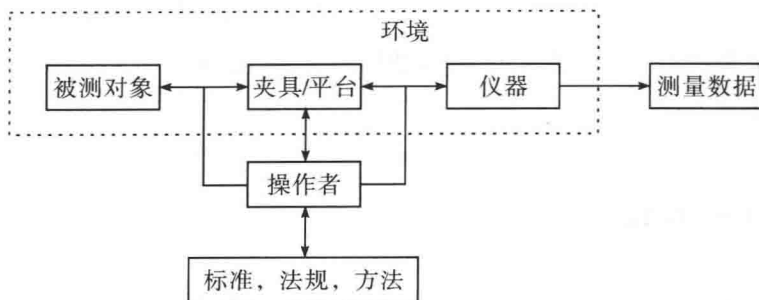


图 1.1 测量系统与测量数据

在图 1.1 所示的测量系统中,传感器可以在仪器内部、也可在夹具式平台上。仪器及传感器的选配与测量方法有关,有的仪器可以完成一部分数据处理功能,如 FFT 功能模块,复杂的数据处理需要专门的处理才能完成。测量模型指观测量与待求变量的数学关系式。

1.2.2 测量技术的应用

测量技术是人们为了对被测对象所包含的信息进行定性地了解 and 定量地掌握所采取的一系列技术措施。测量是科学研究的基础和前提,测量的真实性和精确程度直接决定了研究的成败。从科技发展来看,高精尖的研究更需要高精尖的测量技术。

借助于检测工具对产品进行质量评价是人们十分熟悉的,这是测量技术中最重要的应用领域。测量技术是自动化系统中不可缺少的组成部分。任何生产过程都可以看作是由物流和信息流组合而成的,反映物流的数量、状态和趋向的信息流则是管理和控制物流的依据。为了有目的地进行控制,首先必须通过检测获取有关信息,然后才能进行分析判断以实现自动控制。

另外,随着新型测量技术的不断成熟和发展,它在大型设备安全经济运行和监测中得到了越来越广泛的应用。例如,电力、石油、化工、机械等行业的一些大型设备,通常都在高温、高压、高速和大功率状态下运行,保证这些关键设备的安全运行具有十分重要的意义。为此,通常设置故障监测系统以对温度、压力、流量、转速、振动和噪声等多种参数进行长期动态监测,以便及时发现异常情况,加强故障预防,达到早期诊断以避免严重的突发事件,保证设备和人员安全,这就需要先进的数据处理技术。

1.2.3 测量与科学技术相伴相生

当人类活动领域越过感觉器官极限时,仪器仪表成了一切事业成功的前提。科学进展首先取决于测量技术的进展。

例如,温度计和实用温标的发明(17、18世纪)使温度的概念具有更加准确的科学涵义,成为可以测量和定量计算的基本物理量。直接导致了热力学的诞生,发现了能量守恒定律和热机的一系列基本规律,为欧洲的产业革命奠定了坚实的科学基础。

电流表的发明(19世纪)使电学与磁学研究获得了一个又一个重大的发现,促进了电气时代的来临。

威尔逊云室和众多核物理探测仪器的发明(20世纪)揭开了原子核反应的神秘面纱,逐渐展现出微观世界的真实图景,奠定了原子核物理学与日后原子能利用的基础。近代自然科学是从真正意义上的测量开始的。

许多杰出的科学家都是科学仪器的发明家及新测量方法的创立者。他们留给后世的科学遗产常常包括两个部分,一部分是科学探索的新发现,另一部分是在这种探索过程中创造的新测量技术和仪器仪表。

表 1-1 史上科学家与典型的仪器仪表举例

序号	科学家	仪器仪表
1	伽利略	天文望远镜
2	托里拆利	水银气压表
3	卡文迪什	卡文迪什扭秤
4	安培	电流表
5	基尔霍夫,本生	光谱仪

科学发现和技术发明是从对事物的观察开始的。对事物的精细观察就要借助于仪器,就要测试,特别是在自然科学和工业生产领域更是如此。在对事物的观察、测试基础上经过分析推导,形成认识,上升为事物的描述模型(包括测量模型自身)。

1.2.4 测量过程与方法

依据测量过程,对测量可以有如下分类:

1.2.4.1 直接测量和间接测量方法

(1)直接测量方法

无须进行函数关系的再运算,从实测数据中直接得到被测量值的测量方法。如:直尺测量长度、温度计测温、电压表测电压等。

(2)间接测量方法

直接测量与被测量存在某种关系的各个物理量,通过一定的数学关系获得被测量。如:

流量可通过流速和管道直径测定值而求得。

1.2.4.2 接触测量和非接触测量方法

(1) 接触测量方法

仪表的一部分(传感器)与被测对象相接触并承受对象参数的作用。如:热电偶测温、扭矩式转速计测转速等。

(2) 非接触测量方法

仪表的传感器是不必与被测对象直接接触而给出测量结果的测量方法。如:光学高温计测温、雷达测速仪测量速度等。

1.2.4.3 偏位测量和差示测量方法

(1) 偏位测量方法

被测量作用于仪表的比较装置(测量机构),使其某参量按已知关系变化,当此变化产生的反作用与被测量的作用相平衡时,则测量机构即可显示被测量。如磅秤称重、指针式仪器等。

(2) 差示测量方法

被测量对仪表比较装置的作用由已知量部分或全部抵消,测量机构输出的是已知量与被测量的差值或零,由已知量得出被测量。如托盘天平称重、大象体重称量等。

1.2.4.4 在线测量和离线测量方法

(1) 在线测量方法

在测量过程中,测量机构输出被测量的即时测量值的方法。如:波谱心律仪测心律、柜式电流表、电压表等。

(2) 离线测量方法

取样离线后再行测量的方法。如:岩石成分的取样测定等。

1.2.4.5 静态测量和动态测量方法

(1) 静态测量

测量时被测量物理量处于静止或动平衡状态。

(2) 动态测量

测量时被测物理量处于渐变状态。

1.2.5 测量技术新发展的影响

一是大大提高了被测参数的精度。现代宇航陀螺仪制造,误差控制在“纳米级”以内。超大规模集成电路内部线路间距、物理光栅的刻画,其误差控制级别要求更高。检测技术的新发展为被测参数实现超高精度测量提供了技术保证。

二是极大地扩展了测量的对象和领域。在传统工业、农业、商务物流以及科学实验中,

大型复杂的对象面临多输入参数和多输出参数的综合测量与控制,这离不开新型测量工具和现代测量理论的支持。此外,航空航天、遥感遥测、海洋开发、环境保护、现代化战争的演习等,都离不开新型检测技术的支持。

1.3 测量模型是客观的存在

测量模型是对测量系统或测量方案的科学描述。测量模型来自于人们对被测事物的认识,是自然规律的体现,是一种客观存在。

测量模型有的来自类比推理,如卡文迪什扭称与库伦扭称;再如,知名的摩尔定理称芯片上开关器件的密度每 18 个月翻一番,其实这是在 Taniguchi 于 1974 年提出的制造误差按指数曲线下降的预测基础上得出的。正是精密工程按 Taniguchi 预测的曲线发展,保证了计算机工业的发展,从而保证了摩尔定理的兑现。有的来自对已有体系的修正,如相对论力学体系对经典牛顿力学体系。有的来自于新的理论方法,如突变论、量子物理学。有的来自猜测—验证的结果,如夸克模型。这些模型在一定条件下,与自然现象吻合很好,说明其具有客观性。

测量模型的客观性表现在以下几个方面:

(1) 模型不依人们认识途径而改变

例如,在 19 世纪中期,就有人发现离体肌肉在经过反复刺激后会变酸和变得僵硬。后来人们发现,这种令肌肉变酸的物质是乳酸。有科学家发现,离体肌肉不仅能产生乳酸,还能转化乳酸,后者取决于肌肉是否得到了足够的氧。

1910 年,希尔第一次利用 Blix 设计的热电偶检流计检测肌肉产热的规律。他推测,在肌肉的代谢活动中,乳酸很可能作为一种重要的化学物质参与反应,肌肉中乳酸的产生与供氧没有关联,但乳酸的转化和燃烧则需要氧气的支持。1918 年,梅耶霍夫则从化学角度入手,经过仔细计算和定量实验阐明了乳酸在肌肉活动中的氧化和向糖原的转变^[1]。至此,乳酸在肌肉活动中的规律(模型)被揭示出来。

再如,冯康的有限元和国外的有限元在微分方程数值解方面的研究。虽因封闭而各自为政,最后殊途同归。

(2) 同一模型不因人们表达方式的差异而受到影响

例如,波动力学和矩阵力学是等价的,只是表述方式不同,但是波动力学更易于接受。薛定谔使用波函数描述粒子,并给出的薛定谔方程是偏微分方程,更为当时的物理学家所熟悉;而海森堡把力学量和算符都用矩阵表示,并给出海森堡方程,是一个算符方程,后来薛定谔证明了这两种表述是等价的^[2]。

也有同一个模型,在不同领域名称不同,如统计学中的 lasso 模型和信号处理领域的压缩感知模型,实质是一样的。

(3) 正确的模型是理论与实验的殊途同归

例如,1924 年,德布罗意在 his 博士论文中,提出了电子具有波动性的观点。也正在这一年,美国物理学家戴维孙在一次试验当中,发现电子似乎并不完全具有粒子的性质。因为他观察到,打出来的电子束强度随靶子(锌板)的取向而变化,就好像是一束波绕过障碍物时发生的衍射那样。实际上,他的实验证明了电子的波动性,只是他没有认识到罢了^[3]。1926 年,戴维孙到英国访问,知道了德布罗意的发现;回到美国后,戴维立即再现了试验,证实了这个理论。

(4) 模型具有唯一性

例如,牛顿在物理学上的引力定律与胡克、哈雷、惠更斯、雷恩的发现相“碰撞”;光学上的色散研究与我国明朝方以智的精彩概括相“碰撞”;数学上微积分则与莱布尼兹相“碰撞”^[4]。

迈尔因为和其他科学家同时发现了物质不灭和能量守恒定律而受到焦耳的粗暴攻击,焦耳认为这个定理是他自己证明的,迈尔没有完全证明这个定理。而同时发现这个定理的赫尔姆茨则坚决维护迈尔的声誉,认为这是“迈尔不依赖别人的独立发现”。

日心说模型的天文观测证据是由伽利略(1564—1642)提供的。伽利略手中多了一样第谷所没有的东西:天文望远镜。1610 年 1 月,伽利略发现有 4 颗卫星围绕着木星运行,这对托勒密模型是一大打击。对托勒密模型最致命的打击发生于 1610 年 9 月,伽利略发现金星像月球一样会出现周期性的盈亏变化。金星的位相变化正是日心说所预测的,托勒密模型则完全无法解释。在伽利略公布了这一发现后,大部分天文学家就不得不放弃托勒密模型,改用日心说模型^[5]。

开普勒从小因病眼力受损,没法自己做天文观察。可以说第谷精准的观测为开普勒发现行星运动三大定律,完善日心说模型奠定了基础。为什么开普勒三大定律能够与实测数据符合得很好,根本原因在于日心说模型是客观存在的。更重要的是,开普勒的行星运动三大定律,直接影响到后来万有引力定律的产生。

为了简便起见,这里把行星轨道看作圆形。根据开普勒面积定律,行星应做匀速圆周运动。圆周运动的向心加速度

$$a = v^2/r \quad (1-1)$$

其中, v 是行星运行速度, r 是圆形轨道的半径。

代入牛顿第二定律, $f=ma$,有

$$f = mv^2/r \quad (1-2)$$